PRINCIPLES AND PRACTICE OF GPS SURVEYING

Chris Rizos

MONOGRAPH 17

SCHOOL OF GEOMATIC ENGINEERING

THE UNIVERSITY OF NEW SOUTH WALES SYDNEY NSW 2052 AUSTRALIA

MONOGRAPH 17, 1997

PRINCIPLES AND PRACTICE OF GPS SURVEYING

Chris Rizos

Received: January, 1997

SCHOOL OF GEOMATIC ENGINEERING
THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY NSW 2052
AUSTRALIA

The School of Geomatic Engineering
The University of New South Wales
Sydney NSW 2052
AUSTRALIA

Received January, 1997

National Library of Australia
Card Number and ISBN
0 85839 071 X

Contents

Preface

The Global Positioning System (GPS) is a space-based, microwave, 24-hour, all-weather, global military navigation system designed, deployed, financed and managed by the U.S. military authorities. Since GPS was declared operational in mid-1993 it has had a profound impact on the art and practice of most forms of positioning and navigation. However, GPS has already had a tremendous impact on surveying, initially as a technology for "control surveys", for which purpose it was first introduced into many countries during the early 1980's -- well before the full satellite constellation was available to navigation users. In fact, the application of GPS for control surveys (or so-called "geodetic" surveys) was the first civilian use of GPS which was well beyond that for which GPS was originally intended by its military designers.

Nowadays GPS surveying techniques have completely replaced technologies such as Doppler satellite positioning and even long-range EDM for traditional first (and lower) order control surveys. However, the adoption of GPS is not restricted to control survey applications. More and more GPS is being used for cadastral, topographical and engineering surveying. The constraints that have previously restricted the application of GPS technologies, primarily those of *cost* (capital and running) and *productivity* (measured in terms of the numbers of points which can be coordinated in a day) are being aggressively addressed by the manufacturers, and it is confidently predicted that GPS will shortly be used by the majority of surveyors and geomatic engineers.

These notes are designed to provide the student with an understanding of the principles and practice of GPS surveying. The subject is divided into twelve topics:

- Topic 1: Principles of satellite positioning
- Topic 2: Introduction to GPS positioning
- Topic 3: The GPS signals
- Topic 4: GPS instrumentation
- Topic 5: GPS satellite surveying
- Topic 6: Modelling GPS observations
- Topic 7: Introduction to GPS processing
- Topic 8: GPS baseline processing
- Topic 9: Elements of GPS network processing
- Topic 10: GPS and quality control
- Topic 11: Result transformation and presentation
- Topic 12: Datums and the future of GPS

By working through these topics in a systematic manner the student will obtain a thorough overview of all aspects of the GPS surveying technology. An important objective of these notes is to dispel myths and incorrect perceptions of the capabilities (and shortcomings) of the technology. This is generally the result of people being bewildered by the ever increasing GPS "jargon". Hence it is the intention of this course to, for example, distinguish between the "GPS navigation" techniques based on the relatively imprecise pseudo-range observations (for which GPS was originally designed, and which is still the standard positioning mode for navigation-type applications), and the "GPS surveying" techniques specifically developed for precise positioning applications (which make use of the phase observations of the signal carrier waves).

In addition to focussing on the specialist technology of "GPS surveying", the sub-categories of "conventional GPS surveying" as well as "modern GPS surveying" will also be covered in these notes. *Why is this necessary?* Contributing to the increasing popularity of GPS has been the evolution of precise GPS surveying from a relatively difficult, expensive and complicated technology that could only be used in the so-called "static" mode, to a technique that has tremendous flexibility, including being able to be used in the "kinematic" (moving receiver) mode. This increases the number and range of applications that can be addressed by the GPS surveying technology. To therefore appreciate the directions in which GPS is developing, as well as to be aware of the real (and perceived) constraints on GPS performance, it is necessary to understand the fundamental principles of the GPS hardware, processing algorithms and operational procedures. These notes give this background.

The notes are very comprehensive, and it is unlikely that the student can digest all of the material to the same level of thoroughness. There is therefore considerable scope for "tailoring" a GPS course by giving more emphasis on some topics at the expense of others. However, as a general statement when the student has studied these notes s/he should have a better understanding of:

- the background to the GPS system
- the GPS signals and measurements
- the GPS hardware and software
- the principles of GPS survey planning
- the basic GPS field procedures
- the principles of GPS data processing
- the process of building up networks using GPS results
- the various result transformations which may be necessary
- the issue of quality control as it applies to GPS surveys
- the directions in which the GPS technology is developing

The cartoons and many of the figures in these notes were prepared as part of an Education & Training Foundation grant to the GPS Consortium to develop GPS training materials for surveyors (original members: School of Geomatic Engineering, The University of New South Wales; Research School of Earth Sciences, Australian National University; The Land Information Centre, New South Wales Department of Conservation and Land Management; Building and Construction Division of the New South Wales Department of Technical And Further Education).

Chris Rizos

School of Geomatic Engineering
The University of New South Wales
Sydney, Australia

January 1997

Author

Chris Rizos obtained his Bachelor of Surveying (Hons.1) degree from The University of New South Wales in 1975, and his PhD from the same university in 1980. He was a Fulbright Fellow in 1977, a Rothmans Fellow in 1979, and the holder of an Alexander von Humboldt Fellowship in 1981-83 while a postdoctoral scholar in Munich.

In 1984 Dr. Rizos returned to the School of Surveying (the name was changed to the School of Geomatic Engineering in 1994), as a Research Officer where he joined the team researching the new GPS technology for high precision applications such as the measurement of continental drift. He was appointed to a lectureship at UNSW in 1987, was promoted to Senior Lecturer in 1989, and to Associate Professor in 1996. He is manager of the School's Geodesy Research Laboratory, as well as leader of the Satellite Navigation and Positioning group, which specialises in research for a variety of static and kinematic applications of GPS.

Dr. Rizos has been active GPS education since 1985, having convened and presented at a large number of GPS workshops, seminars and conferences, in Australia and overseas. He regularly speaks at GPS conferences and workshops, and was co-author of the first textbook on GPS surveying in 1985. He is a co-convenor of the Tropical School of Geodesy, a member of the council of the Australian Institute of Navigation, a chairman of a special study group of the International Association of Geodesy, as well as a member of several commissions and sub-commissions relating to geodesy.

He has spent two periods of sabbatical leave in Germany, one at the German Geodetic Research Institute, in Munich, in 1991; and the other at the Geo-Sciences Research Centre (GFZ), in Potsdam, in 1995.

His teaching interests are in GPS surveying and navigation, Law of the Sea, classical geodesy and advanced topics in geodesy. His present fields of research are the GPS technology and software for precise positioning applications.

Chapter 1: Principles of Satellite Positioning

1.1 DESIGNING A SATELLITE POSITIONING SYSTEM

In order to appreciate why the NAVSTAR Global Positioning System (GPS) is configured the way it is, and how it satisfies the needs of a variety of users, it is necessary to understand something of the basic concepts and technological aspects of a satellite-based positioning system. Some general objectives for designers of a *hypothetical* satellite positioning system can be identified:

- ❑ **To devise a system of positioning based on <u>control stations in near-earth orbit</u>, rather than on a conventional network of ground control, that would satisfy the positioning requirements of both stationary and moving platforms.**
- ❑ **To study the impact of various <u>satellite positioning design scenarios</u>, on both the system controllers and the users.**
- ❑ **To investigate the <u>technological aspects</u> of such a system, as they relate to satisfying both moderate and high accuracy requirements.**
- ❑ **Where technological requirements are not entirely met, to <u>devise operational and/or computational strategies</u> that can overcome these deficiencies.**

An additional objective may be included:

To identify those elements of the GPS system which have parallels with the hypothetical satellite positioning system referred to above.

The approach used in this text is to work through the design process in a logical manner, introducing new elements as they are required, and to clearly identify the advantages, problems and impacts of each new system element.

The <u>premises</u> underlying this design process are:

- that satellite positioning technology is worth developing, in order to satisfy a range of navigation and surveying applications.
- that the process can be understood from the viewpoint of conventional notions of positioning.
- that the design of any satellite positioning system is a compromise process between often unreconcilable issues such as economical considerations, technological barriers, and

political imperatives.

The aim of this process is, in effect, to "demystify" GPS, by first understanding the principles upon which it is based, and then through this understanding gain an appreciation of both the *limitations* of the system and its *potential* for accommodating innovative positioning strategies.

1.1.1 DESIGN SPECIFICATIONS

Some design specifications for a hypothetical satellite-based "Global Positioning System":

(1) **Satisfy as large a range of users as possible, military as well as civilian**:

- From those requiring dekametre level accuracy to those seeking millimetre level accuracy,
- From those interested only in position, to those also requiring velocity and time information,
- All air, sea and land users.

without altering the basic system configuration or operation.

☞ *Important for justifying the enormous investment in the system.*

(2) **Relatively low user cost as well as ease-of-operation**.

Hence minimal complexity, low power requirement and minimum bulk for the user equipment.

☞ *Important for gaining user acceptance.*

(3) **Unrestricted access by all users**.

An important attraction over competing positioning technologies.

☞ *Implying global coverage, 24-hour and all-weather operation.*

(4) **Satisfy military positioning requirements**, such as:

- Suitable for all classes of platform: aircraft (jet to helicopter), ship, land (vehicle to handheld units) and space (missiles and satellites).
- Real-time positioning and velocity determination capability.
- Positioning results available on a single geodetic datum.
- Level of real-time accuracy available to non-military users to be controllable.
- Resistant to jamming, and careful attention to other "survivability" issues.
- Non-detectability of the user hence no requirement to transmit signals of any kind.

☞ *Important as the military is the main lobby group for the development of such a system.*

1.1.2 IMPLICATIONS OF OPTING FOR A SATELLITE POSITIONING SYSTEM

There are a number of implications to the use of a satellite-based positioning system:

- ❑ **The satellites must be placed in suitable orbits so as to be "seen" by as many users as possible, from locations all over the earth.**
- ❑ **The reference coordinate datum to which they relate is largely defined by the system controllers.**
- ❑ **Users do not have a direct interest (or control) in the system, and policy decisions regarding the operation of the system tend to be made by the sponsoring agency.**
- ❑ **By replacing monumented ground control stations with satellites, the well-established procedures for terrestrial surveying and navigation may no longer be valid.**
- ❑ **The orbiting control stations are invisible to the user.**
- ❑ **Position determination is most conveniently performed in relation to a three-dimensional datum.**
- ❑ **The coordinates of the orbiting control stations change with time (and must be continually updated).**
- ❑ **Errors in both the coordinates of the satellites (the ephemerides) and in the measurements themselves will affect the final quality of the position determination.**

Some of these are discussed below.

Consequences for the Satellite Deployment Strategy

There are a important consequences in replacing hundreds of thousands of geodetic control stations on the ground (at separations ranging from tens to thousands of kilometres) by "orbiting control stations":

- To provide a global, 24-hour service, a constellation of satellites must be deployed. However, as satellites are expensive, *their total number must be kept to a minimum.*
- *High altitude satellites give good coverage*, but satellite-ground geometry may be weak unless observations are made over very long distances.
- Low altitude satellites most closely mimic the function of ground-based control stations, but too many would be required to ensure global, 24-hour coverage. They also travel at high speed relative to the ground and hence are only visible for short periods of time.
- A compromise needs to be reached as far as *satellite altitude* is concerned, as this most directly impacts the coverage of the satellite system.

- To ensure reliability, redundancy should be built into the system for example by having more than the minimum number of satellites visible to a user at any one time. This requires that more than the minimum number of satellites should be deployed.

Consequences for Geodetic Datum Definition

As the datum is defined by the *coordinates of the satellite tracking stations* :

- *A single global geodetic datum is ensured.*
- National datums have no relevance in satellite positioning, except when the results must be compared with those obtained using conventional ground-based techniques.
- Individual nations have no control over the satellite datum. If the system controllers chose to change the datum in some way, users must accommodate these changes.
- The computation of the satellite ephemerides is an important part of the maintenance of the global satellite datum, and hence is usually the responsibility of the system controllers. The quality of the ephemerides has an effect on the accuracy and reliability of the positioning results obtained by the user.

Consequences of Surrendering National Control

As a satellite-based positioning system is likely to be developed, funded and controlled by one nation (and usually the military agency within that country):

- Operation of the system is firmly vested in the hands of one agency within one country, and the influence of most users is small.
- Decisions that affect the coordinate datum of the system are not necessarily made with the global users in mind.
- If the system controllers are opposed to a certain class of user (for example on political grounds), efforts can be made to restrict system access or performance for such users.

Consequences of Measuring to Satellites

A satellite-based positioning system is likely to have the following effects on the design of user equipment:

- The satellites are invisible, hence the user does not have direct "contact" with the system and its operations are largely of the "black-box" variety.
- In order to increase user acceptance of a satellite-based positioning system, the user equipment should be relatively inexpensive. This has the following consequences:
  - as much of the complexity should be placed in the satellites as at all possible,
  - omnidirectional, small antennas should be used,
  - receivers should be small, low power units, being essentially "passive", receiving the satellite signals, and not required to transmit a signal to the satellite(s), and
  - as many of the computations should be carried out within the user equipment, hence reinforcing the "black-box" nature of these systems.
- The most appropriate technology is that based on *microwave, one-way (passive) ranging*. This also satisfies the requirement for all-weather, night-and-day operation, as well as the military requirement for non-detectability of the user by an enemy.

Consequences to Conventional Notions of Positioning

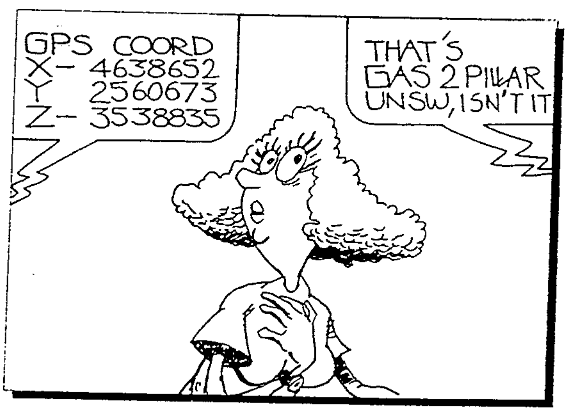
A satellite-based positioning system will have important implications for users:

- Position will be determined in three-dimensions.
- The artificial separation of horizontal and vertical positioning operations will no longer have relevance.
- Design of procedures for position determination will no longer be based on previous notions of ground network "geometry".
- The coordinates of the orbiting control stations must be easily accessible to users. This is most appropriately done by transmitting the satellite ephemerides with the signals themselves.

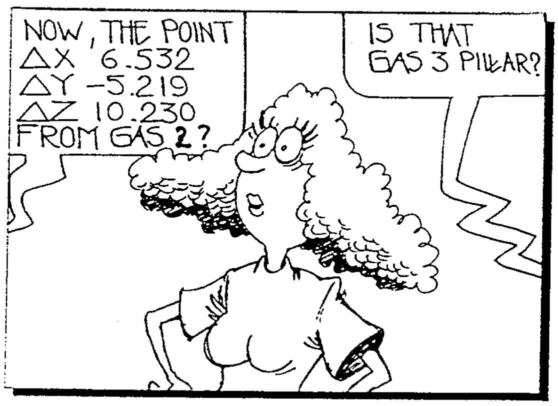
1.1.3 PRINCIPLES OF POSITIONING

By positioning is meant the determination of the spatial location of objects:

- With respect to a coordinate system whose origin is uniquely defined, and generally inaccessible. Positioning in this system is known as **point positioning** or **absolute positioning**.
- With respect to another point, taking that point as the origin of a local coordinate system. This mode of positioning is known as **relative positioning** or **differential positioning**.



Absolute Positioning



Relative Positioning

Positioning Modes

Absolute Positioning

In this mode of positioning the reference system must be rigorously defined and maintained. No direct access to the origin, or the reference axes, is usually possible and total reliance is placed on the *integrity* of the coordinated points within the reference system. In general the origin is the **geocentre**, and the axes of the system are defined in a conventional manner. In classical geodesy, astronomic observations were the only means by which an absolute position (more correctly, the horizontal components of position) could be obtained.

In modern geodesy, satellite tracking offers the means by which 3-D position can be determined with varying degrees of accuracy. Satellite point positioning is the process by which:

- given the **position vector of the satellite** being tracked (in the global system);
- given the **range vector** from the ground tracking station to the satellite being tracked (in the same system);
- determine the **position vector of the ground station**.

This is conceptually illustrated in Figure 1.1-1. The following comments can be made:

(1) Depending on how the range vector is measured, different satellite positioning techniques are possible.
(2) The position vector of the satellite changes with time, and the task of computing the satellite ephemeris requires the special skills of the satellite geodesist.
(3) The ground receiving station may be stationary or in motion.
(4) The "natural" coordinate system for satellite positioning is a **geocentric coordinate system usually realised in the form of an orthogonal Cartesian reference frame**, the primary directions being defined by the rotation axis of the earth and an arbitrarily selected principal direction (either to a point in space if the reference frame is non-rotating, or the intersection of the Greenwich meridian and the equatorial plane in the case of an earth-fixed reference frame).

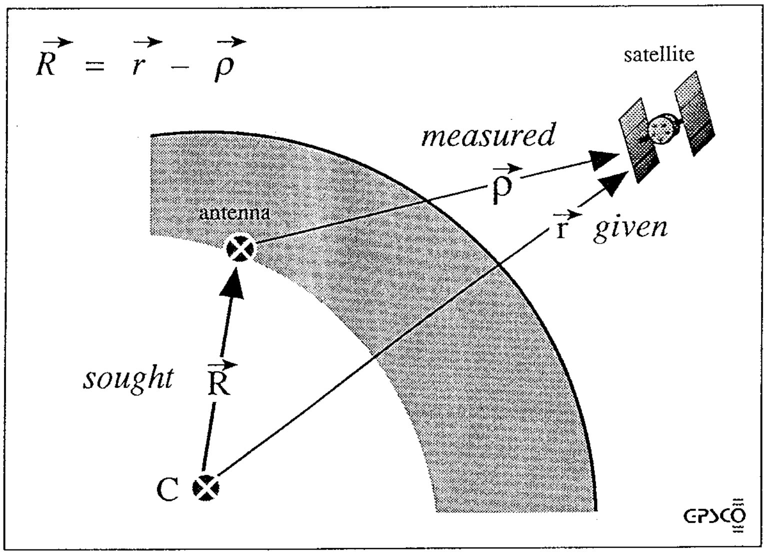


Figure 1.1-1. Basic concept of satellite point positioning.

R is the position vector of the antenna relative to the origin
r is the position vector of the satellite relative to the origin
ρ is the range vector between the antenna and satellite

Some space geodesy technologies can determine the absolute position of a stationary object to a very high accuracy, as for example the Satellite Laser Ranging (SLR) technique. In general, however, the coordinates of a station *in an absolute sense* are determined to a much lower accuracy than the precision of the measurements themselves.

Relative Positioning

This is the mode used in conventional terrestrial geodesy. Although coordinates are expressed in terms of the three coordinate components of a global reference system, they are derived from:

- **observations made to nearby control points whose coordinates are known.**

In **classical geodesy** the absolute coordinates of an "origin" station in a geodetic datum are arbitrarily assigned, and their relation to the geocentre is therefore rather poorly defined. However, as a result of high precision geodetic survey operations, the coordinates of other stations are determined to a comparatively high accuracy, *but only in a relative sense.* An entire family of points can be fixed in this way to construct a network. A network is an efficient means of propagating position information, and given the many possible "pathways" in a network from one station to any other station, the "redundant" information can be utilised in a network "adjustment" to derive the best set of coordinates for all the network control points.

Since conventional terrestrial positioning technologies have in the past been used exclusively to determine the interstation vectors, the links between adjacent network points have been restricted to those which have the property of *station intervisibility.* It is usual to distinguish between a horizontal geodetic network, for which the latitudes and longitudes of control points are determined to a high (relative) accuracy, and a geodetic levelling (or vertical) network comprising points whose heights are known accurately. In general, the horizontal control points have only weakly determined heights, while level benchmarks do not have accurate horizontal position information.

In the case of GPS, absolute position is rather poorly defined (that is, the coordinates relative to the origin of the global satellite datum), but relative positioning of two or more stations can be performed to a very high precision. Conceptually, relative position is the difference between the two position vectors (in the *global system*), expressed in a local reference system with origin at one of the ground stations. Most errors in absolute position are common to both sets of coordinates, and hence largely cancel from the baseline components. In this case the positioning accuracy approaches that of the basic measurement precision itself, and this is therefore the standard **GPS surveying mode** (as well as for *precise* GPS navigation).

Extraterrestrial Observations

Ideally, to determine the position vector of a station, or the baseline vector between a pair of stations, it is preferable to measure both the length of the range vector and its orientation. There is no one system that can provide all this information simultaneously, and to the precision required for surveying and navigation applications. The space positioning techniques that have been developed to date are generally based on the following observation technologies (Figure 1.1-2):

- Measurement of **range** from a ground station to a single satellite, as in the case early microwave systems and the present Satellite Laser Ranging (SLR) and Lunar Laser Ranging (LLR) systems.
- More recently, the measurement of **simultaneous ranges** to a number of GPS satellites.
- Measurement of **range-difference** as in the case of Very Long Baseline Interferometry (VLBI), TRANSIT Doppler and differential GPS.
- Measurement of **directions** to a satellite by optical systems which photograph a satellite against the star background.

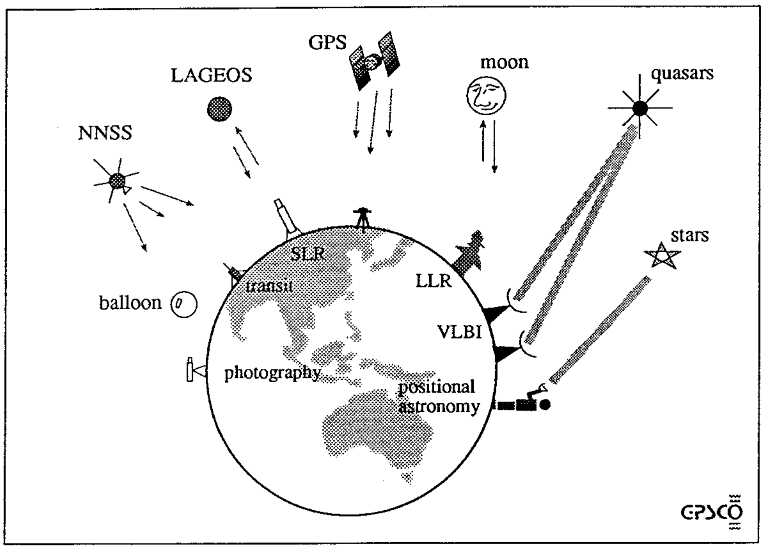


Figure 1.1-2. Extraterrestrial geodetic positioning technologies.

It is beyond the scope of these lecture notes to describe the systems based on these measurement technologies and the reader is referred to textbooks such as SEEBER (1993), and the article by PARKINSON et al (1995), for details. The focus of these notes will be on microwave range and range-rate technologies, which are the basis of the GPS and TRANSIT Doppler satellite positioning systems.

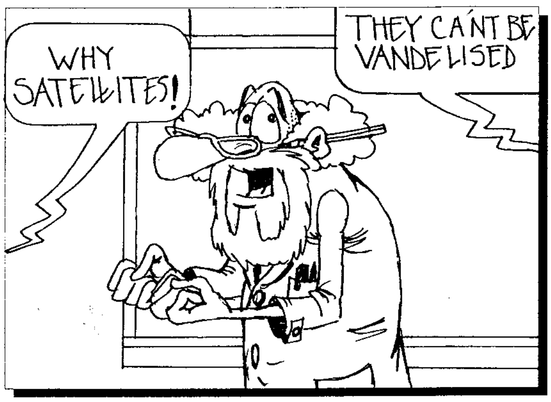
1.2

SYSTEM INGREDIENTS

WHY SATELLITES?

Satellite-based positioning systems are ideal from a number of points of view:

- **They transmit signals that can be "seen" over a far larger area than ground-based systems.**
- **With the appropriate technology they can transmit signals through cloud and rain, and can be used day or night.**
- **They recognise no national boundaries, and hence can be used globally wherever they are visible, on the ground, in the air and at sea.**
- **Satellites are tamperproof.**
- **For most national authorities, there is no investment in the necessary space hardware.**



However, in order to ensure an appropriate design for the satellite positioning system, there are several issues to be considered:

- The **system configuration**: *the "pattern" of satellite deployment (impacting on the coverage of the system), and the reference frame definition.*
- The **measurement technology** to be used: *the nature of the observations, their errors and biases, and suitability for 3-D positioning.*
- The **positioning principles** to be employed: *influenced to a large extent by the above issues.*

1.2.1 SYSTEM CONFIGURATION

Satellites function as "orbiting control stations". There are three issues concerning their use:

- the issue of visibility from ground tracking stations,
- the issue of geodetic datum definition, and (closely related to this), and
- the continuous computation of the coordinates of the satellites.

Satellite Availability and Visibility

As a consequence of their motion satellites "rise" and "set" in a manner similar to stars. There are a number of orbit concepts that are useful for understanding how satellite motion is related to ground station visibility:

(1) PERIOD

This is the time taken (T) for a satellite to complete one revolution. It is related to the semi-major axis of orbital ellipse a according to Kepler's third law:

$$\frac{2\pi}{T} = \sqrt{\frac{GM_e}{a^3}} \quad ; \qquad GM_e \approx 3.986 \times 10^{14} m^3 s^{-2} \qquad (1.2\text{-}1)$$

Hence the *higher* the satellite the *longer* the period, as illustrated in Table 1.2-1.

Table 1.2-1. Orbital period and satellite altitude.

Semi-major axis a (km)	Altitude (km)	Period (min)	Comment
6700	300	90	
7200	800	100	remote sensing satellites
10600	4200	180	
12800	6400	240	
16800	10400	360	
26600	20200	718	GPS satellites
42300	35900	1436*	geostationary satellites

* Length of a sidereal day

(2) Satellite GROUNDTRACK

The trace of the sub-satellite point across the surface of the earth. The angle of the equatorial crossing is a function of the period of the satellite and the rotation rate of the earth. In the case of GPS satellites the groundtracks are very nearly running north-south. *The satellite inclination defines the maximum and minimum latitude of the groundtrack.* If the inclination is less than 90° the orbit is said to be **prograde**, and if it is greater than 90° the orbit is said to be **retrograde**. A plot of the groundtrack of a satellite on a polar plot centred at a tracking station's zenith is known as a "skyplot". Figure 1.2-1 shows the groundtracks for two GPS satellites over one day.

(3) STATION HORIZON

The station horizon, or visibility circle, is the locus of all points around the tracking station which define the sub-satellite points observable at the local horizon. In practice satellites are not tracked down to the horizon, but to some minimum elevation angle (say 15° to 20°). The size of the visibility circles varies with the altitude of the satellite. In Figure 1.2-1 the 20° visibility circles for a station in Australia (Canberra), and a station in the U.S. (Austin) are drawn. Note that the satellites are visible to a ground observer for several hours at a time, but there is no mutual visibility of satellites between Australia and the continental U.S. By comparison, Figure 1.2-2 shows the one day groundtrack and associated 20° visibility circles for a comparatively low altitude (≈1000km) satellite.

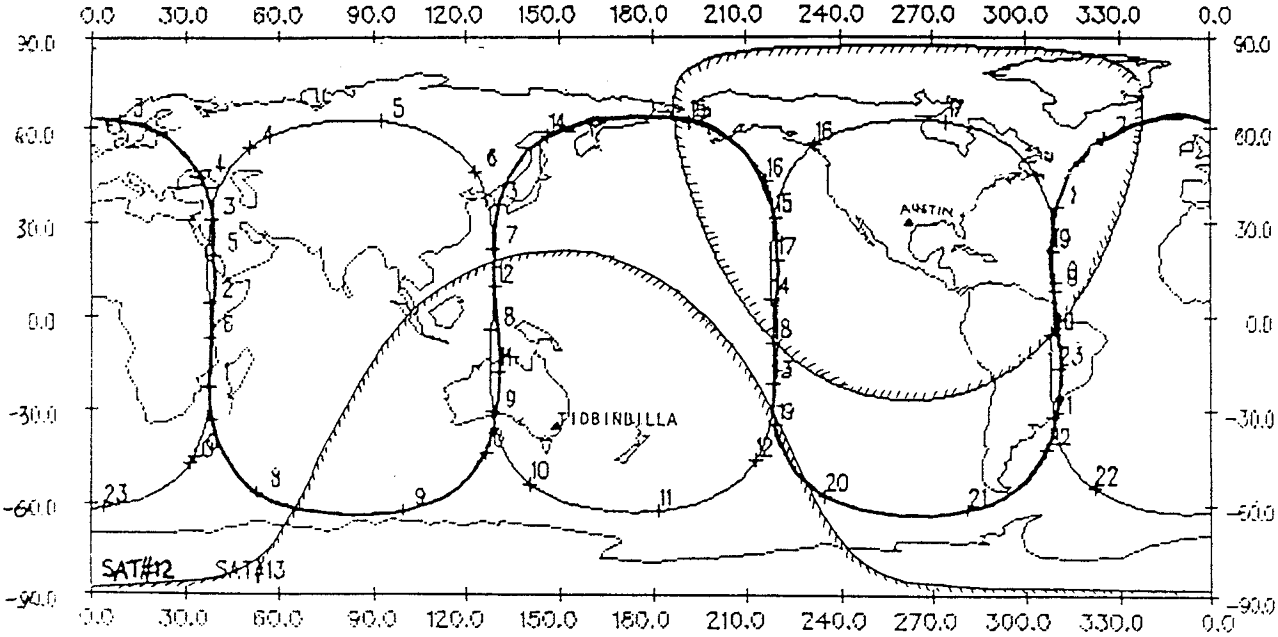


Figure 1.2-1. Groundtracks of two GPS satellites for one day, with 20° station visibility circles for a site in Australia and in the continental U.S.

(4) REPEAT PATTERN

The number of revolutions in a sidereal day:

$$R \approx \frac{1436}{T} \qquad (1.2\text{-}2)$$

where T is in minutes. If the number of daily revolutions is not an integer, the groundtrack repeat period is the number of days needed to complete an integer number of revolutions. For example, $14\frac{1}{3}$ revs/day implies a 3 day repeat period. After 3 days the groundtrack is along the same path as previously. *In the case of GPS, the groundtrack repeats each day to within a few kilometres.*

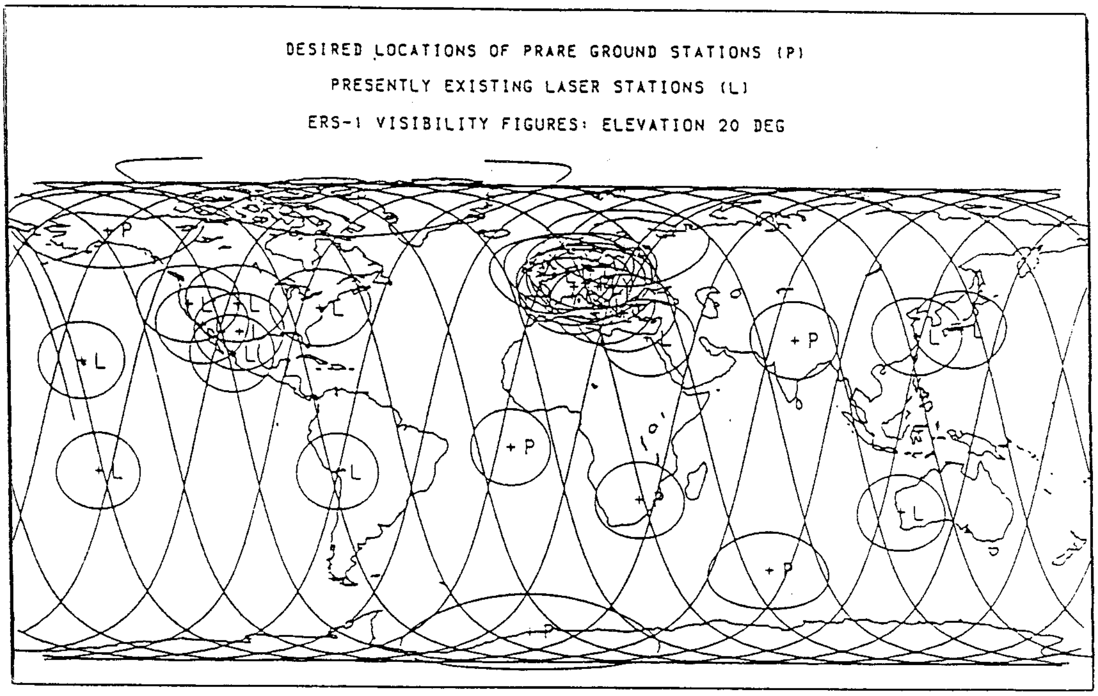


Figure 1.2-2. One day groundtrack for the ERS-1 low altitude remote sensing satellite, with 20° station visibility circles for selected tracking sites around the world.

(5) AZIMUTH and ELEVATION of the Satellite

This can be computed from the coordinates of the ground station and the satellite. The unit vector pointing from the ground station to the satellite is first computed:

$$\mathbf{u} = \frac{\mathbf{x}^s - \mathbf{x}}{\rho} \tag{1.2-3}$$

where ρ is the station-satellite range, $\mathbf{x}^s$ is the satellite position vector (in the earth-fixed reference system), and $\mathbf{x}$ is the station position vector. ϕ and λ are the ellipsoidal latitude and longitude of the station, and $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are the unit vectors of the local coordinate frame given by:

$$\mathbf{i} = \begin{bmatrix} -\sin\phi\cos\lambda \\ -\sin\phi\sin\lambda \\ \cos\phi \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} -\sin\lambda \\ \cos\lambda \\ 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} \cos\phi\cos\lambda \\ \cos\phi\sin\lambda \\ \sin\phi \end{bmatrix} \tag{1.2-4}$$

The zenith angle ζ and the azimuth α follow from the solution to the scalar products (HOFMANN-WELLENHOF et al, 1994):

$$\mathbf{u} \cdot \mathbf{i} = \sin\zeta\cos\alpha \quad , \quad \mathbf{u} \cdot \mathbf{j} = \sin\zeta\sin\alpha \quad , \quad \mathbf{u} \cdot \mathbf{k} = \cos\zeta \tag{1.2-5}$$

The following conclusions can therefore be drawn:

☞ The higher the satellite, the longer they are visible above the horizon (the extreme case is the geostationary satellites).

☞ The higher the satellite, the better the coverage due to a combination of long flyover times and extended visibility of the satellite over large areas.

☞ No satellite can be seen simultaneously from all locations on the earth.

☞ Depending on the measurement type and positioning principles being employed there may be a requirement for observations to be made to several satellites simultaneously from one or more ground stations.

The orbital characteristics of GPS satellites are described in greater detail in §2.2.

Satellite Datums

Satellite datums are the product of the Space Age, having evolved to satisfy the requirement of defining satellite and ground receiver position in a *global sense*. Unlike local geodetic datums, which are essentially defined by parameters associated with one ("fundamental") point, satellite datums are generally defined by:

(a) The **physical models** such as the adopted gravity field of the earth, together with fundamental constants such as GM_e, the rotation rate of the earth, the velocity of light, etc.

(b) The **geometric models** such as the adopted coordinates of the satellite tracking stations used in the orbit determination procedure, and the models for precession, nutation, polar motion, etc., that relate the space-fixed reference system (in which the satellite's ephemeris is computed) to the earth-fixed one (in which the tracking station coordinates are expressed).

In essence, the satellite datum is *defined* by the above models, and *maintained* by the satellite ephemerides expressed in the earth-fixed system. This datum has the following characteristics:

- The datum is **geocentric** (for two reasons: the geocentre is the physical point about which the satellite orbits; and it is preferable to any one local geodetic datum).
- The datum is generally defined as a **Cartesian system**, with axes oriented close to the principle axes of rotation ("z-axis") and the intersection of the Greenwich meridian plane and the equatorial plane ("x-axis").
- There are a number of different satellite datums, each associated with different satellite tracking technology (for example SLR, TRANSIT), and different combinations of gravity field model, earth orientation model, and tracking station coordinates used for orbit computation.

The latter is important in the context of the GPS datum: the World Geodetic System 1984 (§2.1). The relationship between some different satellite datums is given in Table 1.2-2 (taken from SOLER & HOTHEM, 1989). These are different manifestations of earth-fixed reference systems. The differences are characterised by the seven parameters: three origin shift, three small rotation angles and a scale difference (§11.1).

Table 1.2-2. Transformation parameters between various global reference systems.

Coordinate system (datum)	Δx^a (m)	Δy^a (m)	Δz^a (m)	ω^b (")	ϕ^b (")	κ^b (")	ds^c (ppm)
NWL-9D -> WGS72	0	0	0	0	0	-0.26	-0.827
NWL-9D -> WGS84	0	0	+4.5	0	0	-0.814	-0.6
WGS72 -> WGS84	0	0	+4.5	0	0	-0.554	+0.227
BTS87 -> NWL-9D	+0.071	-0.509	-4.666	-0.0179	+0.0005	+0.8073	+0.583
BTS87 -> WGS84	+0.071	-0.509	-0.166	-0.0179	+0.0005	+0.0067	-0.017
BTS87 -> VLBI(NGS)	-0.089	+0.143	-0.016	+0.0043	-0.0093	+0.0033	+0.009
BTS87 -> SLR(GSFC)	0	0	0	+0.0018	-0.0062	+0.0075	0
WGS84 -> WGS84(GPS)	+0.026	-0.006	+0.093	+0.001	0	+0.002	-0.128

[a] origin offsets [b] rotations about the x-, y-, z-axes [c] scale difference

To change the models underlying satellite orbit computation in effect changes the datum. The same is true when an organisation computes its own ephemerides, as is done to support the highest precision geodetic applications by the International GPS Service for Geodynamics (§6.2 and §12.2). In the task of satellite positioning the correct transformation between reference coordinate systems is essential. The generation of a satellite orbit must be performed in a space-fixed reference frame, whereas the station positions are usually expressed in terms of an earth-fixed coordinate system. The commonly used reference coordinate systems for the space-fixed and earth-fixed are the referred to as the Conventional Celestial Reference System (CCRS) and the Conventional Terrestrial Reference System (CTRS), respectively. (Sometimes the word "frame" is substituted for "system" when referring to the *practical implementation* of the reference system -- no such distinction will be made here.)

Earth Motion in Space and Reference System Transformations

Before defining the CCRS and CTRS, the earth's motion in space needs to be briefly reviewed (further details can be found in, for example, SEEBER, 1993). Due to the torques exerted primarily by the gravitational attraction of the moon and sun (and secondarily by other planets), the equator and the ecliptic of the Celestial Sphere precess and nutate. **Precession**, with a period of approximately 26000 years, consists of two components: *luni-solar* and *planetary precession*. The luni-solar effect causes a slow westerly circular motion of the pole of the equator relative to the pole of the ecliptic. The attraction of the planets causes an eastward motion of the equinox by about 12" (arcseconds) per century and a decrease of the obliquity by about 47" per century. Precession causes the equinox to move along with the equator by about 50" per century. **Nutation** is a short-period, irregular motion of the pole with a period ranging from 14 days to 18.6 years, and has a maximum amplitude of about 20". Both are described by the motion of the equator and equinox with respect to the fixed equator and equinox of a given epoch, as shown in Figure 1.2-3.

In order to relate the CCRS to the CTRS, the parameters of polar motion and earth rotation are also needed. The motion of the true celestial pole with respect to the pole of a conventionally selected earth-fixed (terrestrial) reference frame is known as **Polar Motion**. The movement of the true pole is described by means of two parameters, x_p and y_p. Another parameter is the angle between the true equinox-of-date and the Greenwich meridian, known as the **Greenwich Apparent Sidereal Time**, denoted by θ_g (§2.1) The transformation from a space-fixed reference frame to an earth-fixed reference frame is accounted for by the parameter θ_g. The terrestrial reference system transformed from a celestial system at true-of-date is known as the Instantaneous Terrestrial Reference System (ITRS). Therefore, in order to link the true-of-date celestial system and the CTRS, the earth orientation between these two system coordinate axes is defined by the values of x_p, y_p, and θ_g, also known collectively as the **Earth Orientation Parameters**.

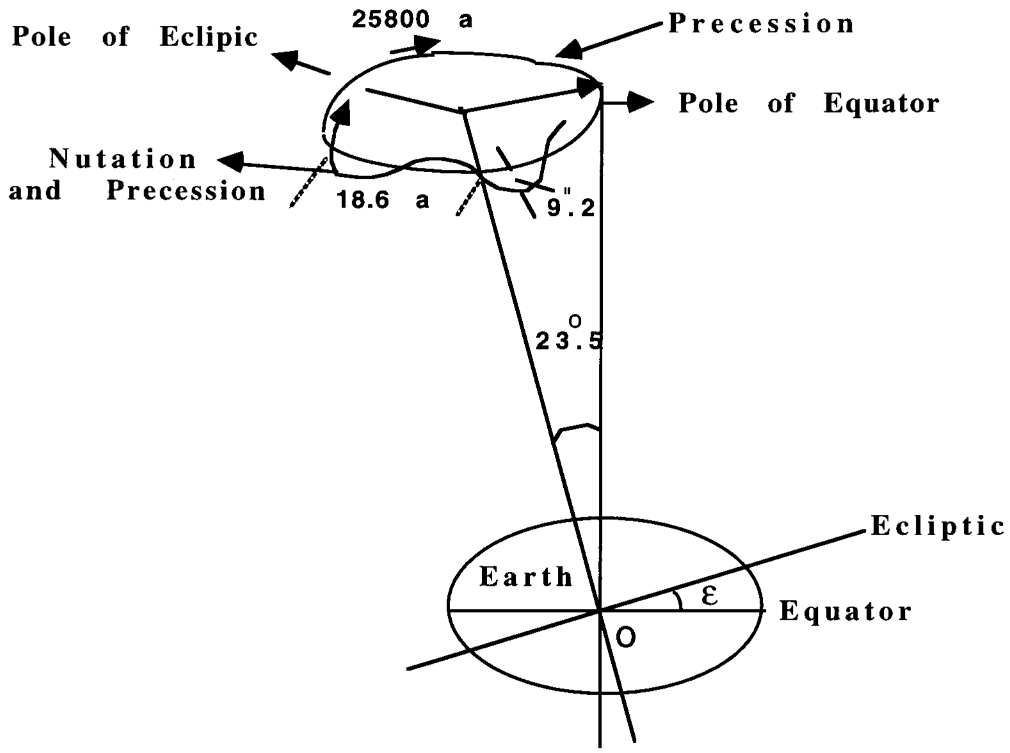


Figure 1.2-3. Precession and Nutation. *(Adapted from TORGE,1980)*

The CCRS is a right-handed coordinate system with its origin at the centre of mass of the earth. It is defined by the *mean equator and equinox of J2000.0.* (The standard epoch J2000.0 is 12hr on January 1, 2000.) The mean equator and equinox is the *fictitious* equator and equinox, derived after removing the effects of nutation. In other words, applying nutation to the mean equator and equinox of a given date yields the *instantaneous* or *true* equator and equinox of that date. If the satellite ephemeris (or station positions) are given in terms of the coordinate system which is defined by the true (or mean) equator and equinox of some reference date (not J2000.0), the reference system is known as the True-of-Date Reference System (or Mean-of-Date Reference System), or simply as the TDRS (or MDRS). It should be noted that the MDRS differs from the CCRS by the variations in time of the directions of the earth's spin axis. This variation is described by the effect of precession.

The CTRS is geocentric, right-handed, and orthogonal; the z-axis of this system is aligned with the mean pole of the earth; the x-axis is in the equatorial plane pointing towards the Greenwich meridian. For historical reasons the mean position of the true celestial pole during the period between 1900 and 1905 has been adopted as the mean pole of the earth, designated as the

Conventional International Origin (CIO) or Conventional Terrestrial Pole (CTP).

The transformation between the CCRS and CTRS is therefore accounted for by the four transformation components: precession, nutation, earth rotation, and polar motion.

(1) *Transformation from the CCRS to MDRS* (in terms of Precession) is defined by:

$$\mathbf{r}'(t_o) = \mathbf{P_r}.\mathbf{r} \qquad (1.2\text{-}6)$$

where

$$\mathbf{P_r} = \mathbf{R_3}(-z_A).\mathbf{R_2}(\theta_A).\mathbf{R_3}(-\zeta_A) \qquad (1.2\text{-}7)$$

and $\mathbf{r}$ is the position vector (x, y, z), expressed as CCRS coordinates,
$\mathbf{r}'(t_o)$ is the position vector (x', y', z'), expressed as MDRS coordinates with respect to reference epoch time t_o,
$\mathbf{R_1}(\varphi)$, $\mathbf{R_2}(\varphi)$, $\mathbf{R_3}(\varphi)$ are the rotation matrices for anticlockwise rotations through an angle φ about the x-, y-, and z-axis respectively, and
z_A, θ_A, ζ_A are the equatorial precession angles (SEEBER, 1993).

(2) *Transformation from the MDRS to TDRS* (in terms of Nutation) is defined by:

$$\mathbf{r}'' = \mathbf{N}.\mathbf{r}' \qquad (1.2\text{-}8)$$

where

$$\mathbf{N} = \mathbf{R_1}(-\varepsilon-\Delta\varepsilon).\mathbf{R_3}(-\Delta\psi).\mathbf{R_1}(\varepsilon) \qquad (1.2\text{-}9)$$

and $\mathbf{r}''$ is the position vector (x", y", z"), expressed as TDRS coordinates,
ε is the obliquity of the ecliptic, the angle between the equator and the ecliptic,
$\Delta\varepsilon$ is the nutation in obliquity, and
$\Delta\psi$ is the nutation in longitude.

(3) *Transformation from the TDRS to ITRS* (in terms of Earth Rotation) is given by:

$$\mathbf{r}''' = \mathbf{E}.\mathbf{r}'' \qquad (1.2\text{-}10)$$

where

$$\mathbf{E} = \mathbf{R_3}(\theta_g) \qquad (1.2\text{-}11)$$

and $\mathbf{r}'''$ is the position vector (x''' y''', z'''), expressed as ITRS coordinates.

(4) Transformation from the ITRS to CTRS (in terms of Polar Motion) is given by:

$$\mathbf{r_e} = \mathbf{P_m}.\mathbf{r'''} \tag{1.2-12}$$

where

$$\mathbf{P_m} = \mathbf{R_2}(-x_p).\mathbf{R_1}(-y_p) \tag{1.2-13}$$

and $\mathbf{r_e}$ is the position vector (x_e, y_e, z_e), expressed as CCRS coordinates.

The **complete transformation from the CCRS to the CTRS** therefore is:

$$\mathbf{r_e} = \mathbf{P_m}.\mathbf{E}.\mathbf{N}.\mathbf{P_r}.\mathbf{r} \tag{1.2-14}$$

The **reverse transformation from the CTRS to the CCRS** is given by:

$$\mathbf{r} = (\mathbf{P_m}.\mathbf{E}.\mathbf{N}.\mathbf{P_r})^T\ \mathbf{r_e} \tag{1.2-15}$$

The various component parameters required to evaluate the Polar Motion $\mathbf{P_m}$, Earth Rotation $\mathbf{E}$, Nutation $\mathbf{N}$, and Precession $\mathbf{P_r}$ matrices are described in detail in SEEBER (1993). The transformation process between the various coordinate systems is summarised in Figure 1.2-4.

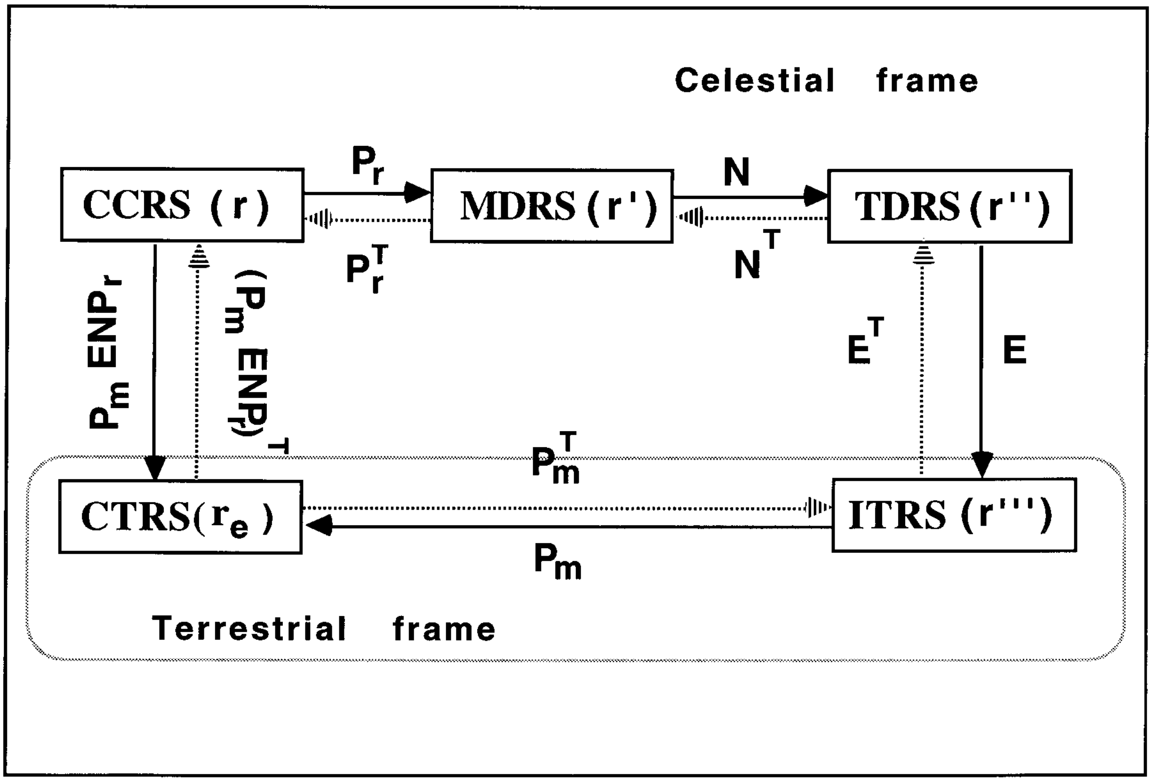


Figure 1.2-4. Coordinate system transformations.

Note, the ITRS is still not very useful for the most precise applications because stations on the surface of the earth move even with respect to this "earth-fixed" reference system due to geodynamic phenomena, the most important of which is **plate tectonic motion**. The ITRS as *defined at some instant in time* is the basis of the International Terrestrial Reference Frame (ITRF -- §2.1).

Satellite Ephemerides

A satellite ephemeris may take a number of forms (in order of decreasing convenience for a user):

- a **list of coordinates** as a function of time,
- a **polynomial representation of the trajectory** in a suitable reference system (for example alongtrack, crosstrack, or radial components), or
- satellite **position and velocity at some reference time**, requiring these parameters to be derived for subsequent times through a solution of the Equations of Motion.

Satellite ephemerides must be computed, and fundamental to this task are the issues of reference systems (see earlier sections) and the forces acting on a satellite in orbit. After a satellite has separated from its carrier rocket it begins orbiting about the earth. The satellite's orbit is determined by its initial position and velocity, and the force fields which are in effect (SEEBER, 1993). In the case of the gravitational field of a spherically symmetric body (a reasonable approximation of the earth to the level of about 1 part in 10^3) this produces an elliptical orbit which is fixed in space -- the **Keplerian ellipse** (Figure 1.2-5).

Due to the effects of *other gravitational and non-gravitational forces* which perturb the orbit, the actual trajectory of the satellite departs from the ideal Keplerian ellipse. In general, the forces that influence satellite motion are (Figure 1.2-6):

- the non-spherical gravitational attraction of the earth,
- the gravitational attractions of the sun, moon, and planets (the so-called "third body" effects),
- atmospheric drag effects,
- solar radiation pressure (both direct and albedo components), and
- the variable part of the earth's gravitational field arising from the solid earth and ocean tides.

In order to determine the motion of a satellite to a high precision these perturbing forces must be accurately modelled. If these forces were known perfectly, and the initial position and velocity of the satellite were given, then the integration of the Equations of Motion would give the satellite's position and velocity at any time in the future:

$$\ddot{\mathbf{r}} = -\frac{GM_e\,\mathbf{r}}{|\,\mathbf{r}\,|^3} + \ddot{\mathbf{r}}_t \qquad (1.2\text{-}16)$$

where $\ddot{\mathbf{r}}$ is the acceleration vector, $\mathbf{r}$ is the satellite position vector, GM_e is the product of the gravitational constant and the mass of the earth, and $\ddot{\mathbf{r}}_t$ are the total perturbing accelerations acting on the satellite, all expressed in the CCRS system.

However, *the models for the perturbing forces are not error-free.* Hence an "orbit computation" is performed in which satellite observations obtained at tracking sites of known position are analysed in order to produce an orbit (by adjusting the initial parameters of the orbit, possibly together with several additional force model parameters) that is a "best fit" to the available observations. *Determining this orbit is the task of the satellite geodesist.*

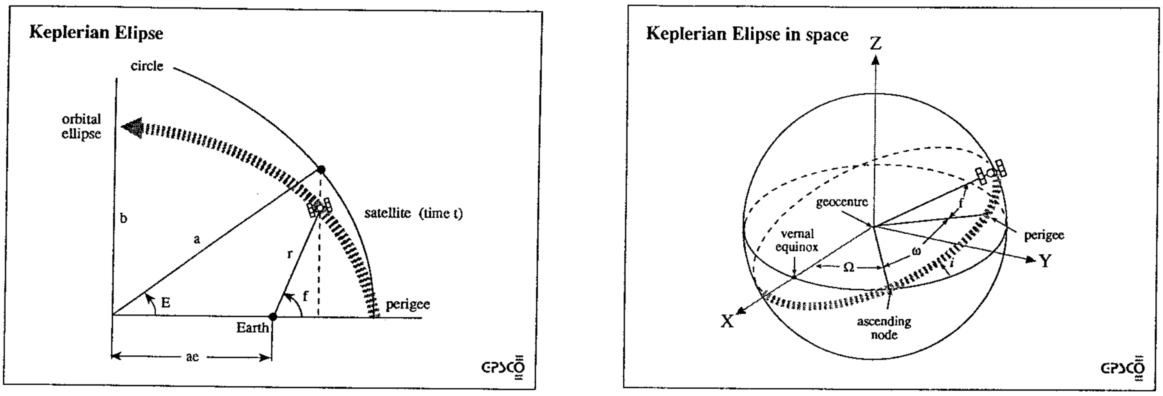


Figure 1.2-5. The Keplerian Ellipse and Keplerian Orbital Elements.

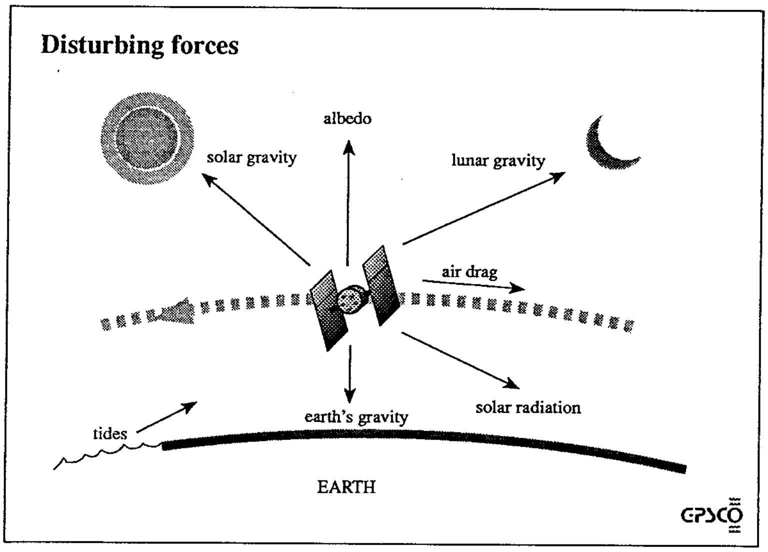


Figure 1.2-6. Perturbing forces acting on a near-earth satellite.

The final orbit may need to be transformed so that the "orbiting control stations" can be used as "targets" by the average user. In particular the "natural" reference system for satellite computations is a non-rotating (CCRS) system, but the reference system required by users is an earth-fixed one (the CTRS). *The transformations were described in an earlier section.*

1.2.2 MEASUREMENT TECHNOLOGY

Only those technologies based on microwave transmissions will be considered because:

- They can penetrate clouds and rain, and can operate day or night.
- They can be most readily implemented within the satellites themselves.
- They are similar to readily understood conventional (terrestrial-based) survey and navigation technologies, such as range and range-difference systems.

It is necessary to further distinguish between one-way and two-way ranging, and then to briefly introduce the measurement modes based on *ranges* and *range-differences.*

One-Way and Two-Way Ranging

Ranging by means of microwave signals can be done in either of two modes (Figure 1.2-7). **Two-way ranging** involves the measurement of the travel time of a signal by one clock. At one end of the line a device (for example, a glass prism, or microwave transponder) reflects the incoming signal back to the transmitter. *The basic measurement is the travel time for the round trip distance.* This is the procedure employed in several EDM and radio-navigation systems. Its major shortcoming however is that it is best suited for a single user system (the reflector-transmitter coupling required for a single measurement tends to be inflexible in a multi-user environment).

One-way ranging involves the measurement of the travel time of a signal from transmitter to receiver through the use of separate clocks. The transmitter clock generates the signal, while the receiver clock detects when the signal arrives. *The difference in transmit and receive time is the travel or "transit" time, hence both clocks must keep the same time.* (An error in synchronisation of the two clocks of 1 nanosecond is equivalent to 30cm in distance.) In general, each clock keeps its own time and the relationship between the clocks may have to be established from the measurements themselves. The main advantage of such a system is that it is "multi-user", each user being a passive "listener".

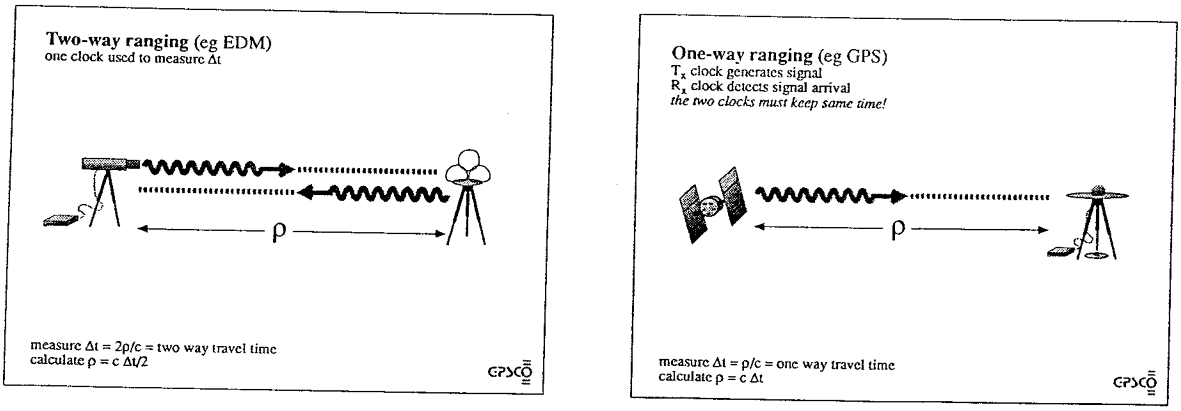


Figure 1.2-7. Two-Way and One-Way Ranging.

Range and Range-Difference Measurements

A range, or distance, measurement can be made to very high precision because it essentially involves the measurement of time delay, or signal transit time. Modern clocks are based on frequency oscillators (§1.3), perhaps based on a quartz crystal, or an atomic "clock" of some sort. Although each clock exhibits different stability characteristics (some are better for measuring long time intervals, others for short intervals), their overall stability is of the order of 1 part in 10^{10} to 10^{12} (Table 1.3-1). The distance is found by multiplying the time delay by the signal's speed of propagation (=299792458m/s in a vacuum). A time delay interval measured to this precision with a single clock translates to a distance precision of a centimetre or better. The use of two clocks, as in two-way ranging, complicates matters for a number of reasons:

- **The time delay measurement is affected not only by the quality of the clocks, but also on how well they are "synchronised".** This itself is influenced by such factors as:
  - how long since the last synchronisation (as the drift of the clocks since synchronisation is the issue, not the short term stability of the clocks for the time delay measurement),
  - how well the synchronisation is carried out in the first place (seldom are the clocks able to be physically brought together and directly compared), and
  - generally this is carried out by comparison with a third time scale, accessible to all clocks (including those in the orbiting satellites), and hence the quality of the definition and maintenance of this "master" time scale is an important issue as well.

- **The "time-of-transmission" information must be made available to the second clock**, and hence the resolution to which this information is available must be adequate (for example, if the digital time signal is defined by a frequency of 10MHz, then the resolution of time is 0.1 microsecond, or the equivalent of 30 metres in range!).

The former can be partly overcome by appropriate modelling of the **biased** range (§1.3). The latter is important for instantaneous range measurements on GPS, but not for range measurements derived from carrier phase observations.

An alternative solution to both of these problems is to create range-differences. This is useful for measurements that contain timing (and other) errors that are linearly correlated, for example if the same bias affects both measurements, then differencing the two will eliminate that common bias. In the case of a single satellite-single station configuration, this approach aims to reduce the measurement to its simplest form, the measurement of a short time delay, so that rather than measuring a one-way range (with all the accompanying problems of synchronisation, etc.) between satellite and station, the change in transit time (or range) is what is important. Hence the exact time-of-transmission is not required, as the range-difference "measurement" is dependent only on the short term stability of both the station and satellite clock.

These "between-epoch" range-differences are the basis of the TRANSIT Doppler satellite measurement. These range-differences (Figure 1.2-8) are inferred from the Doppler frequency shift. As a result of the TRANSIT satellite's movement, the frequency of the satellite signal changes continually. These frequency shifts, when integrated over a given time interval, are functionally related to the change in radial distance (or slant range, as it is sometimes known). The position of the receiver can be determined from these range-differences, as measured using one TRANSIT satellite.

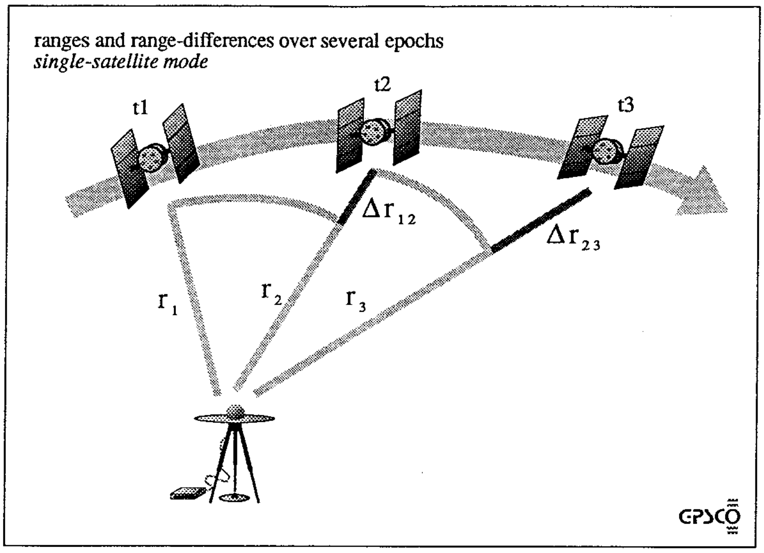


Figure 1.2-8. The definition of range-difference.

There are a number of other types of range-differences that can be formed in the multi-satellite, multi-station scenarios used in GPS positioning (see §6.3).

1.2.3 POSITIONING STRATEGIES

In considering appropriate positioning strategies for the measurements described above it is necessary to make some assumptions, primarily:

- That the satellite coordinates are available.
- That the measurements are not "biased" in any way.

Positioning by Ranging to Satellites

The basic concept of positioning using range data is the same whether it is involves terrestrial distances or satellite measurements. In three dimensions, a measured range to a known point constrains the position in 3-D space to lie on the surface of a sphere centred at the known point. This is the **"surface of position"**. The intersection of three such surfaces describes a unique point in space (Figure 1.2-9). Hence, three ranges are required, to three separated known points, in order to fix position (§1.4).

In 2-D positioning applications, for example as in the case of horizontal geodetic networks or when navigating at sea (position is assumed to be on the surface of the ellipsoid, or a known height above it), the intersection of the surface of position and the ellipsoid is the **"line of position"**, and is approximately a circle (see §1.4). The intersection of two distances defines the location of the point from which the distance measurements were made. In surveying parlance, this is known simply as <u>distance intersection</u>.

If the point is stationary, the two (or three) distances do not need to be measured simultaneously. If the point is moving however, all distances must be measured simultaneously, or over an interval of time during which the point has not moved by an amount greater than the uncertainty of the "fix". Satellite positioning using ranges is the basis of the GPS system for most navigation applications.

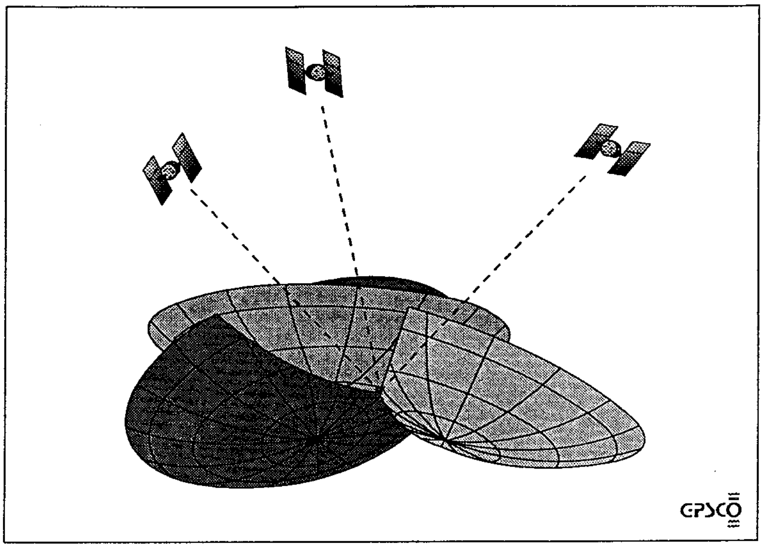


Figure 1.2-9. Intersection of "surfaces of position" based on range observations.

Positioning by Range-Differencing to Satellites

As with ranging, the concept of positioning using range-difference (or "range-rate") data is the same whether based on terrestrial or satellite measurements. Differences in range measured to two known points constrains the position in 3-D to lie on a surface of position which is one half of a hyperboloid of revolution of two sheets. In 2-D the line of position (formed by the intersection of the hyperboloid and the ellipsoid) is approximately one half of a hyperbola (Figure 1.2-10).

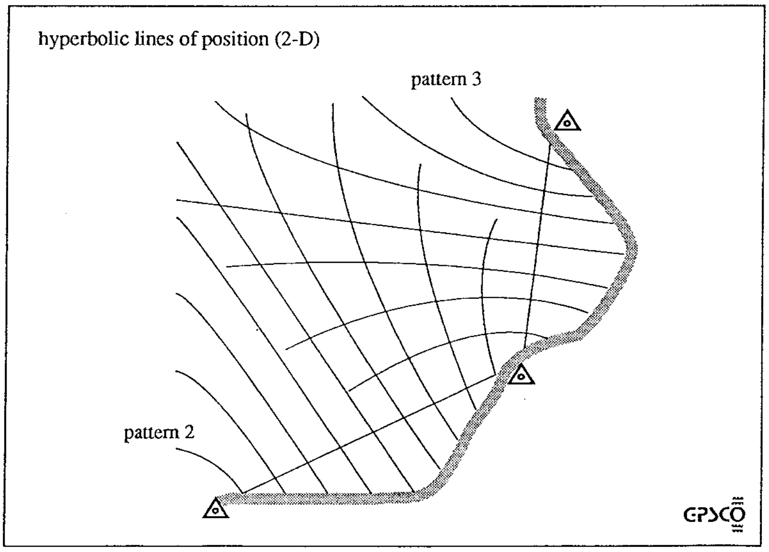


Figure 1.2-10. Hyperbolae and hyperbolic intersection.

In the 3-D case, the intersection of three hyperboloids defines the position of the observer. Each range-difference involves two known points, and since one of the points can be common to other pairs of stations, a minimum of four known points are required for positioning based on range-rate. In the 2-D case, the intersection of two hyperbolas is sufficient for position determination, involving a minimum of three known points. This is the basis of radio-navigation systems such as LORAN-C, operating in the range-difference mode.

The TRANSIT Doppler system operates on a similar principle (see, for example, SEEBER, 1993, for a good overview), except that the two known points are in fact generated by the one moving satellite. The range-difference therefore involves the same satellite (Figure 1.2-8).

As in the case of positioning by range, if the point being positioned is not moving, the measurements need not be made simultaneously. However, if the point is moving, simultaneous measurements are necessary unless the magnitude of the motion over the measurement interval is insignificant in relation to the system errors (Figure 1.2-11).

GPS can also be used in this "Doppler" mode, but this is rarely done, either for navigation or surveying applications.

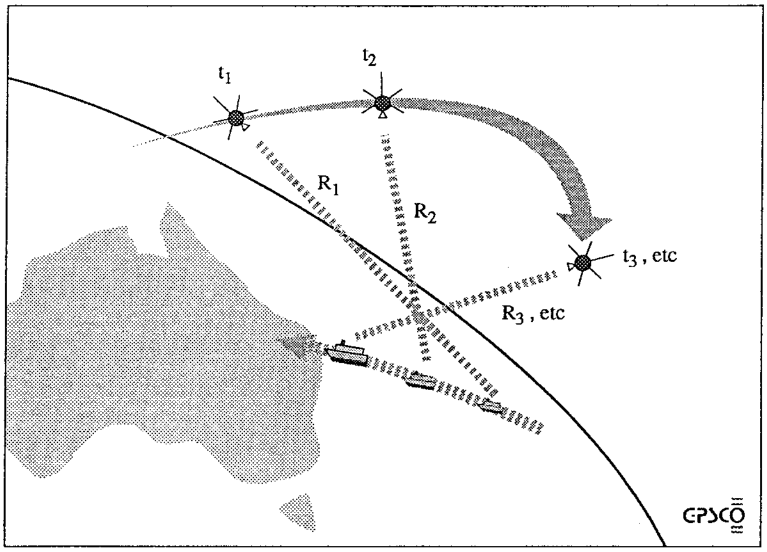


Figure 1.2-9. Range-differences to a single satellite for a moving antenna.

1.2.4 OVERVIEW: Essential Characteristics

From a study of the design requirements of an ideal "Global Positioning System", and as a consequence of the feasible technological features of such a system, the essential characteristics of the system can now be identified.

(1) System Configuration:

- A multi-satellite system at high altitude, but not in a geostationary orbit. *The number of satellites to be visible to a user is dependent upon the observation type to be employed and the positioning strategy adopted.*
- A Control Segment responsible for tracking the satellites, and computing the ephemerides.
- Satellites broadcast their ephemerides to all users.

(2) Satellite Technology:

- The system should concentrate as much complexity into the satellites as possible.
- The system should be passive (one-way) as far as the user is concerned, with the satellites transmitting the signals necessary to support position determination at the user station. *No receiving function is to be performed by the satellites.*

(3) Measurement Technology:

- A one-way ranging system based on microwave transmissions would satisfy the requirements for a listen-only, high precision, simple-to-use positioning system.
- To make such a system work, separate clocks must be used (in the satellites and within the user equipment), and they must be synchronised in some way.
- The satellites should somehow broadcast time-of-transmission information to the user.

What remains to be established is:

☞ the positioning principle to be used (related to the measurement technology and the characteristics of the satellite constellation), and

☞ the residual errors remaining in the system (after application of the best available technology), and the development of strategies to overcome any unacceptably high error sources.

1.3

BIASED RANGES

An important issue is "what is a bias"? The variety of possible GPS positioning modes are all essentially different strategies for accounting for "biases" in GPS measurements. In this text **range biases** are defined as those influences, both instrumental (or "internal") and system origin (or "external"), on the observations which cause the measured distance to be different from the true distance by a *systematic amount.* (In §2.4 and §6.2 the distinction between a "measurement bias" and a "measurement error" will be made clearer.)

There are different types of biases, but the focus shall be those biases that have the following well defined characteristics:

(1) Range biases that affect all measurements taken at a ground station by a similar amount --> *the **station dependent biases**.*

(2) Range biases that affect all measurements made to a particular satellite by a similar amount --> *the **satellite dependent biases**.*

(3) Range biases that are unique to a particular receiver-satellite observation --> *the **observation dependent biases**.*

In the case of GPS measurements, the sources of the biases can be basically partitioned into the above three categories (§6.2). The dominant biases are those due to the station and satellite clocks, and the geometric nature of the high precision measurement used in the GPS survey reductions. They have the following additional features:

- the station and satellite clock errors affect the respective measurements by the **same amount**.
- the observation dependent ambiguity in carrier phase measurements is assumed to be **a constant** for an extended period of uninterrupted tracking.

Before discussing biased ranges further it is necessary to introduce a more precise definition for a "clock" and for the "clock error".

1.3.1 CLOCK BEHAVIOUR

Frequency and Time Stability

A time scale is defined by the period of the basic oscillation of the frequency-determining element (be it the earth's rotation in the case of Sidereal Time, or the oscillation of atoms in the case of atomic time, or of a crystal in the case of quartz clocks) which is measured, and the origin of the time scale, which may be either arbitrarily defined or agreed upon by international convention. Individual clocks maintain their own time scale, however, for one-way (passive) ranging BOTH the ground and satellite clocks need to be synchronised.

How well must this synchronisation be made?

1 nanosecond (10^{-9} sec) $\approx$ 30 centimetres in distance!

To indicate the "absolute" magnitude of clock "error" it is necessary to introduce the notion of "perfect" or "true" time. Hence it is possible to "measure" clock error as an instantaneous "offset" from this perfect time scale. A time-varying clock error $\Delta t(t_i)$ can have the following form:

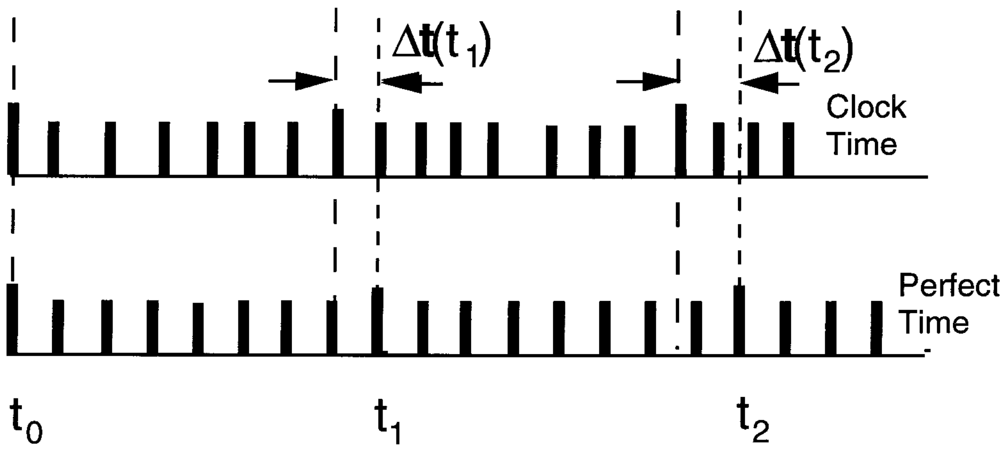


Figure 1.3-1. Time-varying clock error measured against a time scale standard.

Today's high precision clocks are all based on some form of <u>frequency standard</u> or <u>oscillator</u>. In the context of precise ranging systems these belong to one of two classes:

(1) The so-called "atomic" clocks, such as the cesium beam tube, rubidium vapour cell or hydrogen maser oscillators.

(2) The various types of quartz crystal oscillators.

Time intervals are most precisely defined by the cycle counter of a frequency standard. (For example, the second is now defined as 9192631770 cycles of the fundamental resonance of the cesium atom.) Hence it suffices to establish the relationship between frequency and the phase output of an oscillator, and their errors, as the time scale can be directly obtained from such a relationship. The reading on a frequency cycle counter i can be represented by:

$$\Phi(t_i) - \Phi(t_{oi}) = f_i \cdot (t_i - t_{oi}) \qquad (1.3\text{-}1)$$

The wavelength of the phase cycle is (c / f_i), where c is the velocity of electromagnetic radiation (299792458m/s in a vacuum). Substituting the appropriate scaling for the phase change to convert it into a time interval, and defining the origin of the time scale at the arbitrary reference epoch of the cycle counter, the following expression for a "clock reading" is obtained:

$$t_i(t) - t_{oi} = \frac{1}{f_0} \int_{t_o}^{t} f_i(t)\, dt \qquad (1.3\text{-}2)$$

where:

t_o	is the reference epoch,
t_{oi}	is the clock reading at the reference epoch,
$f_i(t)$	is the frequency of the oscillator, and
f_0	is the nominal oscillator frequency.

A standard model for the frequency of an oscillator is:

$$f_i\,(t) = f_0 + \Delta f + \dot{f}\,(t - t_o) + f_r\,(t) \tag{1.3-3}$$

where:

Δf	is the frequency bias,
$\dot{f}$	is the frequency drift, and
$f_r(t)$	are unmodelled random frequency errors.

Substituting eqn (1.3-3) into eqn (1.3-2) gives:

$$t_i\,(t) = t_{oi} + (t - t_o) + \frac{\Delta f}{f_0}(t - t_o) + \frac{\dot{f}}{2f_0}(t - t_o)^2 + \frac{1}{f_0}\int_{t_o}^{t} f_r(t)\;dt \tag{1.3-4}$$

Rearranging terms into a representation of the error of the clock i as a time polynomial:

$$\begin{aligned} \varepsilon_i\,(t) &= t_i - t \\ &= a_o + a_1\,(t - t_o) + \frac{a_2}{2}(t - t_o)^2 + \int_{t_o}^{t} y(t)\;dt \end{aligned} \tag{1.3-5}$$

where:

a_o	is the clock bias term,
a_1	is the clock drift term,
a_2	is the clock drift-rate, and
$\int_{t_o}^{t} y(t)dt$	is the integrated random fractional frequency error.

Expressed as "phase" error this is:

$$\Phi_\varepsilon(t) = f_0\,\varepsilon_i\,(t) \tag{1.3-6}$$

Clock error, whether expressed in terms of phase (eqn (1.3-6)), time (eqn (1.3-5)) or frequency (eqn (1.3-3)) instability, consists essentially of two distinct components:

(1) The systematic part which can be determined (and predicted). This is the explicit polynomial part of eqns (1.3-3) and (1.3-5).

(2) The <u>random part</u>, which may be significant enough that it cannot be ignored.

Therefore, in addition to exhibiting **deterministic deviations** from a "true" time scale, they also undergo **stochastic variations** in both time (or phase) and frequency. An example of a realisation of the time difference between a commercial cesium clock and a time scale generated from an ensemble of atomic clocks is shown in Figure 1.3-2.

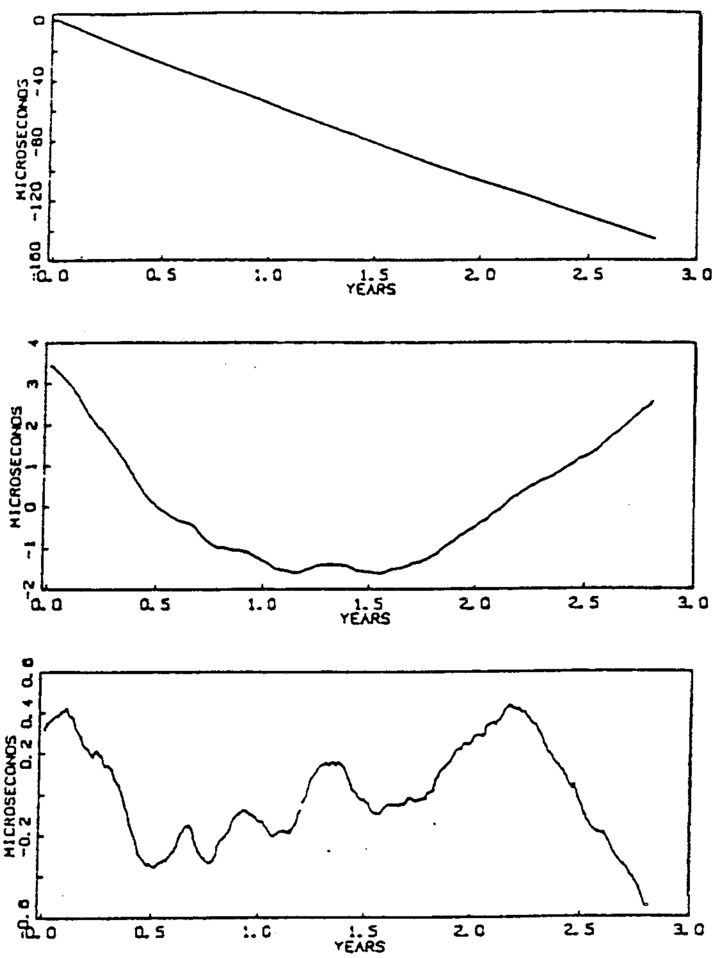


Figure 1.3-2. Time difference between a commercial atomic clock and the U.S. Naval Observatory Time Scale. *(JONES & TRYON, 1987)*
(The middle graph has the linear trend removed, and the bottom graph has a quadratic component removed.)

The top graph shows how the clock has a frequency bias so that the time scale appears to drift linearly away from the "true" time (defined here by the group of atomic clocks). This linear drift (approximately 50 msec/year) does not affect the clock's ability to keep accurate time as long as the rate of the drift (coefficient a_1) is known or can be estimated very well. The middle graph shows the residuals after fitting a straight line to the top graph. The variation is now of the order of 5 msec/year, but the quadratic appearance indicates a higher order effect still present, in this case a significant frequency drift in the clock (the coefficient a_2 representing "ageing"). *That is, the frequency of the clock appears to change linearly with time.* The bottom graph shows the residuals after removing a quadratic function. These are now primarily stochastic variations (a higher order polynomial could be used to remove residual "systematic" trends, but the "stochastic" variations are assumed here to be those that remain after second order polynomial modelling).

The component of eqn (1.3-3) attributable to random error sources is known as the random fractional frequency deviation, or the integrated random fractional frequency error (eqn (1.3-5)). The total fractional frequency deviation (systematic + random), or just the random part, can be analysed. The standard approach is to deal with the sample variance of only the random fractional frequency fluctuations, and model the systematic part by an explicit polynomial-like function. As the frequency count or time difference over some elapsed time interval τ can be measured it is possible to define the mean value of the fractional frequency deviation:

$$y_k = \frac{1}{\tau} \int_{t_k}^{t_k+\tau} y(t)\, dt = \frac{[\Phi(t_k+\tau) - \Phi(t_k)]}{2\pi f_0 \tau} \tag{1.3-7}$$

where $t_{k+1} = t_k + T$, $k=0,1,2,...T$, is the repetition interval for measurements of duration τ, and t is arbitrary.

Now forming the sample variance of $y(t)$:

$$< \sigma_y^2(N,T,\tau) > = < \frac{1}{N-1} \sum_{n=1}^{N} [y_n - \frac{1}{N}\sum_{k=1}^{N} y_k]^2 > \tag{1.3-8}$$

where $< g >$ denotes the infinite time average of "g".

A particular variance measure is chosen so that $N=2$, $T=\tau$. This is the so-called **Allan Variance** (HELLWIG, 1979):

$$\sigma_y^2(\tau) = E\{\sigma_y^2(N=2,T=\tau,\tau)\} = E\{\frac{(y_{k+1} - y_k)^2}{2}\} \tag{1.3-9}$$

This two-sample variance of the fractional frequency error is the standard measure of clock stability. One of its special advantages is its relative simplicity: it is only a function of τ and can be plotted in the form of stability graphs such as those in Figure 1.3-3 (taken from IBID, 1979). The units for $\sigma_y(t)$ are dimensionless. *Note that in Figure 1.3-3 the linear drifts of the quartz crystal and rubidium oscillator of 1 part in* 10^{10} *and 1 part in* 10^{12} *have been removed.*

How can frequency stability plots such as these be interpreted? The clock stability is defined as a function of the time interval between the monitoring of a particular clock. If it is assumed that at the start of the interval the clock is synchronised with (or has been compared to) a "true" time scale, the amount by which the clock has "deviated" (on average) after a certain time interval τ is given by the square-root of the Allan Variance $\sigma_y(\tau)$ times τ:

$$\sigma_x(\tau) = \tau \,.\, \sigma_y(\tau) \tag{1.3-10}$$

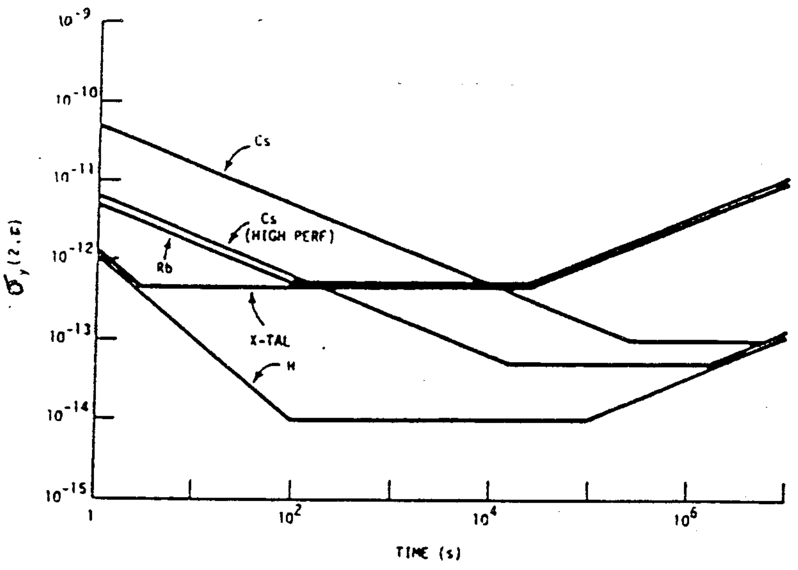


Figure 1.3-3. Square-root of the Allan Variance of typical oscillators (after removal of linear trend for crystal and rubidium oscillators). *(HELLWIG, 1979)*

For example, quartz crystal oscillators are as precise as hydrogen masers if the time intervals are less than approximately 5 seconds. In the short term, up to 10^4 seconds, cesium standards fare badly in comparison with other frequency standards. However, their medium to long term performance is superior to all but the hydrogen maser, which it rivals after approximately 10^6 seconds. The behaviour of the oscillators in Figure 1.3-3 can therefore be characterised by three regimes:

(1) **short-period**, where the Allan Variance decreases as the time interval τ increases, according to the relation:

$$\sigma_y(\tau) = K_1.\tau^{\alpha_1} \tag{1.3-11}$$

where α_1 is positive (=1 for hydrogen maser or crystal oscillator clocks, =0.5 for cesium and rubidium clocks).

(2) **medium-period**, where the Allan Variance is a constant:

$$\sigma_y(\tau) = \sigma_{yF} = \text{constant} \tag{1.3-12}$$

(3) **long-period**, where the Allan Variance increases with increasing time interval τ according to the relation:

$$\sigma_y(\tau) = K_2.\tau^{-0.5} \tag{1.3-13}$$

Representative values of K_1 , σ_{yF} and K_2 are given in Table 1.3-1.

Table 1.3-1. Typical performance data for commercially available oscillators. *(Adapted from HELLWIG, 1979)*

	K_1	σ_{yF}	K_2	Drift (sec/sec)
H (active)	$1x10^{-12}s$	$1x10^{-14}$	$3x10^{-17}\sqrt{s^{-1}}$	10^{-15}
Cs	$5x10^{-11}\sqrt{s}$	$1x10^{-13}$	$3x10^{-17}\sqrt{s^{-1}}$	10^{-15}
	$7x10^{-12}\sqrt{s}$	$5x10^{-14}$	$3x10^{-17})\sqrt{s^{-1}}$	$10^{-15} - 10^{-14}$
Rb	$5x10^{-12}\sqrt{s}$	$5x10^{-13}$	$3x10^{-15}\sqrt{s^{-1}}$	10^{-12}
X-tal	$1x10^{-12}s$	$5x10^{-13}$	$3x10^{-15}\sqrt{s^{-1}}$	10^{-10}

The Allan Variance is a measure of the performance of a clock. The Allan Variance is, however, not based on any physical model of the oscillator, but rather the stability graphs are produced from the results of testing actual oscillators. (For the long-period portion of the graphs, long measurement time periods are required, and hence the results are not as reliable as for the short- and medium-period portions.) Nevertheless it can be used to predict the behaviour of clocks in a positioning system such as GPS over time spans ranging from fractions of a second to several hours, and hence the likely build up of clock error (phase, time or range equivalent). This is important for various aspects of GPS observable modelling (§6.1).

1.3.2 BIASED RANGES

Receiver-Biased Ranges

The satellite clock scale and the receiver clock scale are not synchronised at the instant of measurement (Figure 1.3-4). It is assumed that the satellite-receiver range is affected only by a clock error ($d\tau$) caused by the receiver oscillator (for this discussion satellite time is taken to be "true" time). The relationship between the measured range ρ^* and the true range ρ is:

$$\rho^*(t) = \rho(t) + d\tau(t_r).c \tag{1.3-14}$$

where c is the speed of electromagnetic radiation. Note the time dependence of the range and clock error, and that they are tagged with the time-of-reception. This implies that if the receiver clock is *slow* (positive value of $d\tau$), then the measured range is too *long*; and if the receiver clock is *fast*, then the measured range is too *short.*

If measurements are made simultaneously to several satellites, although the time of transmission of the signals is different for each satellite (and hence the flight time), they will be biased by the same amount.

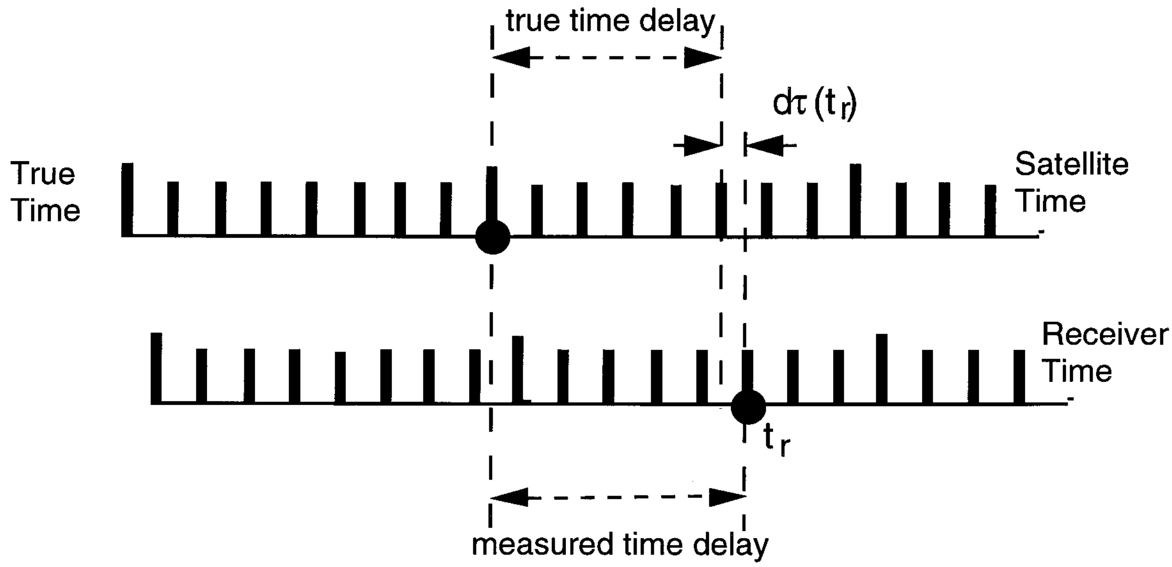


Figure 1.3-4. Receiver clock error in one-way ranging.

A series of observation equations can be constructed:

$$
\begin{aligned}
\rho^{*s1} &= \rho^{s1} + d\tau.c \\
\rho^{*s2} &= \rho^{s2} + d\tau.c \\
\rho^{*s3} &= \rho^{s3} + d\tau.c \\
\rho^{*s4} &= \rho^{s4} + d\tau.c \\
&\cdots\cdots\cdots \\
\rho^{*si} &= \rho^{si} + d\tau.c
\end{aligned}
\tag{1.3-15}
$$

where si refers to the i'th satellite. Figure 1.3-5 illustrates the situation.

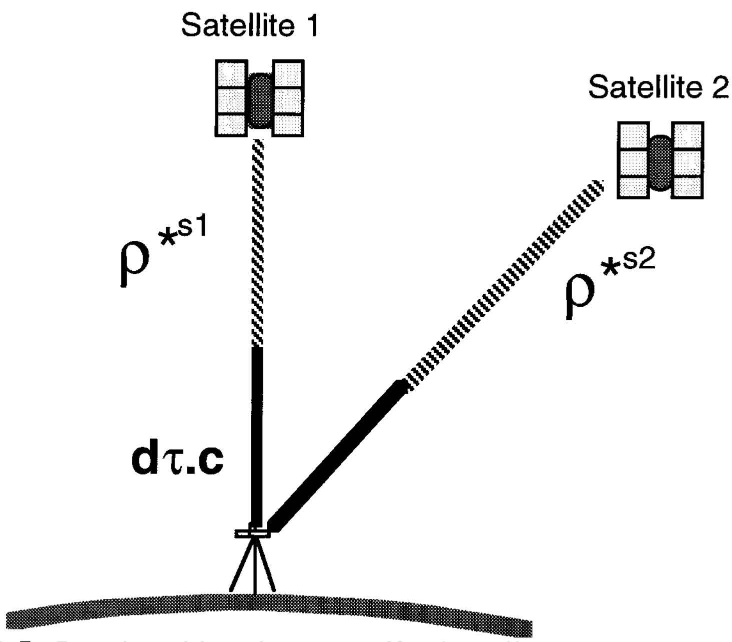


Figure 1.3-5. Receiver-biased ranges affecting all measurements at a receiver.

Satellite-Biased Ranges

The satellite clock scale and the receiver clock scale are not synchronised at the instant of signal transmission (Figure 1.3-6).

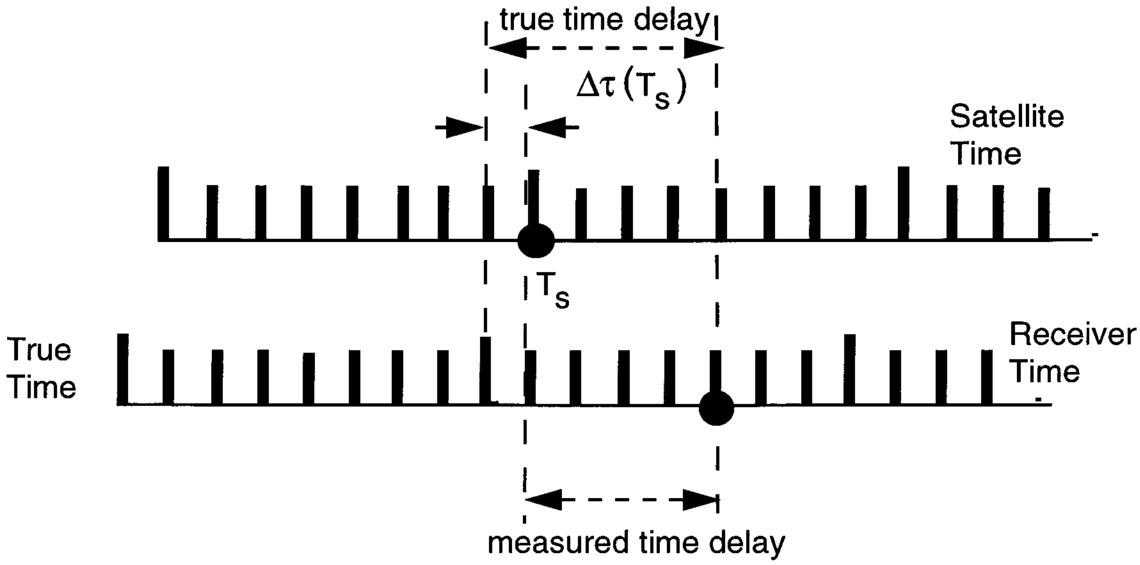


Figure 1.3-6. Satellite clock error in one-way ranging.

Assuming that the satellite-receiver range is affected by only a clock error ($\Delta\tau$) associated with the satellite oscillator (for this discussion receiver time is assumed to be "true" time). The relationship between the measured range ρ^* and the true range ρ is:

$$\rho^*(t) = \rho(t) - \Delta\tau(T_s).c \qquad (1.3\text{-}16)$$

Note the time dependence of the range and clock error, and that the measurements are tagged with the time-of-reception, but the clock error is tagged with the time-of-transmission. This implies that if the satellite clock is *slow* (positive value of $\Delta\tau$), then the measured range is too *short*; and if the satellite clock is *fast*, then the measured range is too *long*.

If measurements are made simultaneously at several ground receivers the time-of-transmission of the signals must have been different, and the measurements will be biased by slightly different amounts:

$$\begin{aligned}
\rho^*_{r1} &= \rho_{r1} - \Delta\tau(T_{r1}).c \qquad (1.3\text{-}17)\\
\rho^*_{r2} &= \rho_{r2} - \Delta\tau(T_{r2}).c\\
\rho^*_{r3} &= \rho_{r3} - \Delta\tau(T_{r3}).c\\
\rho^*_{r4} &= \rho_{r4} - \Delta\tau(T_{r4}).c\\
&\cdots\cdots\cdots\\
\rho^*_{rj} &= \rho_{rj} - \Delta\tau(T_{rj}).c
\end{aligned}$$

where rj refers to the j'th receiver. Figure 1.3-7 illustrates this situation.

How different are the values of $\Delta\tau(T_{rj})$? For receiver separations of the order of 10^3km the maximum difference in arrival time from a satellite at the horizon collinear with the interstation vector is of the order of 0.003 seconds. If the satellite oscillator is a cesium standard, then from Table 1.3-1 it can be seen that $\Delta\tau(T_{rj})$ may vary by up to $5x10^{-11}\sqrt{0.003} \approx 3x10^{-12}$, or approximately one millimetre in range equivalent! Clearly this is negligible, and hence *it can be assumed that the satellite clock error is the same, even widely separated ground receivers.*

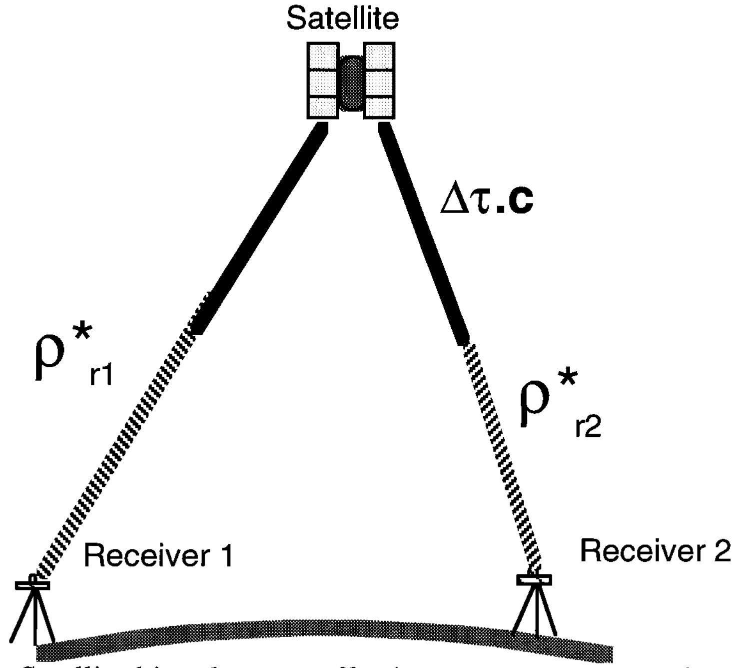


Figure 1.3-7. Satellite-biased ranges affecting measurements made at two receivers.

Ambiguous Ranges

Assume that each measurement made by a receiver to a satellite is ambiguous because only a fraction of a wavelength (in other words, a fraction of the time scale resolution) can be measured (Figure 1.3-8).

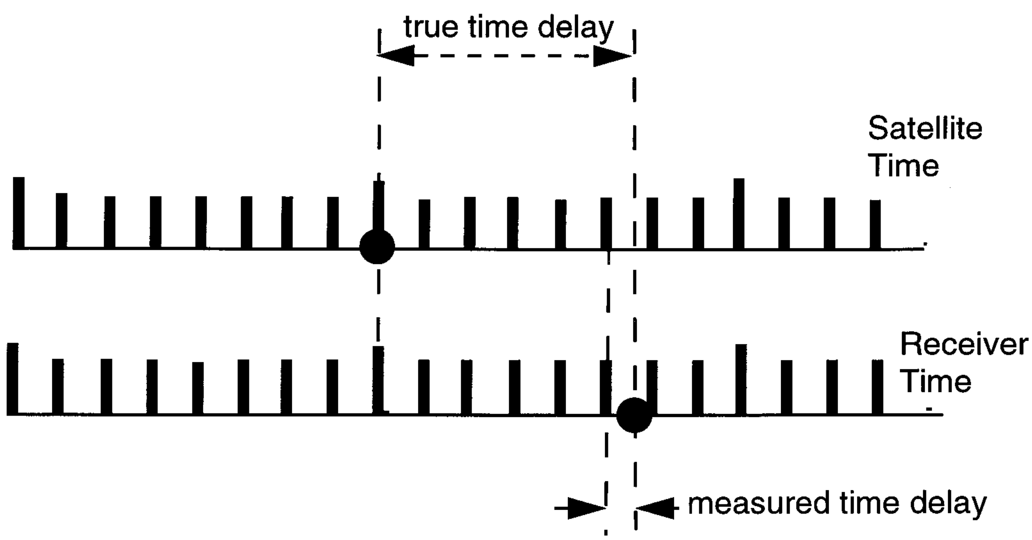


Figure 1.3-8. An ambiguous time delay measurement.

The measurement would be modelled as:

$$\rho^*(t) = \rho(t) - n(T_{rj}^{si}) \tag{1.3-18}$$

where $n(T_{rj}^{si})$ is dependent on the receiver, the satellite and the time. *This is not a very useful measurement!* If it were assumed that the ambiguity n is a *constant over time*, the measurement $\rho^*(t)$ in effect contains the *change in distance* since some initial epoch t_0:

$$\rho^*(t) = \rho(t) - n_{rj}^{si} \tag{1.3-19}$$

Figure 1.3-9 illustrates this situation.

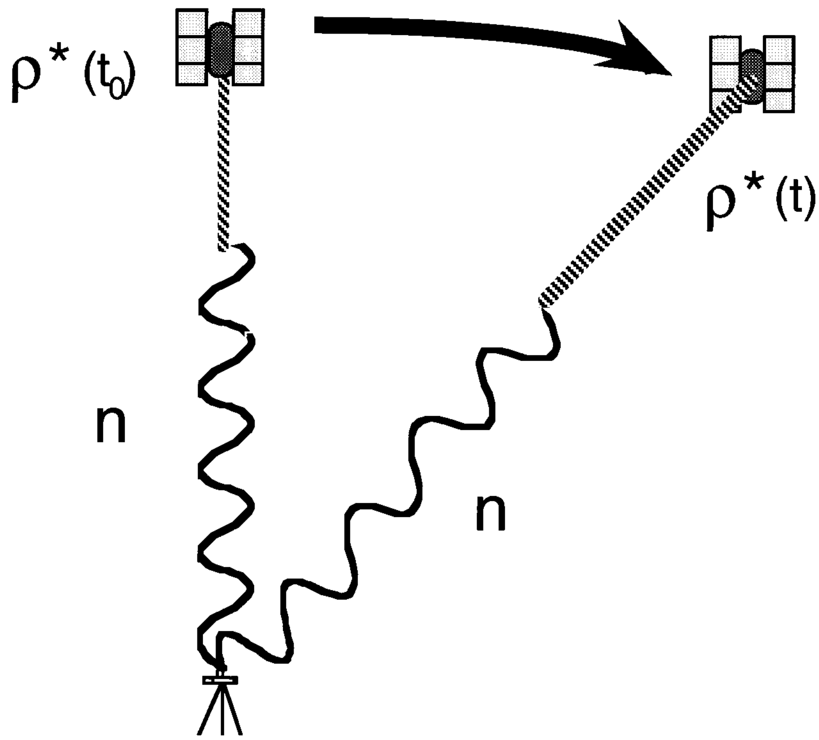


Figure 1.3-9. Ambiguous ranges where the ambiguity is constant with time.

1.4
POSITIONING STRATEGIES

Positioning strategies are first discussed in the context of the measurement technology and the satellite configuration. The geometrical principles of "lines of positions" and "surfaces of positions" are then introduced with particular attention given to the effect of biased ranges, as is the circumstance for GPS positioning. A generalised algebraic description of positioning strategies appropriate for biased ranges is then presented. In particular, the strategies are those which account for the dominant measurement biases: **receiver clock** and **satellite clock** errors, and measurement **ambiguities** (§1.3). In the case of GPS, some of these strategies are also effective for minimising the effect of other important biases. Hence an understanding of the concepts presented here is essential to subsequent discussions on the operational and computational aspects of GPS, both for the navigation and survey modes of positioning.

Measurement types:
- **RANGE**
- **RANGE-DIFFERENCE**

Configurations:
- **Single satellite systems**
- **Multi-satellite systems**

POSITIONING STRATEGIES:

1. Range + single satellite
   ☞ **SLR**
2. Range + multi-satellite
   ☞ GPS ***(point positioning mode)***
3. Range-difference + single satellite
   ☞ **TRANSIT Doppler**
4. Range-difference + multi-satellite
   ☞ GPS ***(surveying or differential mode)***

1.4.1 GEOMETRICAL PRINCIPLES

Lines- and Surfaces-Of-Position

The geometrical principles of positioning can be demonstrated in terms of the intersection of "lines-of-position" (LOP) when considered in two-dimensions, and "surfaces-of-position" (SOP) in the case of three-dimensional positioning.

Figure 1.4-1 illustrates the SOPs for range measurements to a satellite -- *a sphere with radius being a certain distance from the satellite*; and range-difference measurements in the case of two satellites -- *a hyperboloid being the locus of all points a certain distance-difference from two satellites.*

In the case of 2-D positioning the relevant geometry is defined in terms of LOPs. Position can then be defined in terms of the intersection of two sets of LOPs, involving distances to two known points, or distance-differences from three known points as shown in Figure 1.4-2. (The "known points" may be terrestrial control stations, as in the example of offshore positioning in Figure 1.4-2, or satellites.)

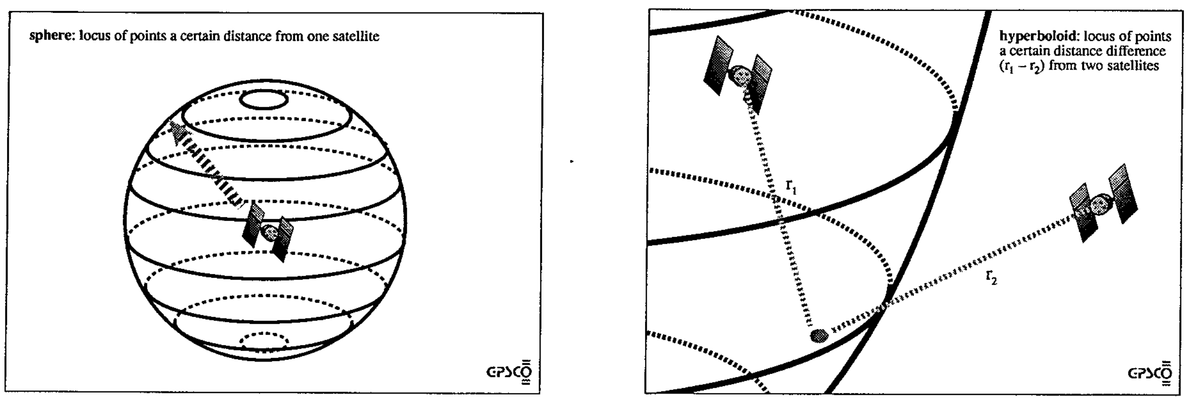


Figure 1.4-1. Surfaces-Of-Position for range and range-difference measurements.

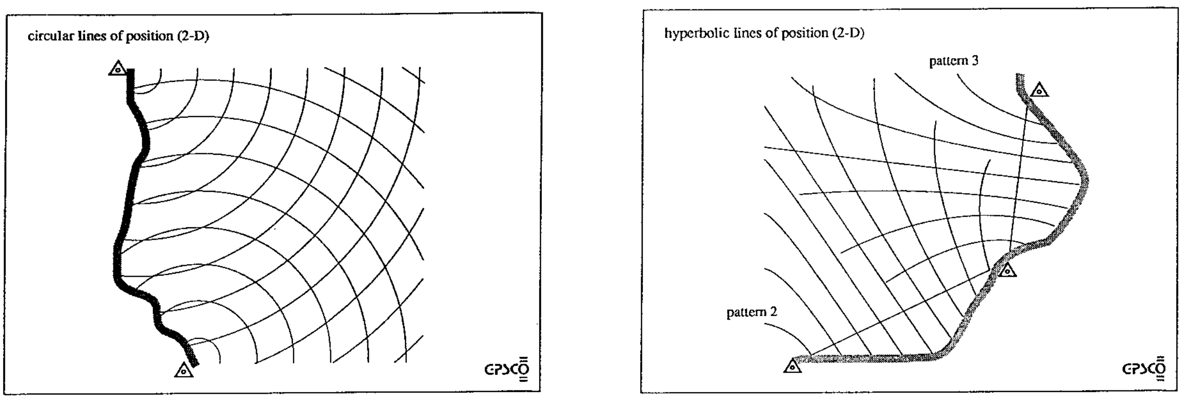


Figure 1.4-2. The intersection of Lines-Of-Position for 2-D positioning.

If attention is restricted to the scenario of range measurements to multiple satellites, the intersection of SOPs necessary for 3-D positioning is illustrated in Figure 1.4-3, in both the geometric and mathematical form. Note that range observations to three satellites (defining three SOPs) will solve the positioning "problem". (There are in fact two solutions to the problem, one of which will be clearly nonsense and can therefore be discarded, see Figure 1.4-4a.)

What if the ranges are "biased" in some way? The situation may arise as in Figure 1.4-4b! If it is assumed that the bias affects all measurements by the same amount (for example, because it has the same physical cause), then receiver-biased range measurements can still be used for positioning as illustrated in Figure 1.4-5. Here the 2-D case is used to illustrate the principle. By assuming that all measurements are biased by the same amount, the addition of a third measurement will result in an *intersection area of a certain size*. The centroid of this error figure would therefore most likely be the correct position. This can of course be extended to the three-dimensional case, which would require the addition of another (biased) range measurement to a fourth satellite.

1.4.2 RECEIVER-BIASED MEASUREMENTS

Receiver-Biased Range Positioning

The observation equation for a receiver-biased range is (eqn (1.3-14)):

$$\rho^*(t) = \rho(t) + d\tau(t_r).c$$

where c is the speed of electromagnetic radiation, $d\tau$ is the receiver clock error caused by the receiver oscillator (satellite time is assumed to be "true" time), ρ^* is the measured range and ρ is the true range. Each observation made by the receiver can be parameterised as:

$$(x^s - x)^2 + (y^s - y)^2 + (z^s - z)^2 = (\rho^* - d\tau.c)^2 \qquad (1.4\text{-}1)$$

where the time argument has been discarded.

If it is assumed that the coordinates of the transmitter (land-based or satellite-based) (x^s, y^s, z^s) are known, then each measurement ρ^* contains four parameters which are unknown: the 3-D coordinates of the receiver (x, y, z) and the receiver clock error $(d\tau)$. If four measurements are made, to four different targets, the following system of equations can be constructed:

$$\begin{aligned}
(x^{s1} - x)^2 + (y^{s1} - y)^2 + (z^{s1} - z)^2 &= (\rho^{*1} - d\tau.c)^2 \\
(x^{s2} - x)^2 + (y^{s2} - y)^2 + (z^{s2} - z)^2 &= (\rho^{*2} - d\tau.c)^2 \\
(x^{s3} - x)^2 + (y^{s3} - y)^2 + (z^{s3} - z)^2 &= (\rho^{*3} - d\tau.c)^2 \\
(x^{s4} - x)^2 + (y^{s4} - y)^2 + (z^{s4} - z)^2 &= (\rho^{*4} - d\tau.c)^2
\end{aligned} \qquad (1.4\text{-}2)$$

This system of equations has a unique solution (see, for example, LANGLEY, 1991c). If more than four measurements are made, the method of Least Squares can be used to obtain the optimal solution. Least Squares provides, in addition to the solution for the unknown parameters, an estimate of the quality of the positioning solution.

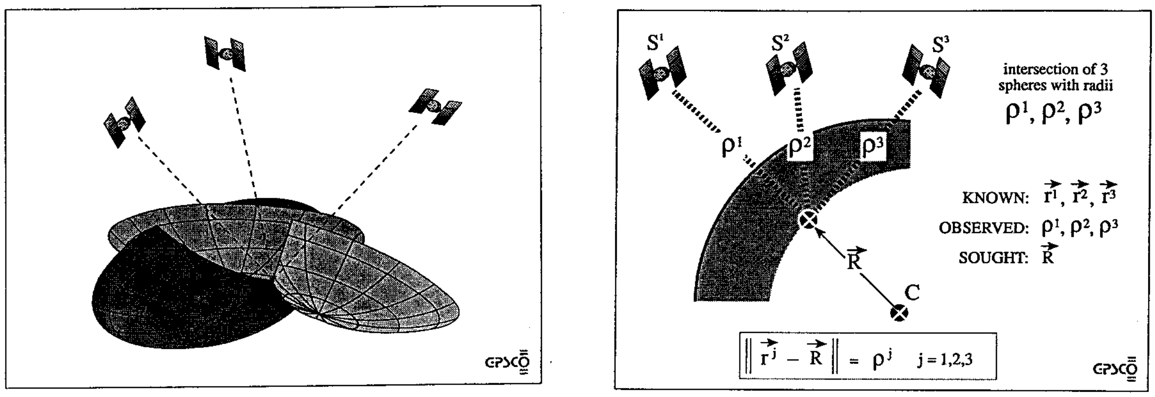


Figure 1.4-3. The geometric and mathematical problem of 3-D positioning from ranges.

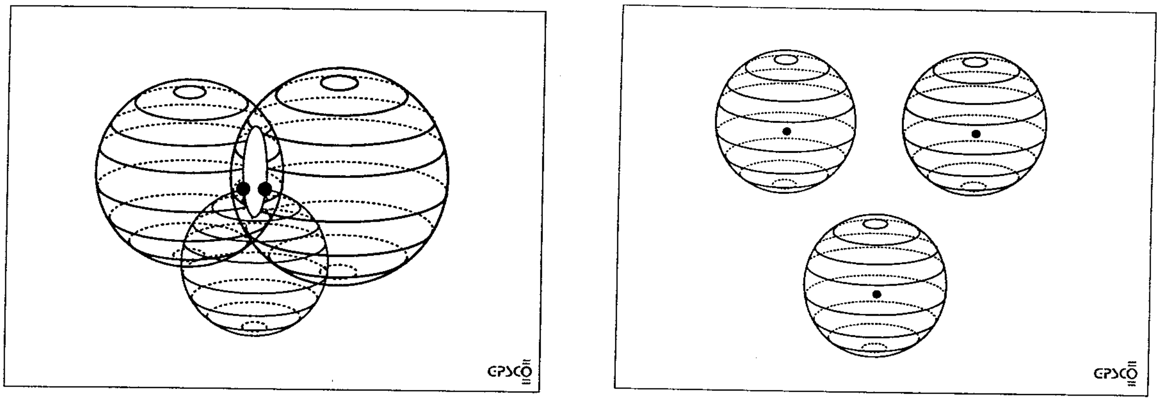


Figure 1.4-4. Intersection of SOPs: unbiased and biased range measurements.

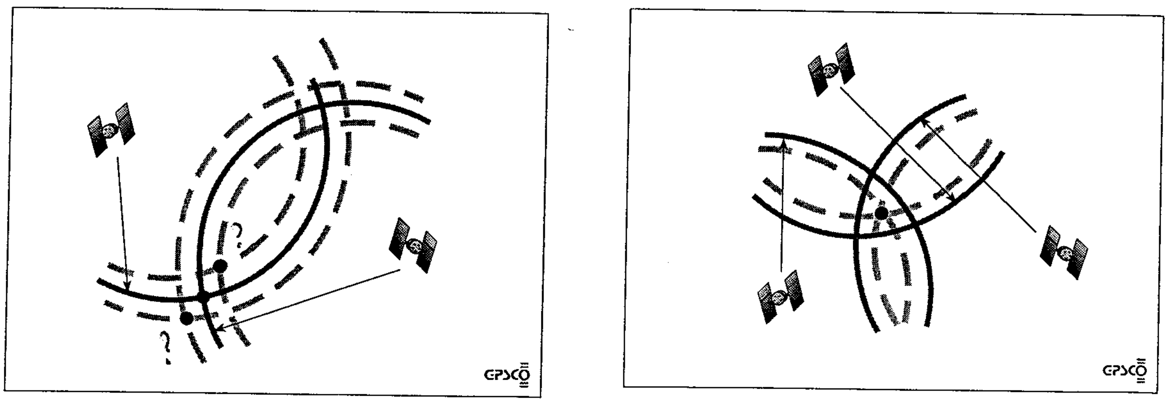


Figure 1.4-5. 2-D positioning with biased ranges: the addition of another measurement.

The steps in a Least Squares solution generally are (see, for example, HARVEY, 1994):

(1) Set up the solution: compute the elements of the design matrix $\mathbf{A}$, containing the partial derivatives of the range observations with respect to the parameters:

$$\frac{\partial \rho}{\partial x} = -\frac{x^s - x}{\rho}$$
$$\frac{\partial \rho}{\partial y} = -\frac{y^s - y}{\rho}$$
$$\frac{\partial \rho}{\partial z} = -\frac{z^s - z}{\rho} \qquad (1.4\text{-}3)$$
$$\frac{\partial \rho}{\partial d\tau} = c$$

(2) Obtain approximate (or apriori) estimates of the parameters: in particular the geodetic parameters $\mathring{\mathbf{x}}$ to be used for the computation of the partial derivatives and the residual quantities (the sum of squares of which are to be minimised):

$$\mathring{\mathbf{v}} = (\boldsymbol{l} - \boldsymbol{f}(\mathring{\mathbf{x}})) \qquad (1.4\text{-}4)$$

where $\boldsymbol{l}$ is the vector of actual observations and $\boldsymbol{f}(\mathbf{x})$ is the functional model for the observations (eqn (1.4-1)).

(3) Specify the quality of the observations: by defining the weight matrix $\mathbf{P}$.

(4) Form the normal matrix: $\mathbf{N} = \mathbf{A}^T\mathbf{P}\mathbf{A}$, and solve the system of equations:

$$\delta\hat{\mathbf{x}} = (\mathbf{N})^{-1}\, \mathbf{A}^T\mathbf{P}\, \mathring{\mathbf{v}} \qquad (1.4\text{-}5)$$

where $\delta\hat{\mathbf{x}}$ are corrections to the apriori values of the parameters $\mathring{\mathbf{x}}$. The quality of the estimated parameters can be gauged from the co-factor matrix $\mathbf{Q}_{\hat{x}\hat{x}} = (\mathbf{N})^{-1}$.

This is the standard mode of **pseudo-range positioning** used in GPS navigation, in which the receiver clock error is treated as an additional unknown. All other biases are assumed to be insignificant (that is, their impact on the quality of the position solution is considered negligible). The GPS satellite clock error can be considered a known quantity, and parameters correcting this bias are transmitted in the Navigation Message (§3.3).

Would it be necessary to solve for the receiver clock error at each epoch? That would depend upon:

- How well the clock error is estimated.
- How often the position solution is carried out.
- The quality of the clock.

A study of the Allan Variance graph (Figure 1.3-3, Table 1.3-1), and assuming: (a) a range measurement precision of about 1 metre due to "noise", and (b) the receiver is equipped with a quartz clock; then the unpredictability of the clock after 30 seconds is as great as the uncertainty of the range measurement. Clearly, once the clock error was determined, it would have to be independently estimated at least every 30 seconds otherwise it would dominate the "equivalent range error". This is one of the reasons why this method could not be used if the measurements were not made "simultaneously" on all transmitters (here, within 30 seconds of each other, so that the clock error can be assumed a constant). In fact the receiver clock can be "reset" to its "true" time on a regular basis, so that the drift of the clock (and the consequent contamination of the range measurements) can be constrained, as shown schematically in Figure 1.4-6.

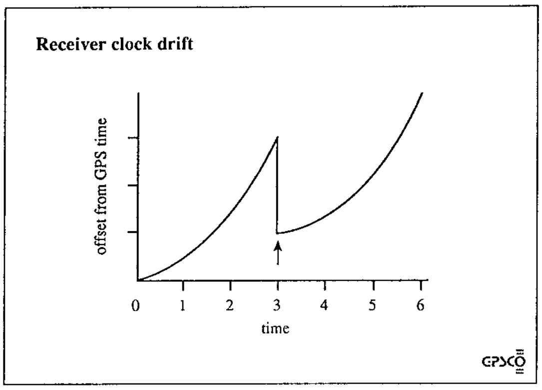


Figure 1.4-6. Clock drift and periodic reset.

Receiver-Biased Range-Difference Positioning

Is there an alternative to this scheme of using four biased ranges to solve for four parameters? The observations can be *differenced* in the following way:

$$\sqrt{(x^{s1}-x)^2+(y^{s1}-y)^2+(z^{s1}-z)^2}-\sqrt{(x^{s2}-x)^2+(y^{s2}-y)^2+(z^{s2}-z)^2}=\rho^{*1}-\rho^{*2}$$

$$\sqrt{(x^{s2}-x)^2+(y^{s2}-y)^2+(z^{s2}-z)^2}-\sqrt{(x^{s3}-x)^2+(y^{s3}-y)^2+(z^{s3}-z)^2}=\rho^{*2}-\rho^{*3}$$

$$\sqrt{(x^{s3}-x)^2+(y^{s3}-y)^2+(z^{s3}-z)^2}-\sqrt{(x^{s4}-x)^2+(y^{s4}-y)^2+(z^{s4}-z)^2}=\rho^{*3}-\rho^{*4} \quad (1.4\text{-}6)$$

This scheme is illustrated in Figure 1.4-7. Three independent "between-satellite" differences have been generated, in which the receiver clock parameter has been eliminated. As before, the observation equations can be linearised and obtain an estimate of the 3-D coordinates can be obtained. The solution would be equivalent (same 3-D coordinate components and variance-covariance matrix) to explicitly estimating the clock error as an additional parameter.

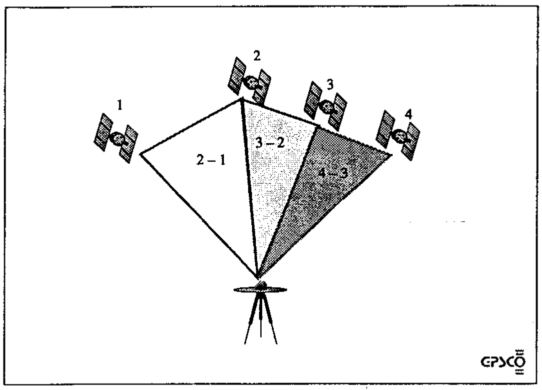


Figure 1.4-7. Between-satellite differencing to eliminate receiver clock error from ranges.

1.4.3 SATELLITE-BIASED MEASUREMENTS

Satellite-Biased Range Positioning

The observation equation for a satellite-biased range is (eqn (1.3-16):

$$\rho^*(t) = \rho(t) + \Delta\tau^s(T_s).c$$

where c is the speed of electromagnetic radiation, $\Delta\tau$ is the satellite clock error caused by the satellite oscillator not being synchronised to "true" time, ρ^* is the measured range and ρ is the true range. Each observation made by the receiver can be parameterised as in equation (1.4-1), except for the replacement of $d\tau$ by $\Delta\tau^s$:

$$(x^s - x)^2 + (y^s - y)^2 + (z^s - z)^2 = (\rho^* - \Delta\tau^s.c)^2 \quad (1.4\text{-}7)$$

Note that the time argument has been discarded.

If it is assumed that the coordinates of the satellite signal transmitter (x^s, y^s, z^s) are known, then there are six unknowns in the system: the 3-D coordinates of the receiver (x_{r1}, y_{r1}, z_{r1}) and the three satellite clock error terms ($\Delta\tau^{si}$). (It is only necessary to observe to three satellites if no receiver bias is present.) Six satellite-biased range observations are therefore required in order to solve this positioning problem.

It is not feasible to simply observe more satellites, as each new satellite observation introduces a new clock parameter. It is possible, however, to take advantage of the fact that all observations made to a particular satellite are biased by the same amount (if made at the same time, or close together so that the satellite clock error can be assumed to have not changed by any appreciable amount). If three range observations are made from another station, whose coordinates are known (x_{r2}, y_{r2}, z_{r2}), then the following system of six equations in six unknowns is obtained:

$$
\begin{aligned}
(x^{s1} - x_{r1})^2 + (y^{s1} - y_{r1})^2 + (z^{s1} - z_{r1})^2 &= (\rho^*{}_{r1}{}^{s1} - \Delta\tau^{s1}.c)^2 \\
(x^{s2} - x_{r1})^2 + (y^{s2} - y_{r1})^2 + (z^{s2} - z_{r1})^2 &= (\rho^*{}_{r1}{}^{s2} - \Delta\tau^{s2}.c)^2 \\
(x^{s3} - x_{r1})^2 + (y^{s3} - y_{r1})^2 + (z^{s3} - z_{r1})^2 &= (\rho^*{}_{r1}{}^{s3} - \Delta\tau^{s3}.c)^2 \\
(x^{s1} - x_{r2})^2 + (y^{s1} - y_{r2})^2 + (z^{s1} - z_{r2})^2 &= (\rho^*{}_{r2}{}^{s1} - \Delta\tau^{s1}.c)^2 \\
(x^{s2} - x_{r2})^2 + (y^{s2} - y_{r2})^2 + (z^{s2} - z_{r2})^2 &= (\rho^*{}_{r2}{}^{s2} - \Delta\tau^{s2}.c)^2 \\
(x^{s3} - x_{r2})^2 + (y^{s3} - y_{r2})^2 + (z^{s3} - z_{r2})^2 &= (\rho^*{}_{r2}{}^{s3} - \Delta\tau^{s3}.c)^2
\end{aligned}
\tag{1.4-8}
$$

for which a unique solution can be obtained.

Satellite-Biased Range-Difference Positioning

It is possible to difference the observations between-receivers / same satellite:

$$
\begin{aligned}
&\sqrt{(x^{s1}-x_{r1})^2 + (y^{s1}-y_{r1})^2 + (z^{s1}-z_{r1})^2} - \sqrt{(x^{s1}-x_{r2})^2 + (y^{s1}-y_{r2})^2 + (z^{s1}-z_{r2})^2} \\
&\qquad = \rho^*{}_{r1}{}^{s1} - \rho^*{}_{r2}{}^{s1} \\
&\sqrt{(x^{s2}-x_{r1})^2 + (y^{s2}-y_{r1})^2 + (z^{s2}-z_{r1})^2} - \sqrt{(x^{s2}-x_{r2})^2 + (y^{s2}-y_{r2})^2 + (z^{s2}-z_{r2})^2} \\
&\qquad = \rho^*{}_{r1}{}^{s2} - \rho^*{}_{r2}{}^{s2} \\
&\sqrt{(x^{s3}-x_{r1})^2 + (y^{s3}-y_{r1})^2 + (z^{s3}-z_{r1})^2} - \sqrt{(x^{s3}-x_{r2})^2 + (y^{s3}-y_{r2})^2 + (z^{s3}-z_{r2})^2} \\
&\qquad = \rho^*{}_{r1}{}^{s3} - \rho^*{}_{r2}{}^{s3}
\end{aligned}
\tag{1.4-9}
$$

These are three independent "between-receiver" differences in which the three satellite clock error parameters have been eliminated. As before, it is possible to proceed and linearise this model and obtain an estimate of the 3-D coordinates. The solution would be equivalent (same 3-D coordinate components and VCV matrix) to explicitly estimating the clock errors as additional parameters in eqn (1.4-8).

1.4.4 AMBIGUOUS MEASUREMENTS

Range Positioning with Ambiguous Ranges

The observation equation for an ambiguous range, *where the bias is a constant,* is (eqn (1.3-19)):

$$\rho^*(t) = \rho(t) - n_{r1}{}^{si}$$

where n is the bias that is specific to a receiver (rj) - satellite (si) pair, ρ^* is the measured range and ρ is the true range. Note that there is no time dependence for n. Each observation made by the receiver can be parameterised by a modification of eqn (1.4-1):

$$(x^s(t) - x)^2 + (y^s(t) - y)^2 + (z^s(t) - z)^2 = (\rho^*(t) - n^{si})^2 \tag{1.4-10}$$

If it is assumed that the coordinates of the satellite signal transmitter $(x^s(t), y^s(t), z^s(t))$ are known, then there are six unknowns in the system: the 3-D coordinates of the receiver (x, y, z) and the three bias terms (n^{si}). (It is only necessary to observe to three satellites if no receiver bias is present.) Therefore six ambiguous range observations are required in order to solve this positioning problem. It is not possible to simply observe more satellites, as each new satellite observation introduces a new ambiguity parameter. It is possible, however, to take advantage of the fact that all observations made from a certain receiver to a particular satellite are biased by the same amount. If three more range observations are made from the receiver to the same three satellites, but at some time interval later, the system of six equations in six unknowns can be obtained:

$$
\begin{aligned}
&(x^{s1}(t) - x)^2 + (y^{s1}(t) - y)^2 + (z^{s1}(t) - z)^2 = (\rho^{*s1}(t) - n^{s1})^2 \\
&(x^{s2}(t) - x)^2 + (y^{s2}(t) - y)^2 + (z^{s2}(t) - z)^2 = (\rho^{*s2}(t) - n^{s2})^2 \\
&(x^{s3}(t) - x)^2 + (y^{s3}(t) - y)^2 + (z^{s3}(t) - z)^2 = (\rho^{*s3}(t) - n^{s3})^2 \\
&(x^{s1}(t+dt) - x)^2 + (y^{s1}(t+dt) - y)^2 + (z^{s1}(t+dt) - z)^2 = (\rho^{*s1}(t+dt) - n^{s1})^2 \\
&(x^{s2}(t+dt) - x)^2 + (y^{s2}(t+dt) - y)^2 + (z^{s2}(t+dt) - z)^2 = (\rho^{*s2}(t+dt) - n^{s2})^2 \\
&(x^{s3}(t+dt) - x)^2 + (y^{s3}(t+dt) - y)^2 + (z^{s3}(t+dt) - z)^2 = (\rho^{*s3}(t+dt) - n^{s3})^2
\end{aligned}
\qquad (1.4\text{-}11)
$$

for which a unique solution can be obtained. Note, **it is assumed that the receiver has not moved between time t and t+dt.**

Range-Differencing of Ambiguous Ranges

It is also possible to difference the observations from the same receiver to the same satellite, but made at different epochs:

$$
\begin{aligned}
&\sqrt{(x^{s1}(t)-x)^2 + (y^{s1}(t)-y)^2 + (z^{s1}(t)-z)^2} - \\
&\qquad \sqrt{(x^{s1}(t+dt)-x)^2 + (y^{s1}(t+dt)-y)^2 + (z^{s1}(t+dt)-z)^2} = \rho^{*s1}(t) - \rho^{*s1}(t+dt) \\
&\sqrt{(x^{s2}(t)-x)^2 + (y^{s2}(t)-y)^2 + (z^{s2}(t)-z)^2} - \\
&\qquad \sqrt{(x^{s2}(t+dt)-x)^2 + (y^{s2}(t+dt)-y)^2 + (z^{s2}(t+dt)-z)^2} = \rho^{*s2}(t) - \rho^{*s2}(t+dt) \\
&\sqrt{(x^{s3}(t)-x)^2 + (y^{s3}(t)-y)^2 + (z^{s3}(t)-z)^2} - \\
&\qquad \sqrt{(x^{s3}(t+dt)-x)^2 + (y^{s3}(t+dt)-y)^2 + (z^{s3}(t+dt)-z)^2} = \rho^{*s3}(t) - \rho^{*s3}(t+dt)
\end{aligned}
\qquad (1.4\text{-}12)
$$

There are three independent "between-epoch" differences in which the three constant satellite-receiver specific biases have been eliminated. As before, it is possible to proceed and linearise this model and obtain an estimate of the 3-D coordinates. The solution would be equivalent (same 3-D coordinate components and VCV matrix) to explicitly estimating the biases as additional parameters in equation (1.4-11).

1.4.5 SOLUTION STRATEGIES FOR BIASED RANGES: SUMMARY REMARKS

Table 1.4-1. Solution strategies for various classes of biased ranges.

	Receiver Bias (A)	Satellite Bias (B)	Receiver-Satellite Ambiguity (C)
Change with Time?	YES	YES	NO
Solution Strategy:			
1. Multi-Satellite?	YES	NO	NO
2. Multi-Receiver?	NO	YES	NO
3. Multi-Epoch?	NO	NO	YES

The following conclusions can be drawn from the above Table:

(1) If only bias "A" is present in the measurements, then the strategy is to observe more than one satellite at the same time. In the case of 3-D positioning problems, the minimum number of satellites to be observed is four. This ensures real-time positioning capability, and is the basis for POINT-POSITIONING (or absolute positioning) using GPS in the **NAVIGATION MODE**. *This mode is the one commonly used for positioning a moving receiver to moderate accuracy levels.* The solution for the 3-D coordinates of a single receiver, and its receiver (clock) bias is known as the "Navigation Solution" (or navigation "fix").

(2) If the biases affecting the range measurements are of type "A + B", then the strategy is to observe multiple satellites simultaneously from two or more receivers and derive the DIFFERENTIAL POSITION. When applied to GPS navigation this is referred to as the **DIFFERENTIAL NAVIGATION MODE** as it is essentially an instantaneous "fix", and can be implemented in real-time if a communication link exists between the receivers.

(3) If the range biases are of the variety "A + B + C", an extended observation "session" is required (as well as multiple satellites tracked simultaneously from two or more GPS receivers) to "resolve" the ambiguity bias. This is the normal GPS **SURVEYING MODE** associated with carrier phase measurements (§3.2).

For GPS, a discussion of the relative magnitudes of these biases and the precisions associated with the range-like measurements possible with GPS receivers is given in §6.2.

1.4.6 THE IMPACT OF SATELLITE GEOMETRY

The accuracy with which positions can be determined is not just a function of the measurement precision, and the appropriate modelling of biases. It is also a function of the satellite(s) - receiver(s) geometry (see LANGLEY, 1991c). Hence, although the systems of eqns (1.4-2), (1.4-6), (1.4-8), (1.4-9), (1.4-11), (1.4-12) are all *theoretically* valid solutions to the positioning problem, geometric considerations may make a certain solution strategy better than another. The simplest case to consider is point positioning using receiver-biased range measurements -- the GPS navigation mode referred to above.

The co-factor matrix $\mathbf{Q}_{XYZ}$ from the Least Squares solution contains the contribution to position error of both the geometry and the random measurement error. While in the surveying discipline the components of the co-factor matrix of parameters are transformed into components of an "error ellipsoid" (orientation and length of the three axes) (§9.1), in the case of the navigation applications the effect of satellite configuration geometry is usually expressed by the **Dilution of Precision** (DOP) factor. DOP is the ratio of the positioning accuracy to the measurement accuracy:

$$\sigma = \mathrm{DOP}.\sigma_o \qquad (1.4\text{-}13)$$

where σ_o is the measurement accuracy, and
σ is the position accuracy.

DOP is always a number greater than unity when there are no redundant observations.

There are a number of different definitions of DOP factors, depending on the coordinate component, or combination of coordinate components, being considered:

$$\begin{aligned}
\mathrm{PDOP} &= \sqrt{\sigma_E^2 + \sigma_N^2 + \sigma_H^2} = \sqrt{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2} \\
\mathrm{HTDOP} &= \sqrt{\sigma_E^2 + \sigma_N^2 + \sigma_T^2} \\
\mathrm{HDOP} &= \sqrt{\sigma_E^2 + \sigma_N^2} \qquad (1.4\text{-}14) \\
\mathrm{VDOP} &= \sqrt{\sigma_H^2} \\
\mathrm{TDOP} &= \sqrt{\sigma_T^2}
\end{aligned}$$

where: σ_E^2, σ_N^2, σ_H^2 are the variances of the east, north and height components,
σ_X^2, σ_Y^2, σ_Z^2 are the variances of the X, Y and Z components, and
σ_T^2 is the variance of the estimated receiver clock error parameter.

are all obtained from the diagonal elements of the co-factor matrix of the Least Squares position solution $\mathbf{Q}_{XYZ}$. (All elements have been divided by the variance of unit weight.) The range solution is likely to be in the form of Cartesian coordinate components (X, Y, Z) -- eqn (1.4-2). The corresponding co-factor matrix for the local geographic components (E, N, H) is obtained as follows:

$$\mathbf{Q}_{ENHT} = \mathbf{R} \,.\, \mathbf{Q}_{XYZT} \,.\, \mathbf{R}^T \tag{1.4-15}$$

or

$$\begin{bmatrix} \sigma_E^2 & \sigma_{EN} & \sigma_{EH} & \sigma_{ET} \\ \sigma_{NE} & \sigma_N^2 & \sigma_{NH} & \sigma_{NT} \\ \sigma_{HE} & \sigma_{HN} & \sigma_H^2 & \sigma_{HT} \\ \sigma_{TE} & \sigma_{TN} & \sigma_{TH} & \sigma_T^2 \end{bmatrix} = \mathbf{R} \,.\, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} & \sigma_{XT} \\ \sigma_{YX} & \sigma_Y^2 & \sigma_{YZ} & \sigma_{YT} \\ \sigma_{ZX} & \sigma_{ZY} & \sigma_Z^2 & \sigma_{ZT} \\ \sigma_{TX} & \sigma_{TY} & \sigma_{TZ} & \sigma_T^2 \end{bmatrix} \,.\, \mathbf{R}^T$$

$$\text{where } \mathbf{R} = \begin{bmatrix} -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi & 0 \\ -\sin\lambda & \cos\lambda & 0 & 0 \\ \cos\phi\cos\lambda & \cos\phi\sin\lambda & \sin\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.4-16}$$

In the case of GPS point positioning, which requires the estimation of four parameters: 3-D position and receiver clock error, the most appropriate DOP factor is the **Geometric Dilution of Precision** (GDOP):

$$\text{GDOP} = \sqrt{(\text{PDOP})^2 + (\text{TDOP})^2)} \tag{1.4-17}$$

GDOP can be interpreted as the reciprocal of the volume of a tetrahedron that is formed from the four satellites and receiver position, hence the best geometric situation for point positioning is when the volume is a maximum, which therefore requires GDOP to be a minimum. Figure 1.4-8 illustrates the situation of good and poor GDOP.

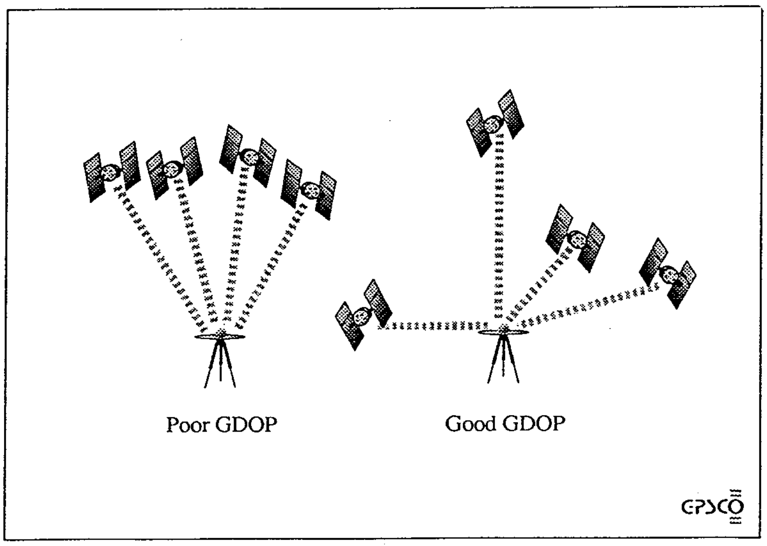


Figure 1.4-8. The relationship between satellite configuration geometry and GDOP.

The following comments can be made regarding DOPs:

- The smaller the value of DOP, the higher the precision of the position results -- *the measurement errors are not as strongly amplified.*
- DOP is usually greater than unity, however, if many satellites are observed (say >8) the value of DOP can be less than unity.
- DOPs can be used as the basis of selecting satellites for solution -- *if GPS receiver cannot track all satellites that are in view.*
- A high DOP (say >10) defines an "outage" -- *a situation where the position solution is too unreliable.*
- DOP varies with time of day and geographic location -- *but the pattern of DOP at a location repeats itself each day because the constellation is unchanged from day-to-day (except it rising approximately four minutes earlier each day), hence it is highly predictable.*
- DOP varies with number of satellites considered-- *due to such factors as elevation cutoff angle used, number of satellites used by receiver to give "fix", etc.*
- DOP can be computed without the need for any measurements -- *only the satellite positions are required (from an appropriate ephemeris) and an approximate receiver position.*
- DOP has only a limited role in differential positioning -- *it may be useful for certain types of GPS survey planning, as well as for quality assurance.*

An example of the variation in PDOP over 24 hours at Sydney, Australia, calculated for all visible satellites above an elevation cutoff angle of 5°, is shown in Figure 1.4-9.

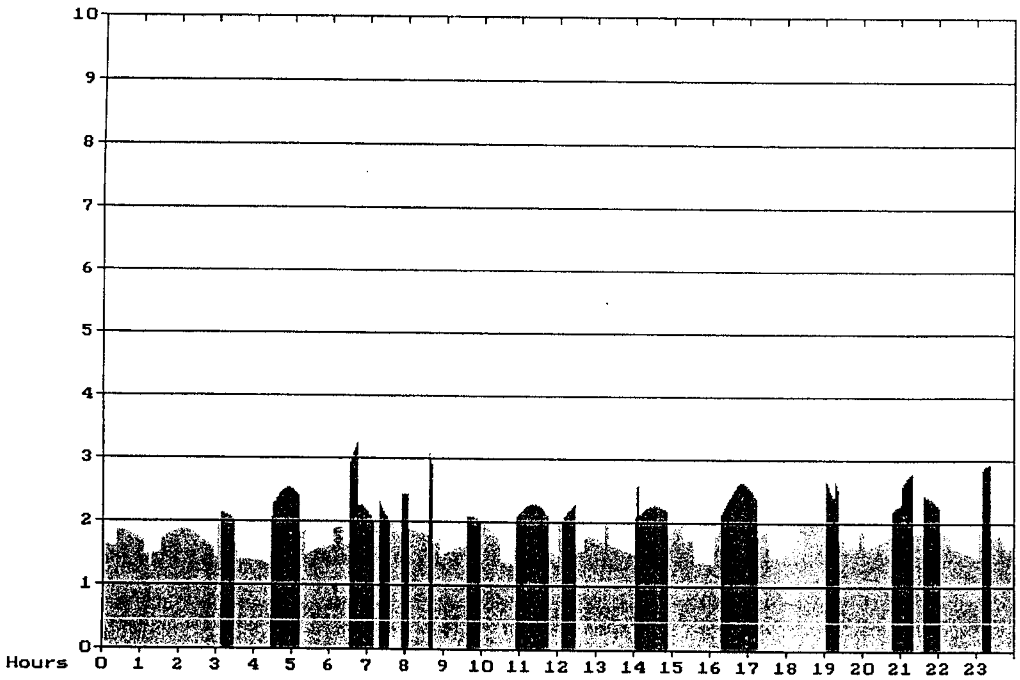


Figure 1.4-9. Variation of PDOP at Sydney, Australia (elevation cutoff 5°).

Chapter 2: Introduction to GPS Positioning

2.1 INTRODUCTION TO GPS

2.1.1 WHAT IS GPS?

The NAVSTAR Global Positioning System (GPS) is a *satellite-based positioning system*, which by virtue of its special characteristics, is revolutionising the tasks of navigation and surveying:

☞ **Relatively high positioning accuracies, from the dekametre to the millimetre level.**

☞ **Determination of velocity and time to an accuracy commensurate with position.**

☞ **Available to users anywhere on the globe: in the air, on the ground, or at sea.**

☞ **Relatively low cost system, with no user charges.**

☞ **All-weather system, available 24 hours a day.**

☞ **The position information is in three dimensions, that is, vertical as well as horizontal information is provided.**

Who Developed GPS?

The Global Positioning System is, in the first instance, a military navigation system designed, financed, deployed and controlled by the U.S. Department of Defense.

Development work on GPS commenced in 1973 as a result of the merger of several R&D programs within the U.S. Department of Defense, namely the Navy's "TIMATION" project, and the Air Force's "621B" project. The first satellite was launched in 1978. (For a background to the development of the GPS program the reader is referred to PARKINSON, 1994, and PARKINSON et al, 1995.) The development and production program for the GPS is managed by the U.S. Air Force (USAF) Systems Command, Space Systems Division, Joint Program Office (JPO) at the Los Angeles Air Force Base, California. The JPO is manned by personnel from the USAF, U.S. Navy, U.S. Army, U.S. Marine Corps, U.S. Coast Guard, U.S. Defense Mapping Agency, NATO nations and Australia.

The aim of the JPO was to develop an all-weather, 24-hour, truly global navigation system to support the positioning requirements for the armed forces of the U.S. and its allies. As such a system was designed to replace the large variety of navigational systems already in use, a great import was placed on the system's utility, reliability and survivability. A number of stringent conditions therefore had to be met. In addition to those listed above, they included:

- ❑ suitable for all classes of platform: aircraft (jet to helicopter), ship, land (vehicle to handheld units) and space (missiles and satellites),
- ❑ able to handle a wide variety of dynamics,
- ❑ real-time positioning and velocity determination capability,
- ❑ the positioning results were to be available on a single global geodetic datum,
- ❑ highest accuracy to be restricted to a certain class of user,
- ❑ resistant to jamming,
- ❑ redundancy provisions to ensure the survivability of the system,
- ❑ low-cost, low power, and hence much of the complexity to be built into the satellite segment, and
- ❑ replace the ageing TRANSIT Doppler system and other navaids.

This led to a design concept based on (§1.1 and §1.2):

- a system using measured ranges,
- the latest technology in clocks and microwave transmission technology,
- multiple satellites, and
- accuracy degradation that was graceful.

The total investment by the U.S. military in the GPS system to date is well over $10 BILLION (US)!

However, *although the primary goal of GPS is to provide land, air and marine positioning capabilities to the U.S. armed forces and its allies, GPS is freely available to all users.* The number of civilian users is already far greater than the military users, and the applications of the positioning technology are growing rapidly (§2.3). The civilian sector therefore represents an important user group that is increasingly lobbying in order to influence official GPS policy. The U.S. military however still operates several "levers" with which they control the performance of GPS (see §2.4). On the other hand, there is tremendous innovation occurring within the civilian sector, with the development of technology and procedures that are increasing making redundant many of the U.S. military procedures intended to restrict GPS performance.

Why a Satellite Positioning System?

Satellite positioning systems are global. The signals can be "seen" over a large area, and are not interfered with by terrain or geography to the same extent as conventional ground-based positioning systems.

GPS was designed to replace the TRANSIT Doppler satellite navigation system which has given good service to the navigation and geodetic community for over 20 years. The advantages of TRANSIT are essentially those of GPS as well. A microwave satellite-based system (§1.1):

- Can transmit signals that can be "seen" over a far larger area than ground-based systems.
- Can transmit signals through cloud and rain.
- Can be used day or night, as long as the transmitting satellite is above the user's horizon.
- Recognises no national boundaries, and refers positions to a global datum uniquely defined for that system.
- The position information is three-dimensional.

However, in addition, the NAVSTAR Global Positioning System has some further advantages over other satellite-based positioning systems:

☞ It is a <u>one-way (listen only) system</u>, in which the satellites transmit signals, but are unaware who is using the signal (no receiving function). The user (or listener) does not transmit a signal, and therefore:
  - cannot be detected by the enemy (military context), and
  - cannot be charged for using the system (civilian context).

☞ As GPS is a <u>multi-satellite system</u>, there is always a number of satellites visible simultaneously anywhere on the globe, and at any time.

☞ GPS <u>can support a number of positioning and measurement modes</u> in order to satisfy simultaneously a variety of users, from those only requiring navigation (dekametre) accuracies to those demanding very high (millimetre - centimetre) accuracies.

2.1.2 SATELLITE-BASED POSITIONING SYSTEMS

Some characteristics of satellite-based positioning systems, operational and proposed, are:

❑ ***TRANSIT Doppler (NNSS)***
- Operational since 1964
- Typically 4-6 useable satellites at 1075km altitude in polar orbits
- U.S. military navigation system, but open to all users (with restrictions)

❑ ***TSIKADA***
- Russian equivalent to TRANSIT Doppler system

❑ ***NAVSTAR Global Positioning System***
- Multi-satellite system to replace TRANSIT
- 24 satellite constellation at 20200km altitude
- U.S. military navigation system, but open to all users (with restrictions)

❑ ***GLONASS***
- Russian equivalent to GPS

❑ ***STARFIX***
- Commercial positioning system for continental U.S.

❑ ***GEOSTAR/LOCSTAR***
- Electronic package for 2-way information transfer on communication satellites
- Limited coverage for subscribers
- U.S. & French initiative

- ***ARGOS***
  - Doppler system from CNES (France) & NOAA (U.S.)
  - Transmitters on ground, receivers in two weather satellites
  - Subscriber service for a variety of platforms
- ***NAVSAT & other LEO systems***
  - ESA proposal for two-way GPS civilian system
  - 18 satellites in mix of HEO & GEO orbits
  - Some may be communication satellites (GEO & LEO)

TRANSIT Doppler System

The history of TRANSIT coincides with the start of the Space Age (4th October 1957). Sputnik I was launched and the Doppler shift of the signals were used to determine the satellite orbit. The method was inverted so that if the orbit were known, the position of the receiver could be determined (PARKINSON et al, 1995). Special features and milestones:

- *Development of TRANSIT began 1959, with first satellite launched in 1961.*
- *Became operative for military use in 1964 and was released for civil use in 1967.*
- *Two transmission frequencies (400MHz & 150MHz), with phase modulation of the navigation message.*
- *Small number of satellites (4-6) and low altitude (1075km) means coverage is not time continuous.*
- *In polar orbit (Figure 2.1-1).*
- *Russian Republic (formerly the USSR) operates a similar system known as Tsikada.*
- *To be phased out in 1997.*

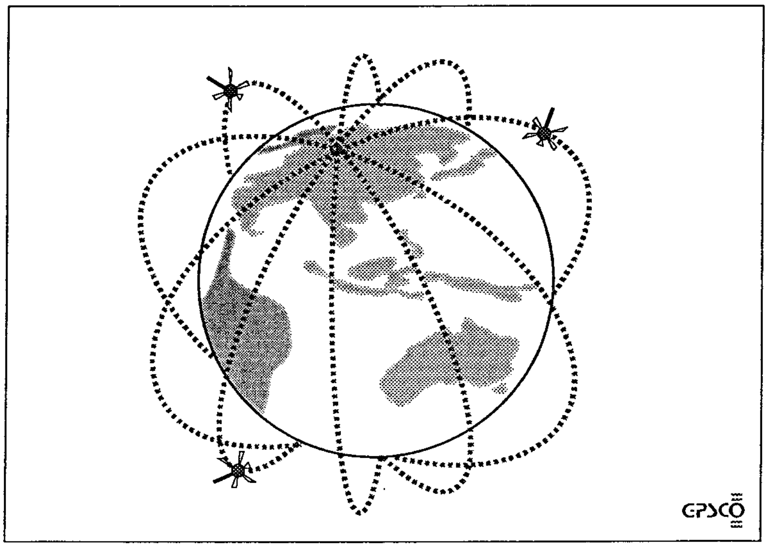


Figure 2.1-1. The TRANSIT Doppler satellite positioning system configuration.

ARGOS System

ARGOS is another satellite system which uses the Doppler principle for positioning. ARGOS is a cooperative project between the French Centre National d'Etudes (CNES), NASA, and the U.S. National Oceanic and Atmospheric Administration (NOAA), and was first deployed in 1978. Transmitters are operated by users on variety of "platforms" (buoys, animal tracking, radiosondes, etc.), and satellites act as receivers (one of two U.S. TIROS weather stations). CNES computes position and velocity of platform, sends information (and the bill!) to user. Other such "subscriber" systems in place include the COSPAS-SARSAT search & rescue system and GEOSTAR. The important distinction is that ARGOS is essentially a *satellite-based tracking system*, while many of the other systems (including GPS) are *self-navigating systems.*

The GLONASS System

The Global Orbiting Navigation Satellite System (GLONASS) is the Russian equivalent to the GPS system, and has the following characteristics (KLEUSBERG, 1990; IVANOV & SALISTCHEV, 1991; PARKINSON et al, 1995):

- *21 satellites + 3 active spares (Figure 2.1-2).*
- *3 planes, 8 satellites per plane.*
- *64.8° inclination, 19,100km altitude (11hr 15min period).*
- *Dual-frequency (1597-1617MHz, 1240-1260MHz).*
- *Each satellite transmits a different frequency.*
- *Spread-spectrum PRN code signal structure.*
- *Global coverage for navigation based on simultaneous pseudo-ranges, with an accuracy of the order of 10-20m (similar to GPS without Selective Availability).*
- *Considered an important complement to GPS for integrity monitoring and quality control purposes.*

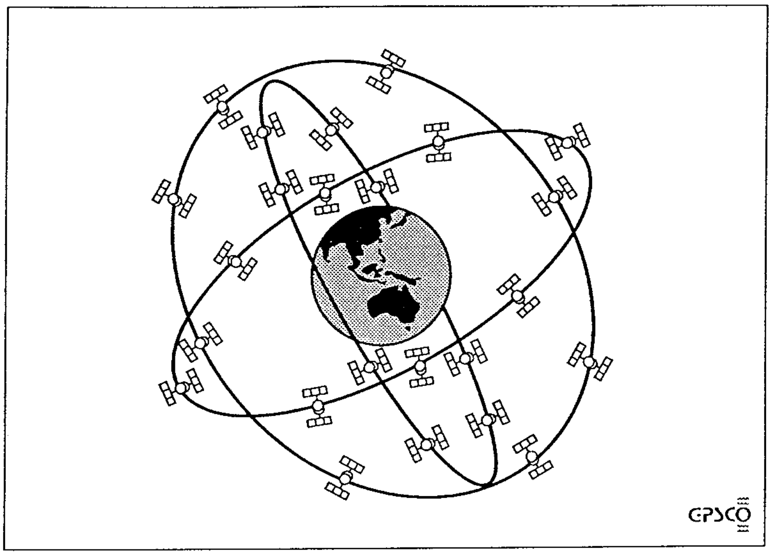


Figure 2.1-2. The GLONASS satellite configuration.

GLONASS was declared operational in 1996, and several GPS-GLONASS receivers have been developed for both surveying and navigation applications.

2.1.3 TECHNOLOGY ADVANCES DRIVING GPS

Rarely have so many seemingly unrelated technological advances been required to make possible such a complex system as GPS. Briefly they are (PARKINSON, 1983):

(1) Space System Reliability: The U.S. space program had by 1973 demonstrated the reliability of space hardware. In particular, the TRANSIT system had important lessons for although the TRANSIT satellites were originally designed to last 2-3 years in orbit, some of the satellites have operated well beyond their design life. TRANSIT has continued to perform reliably for over 25 years.

(2) Atomic Clock Technology: With the development of atomic "clocks" a new era of precise time-keeping had dawned. Before the GPS program however, these precise clocks had never been tested in space. The development of reliable, and stable, space-qualified atomic frequency oscillators (rubidium, and then cesium) was a significant technological breakthrough. The advanced clocks being used on the GPS satellites routinely achieve long-term frequency stability of the order of a few parts in 10^{14} per day (about 1 sec in 3,000,000 years!). This long-term stability is one of the keys to GPS, as it allows for the autonomous, synchronised generation and transmission of accurate timing signals by each of the GPS satellites without continuous monitoring from the ground.

(3) Quartz Crystal Oscillator Technology: In order to keep the cost of user equipment down, quartz crystal oscillators were proposed (similar to those used in modern digital watches), rather than using atomic clocks as in the GPS satellites. Besides their cost, quartz oscillators have excellent short-term stability (Table 1.3-1), and their drift can be corrected as part of the GPS position determination process (§1.4).

(4) Precise Satellite Tracking and Orbit Determination: Successful operation of GPS depends on the precise knowledge and prediction of each satellite's position with respect to the earth. This is the task of the Control Segment. Tracking data collected by the monitor stations is analysed to determine the ephemeris over the period of tracking (typically one week). This reference ephemeris is extrapolated into the future and the data is then up-loaded to the satellites. Prediction accuracies for one day, of the satellite coordinates, in the sub-dekametre range have been demonstrated.

(5) Spread-Spectrum Technology: The ability to track and obtain any selected GPS satellite signal (a receiver will be required to track a number of satellites at the same time), in the presence of a lot of ambient noise is a critical technology. This is now possible using spread-spectrum and pseudo-random-noise coding techniques.

(6) Large-Scale Integrated Circuit Technology: To realise the aim of low cost, low power and small size for much of the user equipment, the GPS program relies heavily on the successful application of VLSI circuits, and the powerful computing capabilities built onto them.

The GPS system is possible because of advances in:

- ☞ **Space system reliability.**
- ☞ **Clock technology for receivers and satellites.**
- ☞ **Precise satellite tracking and orbit determination capabilities.**
- ☞ **Spread-spectrum and VLSI technology.**

2.1.4 THE GPS REFERENCE SYSTEMS

The WGS84 System

World Geodetic System 84 is defined and maintained by the U.S. Defense Mapping Agency as a *global geodetic datum* (D.M.A., 1991). It is the datum to which all GPS positioning information is referred by virtue of being the reference system of the Broadcast Ephemeris. (Prior to January 1987, the system in use was WGS72.) The *realisation* of the WGS84 satellite datum is the catalogue of coordinates of over 1500 geodetic stations (most of them active or past tracking stations) around the world. They fulfil the same function as national geodetic stations, they provide the means by which a position can be related to a datum. WGS84 is an earth-fixed Cartesian coordinate system with:

- Its origin at the earth's centre of mass, the **geocentre** (for two reasons: the geocentre is the physical point about which the satellite orbits; and it is preferable to any local geodetic datum).
- Its "z-axis" is aligned parallel to the direction of the Conventional Terrestrial Pole (CTP) for polar motion, as originally defined by the Bureau International de l'Heure (BIH), and since 1989 by the International Earth Rotation Service (IERS).
- Its "x-axis" is the intersection of the WGS84 Reference Meridian Plane and the plane of the CTP Equator (the Reference Meridian being parallel to the Zero Meridian defined by BIH/IERS).
- Its "y-axis" completes a right-handed, earth-centred, earth-fixed (ECEF) orthogonal coordinate system, measured in the plane of the CTP Equator, 90° east of the x-axis.

The four defining parameters of the **WGS84 ellipsoid** are:

- Semi-major axis (a): 6378137m.
- Ellipsoid flattening (f): 1/298.257223563 (derived from the value of the normalised second degree zonal harmonic coefficient of the gravitational field: $-484.16685 \times 10^{-6}$).
- Angular velocity of the earth (ω): 7292115×10^{-11} rad/sec.
- The earth's gravitational constant (atmosphere included) (GM): 3986005×10^{-8} m^3/sec^2.

The relationship between WGS84, as well as other global datums, and local geodetic datums have been determined empirically, and transformation parameters of varying quality have been derived (see SOLER & HOTHEM, 1989; STEED, 1990; and Table 1.2-2). Reference systems are periodically redefined, for various reasons, such as when the primary tracking technology changes (for example when the TRANSIT system was superseded by GPS), or if the configuration of ground stations alters radically enough to justify a recomputation of the global datum coordinates. The result is generally a small refinement in the datum definition, and a change in the numerical values of the coordinates. For example, there is a small difference in the definition of WGS84 from TRANSIT and GPS (see Table 1.2-2): origin offsets of approximately 10cm in the z-direction and a scale difference of about 0.1ppm.

However, with dramatically increasing tracking accuracies another phenomenon impacts on datum definition and its maintenance: *the motion of the tectonic plates across the earth's surface*, or "continental drift" as it is often known (it is assumed there is relatively little vertical motion). This motion is measured in centimetres per year, with the fastest rates being over 10cm/year. Nowadays this motion can be monitored and measured to centimetre accuracy, on a global annual-average basis. In 1994 the GPS reference system underwent a subtle change to WGS84(G730) to bring it into the same system as used by the International GPS Service for Geodynamics to produce precise GPS ephemerides (§6.2 and §12.2).

The International Terrestrial Reference Frame

The WGS84 system is the most widely used global reference system because it is the system in which the GPS satellite coordinates are expressed in the Navigation Message (§3.3). Other satellite reference systems have been defined but these have mostly been for "scientific" purposes. However, since the mid 1980's, geodesists have been using GPS to measure crustal motion, and to define more precise satellite datums. The latter were essentially by-products of the sophisticated data processing, which included the computation of the GPS satellite orbits. These GPS surveys required coordinated tracking by GPS receivers spread over a wide region during the period of GPS "campaigns". Little interest was shown in these alternative datums until:

- The network of tracking stations evolved into a *global one*, rather than being concentrated about the region of interest of the GPS campaign.
- The network of tracking stations was maintained on a *permanent basis*, rather than operating intermittently.
- The scientific community initiated a *project to define and maintain a datum at the highest level of accuracy*.
- The policy of Selective Availability (§2.4) was announced, whereby the WGS84 datum as realised by the contents of the GPS Navigation Message could be corrupted.

In 1991, the International Association of Geodesy decided to establish the International GPS Service for Geodynamics (IGS) to promote and support activities such as the maintenance of a permanent network of GPS tracking stations, and the continuous computation of the satellite orbits and ground station coordinates (DIXON, 1995; ZUMBERGE et al, 1995). Both of these were preconditions to the definition and maintenance of a new satellite datum independently of the DMA network (used to maintain the WGS datum) and the Control Segment monitor station network used to provide the data for the operational computation of the GPS broadcast ephemerides. After a test campaign in 1992, routine activities commenced at the beginning of 1994. The network now consists of about 40 core tracking stations located around the world, supplemented by more than 100 other stations (some continuously operating, others only intermittently). The precise orbits of the GPS satellites are available from the IGS with several days delay. See §12.2 for further details concerning the IGS.

The definition of the reference system in which the coordinates of the tracking stations are expressed and periodically redetermined is the responsibility of the International Earth Rotation Service (IERS). The reference system is now known as the **International Terrestrial Reference System** (ITRS), and its definition and maintenance is dependent on a suitable combination of Satellite Laser Ranging, Very Long Baseline Interferometry and GPS coordinate results (BOUCHER & ALTAMIMI, 1996). (Increasingly it is the GPS system that is providing most of the data.) Each year a new combination of precise tracking results is performed, and the resulting new coordinates of SLR, VLBI and GPS tracking stations constitutes a new **International Terrestrial Reference Frame** (ITRF) or "ITRF datum" which is referred to as "ITRFxx", where "xx" is the year. A further characteristic that sets the ITRS series of datums apart from the WGS, is that the definition not only consists of the station coordinates, but also their *velocities* due to continental and regional tectonic motion. Hence, it is possible to determine station coordinates within the datum, say ITRF92, at some "epoch" such as 1995, by applying the velocity information and predicting the coordinates of the station at any time into the future (or the past). **The WGS84(G730) reference system is identical to that of ITRF91 at epoch 1994.0.**

Such ITRS datums, initially dedicated to geodynamical applications requiring the highest possible precision, have been used increasingly as the fundamental basis for the redefinition of many nations' geodetic datums. For example, the new Australian datum, known as the Geocentric Datum of Australia (MANNING & HARVEY, 1994), is defined as ITRF92 at epoch 1994.0 (§12.1). Of course other countries are free to chose any of the ITRS datums (it is usually the latest), and define any epoch for their national datum (the year of GPS survey, or some reference date in the future, such as the year 2000). Only if both the ITRS datum (ITRFxx) and epoch are the same, can it be claimed that two countries have the same geodetic datum. *Note, the recent redefinition of the WGS84 datum was made in order to bring it into line with this new international approach.*

2.1.5 TIME SYSTEMS - SOME DEFINITIONS

Time Scales

In order to appreciate the role of time in GPS data analysis it is necessary to review briefly the various *time systems* involved, and their associated *time scales*. Some of these definitions are standard and inherent to all space positioning technologies, while others are particular to the GPS system. In general there are three different time systems that are used in space geodesy (KING et al., 1987; LANGLEY, 1991d; SEEBER, 1993):

☞ Dynamical time

☞ Atomic time

☞ Sidereal time

Dynamical Time is the uniform time scale which governs the motions of bodies in a gravitational field: that is, the independent argument in the Equations of Motion for a body according to some particular gravitational theory, such as Newtonian Mechanics or General Relativity. **Atomic Time** is time defined by atomic clocks, and is the basis of a uniform time scale on the earth. **Sidereal Time** is measured by the earth's rotation about its axis, and although sidereal time was once used as a measure of time it is much too irregular by today's standards. Rather, it is a measure of the *angular position* of a site on the earth with respect to a celestial body (though in keeping with traditional practice, its units are seconds of time rather than seconds of arc). (When the celestial body is the sun, the time scale can also be referred to as Solar Time.) Within each of these broad categories there are specific *measures* of time, or time scales, that are commonly used in space geodesy and astronomy.

Some time scales have a special importance because they provide the "benchmark" or reference scale within a particular time system. This often occurs by international convention. Often, however, the time scales to which we have access are merely *realisations* of the "true" or definitive reference time scale (or scales) associated with each time system.

Dynamical Time

Dynamical time is required to describe the motion of bodies in a particular reference frame and according to a particular gravitational theory. Today, General Relativity and an inertial (non-accelerating) reference frame are fundamental concepts. The most nearly inertial reference frame to which we have access through gravitational theory has its origin located at the centre-of-mass of the solar system (the barycentre). Dynamical time measured in this system is called **Barycentric Dynamical Time** (TDB -- the abbreviation for this and most other time scales reflects the French order of the words). A clock fixed on the earth will exhibit periodic variations as large as 1.6 milliseconds with respect to TDB due to the motion of the earth in the sun's gravitational field. However, in describing the orbital motion of near-earth satellites we need not use TDB, nor account for these relativistic variations, since both the satellite and the earth itself are subject to essentially the same perturbations.

For satellite orbit computations it is common to use **Terrestrial Dynamical Time** (TDT), which represents a uniform time scale for motion within the earth's gravity field and which has the same rate as that of an atomic clock on the earth, and is in fact defined by that rate (see below). In the terminology of General Relativity, TDB corresponds to **Coordinate Time**, and TDT to **Proper Time**. The predecessor of TDB was known as **Ephemeris Time** (ET).

Atomic Time

The fundamental time scale for all the earth's time-keeping is **International Atomic Time** (TAI). It results from analyses by the Bureau International des Poids et Mesures (BIPM) in Sévres, France, of data from atomic frequency standards (atomic "clocks") in many countries. (Prior to 1 January, 1988, this function was carried out by the Bureau International de l'Heure.) TAI is a continuous time scale and serves as the *practical definition of TDT*, being related to it by:

$$\text{TDT} = \text{TAI} + 32.184 \text{ seconds} \tag{2.1-1}$$

The fundamental unit of TAI (and therefore TDT) is the SI second, defined as "the duration of 9192631770 periods of the radiation corresponding to the transition between two hyperfine levels of the ground state of the cesium 133 atom". The SI day is defined as 86400 seconds and the Julian Century as 36525 days.

Because TAI is a continuous time scale, it has one fundamental problem in practical use: the earth's rotation with respect to the sun is slowing down by a variable amount which averages, at present, about 1 second per year. Thus TAI would eventually become inconveniently out of synchronisation with the solar day. This problem has been overcome by introducing **Coordinated Universal Time** (UTC), which runs at the same rate as TAI, but is incremented by 1 second jumps (so-called "leap seconds") when necessary, normally at the end of June or December of each year. *During the period mid-1994 to the end of 1995, one needed to add 29 seconds to UTC clock readings to obtain time expressed in the TAI scale.*

The time signals broadcast by the GPS satellites are synchronised with atomic clocks at the GPS Master Control Station, in Colorado Springs, Colorado. These clocks define **GPS Time** (GPST), and are in turn periodically compared with UTC, as realised by the U.S. Naval Observatory in Washington D.C. GPST was set to UTC at 0hr on 6 January, 1980, and is *not incremented by leap seconds*. As a result there will be integer-second differences between the two time scales. For example, in December 1994 clocks running on GPST were offset from UTC by 10 seconds. There is therefore a *constant offset of 19 seconds* between the GPST and TAI time scales:

$$\text{GPST} + 19 \text{ seconds} = \text{TAI} \tag{2.1-2}$$

Solar and Sidereal Time

A measure of earth rotation is the angle between a particular reference meridian of longitude (preferably the Greenwich meridian) and the meridian of a celestial body. The most common form of solar time is **Universal Time** (UT) (not to be confused with UTC, which is an atomic time scale). UT is defined by the Greenwich hour angle (augmented by 12 hours) of a *fictitious sun* uniformly orbiting in the equatorial plane. However, the scale is not uniform because of oscillations of the earth's rotational axis. UT corrected for *polar motion* (§1.2) is denoted by **UT1**, and is otherwise known as **Greenwich Mean Time** (GMT). The precise definition of UT1 is complicated because of the motion both of the celestial equator and the earth's orbital plane with respect to inertial space, and the irregularity of the earth's polar motion. UT1 is derived from the analysis of observations carried out by the IERS, and can be reconstructed from published corrections (ΔUT_1) to UTC:

$$UT1 = UTC + \Delta UT_1 \tag{2.1-3}$$

A measure of sidereal time is **Greenwich Apparent Sidereal Time** (GAST), defined by the Greenwich hour angle of the intersection of the earth's equator and the plane of its orbit on the Celestial Sphere (the vernal equinox). Taking the mean equinox as the reference leads to Greenwich Mean Sidereal Time (GMST). The conversion between mean solar time corrected for polar motion (UT1) and GAST is through the following relation:

$$\theta_g = 1.0027379093.UT1 + \theta_o + \Delta\psi.\cos\varepsilon \tag{2.1-4}$$

where $\Delta\psi$ is the nutation in longitude, ε is the obliquity of the ecliptic and θ_o represents the sidereal time at Greenwich midnight (0hr UT). The omission of the last term in eqn (2.1-4) permits the GMST to be determined. θ_o is represented by a time series:

$$\theta_o = 24110.54841^s + 8640184.812866^s.T_o + 0.093104^s.T_o^2 - 6.2^s.10^{-6}.T_o^3 \tag{2.1-5}$$

where T_o represents the time span expressed in Julian centuries (of 36525 days of 86400 SI seconds) between the reference epoch J2000.0 and the day of interest (at 0hr UT).

Figure 2.1-3 illustrates the relationship between the various time scales discussed above. The vertical axis indicates the relative offsets of the origins of the time scales, and the slope of the lines indicate their drift. Note that with the exception of UT1 (or GAST) all time scales (nominally) have zero drift as defined by TAI.

Calendar Definitions

The **Julian Date** (JD) defines the number mean solar days (each of which is 86400 SI second in length) elapsed since the epoch 1.5 (midday) January, 4713 B.C. The **Modified Julian Date** (MJD) is obtained by subtracting 2400000.5 from the JD. *(MJD therefore commences at midnight.)* The standard epoch for GPS Time (0hr 6 January, 1980) is therefore MJD44244.0.

The date conversions described below are taken from HOFMANN-WELLENHOF et al. (1994), and are valid for the period March 1900 to February 2100. The JD can be computed from the year number Y (a full four digit integer), integer month number M, integer day number D, and the real-valued time in hours H:

$$JD = Int[\, 365.25y \,] + Int[\, 30.6001(m+1) \,] + D + H / 24 + 1720981.5 \quad (2.1\text{-}6)$$

where Int denotes the integer part of the number, and:

$$\begin{array}{llll} y = Y - 1 & \text{and} & m = M + 12 & \text{if } M \leq 2 \\ y = Y & \text{and} & m = M & \text{if } M > 2 \end{array}$$

The reverse conversion is carried out stepwise by first defining the quantities:

$$\begin{aligned} b &= Int[\, JD + 0.5 \,] + 1537 \\ c &= Int[\, (b - 122.1) / 365.25 \,] \\ d &= Int[\, 365.25c \,] \\ e &= Int[\, (b - d) / 30.6001 \,] \end{aligned} \quad (2.1\text{-}7)$$

The date parameters are then obtained:

$$\begin{aligned} D &= b - d - Int[\, 30.6001e \,] + Frac[\, JD + 0.5 \,] \\ M &= e - 1 - 12.Int[\, e / 14 \,] \\ Y &= c - 4715 - Int[\, (7 + M) / 10 \,] \end{aligned} \quad (2.1\text{-}8)$$

where Frac denotes the fractional part of a number.

A further useful relation is between JD and the **GPS week number**:

$$GPSWeek = Int[\, (JD - 2444244.5) / 7 \,] \quad (2.1\text{-}9)$$

The GPS week starts on Saturday midnight (Sunday morning), and runs for 604800 seconds.

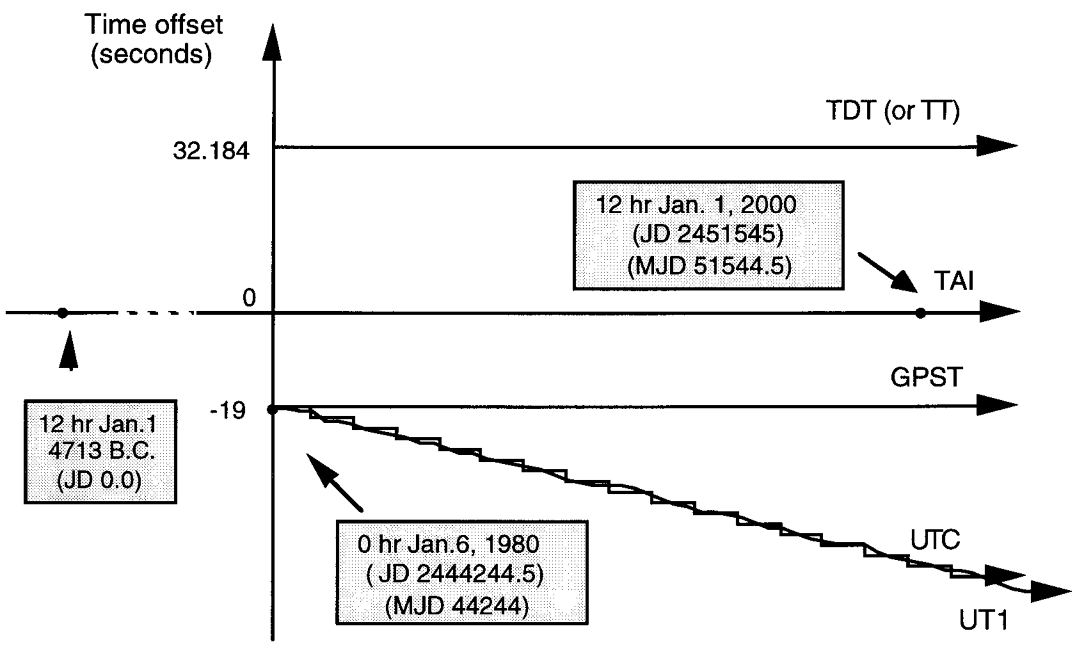


Figure 2.1-3. Time scale relationships.

2.2

THE GPS SYSTEM

The GPS system can be conveniently partitioned into *three segments* (Figure 2.2-1):

☞ **The SPACE SEGMENT: comprising the satellites themselves, transmitting the signals necessary for the system to operate.**

☞ **The CONTROL SEGMENT: the ground facilities carrying out the task of satellite tracking, orbit computations, telemetry and supervision necessary for the daily management of the Space Segment.**

☞ **The USER SEGMENT: the entire spectrum of applications equipment and computational techniques that provide the users with the position results.**

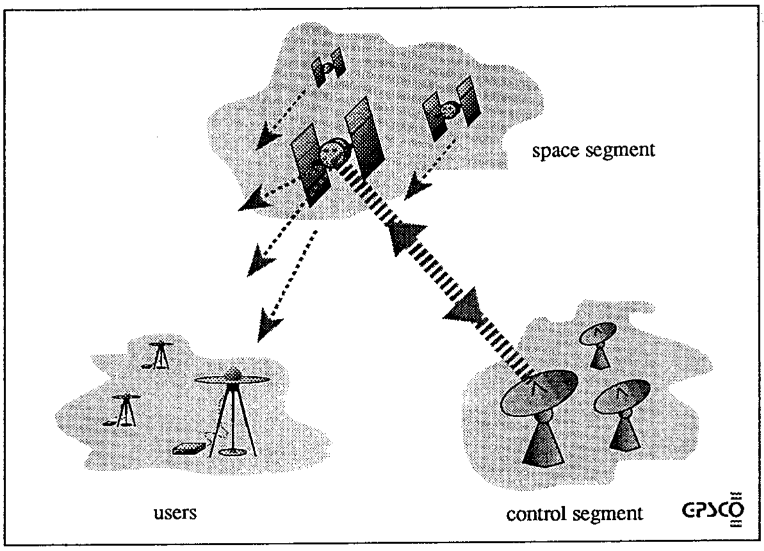


Figure 2.2-1. The GPS system elements.

Each of these segments is briefly described in this chapter. The satellite signal structure is discussed in greater detail in §3.1. A excellent "engineering" reference book on the GPS technology is KAPLAN (1996).

2.2.1 SYSTEM COMPONENTS

Overall operation of the Control and Space Segments is the responsibility of the U.S. Air Force Space Command, Second Space Wing, Satellite Control Squadron at the Falcon Air Force Base, Colorado.

The Space Segment

The basic functions of the satellites are to (PARKINSON et al, 1995):

- **Receive and store data** transmitted by the Control Segment.
- **Maintain accurate time** by means of several onboard atomic frequency standards.
- **Transmit information and signals to users** on one or both L-band frequencies.

The Block I satellites were built by the Rockwell International Corporation. The follow-on operational satellites are separated into four series: II, IIA, IIR and IIF. The Block II and IIA satellites were also built by Rockwell International and are massive, weighing over 900kg each. 28 Block II/IIA satellites have been built, of which 26 have been launched and activated, with the remainder to be launched between now and the end of the century. The replacement Block IIR satellites are being built by the General Electric Corporation (now the Lockheed-Martin Corporation), and the entire constellation of 21 satellites will be launched from 1997 through to 2001. The Block IIF series are still in the design phase and may, for example, incorporate an additional civilian transmission frequency (LACHAPELLE, 1995). Rockwell International was awarded the contract to build the satellites, which are to be launched between 2001 and 2015.

The Block II satellites are maintained on a once per day upload, while the Block IIA satellite incorporate features that permit the navigation message to be up to 180 days in length (compared to the Block II's maximum of 14 days). The Block IIR satellites incorporate the autonomous (180 day) navigation feature based on crosslink ranging between satellites. The Block IIF will have the same capability.

The design life of the Block II/IIA GPS satellites is 7.5 years (compared with the intended life of Block I satellites of 5 years, and the planned life of 10 years for the Block IIR satellites). The status of the present constellation is shown in Table 2.2-1. Details include launch and official commissioning date, the orbital plane and position within the plane, and the satellite I.D. number(s). There are a number of satellite numbering conventions in use, the two most widely used are the:

- **NAVSTAR (or SVN) number**. The Block II satellites are numbered SVN 13 through SVN 21, the Block IIA satellites are designated SVN 22 through SVN 40, and the Block IIR will be designated SVN 41 and above.
- **PRN number**, being the number of the segment of the available 37 seven-day sections of the P-code pseudo-random-noise (PRN) code assigned to each satellite (§3.1).

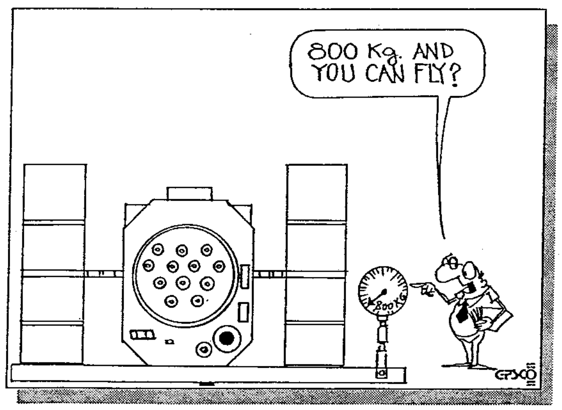


Table 2.2-1. Status of the GPS satellite constellation (November 1996).

SVN	PRN	Launch Date	Orbit Plane Position	Useable
Block I				
1	4	22-02-78		no
2	7	13-05-78		no
3	6	6-10-78		no
4	8	10-12-78		no
5	5	9-02-80		no
6	9	26-04-80		no
7		launch failure		
8	11	14-07-83		no
9	13	13-06-84		no
10	12	8-09-84		no
11	3	9-10-85		no
Block II				
14	14	14-02-89	E1	15-04-89
13	2	10-06-89	B3	10-08-89
16	16	18-08-89	E5	14-10-89
19	19	21-10-89	A4	23-11-89
17	17	11-12-89	D3	6-01-90
18	18	24-01-90	F3	14-02-90
20	20	26-03-90		no
21	21	2-08-90	E2	22-08-90
15	15	1-10-90	D2	15-10-90
Block IIA				
23	23	26-11-90	E4	10-12-90
24	24	4-07-91	D1	30-08-91
25	25	23-02-92	A2	24-03-92
28	28	10-04-92	C5	25-04-92
26	26	7-07-92	F2	23-07-92
27	27	9-09-92	A3	30-09-92
32	1	22-11-92	F1	11-12-92
29	29	18-12-92	F4	5-01-93
22	22	3-02-93	B1	4-04-93
31	31	30-03-93	C3	13-04-93
37	7	13-05-93	C4	12-06-93
39	9	26-06-93	A1	20-07-93
35	5	30-08-93	B4	28-09-93
34	4	26-10-93	D4	22-11-93
36	6	10-03-94	C1	28-03-94
33	3	28-03-96	C2	9-04-96
40	10	16-07-96	E3	15-08-96
30	30	12-09-96	B2	1-10-96

The orbital planes are identified by the letters A through F, and the satellite "slots" are numbered 1 to 4. Note that according to the Table there are two spare satellites in slots 5 and 6.

The Block I satellite orbits are different to those of the Block II satellites. Most significantly they have been launched into orbits of different inclination: 63° for the Block I satellites, 55° for the Block II satellites. There are four series of operational satellites: II, IIA, IIR and IIF (none launched yet of the last two series).

The Control Segment

The Control Segment consists of facilities required for satellite health monitoring, telemetry, tracking, command and control, ephemeris computations and uplinking. There are five ground facility stations: Hawaii, Colorado Springs, Ascension Is., Diego Garcia and Kwajalein (Figure 2.2-2). They perform the following functions:

- All five stations are **monitor stations**, tracking the satellites and sending the tracking data to the Master Control Station.
- Falcon Air Force Base, Colorado Springs, is the site of the **Master Control Station** (MCS), where the tracking data is processed in order to compute the satellite ephemerides and satellite clock corrections. It is also the station that initiates all operations of the Space Segment, such as satellite manoeuvring, signal encryption, satellite clock keeping, etc. The MCS is managed by the 2nd Operations Squadron, 50th Space Wing, U.S. Air Force.
- Three of the stations (Ascension Is., Diego Garcia, and Kwajalein) are **upload stations** allowing for the uplink of data to the satellites. The data includes the ephemerides and clock correction information transmitted within the Navigation Message, as well as command telemetry from the MCS.

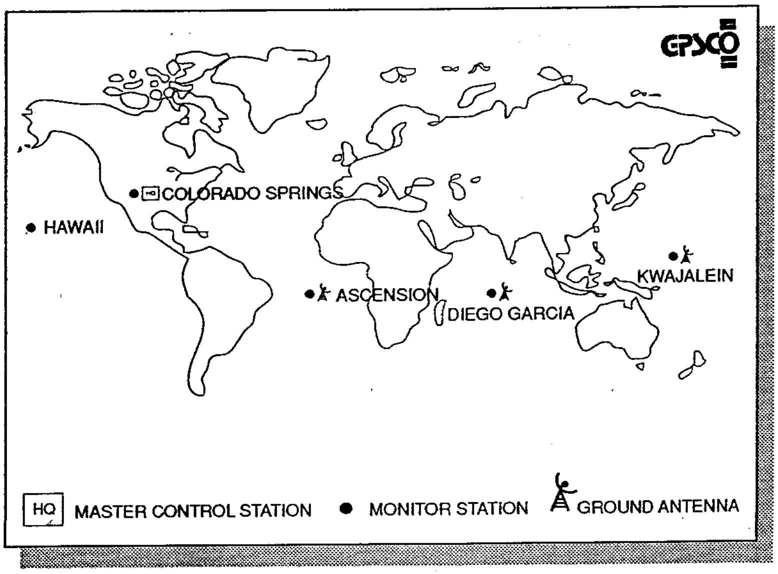


Figure 2.2-2. The GPS Control Segment.

Each of the upload stations can view all the satellites once a day. All satellites can therefore be viewed by an upload station three times a day. New Navigation Messages and command telemetry can be transmitted to GPS satellites approximately every 8 hours, if necessary. At present the upload rate is once (and sometimes twice) per day.

An important, latent function of the Control Segment is to maintain the WGS84 reference system (§2.1). This reference system is accessible to the GPS user via the satellite ephemerides computed by the MCS, from data collected at the monitor stations. If any organisation were to compute its own satellite orbits (for example, from the post-processing GPS tracking data acquired by its own ground receiver network), the resulting reference system would be defined by the datum (or fixed) tracking stations in that system. This may not be the same as WGS84, though generally it is very close to it. The basis of the International GPS Service for Geodynamics (IGS) post-processed ephemerides is the International Terrestrial

Reference System (ITRS) series of International Terrestrial Reference Frame datums, which is very close to the WGS84 system (to within one metre) (§6.2 and §12.2).

As the GPS system matures the satellites will operate with greater independence from the ground-based Control Segment, without significant degradation in performance. The Block IIR and IIF satellites will have a crosslink enabling between-satellite communication and ranging. The data would be processed to produce the ephemeris information within the Space Segment itself. This fact, together with the planned expansion of the Control Segment to 11 monitor stations and more frequent satellite uploads, should lead to improved GP performance for the military users.

The User Segment

GPS user equipment has undergone an extensive program of development, both in the military and civilian area. In this context, GPS "equipment" refers to the combination of:

- **hardware** (signal tracking and measurement),
- **software** (positioning algorithms, user interface), and
- **operational procedures** (influenced by accuracy, functionality, etc.).

There is a wide variety of GPS applications (§2.3), which is matched by a similar diversity of user equipment. Nevertheless, the most basic classification system is provided by the type of observable that is tracked: (a) civilian navigation receivers using the satellite ranging code on the L1 frequency, (b) military navigation receivers using the satellite ranging codes on both L-band frequencies, (c) single frequency (L1) carrier phase tracking receivers, and (d) dual-frequency carrier phase tracking receivers. §3.1 describes how pseudo-range and carrier phase measurements are made. GPS instrumentation is discussed in §4.1, §4.2 and §4.3. §6.2 describes the biases and errors that affect all GPS measurements.

While military R&D programs have concentrated on achieving a high degree of miniaturisation, modularisation and reliability, the civilian user equipment manufacturers have, in addition, sought to bring down costs and to develop features that can enhance the capabilities of a system that was not optimised for many groups of civilian users, principally in terms of accuracy and reliability. *This is particularly true of the survey user seeking levels of accuracy several orders of magnitude higher than that of the navigation user.* It is fair to say that GPS user technology is being driven by the precise positioning market -- in much the same way that automotive technology benefits from car racing.

Yet another major influence on the development of GPS equipment has been the increasing variety of civilian applications (§2.3). For although there may exist a similar positioning accuracy requirement across many user applications, to address a particular application in the most satisfactory manner, a specific combination of hardware and software features is often required. As the system matures it is likely that several trends in instrument development will emerge, some designed for high precision, while others will emphasise other attributes. There are at present over 100 manufacturers of GPS instruments of varying kinds.

It is important to emphasise that *GPS was designed for navigation, and not for surveying.* Navigation requires real-time point positioning at the several dekametre (tens of metres) level, while surveying aims to achieve a high relative accuracy at the centimetre level, over distances of a few tens of kilometres. *Special observation, measurement and data reduction procedures have been developed to enable these surveying accuracies to be achieved.*

2.2.2 THE GPS SATELLITE CONSTELLATION

There are essentially three constellations at present (Table 2.2-1):

- The **Block I satellites** were the *experimental satellites* launched between 1978 and 1985 used to test the system (Figure 2.2-3). Eleven were launched, none of which are functioning any longer.
- The **Block II satellites** comprise the first nine spacecraft of the *operational series.*
- The **Block IIA satellites** comprise the second 19 spacecraft of the *operational series.*

There are 26 Block II and IIA satellites in orbit (Figure 2.2-3). (In essence 21 satellites are placed in a particular orbit pattern, while additional active satellites are deployed as "spares" so that they can quickly be moved to replace any of the 21 primary satellites when they become unserviceable.)

The Block I satellites are not identical to the Block II/IIA satellites (HOFMANN-WELLENHOF et al., 1994), however the most important difference is that they have been launched into orbits of different inclination: 63° for the Block I satellites, 55° for the Block II/IIA satellites.

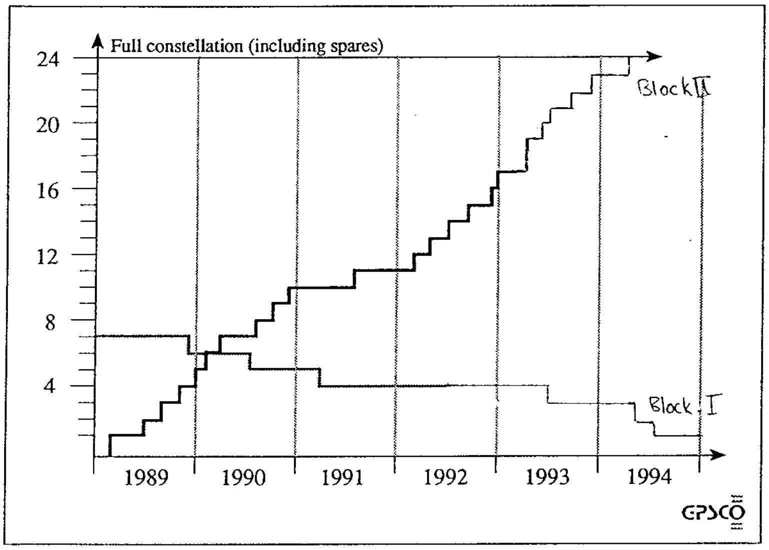


Figure 2.2-3. The GPS satellite system deployment schedule.

The operational system was to be fully deployed by the late 1980's. However, a number of factors, the major one being the Space Shuttle Challenger disaster (28 January, 1986), have meant that the full GPS system was only deployed in 1994 (24 Block II/IIA satellites). 24 orbiting satellites is considered sufficient to ensure that there will always be *at least 4 satellites visible,* at all sites on the globe, and at all times. Figure 2.2-4 illustrates the visibility of satellites through the day for an observer in Sydney, Australia. Note that the graph illustrates visibility down to the local horizon. *At certain times of the day there are up to 12 satellites visible simultaneously.*

Initial Operational Capability (IOC) was declared in July, 1993 (24 Block I, II, or IIA satellites), and **Full Operational Capability** (FOC) was declared on 17 July, 1995, (24 Block II or IIA satellites operating satisfactorily). The U.S. DoD will guarantee 24 satellite coverage 70% of the time, and 21 satellite coverage approximately 98% of the time. There could therefore be occasional periods of degraded satellite coverage, even under FOC conditions.

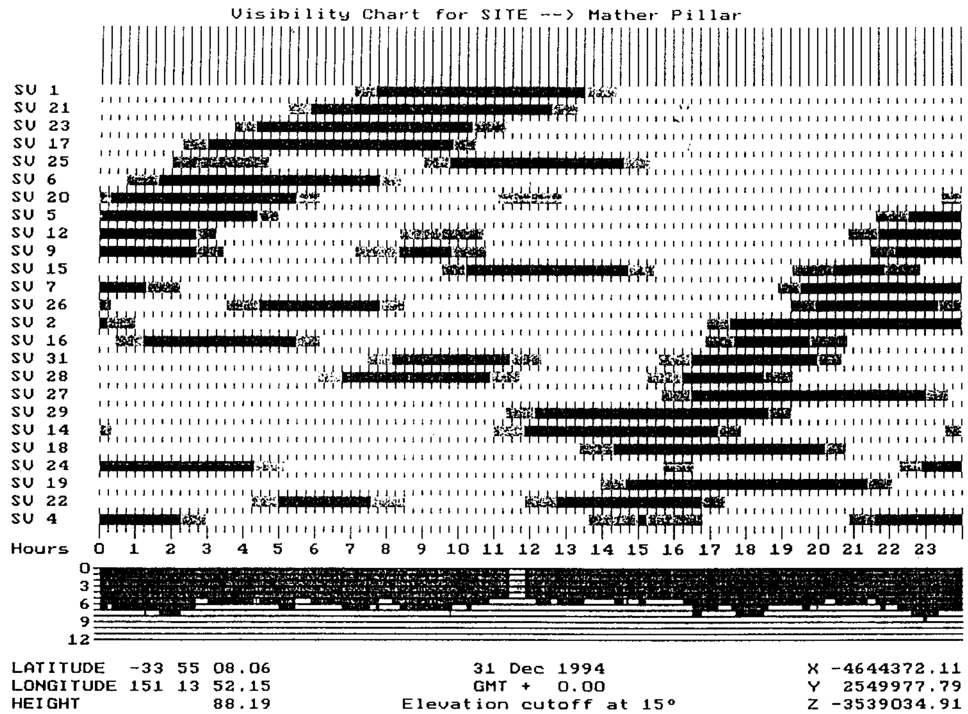


Figure 2.2-4. Typical 24hr GPS coverage at Sydney, Australia (December, 1994).

GPS Orbit Characteristics

> **The Block II/IIA GPS satellites have been deployed (Figure 2.2-5):**
>
> ☞ **In 6 nearly circular orbital planes.**
>
> ☞ **With 4 satellites equally spaced within the plane (the in-orbit spares are spread across different orbital planes).**
>
> ☞ **In orbital planes at an inclination of 55°.**
>
> ☞ **At altitudes of approximately 20,200km above the earth.**

As the GPS satellites are in nearly circular orbits, at an altitude of approximately 20,200km above the earth, this has a number of immediate effects which make the prediction of satellite location comparatively easy:

- Their orbital period is approximately 11hrs 58mins, so that each satellite makes two revolutions in one sidereal day (the period taken for the earth to complete one rotation about its axis with respect to the stars).
- At the end of a sidereal day (approximately 23hrs 56mins in length) the satellites are again over the same position on earth.

- Reckoned in terms of a solar day (24hrs in length), the satellites are in the same position in the sky about four (4) minutes earlier each day.
- The orbit groundtrack approximately repeats each day, except that there is a small drift of the orbital plane to the west (-0.03° per day).

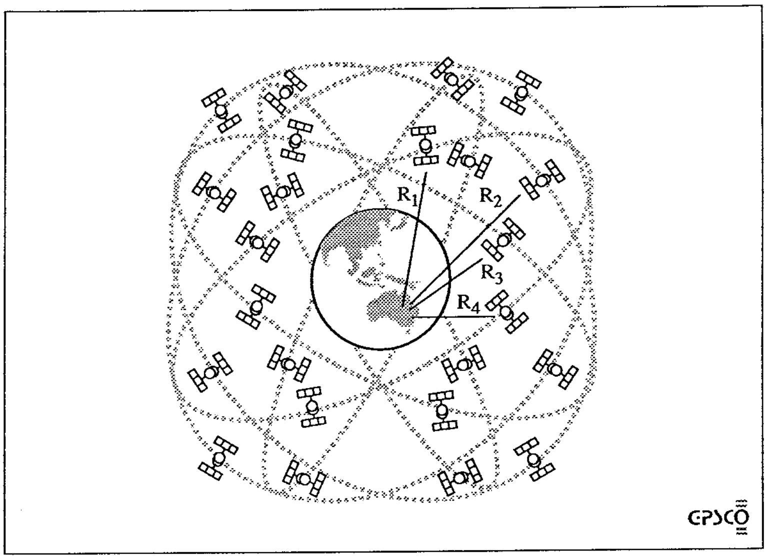


Figure 2.2-5. The GPS constellation "birdcage".

The Block II/IIA satellites have been deployed in six orbital planes at 60° intervals about the equator, with each containing 3 or 4 of the primary satellites equally spaced in the orbital plane (Table 2.2-1). The orbital planes are at an inclination of 55° relative to the equatorial plane. The 3 active spare satellites are spread evenly within the primary 21 satellite configuration. Figure 2.2-6 shows the ground tracks over one hour of a number of Block II/IIA satellites, together the visibility circles at three locations on the earth. *The visibility circles are for a minimum elevation above the local horizon of 15°.*

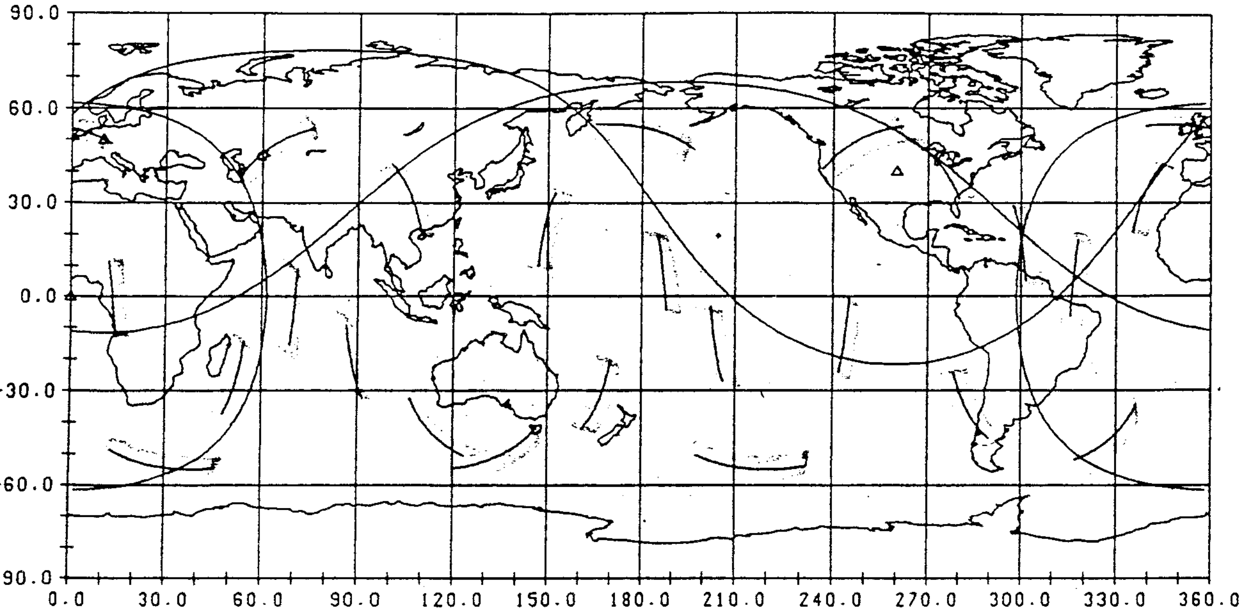


Figure 2.2-6. Block II/IIA GPS constellation over 1 hour, with 15° visibility circles for sites in North America, Europe and at the intersection of the Equator and the Greenwich meridian.

GPS Satellite Coverage

The Block II/IIA GPS satellite coverage provides:

- ☞ **From 4 to over 12 satellites above an observer's horizon.**
- ☞ **Satellites will be visible for many hours above an observer's horizon.**
- ☞ **There may be short periods of degraded satellite geometry which will affect navigation users, though not surveyors.**

The satellites are in near circular orbits at an altitude more than three times the earth's radius, a satellite will be visible above an observer's horizon for many hours, perhaps up to 6-7 hours. Figure 2.2-7 illustrates the change of range, range-rate and other geometric quantities for a GPS satellite passing overhead. At various times of the day, and at various locations on the surface of the earth, the number of satellites and the length of time they are visible above an observer's horizon will vary. Figure 2.2-8 shows the coverage at four locations on the globe for the December 1993 constellation of 26 satellites. The satellite availability is expressed in the form of satellite visibility graphs (Figures 2.2-8a and 2.2-8b), and "skyplots" showing the satellite tracks on a polar projection relative to a site's zenith (Figure 2.2-8c).

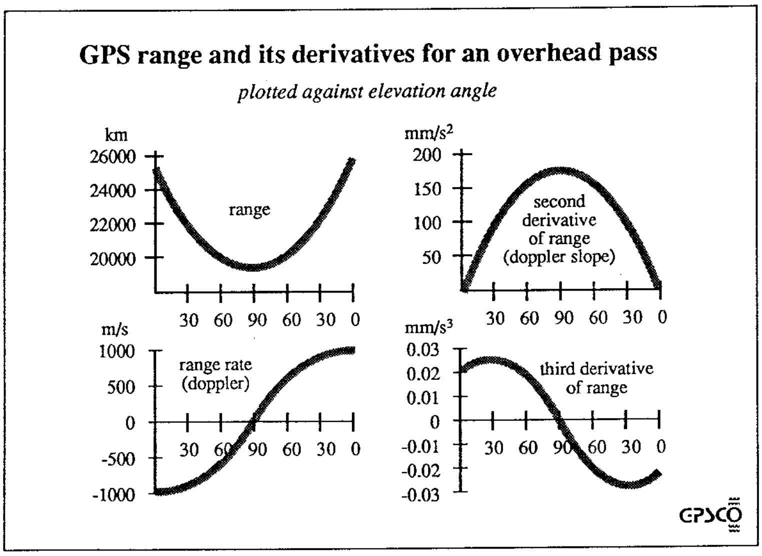
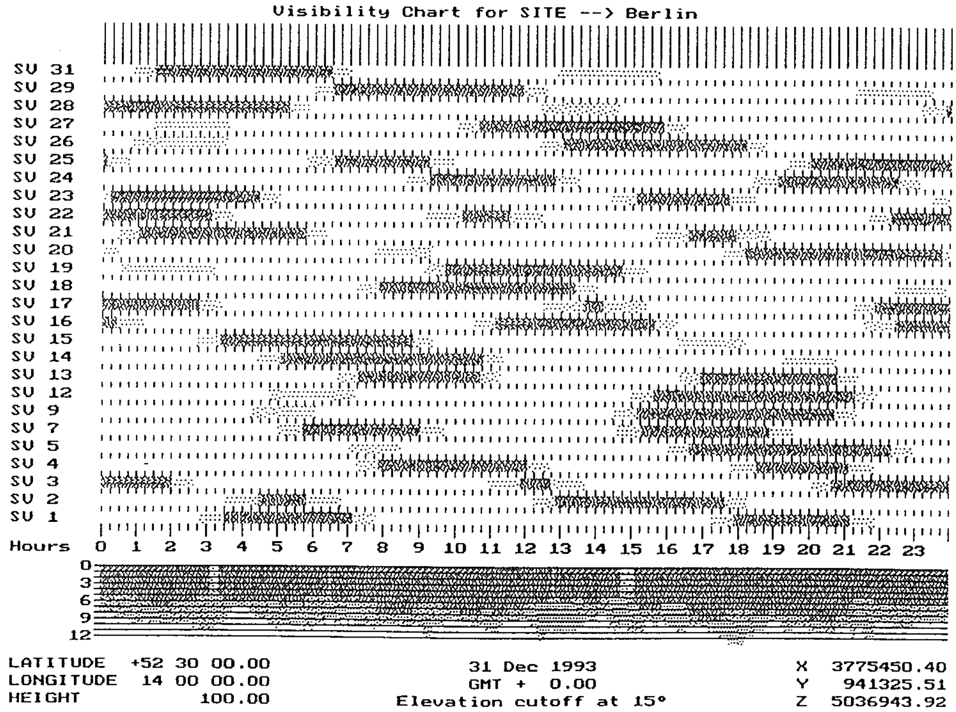
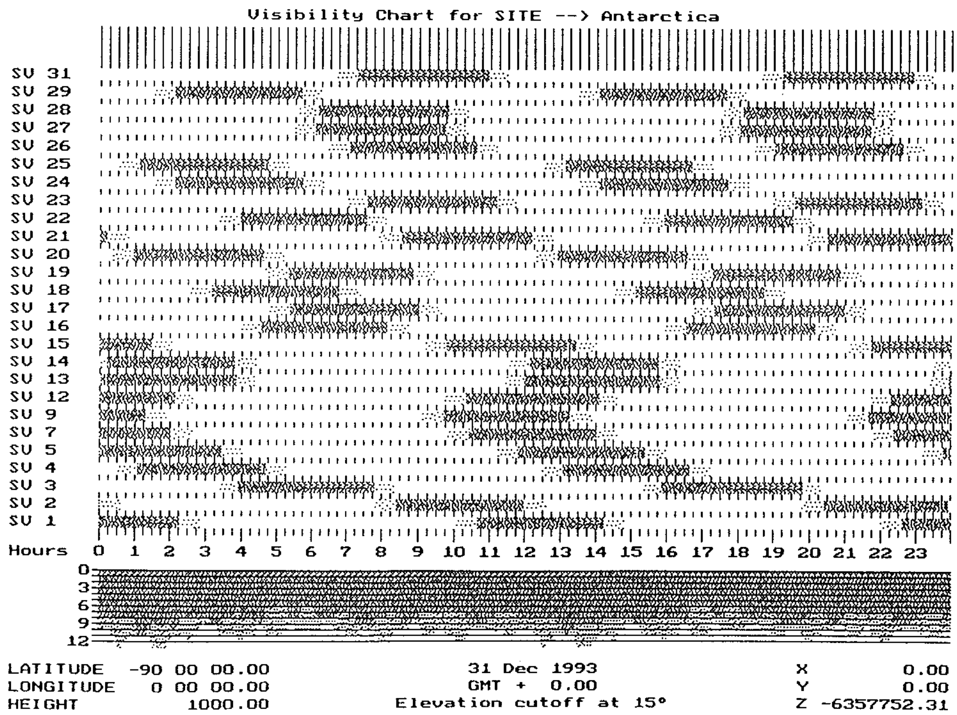


Figure 2.2-7. Effect of GPS geometry on range, range-rate, etc.

With the Block II/IIA constellation, there are always at least four satellites visible, no matter where the observer is located, and sometimes more than ten. There may, however, be times when some satellites are not healthy and the satellite configuration results in rather poor geometry for navigation users (see §1.4 for a discussion of GDOP). These periods of degraded geometry are known as **outages**. To decrease the occurrences of outages, and to improve the system reliability, a number of technological solutions have been proposed, including the integration of GPS and GLONASS (the Russian equivalent of the GPS satellite system) so that receivers can track both types of satellites. This will significantly increase the number of mutually visible satellites. See Table 2.2-2 for a comparison of the two systems.

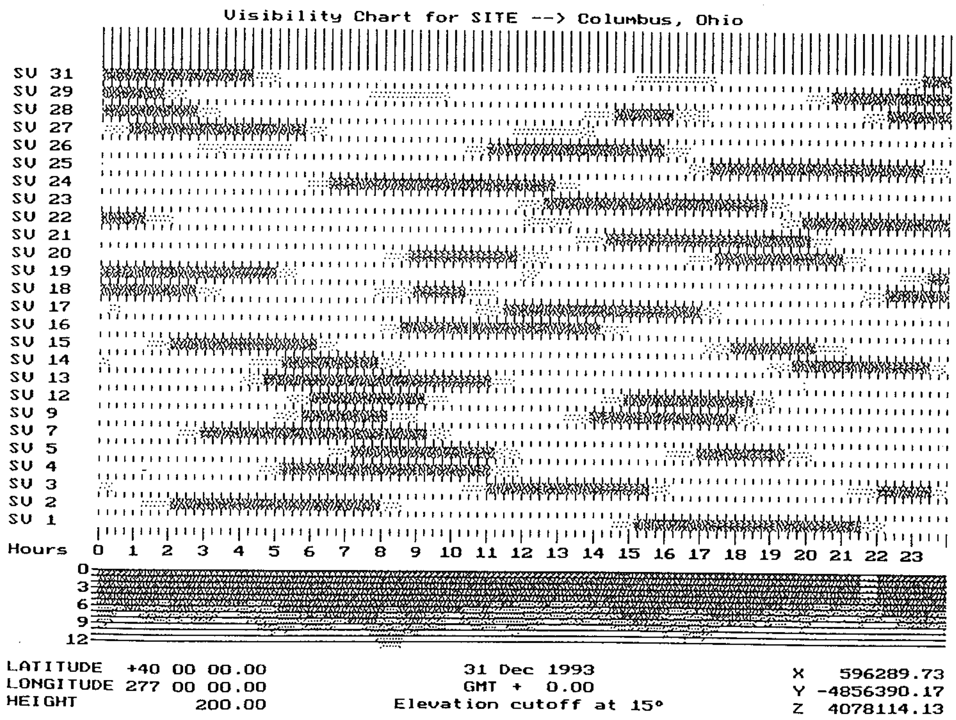


Mid-Europe Site

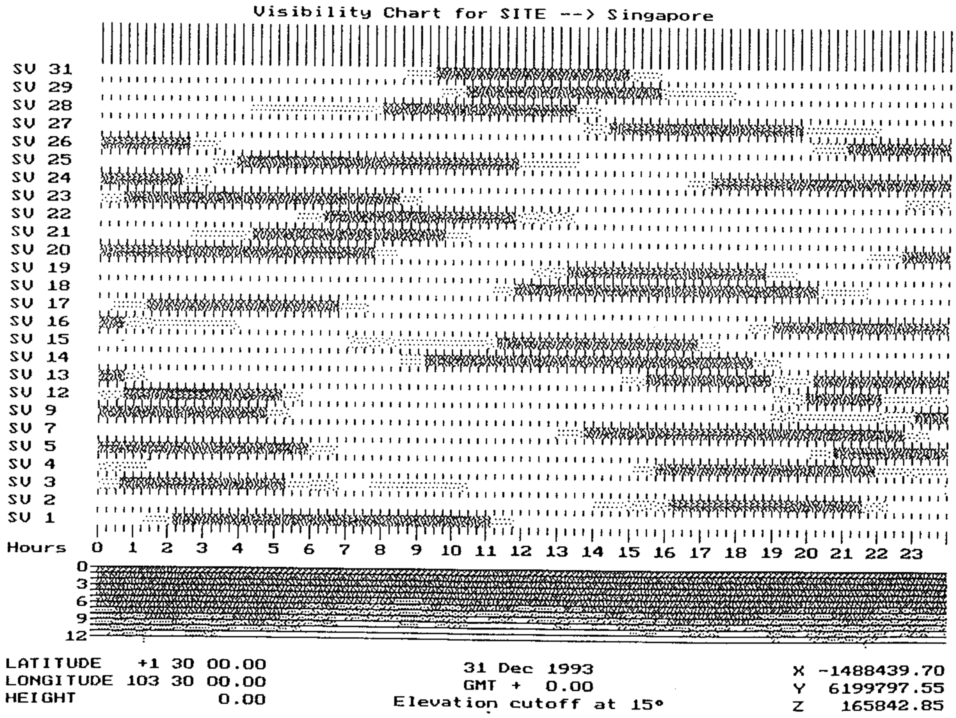


Antarctica Site

Figure 2.2-8a. December 1993 GPS satellite availability.



Mid-U.S.A Site



Equatorial Site

Figure 2.2-8b. December 1993 GPS satellite availability.

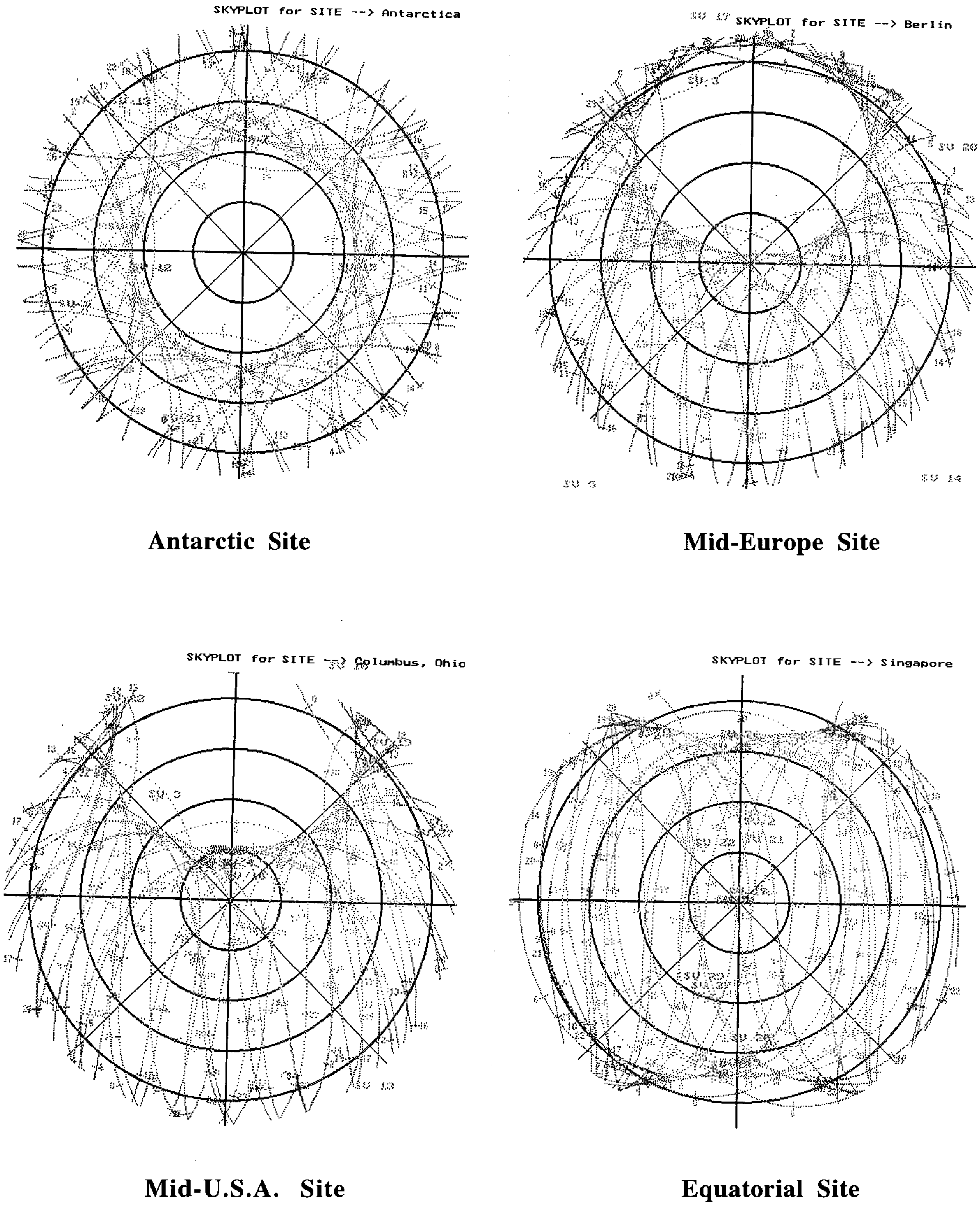


Figure 2.2-8c. Skyplots for December 1993 GPS satellite constellation.

Table 2.2-2. Comparing GPS and GLONASS.
(IVANOV & SALISTCHEV, 1991)

Parameter	GLONASS	GPS
Ephemeris information presentation method	9 parameters of s/c motion in the geocentric rectangular rotated coordinate system	Interpolation coefficients of satellite orbits
Geodesic coordinate system	SGS 85	WGS 84
Referencing of the ranging signal phases	To the timer of GLONASS system	To the timer of GPS system
System time corrections relative to the universal coordinated time (UTC)	UTC [SU]	UTC [USNO]
Duration of the almanac transmission	2.5 min	12.5 min
Number of satellites in the full operational system	21 + 3 spares	21 + 3 spares
Number of orbital planes	3	6
Inclination	64.8°	55°
Orbit altitude	19,100 km	20,180 km
Orbital period	11 h 15 min	12 h
Satellite signal division method	Frequency division	Code division
Frequency band allocated	1602.5625–1615.5 ± 0.5 MHz	1575.42 ± 1 MHz
Type of ranging code	PRN–sequence of maximal length	Gold code
Number of code elements	511	1023
Timing frequency of code	0.511 MHz	1.023 MHz
Crosstalk level between two neighboring channels	−48 dB	−21.6 dB
Synchrocode repetition period	2 sec	6 sec
Symbol number in the synchrocode	30	8

GPS Satellite Health Monitoring

There are a number of sources of information on GPS satellite health, and the general status of the system (including notification of satellite manoeuvres, status of Selective Availability, launch program, etc.). They include:

- "Health status" flags transmitted as part of the Navigation Message (§3.3). These are monitored automatically by the GPS receiver, and the necessary action may be initiated. Note, however, that although a satellite may be "unhealthy" for navigation purposes, it may still be suitable for phase tracking and post-processing in the GPS survey mode.
- Several Bulletin Board Services, and other electronic GPS information sources on the Internet (§3.4).
- Specialist communication services for certain classes of users, for example, real-time differential GPS (DGPS) users are transmitted health messages within the standard data stream, aviation users will be able to receive NANUs (Notice Advisory to Navigational Users) directly via radio links, etc. It is likely that these specialist "integrity monitoring" services will expand rapidly in the coming years and be offered to survey users as well.

2.3 APPLICATIONS OF GPS

GPS applications can be classified as follows:

- ☞ **Surveying and Mapping, on land, at sea and from the air. The applications are of relatively high accuracy, for positioning in both the stationary and moving mode. Includes geophysical and resource surveys, GIS data capture surveys, etc.**
- ☞ **Land, Sea and Air Navigation, including enroute as well as precision navigation, cargo monitoring, vehicle tracking, etc.**
- ☞ **Search and Rescue Operations, including collision avoidance and rendezvous functions.**
- ☞ **Spacecraft Operations.**
- ☞ **Military Applications.**
- ☞ **Recreational Uses, on land, at sea and in the air.**
- ☞ **Other specialised uses, such as time transfer, attitude determination, automatic operation, etc.**

Land Applications

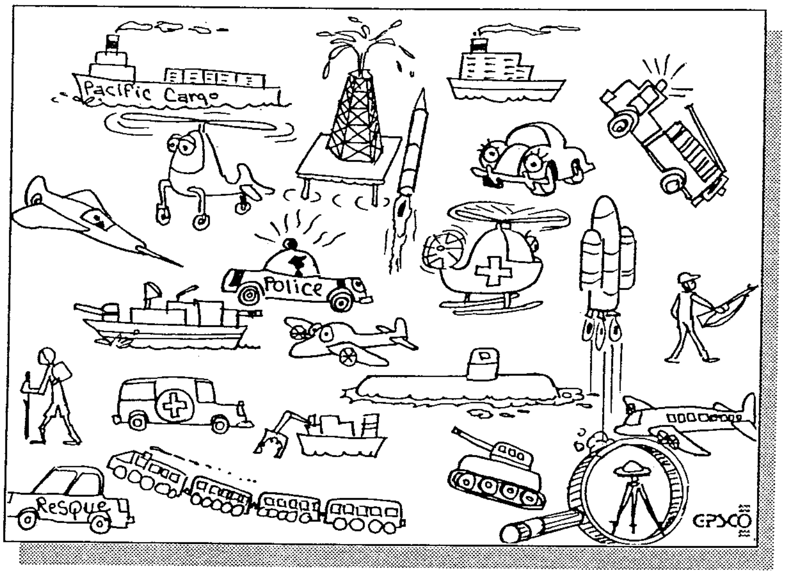
- ❑ *Surveying and mapping*, including cadastral and urban networks, data capture surveys for Geographic Information Systems (GIS), engineering surveys, photogrammetrical control (airborne and terrestrial), and geophysical resource surveys.
- ❑ *Geodetic applications*, including the establishment of control networks over regional and continental extent, height and geoid determination, precise engineering and subsidence monitoring surveys.
- ❑ *Geodynamic applications*, for measuring the relative position of a regional network at regular intervals in order to infer horizontal and vertical crustal motion.
- ❑ *Land navigation*, to support emergency vehicles (police, search & rescue, etc.) and for the monitoring of cars, taxis, dangerous and valuable cargoes, trucks and railways.
- ❑ *Transportation and communication*, to support aids for navigation for land, sea and air users, land operations taking advantage of permanent GPS stations, time transfer operations, etc.
- ❑ *Recreational uses*, for hiking, orienteering, etc.

Air Applications

- ❑ *Airport approach and landing*, as an aid for all categories of landing including full instrument landings.
- ❑ *Domestic and intercontinental enroute navigation*, including helicopter operations.
- ❑ *Air Traffic Control operations*, dynamic routing, new airport approaches.
- ❑ *Search and rescue operations*, including coordinated search operations, collision avoidance, and rendezvous.
- ❑ *Aerial photogrammetry*, including laser profiling and radar imaging.
- ❑ *Airborne geophysical surveys*, including position and attitude determination for gravimeter, magnetometer and remote sensing operations.
- ❑ *Recreational applications*, glider, ballooning, light aircraft, and parachuting operations.

Marine Applications

- ❑ *Open ocean and coastal navigation*, with or without aid of electronic charts.
- ❑ *Harbour and inland waterway navigation*, including when visibility is low.
- ❑ *Search and rescue operations*, including coordinated search operations and rendezvous.
- ❑ *Ship Monitoring Systems* for collision avoidance, remote piloting, and tracking of fishing fleets and dangerous cargoes.
- ❑ *Offshore geophysical surveys*, including seismic and gravity surveys.
- ❑ *Engineering applications* such as drill rig, pipeline and other offshore structure positioning.
- ❑ *Hydrographic surveys*, including surveys to support oceanographic research.
- ❑ *Recreational applications*, such as fishing, snorkelling, sailing.



Space Applications

- *Spacecraft launch and landing.*
- *Navigation, inflight, re-entry and rendezvous*, for earth-orbiting and interplanetary missions.
- *Orbit determination*, for many earth resource and scientific missions.
- *Sounding the atmosphere*, including GPS "meteorology", and ionospheric studies.

Military Applications

- *Enroute and low-level navigation* for tactical and strategic operations.
- *Target acquisition*, including forward observation, covert operations.
- *Photo reconnaissance and intelligence gathering.*
- *Remotely operated vehicles*, for reconnaissance and targeting.
- *Weapon guidance and control*, "smart" bombs, cruise missiles, etc.
- *Command and control*, the C^3 applications, tracking of battle elements.
- *Updating inertial navigation systems*, at sea, in the air and on land.
- *Fleet, air and land operations*, the "electronic battlefield", battle element manoeuvring.

2.3.1 GPS SATELLITE SURVEYING

Adopting the broadest definition of "GPS surveying", the following classes of surveys can be identified:

☞ **Land Survey**: applications which for the most part are associated with surveying and mapping operations.

☞ **Marine Survey**: applications related to hydrographic, oceanographic and exploration geophysics survey operations.

☞ **Airborne Survey**: applications associated with aerial mapping, scientific and exploration geophysics surveys.

What are the criteria for deciding if an application belongs to "surveying", "navigation", or "other"? This is not as easy as it may first appear. In general (and there are exceptions), a "surveying" application is a positioning task that:

- Is of *comparatively high accuracy.* This is, of course, a subjective judgement, but in general "high accuracy" is interpreted as being of an accuracy beyond that originally intended for a positioning system operated in a standard routine manner. As GPS is a navigation system which is intended to deliver only 100m absolute positioning accuracy, there is the potential for applications requiring a higher accuracy than this to be legitimately regarded as "survey" applications. However, "differential navigation", or DGPS, can deliver accuracies of several metres. Hence the accuracy threshold for "surveying" may be arbitrarily set at the sub-metre level.

- Requires the use of *unique observation procedures, measurement technologies and data analysis.* The development of specialised procedures, instrumentation and sophisticated software is the hallmark of GPS "surveying". There are two main types of observations that can be made on the GPS signals, each with very different "noise" characteristics:

measurements of phase on the L-band carrier signals have millimetre random error, while pseudo-range measurements made with the aid of the time signals modulated on the carrier waves are between 100 and 1000 times noisier.

- Is *not required "urgently"*. That is, it is an application in which if positioning information is not available in real-time, a tragedy is unlikely to occur. "Navigation" is concerned with the safe passage of vehicles, ships and aircraft.
- In general permits *post-processing of data* to obtain the highest accuracy possible.
- Has as its reason d'entre, the *production of a map,* or the establishment of a *network of coordinated points* which support the traditional functions of the surveying discipline, as well as new applications such as GIS.
- Is generally a *static positioning task.* This is, of course, not the case with marine and airborne surveying operations. (There is however a strong trend towards "kinematic" GPS surveying as described in §5.5.)

In the case of **land surveying applications**, the characteristics of GPS satellite surveying are less contentious:

(1) The points being coordinated are generally *stationary.*

(2) Depending on the accuracy sought, GPS *data are collected over some "observation session"*, ranging in length from several seconds to several hours, or even days.

(3) Restricted to *relative positioning modes of operation* (those applications that can be satisfied with GPS operated in the single point positioning mode are therefore considered to belong to the "land navigation" category).

(4) In general (depending on the accuracy sought) the measurements used for the data reduction are those made on the transmitted *L-band carrier wave*, and not on the timing signals modulated on the carrier waves -- hence the requirement for specialised hardware and software.

(5) Mostly associated with the *traditional surveying and mapping functions*, but accomplished using GPS techniques in less time, to a higher accuracy (for little extra effort) and with greater efficiency.

Land Surveying Applications for GPS

A convenient approach is to adopt an applications classification on the basis of accuracy requirements. Three classes of applications can be identified on this basis, for which a range of relative accuracies (it is assumed that single receiver point positioning is not accurate enough to satisfy these applications) ranging from low-to-moderate, 1 part in 10^4, through to the ultra-high 1 part in 10^7 or better accuracies:

- **Category A (Scientific)**: better than 1 ppm
- **Category B (Geodetic)**: 1 to 10 ppm
- **Category C (General Surveying)**: lower than 10 ppm.

Category A surveys primarily encompass those surveys undertaken for precise engineering, deformation analysis, and geodynamic applications. Category B surveys include geodetic surveys undertaken for the establishment, densification and maintenance of control networks to support mapping. Category C surveys primarily encompass lower accuracy surveys, primarily undertaken for urban, cadastral, geophysical prospecting, GIS and other general purpose mapping applications. Users in the latter two categories form the majority of the GPS user community, while category A users often provide the primary "technology-pull" for the development of new instrumentation and processing strategies, which may ultimately be adopted by the category B and C users.

Note that this classification scheme is entirely *arbitrary*, and does not reflect any "order" of survey as defined by Survey Authorities. It does, however, provide a convenient breakdown of GPS survey "type", enabling the similarities and differences between the categories to be highlighted. Below are listed the advantages and disadvantages of the GPS technology (in the context of land survey applications) *in broad-brush terms only.*

Advantages of GPS Over Conventional Surveying Methods

There are several **advantages** of the GPS satellite surveying technique:

- Intervisibility between stations is not necessary.
- Because GPS uses radio frequencies to transmit the signals, the system is independent of weather conditions.
- If the same field and data reduction procedures are used, position accuracy is largely a function of interstation distance, and not of network "shape" or "geometry".
- Because of the generally homogeneous accuracy of GPS surveying, geodetic network planning in the classical sense is no longer relevant. The points are placed where they are required (for example, in a valley), and need not be located at evenly distributed sites atop mountains to satisfy intervisibility, or network geometry, criteria.
- Because of the two advantages of not requiring intervisibility of stations, or following a conventional network design strategy, GPS surveying is more efficient, more flexible and less time consuming a positioning technique than using terrestrial survey technologies.
- GPS can be used around-the-clock.
- GPS provides three-dimensional information.
- High accuracies can be achieved with relatively little effort, unlike conventional terrestrial techniques. The GPS instrumentation, and to some extent the data processing software, is similar whether accuracies at the 1 part in 10^4 or 1 part in 10^6 level are sought.

Shortcomings of GPS Surveying

It would be remiss not to also mention the **shortcomings**, some of which will no doubt be overcome in time, others with some additional effort, while others cannot be dismissed so easily:

- High efficiency has its price. Efficient use of GPS requires that travel times between stations are cut in order to match the savings in on-site time.
- Because station intervisibility is not necessary GPS is a particularly attractive technology for use in rugged, inhospitable terrain. However, the logistical problems of transporting and supporting several field parties are still formidable (and would have been even if conventional terrestrial techniques were used). If helicopters are necessary, the costs of the survey will rise substantially.
- GPS requires that there be no obstruction to the signals by overhanging branches or

structures (though the antenna can be raised above the obstruction). *It cannot, of course, be used underground, and may have limited application in densely settled urban areas.*

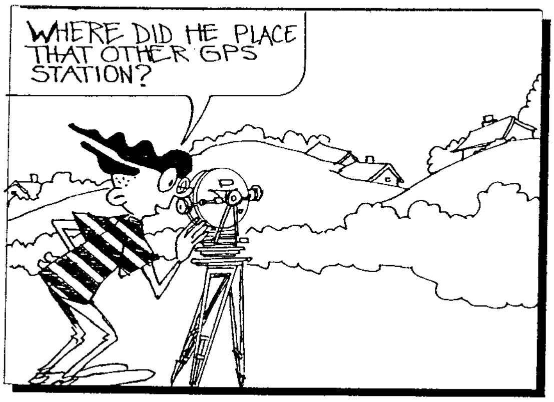
- Because GPS surveys can be "optimised" (by appropriate selection of sites) to satisfy the specific needs of the particular survey, these may not be useful for other applications in the same area. Further GPS surveys may need to be carried out, as the need arises, for new applications. The extreme of this would be to dispense with a permanently monumented control network altogether, and to require the re-establishment of coordinates by GPS each time the need arises.
- Two intervisible stations would have to established by GPS in order to satisfy the requirement for azimuth data for use by conventional (line-of-sight) survey methods.
- GPS coordinates are provided in the earth-centred, earth-fixed coordinate system defined by the GPS satellite ephemerides (the WGS84 system when the broadcast ephemeris is used). *Results may need to be transformed into a local geodetic system before they can be integrated with results from conventional surveys.*
- GPS results are, in general, more accurate than the surrounding control marks established by terrestrial techniques over time. *Comparison of GPS and terrestrial results will be the source of confusion, controversy and conflict for many years to come.*
- GPS vertical information is not provided in the height system generally required. *The GPS heights have to be reduced to a sea level datum (more precisely, the geoid).*
- The GPS instrumentation is still comparatively expensive. Although the price of one receiver is likely to soon match that of a theodolite-EDM instrument, a minimum of two are required for survey work.
- GPS requires new skills to be learned, and new procedures and strategies for planning, field operation and data analysis to be developed. In addition, an understanding of how GPS results can be integrated with conventional horizontal and vertical networks is required.

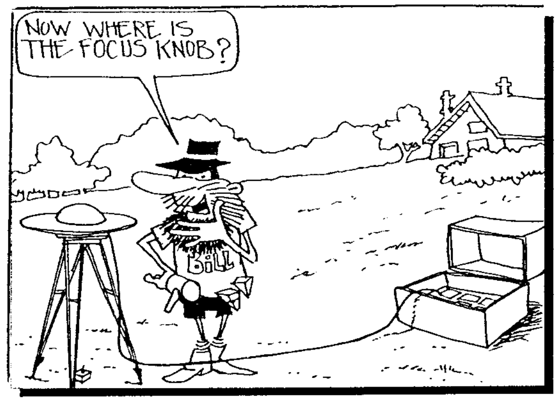
Concluding Remarks

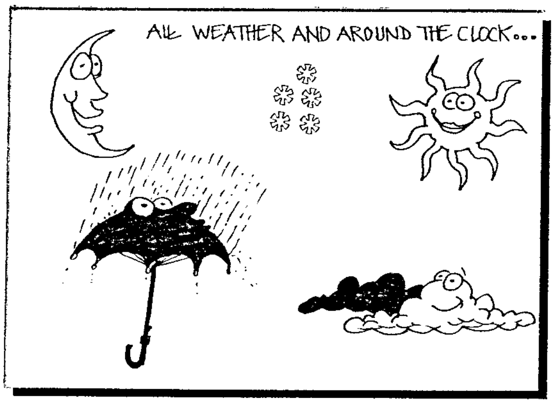
The prospect for increased acceptance of GPS satellite surveying is very good, particularly as the cost of GPS systems drops and new higher productivity techniques are developed. Although GPS was initially used for high-order geodesy and geodetic control surveys on the one hand, and geophysical exploration surveys on the other, adoption of the GPS technology for applications such as lower-order control densification, and even cadastral, engineering and detail surveys, has already commenced.

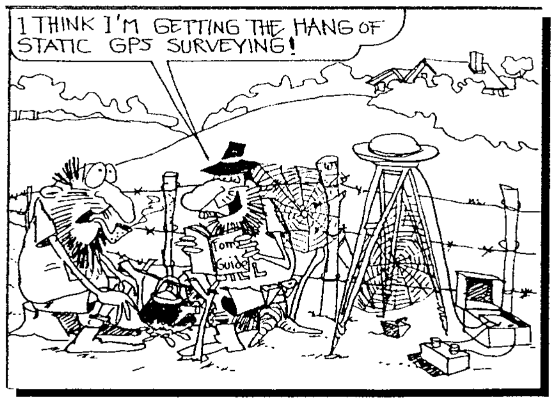
However, for all its technical advantages, there remain a number of significant differences between the results of the GPS surveying technology and that of conventional terrestrial techniques. To reconcile these differences, and in order to ensure that GPS will complement these other technologies (and hence maintain compatibility with the geodetic framework established in many countries over a long period of time), a significant amount of post-processing of GPS results is necessary. This tends to make the GPS technology less attractive, and has the effect of raising the threshold of acceptance slightly higher than it would otherwise have been.

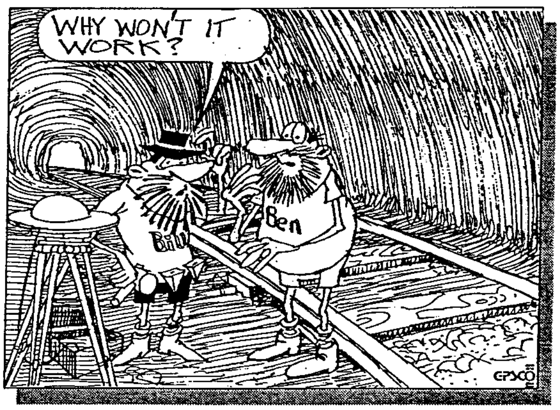
In addition there is an investment in human resource development that must be taken into account. GPS manufacturers are striving to make equipment that is ever more "user-friendly", which will mean that many other professionals apart from "qualified professional surveyors" will be able to carry out high accuracy GPS surveys. The challenge, however, to surveyors is to maintain the "edge", by seeking to use their best judgement and skills not only to achieve high GPS accuracy, but also to ensure that the *quality and reliability of the results* are at the level demanded by the client. A further advantage that surveyors enjoy over other professionals is that they often are the only ones skilled in integrating GPS results into previously coordinated networks.











2.4
HOW GOOD IS GPS?

To appreciate the enormous potential that GPS has to address a wide range of applications it is necessary to analyse the overall *performance* of GPS. There are of course many aspects to GPS performance, including:

- **Accuracy**, a certain level of performance when the appropriate hardware, software and operational procedures are followed.
- **Availability** to users, across the globe and through the day as it is needed.
- **Reliability** of the system and results (a certain "repeatability" of the positioning performance).
- **Integrity**: a management issue concerned with maintaining a specified level of performance.
- **Cost**, of hardware and software as well as indirect operational costs.
- **Competitive technologies**: do they exist? what do they offer in terms of superior performance?

2.4.1 FACTORS INFLUENCING GPS ACCURACY

Although GPS can claim to excel with regard to many of these performance measures, the most important for most users is usually **accuracy**. The main factors influencing GPS positioning accuracy are:

- ❑ **Measurement errors and biases.**
- ❑ **Absolute or differential positioning mode.**
- ❑ **Satellite-receiver geometry.**
- ❑ **Processing algorithms, operational mode and other enhancements.**

All GPS measurements: pseudo-range, carrier phase or Doppler frequency, are affected by **biases** and **errors** (§6.2). Their combined magnitudes will affect the accuracy of the positioning results. Biases may be defined as being those *systematic errors* that cause the *true measurements* to be different from *observed measurements* by a "constant, predictable or systematic amount", such as, for example, all distances being measured too short, or too long. Biases must somehow be accounted for in the measurement model used for data processing if high accuracy is sought. There are several sources of biases with varying characteristics, such as magnitude, periodicity, satellite or receiver dependency, etc. Biases may have physical bases, such as the atmosphere effects on signal propagation, but may also enter at the data processing stage through imperfect knowledge of constants, for example any "fixed" parameters such as the satellite orbit, station coordinates, velocity of light, etc. A useful way of considering biases is as errors which are **correlated** in space or time *Residual biases* may therefore arise from incorrect or incomplete observation modelling and hence they will be treated as *random errors.*

There are two GPS positioning modes which are fundamental to considerations of (a) *error/bias propagation* into (and hence accuracy of) GPS results, and (b) the *datum* to which the GPS results relate. The first is **absolute or point positioning**, with respect to a well-defined coordinate axis system. The coordinate system generally associated with GPS positioning is the earth-centred WGS84 Cartesian reference system. This coordinate system is *realised* via the coordinates of the monitor stations (of the Control Segment), and subsequently *transferred* to users via the (changing) coordinates of the GPS satellites. As the satellite coordinates are essential for the computation of user position, any error in these values, as well as the presence of other biases, will directly affect the quality of the position determination. The **satellite-receiver geometry** will also play an important role in error/bias propagation into GPS positioning results (§1.4, and LANGLEY, 1991c).

Higher accuracies are possible if the relative position of two GPS receivers, simultaneously tracking the same satellites, is derived. Because many errors will affect the absolute position of two or more GPS users to almost the same extent, these errors largely cancel when **differential or relative positioning** is carried out. There are different implementations of the differential positioning procedures, but all share the characteristic that the position of the GPS receiver of interest is derived *relative* to another fixed, or reference, receiver whose absolute coordinates in the satellite datum are assumed to be known. One of these implementations, based on combining the data from the two receivers *before* processing, is the standard mode for GPS surveying. GPS surveying is therefore essentially concerned with the measurement of the *baseline components* between simultaneously observing receivers. (The effect of satellite-receiver geometry in differential positioning is more complex than in the case for point positioning.)

Finally, GPS accuracy is also dependent on a host of other **operational, algorithmic and other factors**:

- Whether the user is moving or stationary. Clearly repeat observations at a stationary station would permit an improvement in precision due to the effect of averaging over time. A moving GPS receiver does not offer this possibility.
- Whether the results are required in real-time, or if post-processing of the data is possible. Real-time positioning requires a "robust" but less precise technique to be used. The luxury of post-processing the data permits more sophisticated modelling and processing of GPS data to minimise the magnitude of residual biases and errors.
- The level of measurement noise has a considerable influence on the precision attainable with GPS. Low measurement noise would be expected to result in comparatively high accuracy. Hence carrier phase measurements are the basis for high accuracy techniques, while pseudo-range measurements are used for low accuracy applications.
- The degree of redundancy in the measurements. For example, the number of tracked satellites (dependent upon the elevation cutoff angle, the number of receiver tracking channels, satellites apart from GPS such as GLONASS and pseudolites, etc.), the number of observations (dual-frequency carrier phase, dual-frequency pseudo-range data).
- The algorithm type may also impact on GPS accuracy. For example, "exotic" data combinations are possible (carrier phase plus pseudo-range), Kalman filter solution algorithms, more sophisticated phase processing algorithms.
- Techniques of data enhancements and aiding may be employed. For example, the use of carrier phase smoothed pseudo-range data, external data such as from Inertial Navigation Systems (and other such devices), additional constraints, etc.

Accuracy versus Positioning Mode

Figure 2.4-1 illustrates the different positioning accuracies associated with the different GPS positioning modes. The following comments may be made to this diagram:

- The top half refers to point positioning, the lower half to the relative positioning mode.
- The basic GPS "services" envisaged by the system operators, the U.S. Department of Defense, are the Standard Positioning Service (SPS) and the Precise Positioning Service (PPS).
- There is a large range of potential accuracies associated with the point positioning mode:
  - 100m level accuracy SPS positioning with SA on, *as a result of an artificial degradation of the system.*
  - 20-30m level accuracy standalone SPS positioning (no SA), representing the "natural" accuracy ceiling when using the most basic navigation-type GPS receiver.
  - 2-15m level accuracy PPS positioning.
  - 10-20% improvement is possible using dual-frequency SPS receivers that eliminate the ionospheric delay effect (§6.2).
  - improvements in SPS receiver technology (and dual-frequency tracking capability) could deliver a 50% reduction in the standalone (no SA) SPS accuracy.
  - the above improvements, coupled with higher accuracy satellite clock and ephemeris data, could deliver, in theory, a further 50% reduction in accuracy to the PPS level.
- The carrier phase-based procedures can only be applied in the relative positioning mode.
- The accuracy of carrier phase-based positioning procedures is a function of receiver separation, and is significantly better than code (or range) based procedures.

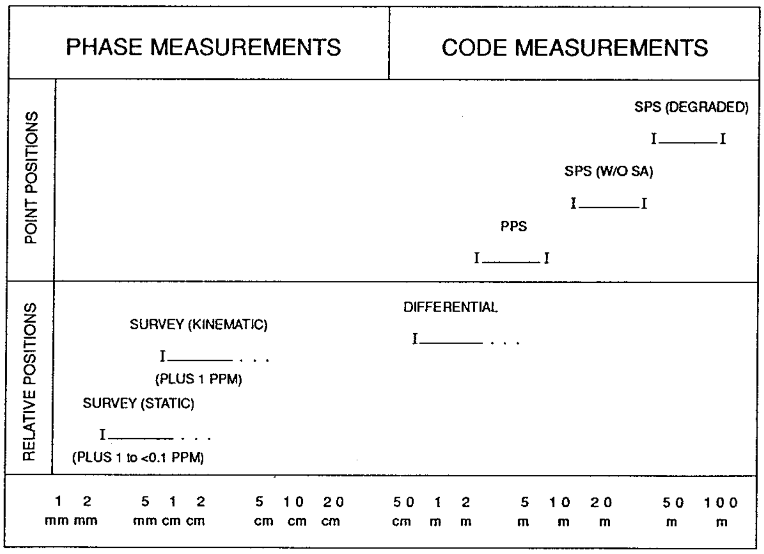


Figure 2.4-1. GPS accuracies and positioning modes.

<u>*Note:*</u> SA -- Selective Availability; ppm -- parts per million; SPS -- Standard Positioning Service; PPS -- Precise Positioning Service

It should be emphasised that GPS was designed to provide accuracies of the order of a dekametre or so in the absolute positioning mode, and is optimised for real-time operations. All other developments to improve this basic accuracy capability must be viewed in this context. As a general axiom of GPS positioning, **the higher the accuracy sought, the more**

effort (in time, instrumentation and processing sophistication) is required.

GPS Positioning Modes

From the discussions above, the main positioning modes can now be identified (Figure 2.4-2):

- **ABSOLUTE or POINT positioning**: coordinates in relation to a well-defined global reference system.
- **DIFFERENTIAL or RELATIVE positioning**: coordinates in relation to some other fixed point. In GPS surveying this is referred to as baseline determination.
- **STATIC positioning**: coordination of stationary points, either in the absolute or relative mode. Relative static positioning is generally synonymous with the SURVEYING mode of operation, and is based on the analysis of carrier phase observations.
- **KINEMATIC positioning**: coordination of moving points, either in the absolute or relative mode. This is the NAVIGATION mode of positioning, and is based on the use of pseudo-range observations.

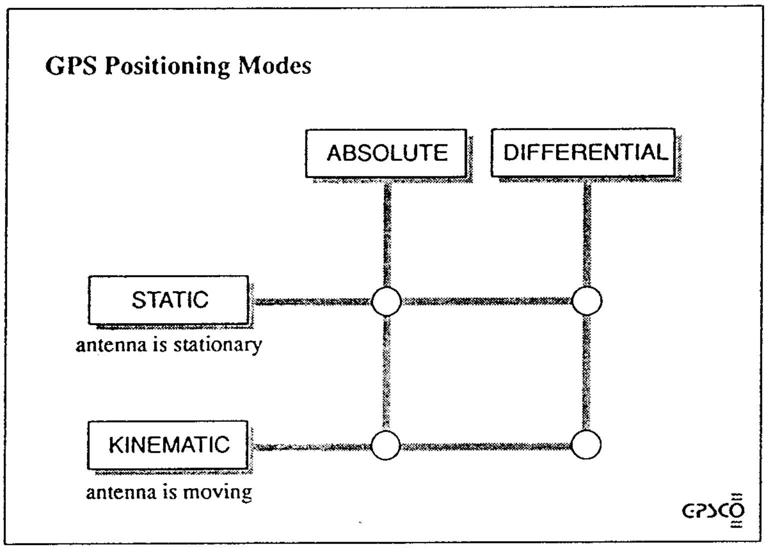


Figure 2.4-2. The basic GPS positioning modes.

Comment:

- Real-time positioning in the kinematic mode is the type of positioning for which GPS was explicitly developed. Only low to moderate navigation accuracies are possible.
- Pseudo-range data can be used for both absolute and differential positioning. Carrier phase data on its own is used in precise differential positioning. "Smoothing" the relatively noisy pseudo-range data with carrier phase is a useful technique for enhancing absolute or differential pseudo-range results.
- With respect to differential positioning, the dividing line between the two extremes of static positioning with carrier phase data (conventional "GPS surveying"), and kinematic positioning using pseudo-range data (conventional "precise navigation") is

becoming increasingly blurred. It is possible to now speak of "rapid" static positioning, or high precision phase-based kinematic positioning (§5.5).

Errors and Quality Measures

All measurements are subject to errors. Traditionally errors have been classified into three categories:

(1) **Random Errors**

(2) **Systematic Errors**

(3) **Gross Errors**

Random measurement errors are, as their name implies, essentially *unpredictable* (in magnitude and sign) and are basically due to: (a) the "resolution" of the measurement scale (or its "least count"), (b) random internal instrumental effects, and (c) some external but very local effects such as micro-meteorological conditions, electrical connectors and antenna quality, local signal interference, etc. All of these, in essence, define the level of "instrumental noise". In classical statistics the behaviour of random errors can be studied using probability theory, and for most cases in navigation and surveying such errors are assumed to be "white", or have a Gaussian distribution (see HARVEY, 1994, for further reading on this topic).

Precision is the degree of *repeatability* (or closeness) that repeated measurements of the same quantity display, and is therefore a means of describing the quality of the data with respect to random errors. Precision is traditionally measured using the standard deviation. Hence measurements (or position results) that are *closely grouped* are said to have a *high precision* because their random errors (or the propagation of measurement random errors into the positioning results) are *small* (Figure 2.4-3). Conversely, measurements (or position results) that are *widely spread or dispersed* are said to have a *low precision* because their random errors, or the propagation of measurement random errors into the positioning results, are *large* (Figure 2.4-3). *Of course, the distinction between high and low precision is merely an arbitrary judgement.*

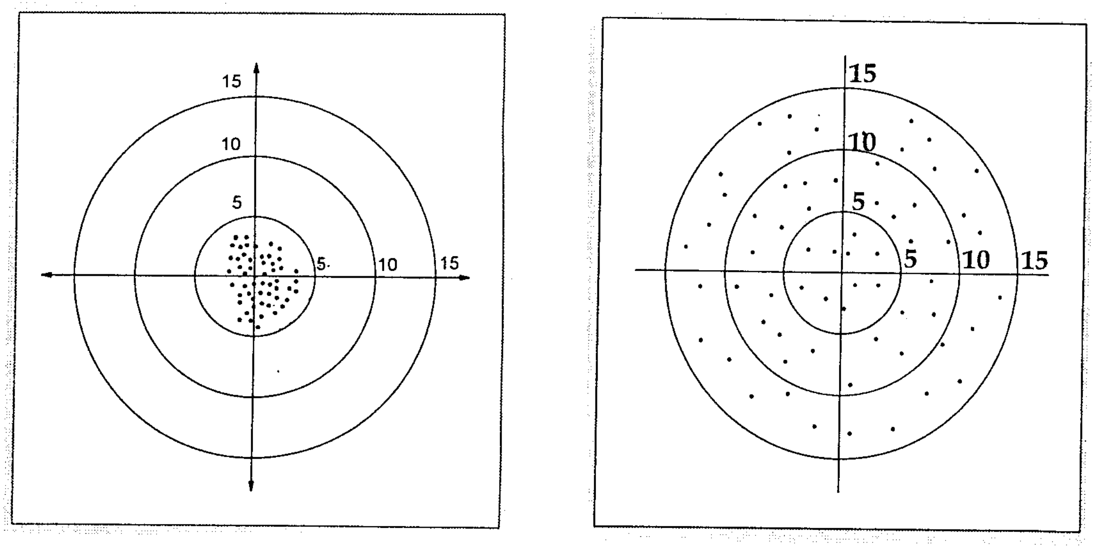


Figure 2.4-3. Quality Indicator: PRECISION of measurement or position result.

In the context of GPS data processing, under the heading of random errors are also lumped any unmodelled or residual biases which are not accounted for in the mathematical modelling of the observation(s) -- see below. For this reason, in these notes the term "error" has a special significance because it is not simply a catalogue of random measurement disturbances, but also of biases that may be of equivalent magnitude to the measurement "noise". GPS errors can therefore no longer be considered to be "white".

Systematic errors occur according to some pattern, for example:

- They may be of constant magnitude and sign.
- They may be induced by the instrument, the observer, the physical or environmental conditions.
- They may be present in the observation model due to entry of an incorrect parameter or constant.
- They may be the result of incorrect application of calibration data.

Accuracy is the degree of closeness (or conformity) of a measurement, or position result, *to its true value.* Although the accuracy indicator accounts for all types of error it is particularly useful for describing systematic errors. For example, in Figure 2.4-4, the quantities in question are *very precise* but *inaccurate.* Such a situation is undesirable as the *potential* quality is high (due to the low intrinsic noise), but the *actual* quality is poor because the value of the quantity is *biased.*

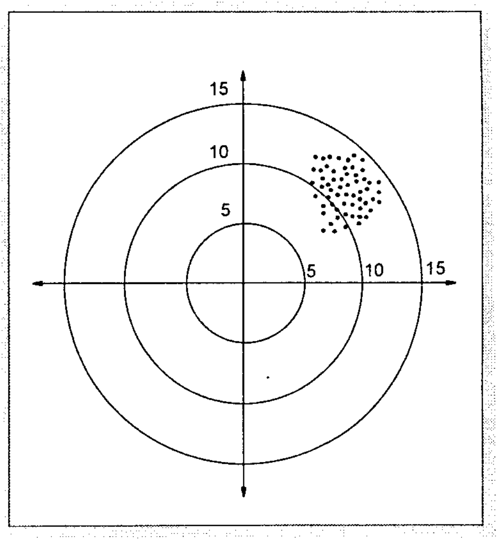


Figure 2.4-4. Quality Indicator: ACCURACY of measurement or position result.

In the context of GPS data modelling systematic errors are generally termed "biases" (§6.2), because they may lead to results as illustrated in Figure 2.4-4. Dealing with the various biases in GPS data is a considerable problem and various strategies have been developed to account for them:

- The biases can be *estimated* as explicit (additional) parameters.
- Those biases linearly correlated across different datasets can be *eliminated by differencing.*
- Certain biases can be *eliminated by constructing linear combinations of different data types.*
- The biases can be *directly measured.*
- The biases can be considered known or adequately *modelled.*

It must be stressed that biases larger than the noise level of the measurement, which are likely to cause the quality of positioning to fall below some specified level of accuracy, should be accounted for in some way. In the case of pseudo-range measurements, the magnitude of most of the biases is below the noise level of the observations and can therefore be simply ignored (they become absorbed into the "error" as discussed earlier). However, in the case of carrier phase measurements all biases are potentially a matter of concern. In other words, **different GPS applications require different levels of GPS accuracy, hence there is the possibility of a different partitioning of "biases" and "errors".**

Gross errors are the result of blunders or mistakes. If these errors have large magnitudes they are usually easy to identify, and hence can be easily removed. (Another name for gross errors is **outliers**, and *outlier detection* is an important step in any navigation or survey data processing procedure.) However, not all gross errors may be large enough to be noticed and hence they can still contaminate the final results.

(Unmodelled or residual biases referred to earlier as belonging to the category of "errors" could be considered as being a type of *gross* error. Although an unambiguous definition is not possible, in these notes the following convention will be adopted: unmodelled or residual biases which are *smaller* than the magnitude of the measurement noise are "errors", while those which are *larger* than the magnitude of the measurement noise will be considered as being "gross errors".)

Reliability is often used in the context of sensitivity to gross errors. It is a measure of the *ease* with which gross errors can be detected. A *highly reliable position result* is one in which even quite *small outliers will be noticed.* Conversely, an *unreliable position result* is one where *large outliers will go unnoticed.* Extra (or redundant) observations is an effective way of improving reliability. For example, Figure 2.4-5 illustrates the situation in which the distribution implies high precision, except that there are several quantities which are significantly different. It could be concluded that these quantities are outliers and as such they should be rejected.

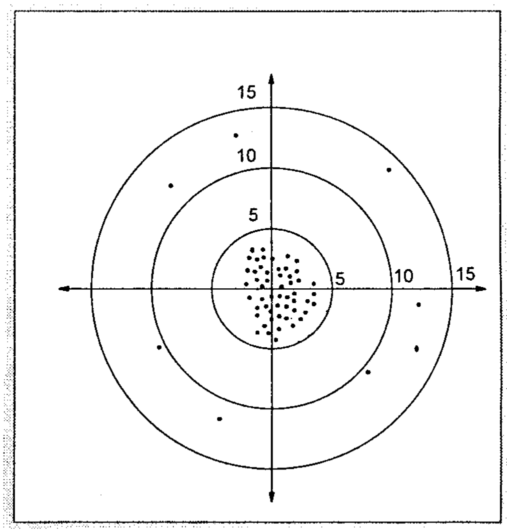


Figure 2.4-5. Quality Indicator: RELIABILITY of measurement or position result (presence of outliers or gross errors).

If systematic and gross errors have been removed, then the accuracy will only be a function of the magnitude of random errors.

2.4.2 POINT POSITIONING ACCURACY PERFORMANCE

Measures of GPS Point Positioning Accuracy

The uncertainty in position can be expressed as the probability that the error will not exceed a certain amount. Under the assumption that position errors follow the normal (or Gaussian) error distribution (for arguments sake, there are only random errors being propagated into the position results), this probability can be related to the magnitude of the standard deviation (HARVEY, 1994). For example, in the case of a linear (one-dimensional) accuracy measure, one standard deviation (one sigma) would correspond to a 68.27% *confidence interval.* That is, it is assumed that: (a) the mean value of an infinitely large sample of position results is the correct result, and (b) the standard deviation of this sample defines the interval on either side of the mean (or correct) quantity that contains 68.27% of all the results. 31.73% of the results will therefore be outside this range, and if the one sigma quantity is taken as the *accuracy measure*, then 68.27% of the results will be deemed acceptable and the remainder will be outside the accuracy "specification". The probability of the result being in the interval two standard deviations on either side of the mean is 95.45%. In general, the 95% confidence level is taken as the measure of adequate accuracy, and this corresponds to 1.96 standard deviations (but is generally approximated by two standard deviations, or two sigma). (The probability corresponding to three sigma is 99.73%, which is inclusive of almost all position results.) Vertical uncertainty can be expressed in this one-dimensional form.

This concept can be extended to two dimensions, so that areas can be constructed corresponding to distinct error probabilities such as 50%, 95%, etc. These zones are centred at the correct or true position. In general these zones are *elliptical* in shape, and they are known as **error ellipses** (see IBID, 1994; and §9.1). However, the error ellipse constructed from the standard deviations of the two dimensional quantities (for example, east and north position components), and the correlations between these two quantities, contains only 39% of the position results and the error ellipse's axes are generally inflated by a factor of approximately 2.45 to create the 95% error ellipse. Surveyors are well acquainted with the concepts of error ellipses. (In three dimensions, the error figure is an **error ellipsoid**.)

Traditionally navigation users have expressed horizontal position uncertainties in the form of *circles* and 3-D position uncertainties as *spheres*. This simplification of the error distribution requires the definition of the **radial or distance "root mean square" error**, which can be determined for the 2-D case from the horizontal component standard deviations (the 3-D case would involve three standard deviation quantities) σ_E and σ_N:

$$d_{RMS} = DRMS = \sqrt{\sigma_E{}^2 + \sigma_N{}^2} \qquad (2.4\text{-}1)$$

The probabilities described by 1.DRMS and 2.DRMS are defined as the *typical* 68.27% and 95.45% values respectively, associated with the 1-dimensional distribution. (2.DRMS refers to TWICE the distance root mean square, irrespective of whether it is the 2-D or 3-D case.) <u>These probabilities are not constant, but are dependent on the geometry of the position solution</u>. For example, if the geometry of the solution is very poor then the 95% error ellipse is very elongated, and the probability associated with an error circle of radius 2.DRMS may not be 95%. Conversely, if the error ellipse is almost circular, then the probabilities of the ellipse and the circle are also almost identical. In the case of GPS it has been found that the probability associated with the circle of radius 2.DRMS ranges from 95.4% to 98.2%.

This accuracy measure is also closely tied to *Dilution Of Precision* analysis (§1.4) by:

$$2.DRMS_{Hor} = 2.HDOP.\sigma_0$$
$$2.DRMS_{Ver} = 2.VDOP.\sigma_0 \quad (2.4\text{-}2)$$
$$2.DRMS_{3\text{-}D} = 2.PDOP.\sigma_0$$

where σ_0 is the standard deviation of the range measurement errors, $DRMS_{Hor}$ refers to the horizontal components, $DRMS_{Ver}$ refers to the vertical component, and $DRMS_{3\text{-}D}$ refers to the 3-D position. HDOP, VDOP and PDOP refer to Horizontal DOP, Vertical DOP and Position DOP respectively.

An alternate measure is the **Circle Error Probable** (CEP), sometimes also referred to as the "Circle Error Probability" or "Circle of Equivalent Probability". This defines the radius of a circle inside which there is a *50% probability* of the position being located. How is CEP related to $DRMS_{Hor}$? There is no exact relation, but an approximate relation defined for GPS is:

$$2.DRMS_{Hor} \approx 2.45 \,.\, CEP \quad (2.4\text{-}3)$$

Other CEP's for different probability radii can be obtained by multiplying the above 50% CEP by the appropriate factors. Figure 2.4-6 illustrates the accuracy circles for 2-D positioning. In the 3-D case the equivalent accuracy measure is the **Spherical Error Probable** (SEP), and it is 60m and 16m for the SPS and PPS respectively.

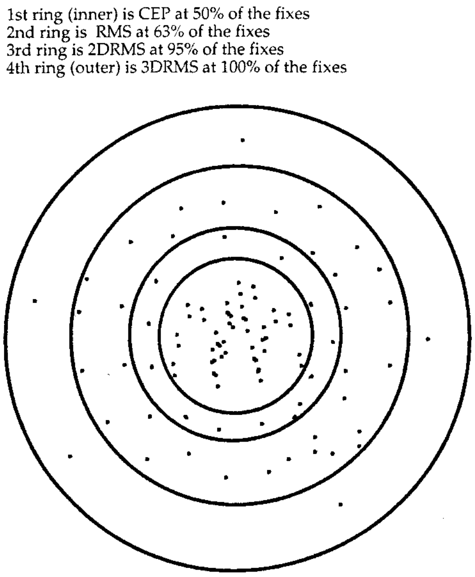


Figure 2.4-6. Rings of accuracy used to describe GPS horizontal point positioning accuracy. *(CLARKE, 1994)*

If the accuracy performance of a positioning system is not known apriori, then the radius of the 95% confidence interval zone, as well as for other probabilities, has to be determined *empirically* from a large sample of positioning results (Figure 2.4-6). On the other hand, if the positioning system controllers wish to guarantee a certain level of accuracy performance, then the quantity must be specified apriori and all effort applied to reduce the level of total error in the system so that 95% of position results are within the specified zone. **Hence, the SPS**

service is designed to support 2-D (horizontal) positioning so that at least 95% of position results are within a circle of radius 100m centred at the "true" position (which can be determined from the mean of a very large sample of results). In this case σ_0, referred to as the User Equivalent Range Error (UERE), is maintained, on average, to be 32m, while the average HDOP is 1.6. Further details of the SPS performance can be found in NAVSTAR (1993).

The Civilian - Military Relationship and its Impact on GPS Performance

Although GPS is a military navigation system, the civilian sector represents an important (and rapidly growing) user group that has increasingly lobbied the U.S. Government in order to influence: (a) the direction of GPS system development; (b) official GPS policy concerning enhancement and control; and (c) the follow-on systems to GPS for the 21st century.

Several policy decisions have already been made which impact on GPS performance. Some of these actions were agreed to during the early system design phase, while others were invoked only after much of the present system had been deployed:

(1) *Two PRN ranging codes are implemented* (§3.1). The C/A code is intended for general, civilian use, while the P-code was reserved for U.S. military and other authorised users. The important distinction between the codes was that as a result of the P code's tenfold higher measurement resolution, it was expected that the accuracy of positioning using the P code would be better than that possible using the C/A code by a similar margin. However, to everyone's surprise, the performance of C/A code positioning was often no worse than by a factor of two compared to P code positioning. (The latest C/A code tracking technology promises ranging quality almost as good as P code ranging.)

(2) *Two levels of positioning performance were therefore designed into the GPS system from the very beginning.* The positioning service based on using C/A code ranging data is the Standard Positioning Service (SPS), while the service based on P code ranging data is the Precise Positioning Service (PPS).

(3) As a result of the surprisingly good level of SPS accuracy, the policy of **Selective Availability** (SA) was endorsed in order *to artificially widen the gap between the two positioning services* (GEORGIADOU & DOUCET, 1990). SA is an intentional degradation of the accuracy of GPS horizontal positioning to 100m (at the 95% confidence level), and height determination to 150-170m (at the 95% confidence level), for SPS users (see Figure 2.4-7). SA has been implemented since 25 March, 1990, through: (a) a corruption of the transmitted ephemeris data within the Navigation Message (the so-called "epsilon" effect), and (b) the satellite clock is "dithered" (the so-called "delta" effect). SA does not affect PPS users who have GPS equipment able to decipher the correct ephemeris and clock error data.

(4) *Prior to 1994 any GPS hardware manufacturer was able to develop a P code ranging receiver* (the P code PRN generation algorithm having been published). The non-military market for P code receivers was always assumed to be very small, and one that could be controlled by the U.S. Government through the issuance of "export licences", etc. However, a significant demand by the surveying market for dual-frequency phase tracking GPS receivers led to an expansion in the production of P code capable positioning equipment.

(5) Under another policy known as **Anti-Spoofing** (AS), *access is denied to the P code modulated on both L-band frequencies.* AS was implemented on 31 January, 1994, through the encryption of a further secret "W code" onto the P code. The rationale behind

this decision was that by keeping the military PRN code secret, an enemy of the U.S. could not jam the signal using a ground-based transmitter, nor "spoof" military GPS receivers by transmitting a false P code signal from a satellite. However, several GPS receiver manufacturers have developed proprietary techniques for making dual-frequency measurements even in the presence of AS (§4.2).

(6) Dual-frequency observations will lead to more accurate positioning results than single frequency observations because the ionospheric bias can be eliminated from the code range measurements. *The fact that the C/A code is only modulated on the L1 carrier is therefore an intentional design decision to ensure that the SPS service cannot approach the quality of the PPS service, even with SA off.*

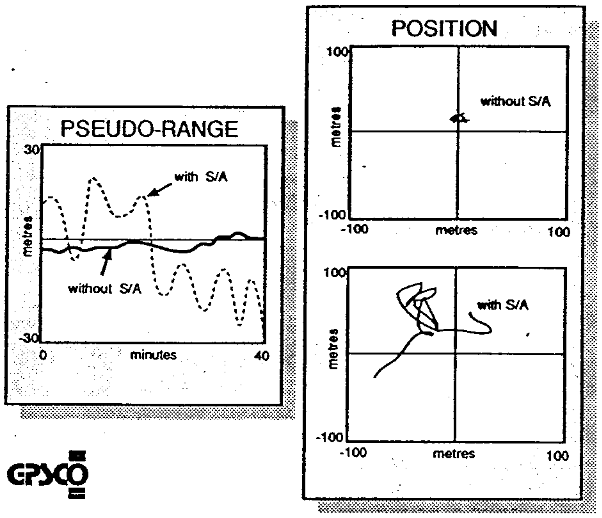


Figure 2.4-7. Range measurement errors (left) and point position errors (right) at a static site, before and after SA was implemented.
(After GEORGIADOU & DOUCET, 1990)

SELECTIVE AVAILABILITY:

- ❑ **Policy of intentional degradation of SPS navigation accuracy introduced on 25 March, 1990.**
- ❑ **100m horizontal accuracy (95% confidence level), 150-170m height accuracy.**
- ❑ **Horizontal accuracy better than 300m 99.99% of the time.**
- ❑ **60m *Spherical Error Probable.***
- ❑ **0.3 m/sec velocity accuracy and approximately 340nsec time transfer accuracy.**
- ❑ **Classified algorithm and characteristics (periodicities, etc.).**
- ❑ **Military receivers are able to overcome SA.**
- ❑ **Two effects: (a) "dithering" of the satellite clock *(the "δ" component)*, and (b) errors in the broadcast ephemerides *(the "ε" component)*.**
- ❑ **Can be overcome by using differential GPS positioning mode.**

It must be emphasised that the situation concerning the SA and AS policies are under almost continuous review. The "tug-a-war" situation between the GPS system controllers and civilian users needs to be resolved, and the reader is referred to LACHAPELLE (1995) and SANDLIN et al (1995) regarding the outcome of studies into alternative models of joint civilian-military GPS operation and recommendations concerning SA and AS. *In particular, it has been recommended that SA be deactivated within the next few years, and that another frequency be transmitted to support civilian dual-frequency operations.*

SA also biases carrier phase measurements (§6.2). However, the policy of SA is explicitly aimed at the real-time point positioning (that is, the navigation) user, and has only a marginal effect on those users that have adopted differential GPS techniques, in either the real-time (precise navigation) or the post-processed mode (in general, surveying), based on pseudo-range and/or carrier phase data.

ANTI-SPOOFING:

- ❑ **Satellite signal design intended to prevent military receivers from being "spoofed" (by tracking "false" GPS satellites).**
- ❑ **P code is replaced by a secret Y code.**
- ❑ **P code receivers will not function when AS is on.**
- ❑ **Only receivers with the appropriate "key" can use the Y code for ranging.**
- ❑ **Implemented on 31 January, 1994.**

2.4.3 RELATIVE POSITIONING ACCURACY PERFORMANCE

Correlated GPS Measurement Biases

Relative positioning is the most effective means of accounting for many of the troublesome GPS measurement biases, and hence is the basis for all high precision GPS positioning techniques. The **correlated** nature of biases is best illustrated by an examination of the basic GPS pseudo-range measurement model (§6.2):

$$\mathrm{p}_i{}^p = \rho_i{}^p + \varepsilon_{rc(i)} + \varepsilon^{sc(p)} + \varepsilon^{orbit(i,p)} + \varepsilon_{atmos(i,p)} + \nu_i{}^p \qquad (2.4\text{-}4)$$

The subscript in brackets refers to the GPS station "i", and the superscript in brackets refers to the satellite "p". p is the measured pseudo-range, ρ is the true geometric range from one receiver to one satellite, ε_{rc} is the receiver clock error, ε^{sc} is the satellite clock error, $\varepsilon^{orbit(i,p)}$ is the satellite orbit error mapped into the range, $\varepsilon_{atmos(i,p)}$ is the atmospheric refraction error, and $\nu_i{}^p$ are the remaining errors and biases not explicitly accounted for in the above observation model. Although the time argument has been dropped for the sake of clarity, all quantities in eqn (2.4-4) vary with time, and hence the equation represents a "snapshot" of a GPS pseudo-range measurement at a single epoch, or instant of time.

The spatially correlated nature of many GPS errors is obvious when an observation from another station "k" to the same satellite, at the same epoch, is modelled as in eqn (2.4-4):

$$p_k^p = \rho_k^p + \varepsilon_{rc(k)} + \varepsilon^{sc(p)} + \varepsilon^{orbit(k,p)} + \varepsilon_{atmos(k,p)} + v_k^p \qquad (2.4\text{-}5)$$

The following comments can be made with regard to eqns (2.4-4) and (2.4-5):

- The *receiver clock error* $\varepsilon_{rc(i)}$ systematically affects all measurements made at station "i" to all satellites by exactly the same amount. It is completely unrelated to the value of $\varepsilon_{rc(k)}$ at the same measurement epoch. $\varepsilon_{rc(i)}$ and $\varepsilon_{rc(k)}$ across different epochs may be considered uncorrelated.
- The *satellite clock error* $\varepsilon^{sc(p)}$ systematically affects all measurements made to satellite "p" by any GPS receiver making a measurement at the same time-of-transmission. Hence $\varepsilon^{sc(p)}$ is spatially correlated (across different receivers) at an epoch, and this property can be exploited to overcome the effect of this error/bias.
- The *satellite orbit error* $\varepsilon^{orbit(i,p)}$, although explicitly associated with satellite "p", is mapped differently as a range error in the case of measurements made by station "i" compared to those made by station "k" (hence the use of the double index identifying both the receiver and satellite involved in the measurement). However, if the two stations are close together the residual bias $(\varepsilon^{orbit(i,p)} - \varepsilon^{orbit(k,p)})$ will be very small. $\varepsilon^{orbit(i,p)}$ across different epochs will change comparatively slowly (hence there is a high temporal correlation).
- The *atmospheric refraction error* $\varepsilon_{atmos(i,p)}$ expressly tags the measurement made to satellite "p" by station "i". However, if the two stations are close together the atmospheric conditions along the signal line-of-sight to the satellite can be expected to be very similar, and hence the residual bias $(\varepsilon_{atmos(i,p)} - \varepsilon_{atmos(k,p)})$ will be quite small in magnitude. The atmospheric refraction effect has an ionospheric and tropospheric component, each with their own spatial and temporal characteristics. $\varepsilon_{atmos(i,p)}$ across different epochs will change comparatively slowly (hence a high temporal correlation).
- The *residual error term* v_k^p contains the random effects of measurement noise (which will vary according to satellite, receiver and measurement epoch), disturbing biases which are not spatially correlated (their effects are too dissimilar at different stations), and any other biases not explicitly included in eqns (2.4-4) and (2.4-5).

Characterising Accuracy of Differential Positioning

The accuracy of a relative position has two components, due to the two classes of errors contained within the v term:

(a) *The random measurement noise DOES NOT influence accuracy as a function of receiver separation.* The magnitude of relative position error of two stations due to the random measurement noise is a function of:
  - The <u>type of measurement</u>, therefore carrier phase measurements are likely to contribute just a few millimetres, whereas in the case of pseudo-range measurements this may range from several decimetres to a few metres.
  - The <u>degree of redundancy</u>, influenced by such factors as the number of satellites tracked, the number of observation epochs, the degree of freedom of the solution (kinematic positioning has far less degrees of freedom than static positioning), etc.
  - The <u>quality of the antenna centring</u> over the physical station markers, including the stability of the electrical antenna centre.

- The type of carrier phase solution, ranging from the millimetre level in the case of the best "ambiguity-fixed" solution, to several centimetres or a few decimetres where the carrier phase ambiguity is not fixed to an integer, or it is eliminated from the observation model through between-epoch differencing (see §8.1).

(b) *The residual biases that remain will largely influence accuracy as a function of receiver separation.* These residual biases arise mainly because the satellite orbit errors ε^{orbit} and the atmospheric biases ε_{atmos} are not eliminated when observations from two receivers are combined. Their effect on relative position determination is greater for long baselines than for short baselines. This error signature is usually expressed as the ratio of the magnitude of the error to baseline length, as some many "parts per million" (ppm). Figure 2.4-8 illustrates the relationship between relative error (ppm), relative error (centimetres), and baseline length. The dependency of relative error to baseline length has only been observed in GPS surveying results using carrier phase observations because the noise of the measurements is generally very much lower than the residual biases.

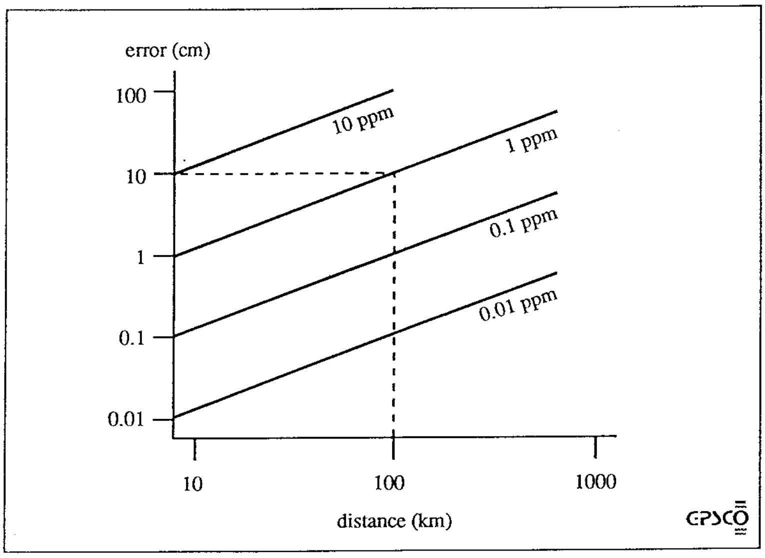


Figure 2.4-8. GPS relative accuracy as a function of baseline length.

Gross errors in the observations will largely propagate in a similar fashion to the random error effects. **Independent testing of GPS surveying systems over a variety of baseline lengths has estimated that relative positioning error between static receivers typically is 1-10ppm, with a small constant term of 0.3-1cm** (§4.4).

Implementations of Differential GPS Positioning

Clearly the use of two GPS receivers, simultaneously tracking the same satellites, is an effective means of overcoming the effect of spatially correlated biases. There are essentially two ways in which measurements from two receivers are used to account for biases, and hence improve accuracy:

(1) Differencing measurements between receivers (subtracting eqn (2.4-4) from eqn (2.4-5)) leads to an observable that is free of biases (or at least substantially reduced if the receivers are not too far apart). *This is the GPS surveying mode of differential*

positioning using carrier phase data.

(2) Each set of measurements at a receiver are independently used to derive a position which is in error by more or less the same amount. *This is the DGPS procedure implemented for precise navigation applications using pseudo-range data.*

Relative positioning accuracies are mostly dependent on the differential technique used:

☞ **Differential GPS navigation, relative positions derived by differencing two absolute position determinations, or application of range "corrections" to measurements.** ***Typically accurate at the few metre level, though several new procedures deliver sub-metre accuracy.***

☞ **GPS surveying, where baseline components are derived from the simultaneous processing of data collected over an extended period of time (static mode), or under certain constraints (kinematic mode).** ***Accuracy expressed as a ratio to baseline length, typically a few ppm, with a small constant term of a centimetre or less.***

Several distinct categories of differential GPS positioning performance can be identified, each a function of data type (pseudo-range or carrier phase) and receiver mode (static or kinematic).

Differential Navigation Based on Real-Time Point Positioning

The distinguishing characteristic of navigation is the urgency with which positioning information is required in order to ensure safe passage from port to port. (The situation with regards to "land navigation" is slightly different from that of marine or air navigation, but the notion of keeping track of where the user is going, or has gone, is still important.) Relative positioning information is derived from two separate navigation "fixes" based on the processing of pseudo-range observations (see §1.4). One of these is at a fixed station of known position (the "base station"), while the other is at an unknown location (the "remote station", or if it is moving, the so-called "mobile station"). Two implementations are possible:

(1) Differential positioning can be accomplished by the continuous transmission of the coordinate solution from the "base station" to the "remote station" as illustrated in Figure 2.4-9. This block shift technique is the easiest to implement (although it does have certain limitations):
   (a) Base receiver at known point -- *WGS84 or local datum coordinates.*
   (b) Compare known position and instantaneously computed position.
   (c) Generate correction ΔX, ΔY, ΔZ.
   (d) Transmit correction to remote receiver for immediate correction of "raw" field coordinates (or saved to file at base for later correction of field receiver coordinates).

It is important that both the remote and base receivers use the same satellite constellation to generate their point solutions, otherwise severe errors can result, possibly worse than those of the (uncorrected) point positioning. As simple as this may sound, it can be difficult to implement in practice!

(2) A popular real-time DGPS strategy is the method of range corrections. Rather than making corrections to the coordinates, the ranges *before* computation of the receiver position are corrected (Figure 2.4-10). This is achieved by a process similar in many respects to that of block shift method:
   (a) Base station at known point --> *WGS84 or local datum coordinates.*
   (b) Using known position compute "true" range.
   (c) Generate corrections to all pseudo-range data by comparing "true" to "observed" range.
   (d) Transmit correction data to remote receiver for correction of ranges before solution carried out.

The technique is far more flexible because the correction is made to the pseudo-ranges and hence the remote GPS receiver can use any combination of corrected ranges to obtain a solution, and not just the satellite set used at the base station.

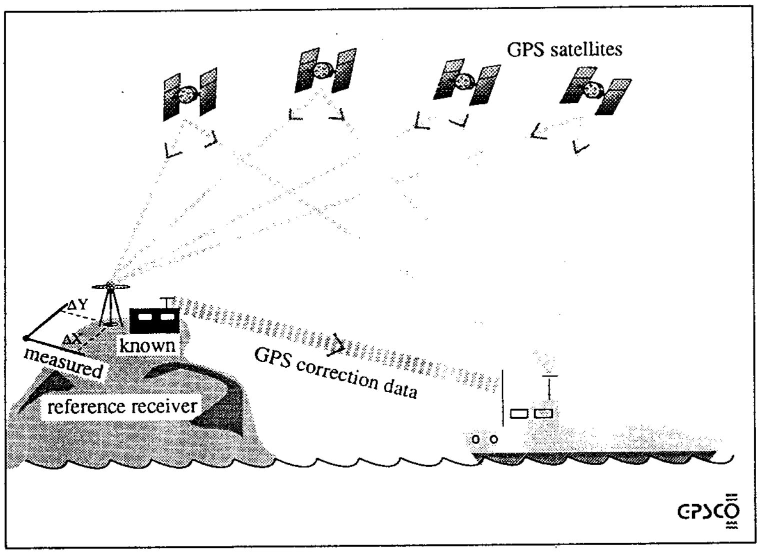


Figure 2.4-9. Principle of DGPS using the block shift method.

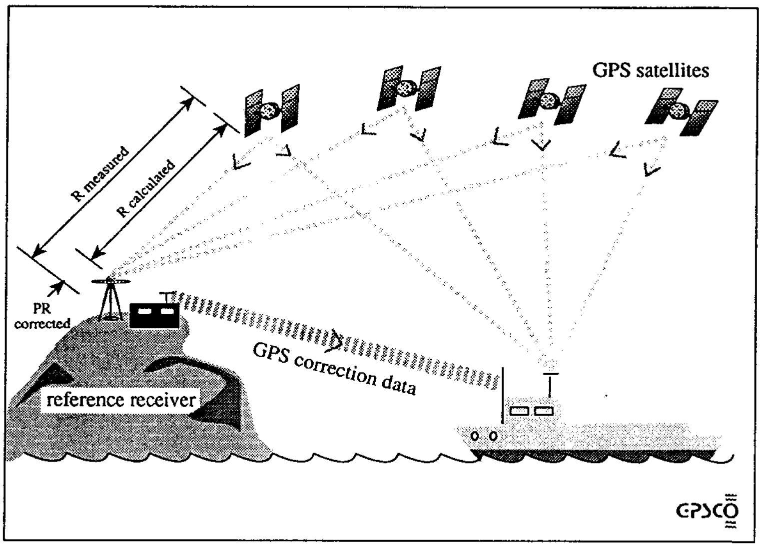


Figure 2.4-10. Principle of DGPS using the range corrections method.

The relative accuracies expected in such schemes are of the order of 1-10m, though enhancements have been introduced that appear to deliver submetre accuracy.

Baseline Surveying via the Combined Processing of Data from Two Phase Tracking Receivers

The distinguishing characteristics of the survey mode of positioning are:

- In one scenario the data is collected simultaneously over an extended observation period, at two or more sites, whose positions do not change with time (static baseline).
- An alternative scenario involves data collected in kinematic mode, at two receivers, where the remote receiver if moving and the base receiver is stationary.
- The carrier phase data is used because of its very high measurement precision.
- The data processing techniques involve the comparatively sophisticated modelling of the systematic errors (or biases) in the GPS system.
- The vertical error is greater than the horizontal components by a factor of 2 or 3.

In tests carried out by the U.S. Federal Geodetic Control Sub-Commission (FGCS) on GPS surveying systems (§4.4), the internal consistency (or repeatability) for static GPS baseline results were characterised by a one-sigma base positional uncertainty in each component that ranged from 3 to 10mm. Added to the base error in each component was a one-sigma line-length dependent uncertainty of 1-2 ppm. Possible sources for the base error included: phase centre variation, effects of antenna multipath, errors in centring and measuring height of antenna phase centre. The line-length dependent error was caused mostly by errors in the orbital ephemeris data and atmospheric effects. During periods of peak ionospheric activity, leading to degraded single frequency baseline results, it likely that the line-length error budget may have to be increased slightly.

Much higher accuracies are possible when "GPS geodesy" procedures are applied. Typically relative accuracies of the order of 0.1 to 0.01ppm are obtained.

Kinematic phase-based GPS positioning is a relatively recent technique that can deliver accuracies perhaps only a few factors worse than static baseline techniques. However, there are a number of constraints:

- The interstation distances must be comparatively short (generally less than 20km).
- The ambiguous carrier phase data must be converted to precise "carrier-range" by the determination of the integer ambiguities *before* the GPS survey can commence.
- In general, the most sophisticated hardware (dual-frequency receivers) is used.
- The dynamics of the moving receiver must not be too great to cause significant periods of loss-of-lock on the satellite signals.

Kinematic phase-based positioning techniques are discussed in §5.5.

Chapter 3: The GPS Signals

3.1 GPS SATELLITE SIGNALS

To understand the GPS measurement process it is necessary to know something of how the signals are generated, transmitted, received and subsequently processed.

Each GPS satellite transmits a unique navigational signal centred on two L-band frequencies of the electromagnetic spectrum: L1 at 1575.42MHz and L2 at 1227.60MHz. At these microwave frequencies the signals are highly directional and hence are easily blocked, as well as reflected, by solid objects and water surfaces. However, clouds are easily penetrated, but the signals can be blocked by dense or wet foliage. The satellite signals basically consist of (Figure 3.1-1):

- The **two L-band carrier waves.**
- The **ranging codes** modulated on the carrier waves.
- The **Navigation Message.**

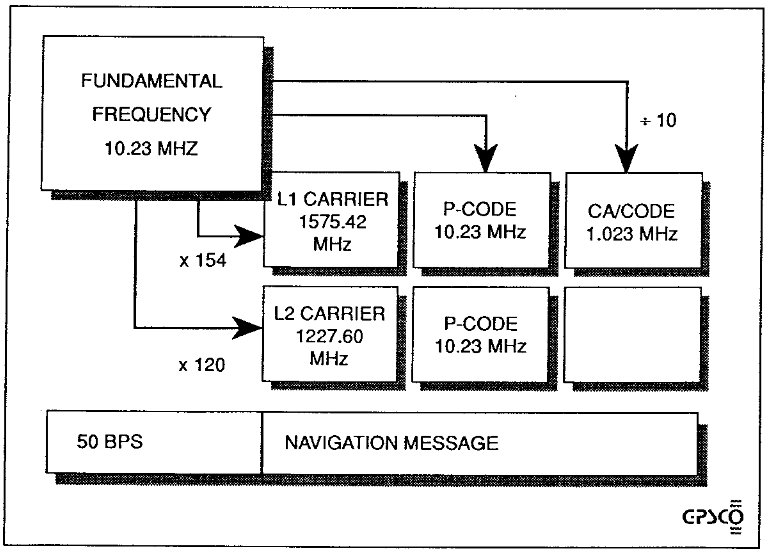


Figure 3.1-1. GPS satellite signal components.

As the name implies, the carrier waves provide the means by which the ranging codes and Navigation Message is transmitted to earth (and hence to the user). The primary function of the ranging codes is to permit the **signal transit time** (from satellite to receiver) to be determined. (This quantity is also sometimes referred to in the navigation literature as the *time-of-arrival* -- TOA.) The transit time when multiplied by the speed of electromagnetic radiation (= 299792458m/s in a vacuum) gives the receiver-satellite range. The Navigation Message is modulated on both carrier frequencies and contains the satellite ephemeris, satellite clock parameters, and other pertinent information such as general system status messages and an ionospheric delay model, necessary for real-time navigation to be performed (§3.3). Each of these signal components are described below. (Many of the following diagrams are adapted from TALBOT, 1987.)

All signal components are derived from the output of a highly stable atomic clock (Figure 3.1-2). In the operational (Block II/IIA) GPS system each satellite is equipped with two cesium and two rubidium atomic clocks. (The Block IIR satellites will be equipped with a space-qualified hydrogen maser.) The clocks generate a pure sine wave at a frequency f_0 = 10.23MHz, with a stability of the order of 1 part in 10^{13} over one day. This is referred to as the **fundamental frequency**.

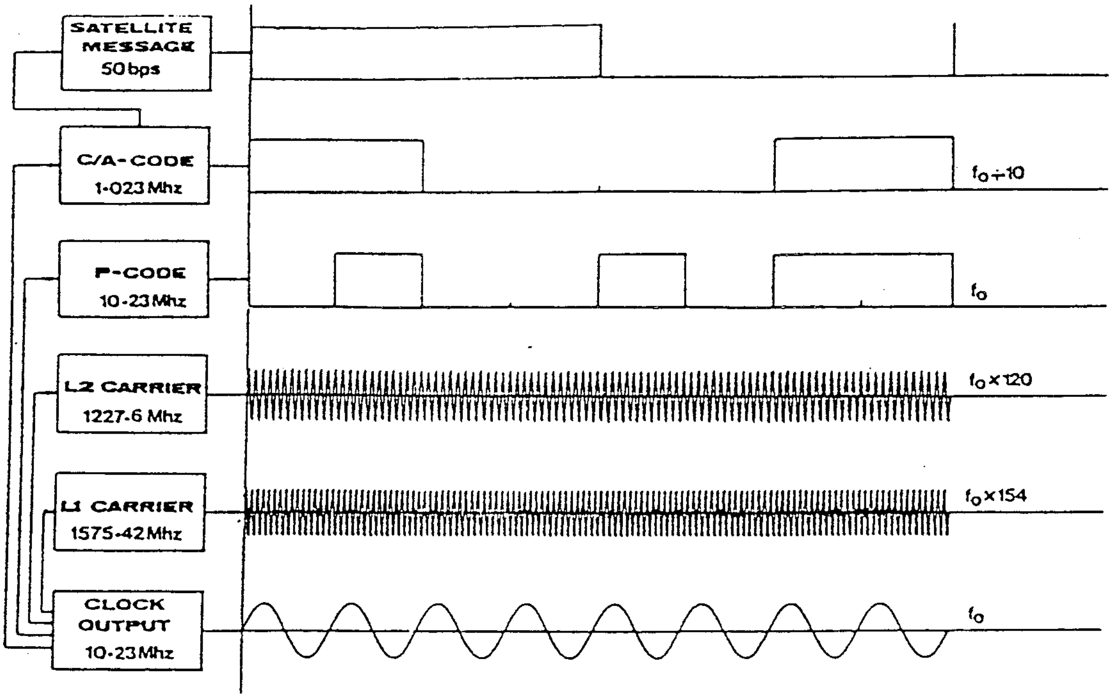


Figure 3.1-2. GPS signal component frequencies.

L-Band Carrier Waves

Multiplying the fundamental frequency f_0 by integer factors yields the two microwave L-band carrier waves L1 and L2 respectively (Figures 3.1-1 and 3.1-2). The frequency of the two waves are obtained as follows:

$$f_{L1} = f_0 \times 154 = 1575.42\text{MHz}$$
$$\text{equivalent wavelength: } \lambda_{L1} = c / f_{L1} \approx 19\text{cm}$$
$$f_{L2} = f_0 \times 120 = 1227.60\text{MHz}$$
$$\text{equivalent wavelength: } \lambda_{L2} = c / f_{L2} \approx 24\text{cm}$$

These are righthand circularly polarised radio frequency waves capable of transmission through the atmosphere over great distances, *but they contain no information.* All satellites broadcast the same frequencies (though the *received* frequencies are slightly different because of the Doppler shift). In order to give the carriers information they must be modified, or *modulated,* in some way. In the Global Positioning System there are two distinct codes used to modulate the L-band carriers, namely the ranging codes and the Navigation Message.

The L1 carrier was designed to be modulated with two codes, one intended for civilian use and the other reserved for the military, whereas the L2 carrier is modulated only with the military code. Both carriers also contain the Navigation Message (see Figure 3.1-1).

PRN Ranging Codes

Two ranging codes are used:

- ☞ The **C/A code**, the "clear/access" or "coarse/acquisition" code (sometimes also referred to as the "S code").
- ☞ The **P code**, the "private" or "precise" code, which under AS is replaced by the "Y" code (§2.4).

The C/A and P (or Y) codes can be considered as the measuring rods -- they provide the means by which a GPS receiver can measure one-way distances to the satellites. Both codes have the characteristics of *random noise*, but are in fact binary codes generated by mathematical algorithms and are therefore referred to as "pseudo-random-noise" (or PRN) codes.

Figure 3.1-3 illustrates the C/A code generation procedure based on "Gold Codes". **Tapped Feedback Shift Registers** are used to generate a sequence of "0"s and "1"s at the clock rate of 1.023 MHz. At each clock pulse the bits in the registers are shifted to the right where the contents of the rightmost register is read as output. A new value in the leftmost register is created by the modulo-2 addition (or binary sum) of the contents of a specified group of registers. In the case of the C/A code two 10-bit TFSRs are used, each generating a Gold Code: (1) the G1 (represented here as the polynomial: $1 + X^3 + X^{10}$), and (2) the G2 (represented here as the polynomial: $1 + X^2 + X^3 + X^6 + X^8 + X^9 + X^{10}$). The output of the G1 TFSR (rightmost register) is modulo-2 added to the register contents of the G2. Different combinations of the outputs of the registers of G2 (or "taps" from the register) when added to the output of the G1 code lead to different PRN codes. There are 36 unique codes that can be generated in such a straightforward manner. Figure 3.1-3 shows the first three PRN taps: PRN1 taps the contents of register 2 and 6, and adds it to the output of the G1 TFSR, PRN2 taps the contents of register 3 and 7, PRN3 taps the contents of register 4 and 8, and so on (see NAVSTAR, 1993, for a full list of PRN generating sequences).

To measure one-way range, a knowledge of the codes is required by the GPS receiver's computer. Hence, knowing which PRN code is being transmitted by a satellite means that a receiver can generate a local replica of the same code sequence. These PRN codes possess a very important attribute: *a given C/A (or P or Y) code will correlate with an exact replica of itself only when the two codes are aligned.* Furthermore, without knowledge of the ranging code sequence, the Navigation Message cannot be recovered.

The C/A codes are 1023 "chip" long binary sequences, which are generated at a rate of 1.023 million chips per second, that is at a frequency of 1.023MHz (Figure 3.1-2). Hence *the entire C/A code sequence repeats every millisecond.* The "wavelength" of the code (length of the chip) is approximately 300m, and the total sequence is therefore about 300km long. Each GPS satellite is assigned a unique C/A code (see Table 2.2-1).

The P code is a far more complex binary sequence, being approximately 266.4 days long with a chipping rate at the fundamental frequency f_0 = 10.23MHz. It is generated in an analogous manner to the C/A code, using two TFSRs. The "wavelength" of this code (length of the P code chip) is approximately 30m, *ten times the resolution of the C/A code* (Figure 3.1-4). Instead of assigning each satellite a unique code of its own, as is the case with the C/A code, the P code is allocated such that each satellite transmits a one week portion of the 266.4 day long sequence (restarting on Saturday midnight). Further details on how PRN codes are generated are given in, for example, WELLS et al (1987), NAVSTAR (1993) and KAPLAN (1996).

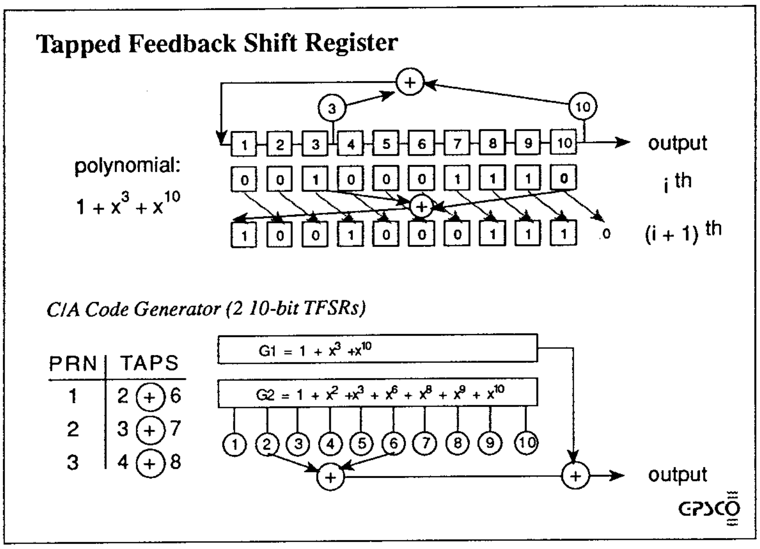


Figure 3.1-3. Generating PRN codes using two Gold Code registers.

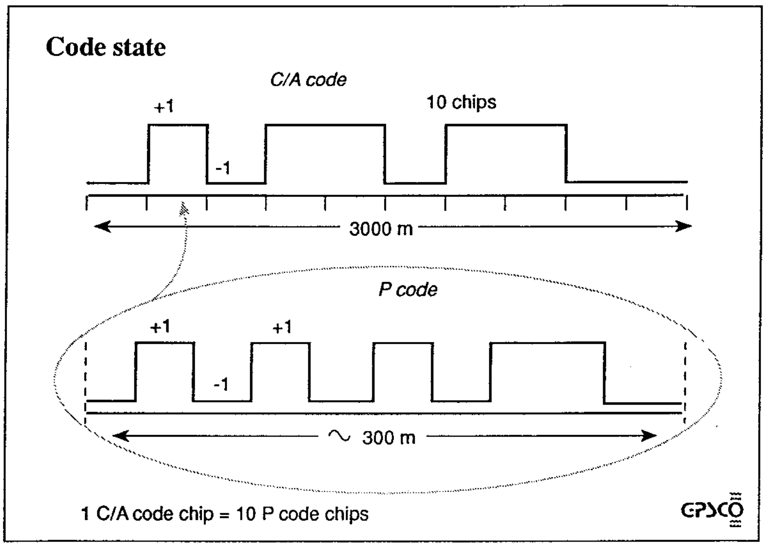


Figure 3.1-4. Examples of the C/A and P code chip sequences.

Under Anti-Spoofing the P code is encrypted through the modulation of a further secret code -- the "W code". The sum, referred to as the "Y code", is then modulated in the normal way onto the L1 and L2 carrier waves. The same P (or Y) code is modulated on both carrier waves, and any difference in signal transit time of the same PRN sequence is due to the retardation of the two L-band signals by a different amount as they travel through the ionosphere. (The effect of the ionosphere on signal propagation is essentially a function of signal frequency -- see §6.2.) **The effect of the ionosphere is to *retard* the PRN sequence, and to *advance* the carrier phase**. An approximate ionospheric delay model is provided within the Navigation Message (§3.3).

Navigation Message

In order for a GPS navigator to derive real-time position (and to make the task of the GPS surveyor easier when he comes to reduce his data), a Navigation Message is transmitted on both L-band frequencies, containing the following information (§3.3):

- ☞ Predicted **satellite ephemerides**.
- ☞ Predicted **satellite clock correction model** coefficients.
- ☞ GPS **system status** information.
- ☞ The GPS system **ionospheric model**.

The Control Segment (via the Upload Stations) uplinks this information into each satellite for subsequent transmission to all users on a regular (nominally daily) basis. The satellite message is in a binary form, like the ranging codes, but the sequence is *not* random. The message is transmitted at a rate of one bit ("0" or "1", as in a computer) every 20 repetitions of the C/A code. This corresponds to a rate of 50bps (bits per second). The entire message length is 1500 bits.

WHY IS THE SIGNAL SO COMPLICATED?

MULTI-USER SYSTEM
- ❑ One-way measurements -- *a listen-only system*

REAL-TIME POSITIONING
- ❑ Simultaneous measurements to many satellites -- *need to identify different signals*
- ❑ Unambiguous range measurements -- *need to determine signal delay*
- ❑ Satellite positions needed -- *broadcast ephemerides*

HIGH ACCURACY POSITIONING
- ❑ High frequency modulation -- *P code at 10 MHz*
- ❑ Dual-frequency -- *for ionospheric delay*
- ❑ Microwave carrier frequency -- *1.2 to 1.6 MHz*

ANTI-JAMMING REQUIREMENT
- ❑ Spread spectrum technique

MILITARY AND CIVILIAN USERS
- ❑ Need two codes and restriction on dual-frequency use -- *P and C/A codes*

3.1.1 THE TRANSMITTED SIGNAL

The signal that actually leaves a GPS satellite antenna is a combination of the three components: carrier wave, ranging codes and Navigation Message. The signal is transmitted with enough power to ensure a minimum signal power level of -160dBw at the earth's surface (the maximum it is likely to reach is about -153dBw -- NAVSTAR, 1993).

To compensate for relativistic effects, the output of the satellite's frequency generating clock -- as it would appear to an observer located at the satellite -- is 10.23MHz offset by a small amount. This frequency offset results in an actual clock output of 10.22999999543MHz (SPILKER, 1980; KAPLAN, 1996).

The generation of the signal to be transmitted is carried out in a number of steps, and relies on the fact that all the components are generated in synchrony (coherent in time), that is, the frequencies are all derived by multiplying or dividing the fundamental frequency (Figure 3.1-2). For example, the P code and C/A code transitions (from "0" to "1", or "1" to "0") occur simultaneously, to within an accuracy of 10 nanoseconds (Figure 3.1-4). Two distinct procedures for the combination of signal components can be identified.

Binary-to-binary modification of codes, whereby the binary data of the Navigation Message is modulo-2 added to the C/A code, as shown in Figure 3.1-5. This has the effect of inverting 20 C/A code repetitions whenever the data bit of the Navigation Message is equal to "1". Conversely, when the data bit is "0" the C/A code sequence remains unaffected. The same satellite message is also modulated onto the P code sequence using this modulo-2 addition procedure. The C/A code is *not*, however, modulated on the P code.

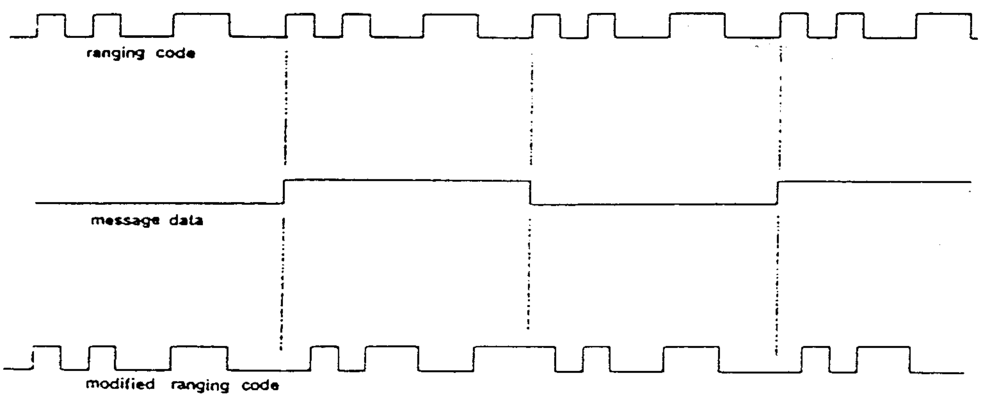


Figure 3.1-5. Ranging code modification using message signal.

Bi-Phase Shift Key Modulation (BPSK) is the technique used to added a binary signal to a sine wave carrier. This amounts to causing a 180° phase shift in the carrier at a distinct wave "trough" or "crest" each time the binary sequence undergoes a transition from "0" to "1", or "1" to "0". This is illustrated in Figure 3.1-6. The modified P code (P code plus Navigation Message) is used to modulate both the L1 and L2 carriers, and the modified C/A code (C/A code plus Navigation Message) is only used to modulate the L1 carrier. This creates a **spread spectrum ranging signal**.

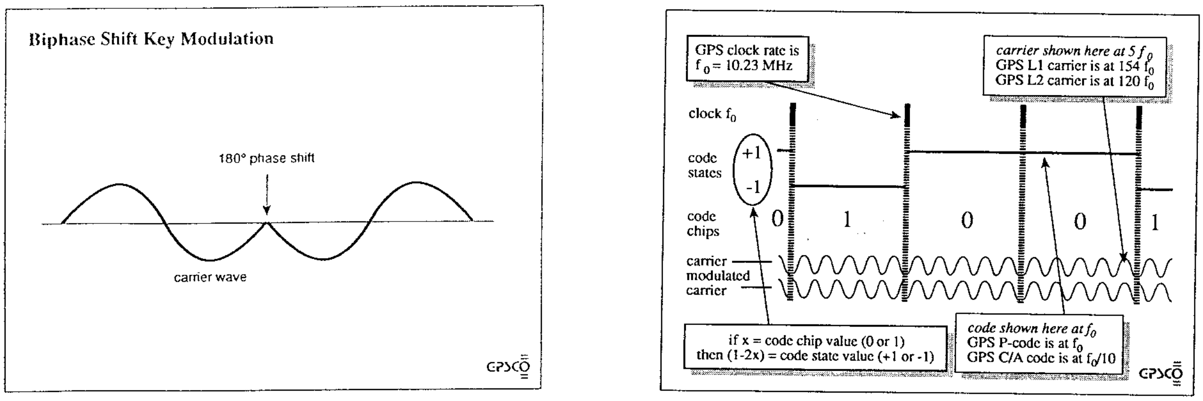


Figure 3.1-6. Bi-phase shift key modulation of code states on carrier.

Phase shifting of the carrier results in a spreading of power between ±10.23MHz of centre frequency due to the P code BPSK, and ±1.023MHz due to the C/A code BPSK (the resulting waveform is shown in Figure 3.1-7).

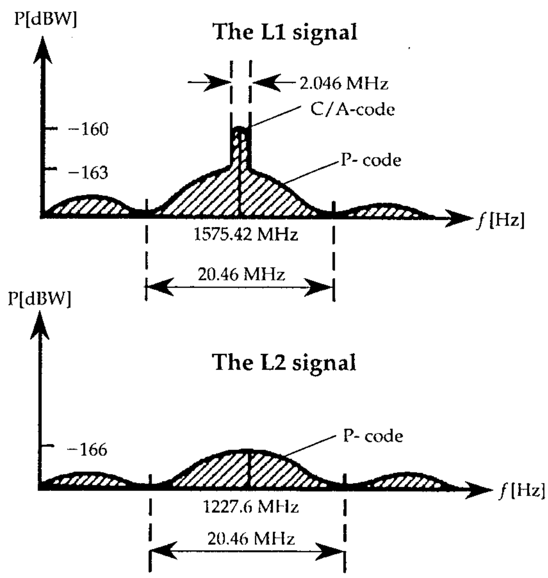


Figure 3.1-7. The GPS signal frequency spectrum.
(CLARKE, 1994)

The signals on L1 are more complex, as the L1 wave modulated with the modified C/A code is in phase quadrature (at 90°) to the L1 carrier used for the P code modulation. Figure 3.1-8 illustrates the situation with regard to the L1 signal.

Mathematically, the complete signal leaving the satellite antennas can be represented by:

$$A_cC(t)D(t)\sin(2\pi f_{L1}+\phi_c) + A_pP(t)D(t)\sin(2\pi f_{L1}+\phi_{p1}) + A_pP(t)D(t)\sin(2\pi f_{L2}+\phi_{p2})$$

where:

- A_c and A_p are the amplitudes of the C/A and P code modulations,
- $C(t)D(t)$ and $P(t)D(t)$ are the modified C/A and P code PRN sequences, and
- ϕ_c, ϕ_{L1} and ϕ_{L2} are the phases (0 to 2π) of the C/A code signal, the L1 P code signal and the L2 P code signal, respectively. These are sometimes known as the *"code-phase"*.

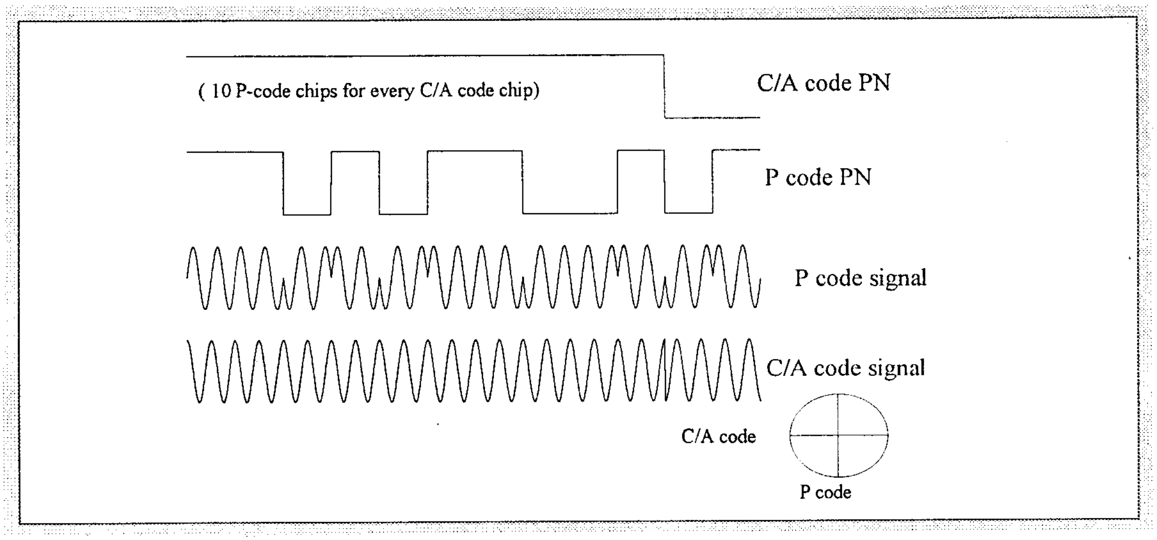


Figure 3.1-8. The PRN modulations on the L1 signal.

Note, the first component is the modified C/A code signal, the second component is the modified P code on the L1 carrier, and the third component is the modified P code on the L2 carrier. The final transmitted signals are illustrated in Figure 3.1-9 (note: under AS the additional secret W code signal is modulated on the P code to create the Y code).

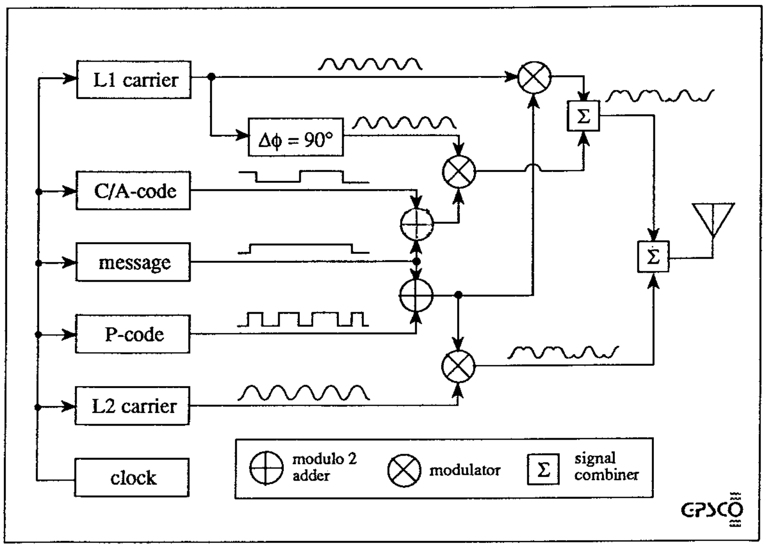


Figure 3.1-9. The composite GPS signal transmitted by the satellite.

More details of the signal structure can be found in SPILKER (1980) and KAPLAN (1996).

3.2

THE GPS MEASUREMENTS

3.2.1 GPS RECEIVER OPERATION

There are two range-type measurements that can be made on the GPS signals (LANGLEY, 1993):

- ☞ **Pseudo-ranges**, and
- ☞ **Carrier phase** observations.

Both are a product of the *operation* of the GPS receiver (that is, the acquisition and maintenance of signal tracking), both are used for GPS *navigation* (position, velocity and time -- PVT -- determination), and both have a role in the specialised data processing that characterises GPS *surveying*. Before studying these measurements it is useful to consider the overall GPS hardware tracking operation (in a much abbreviated form!).

The received satellite signal level is actually less than the background noise level, hence correlation techniques are used to obtain the satellite signals. A typical satellite tracking sequence begins with the receiver determining which satellites are visible above the horizon. Satellite visibility is estimated from predictions of present PVT, and on the stored satellite almanac information residing within the receiver. (If no stored almanac information exists, or only a very poor estimate of PVT is available, the receiver will carry out a "sky search", attempting to randomly locate and lock onto a signal. The receiver will then decode the Navigation Message and read the almanac information about all the other satellites in the constellation.) A carrier-tracking loop is used to track the carrier frequency while a code-tracking loop is used to track the C/A and/or P code signals. The two tracking loops have to work together in an iterative manner, aiding each other in order to acquire and track the satellite signals.

The receiver's carrier-tracking loop will locally generate an L1 carrier frequency (or L2 if the receiver is capable of tracking this frequency) which differs from the received carrier signal due to a Doppler offset of the carrier frequency. This Doppler offset is proportional to the relative velocity along the line-of-sight to the satellite. In order to maintain lock on the carrier, the carrier-tracking loop must, in effect, adjust the frequency of the receiver-generated carrier until it matches the incoming carrier frequency. The amount of this offset is the "beat" frequency which can be processed to give a periodic carrier phase measurement. The derivative of this phase measurement is the "Doppler" measurement, which is used to determine the receiver's velocity.

What role does the code-tracking loop play in this process? In order for the carrier-tracking loop to acquire the incoming satellite signal in the first place the carrier signal must be made visible above the background noise. This is generally done by the code-tracking loop using the **code-correlating technique** to "reconstruct" the carrier wave (see discussion below under "Carrier Phase Measurements"). A by-product of code-tracking are the pseudo-range measurements.

3.2.2 PSEUDO-RANGES

Ranging with the PRN Codes

Consider for a moment a perfect system where all satellite clocks are synchronised to the same time system: *GPS Time*. Furthermore, the ground receiver's clock also maintains the same synchronisation, and none of the clocks drift with respect to the GPST scale. Now suppose the satellite starts transmitting its L1 carrier (modulated with the combined C/A code and navigation data), and at the same instant the receiver begins generating the C/A code corresponding to that particular satellite (see Figure 3.2-1). Under these circumstances, the satellite and receiver generated C/A codes would be output in unison. However, when the satellite signal is received it will be lagging the receiver generated code due to the **signal transit time**. Multiplying the time offset required to align the two code sequences within the code-tracking loop (one from the received satellite signal and the other an internally generated code) by the speed of electromagnetic radiation yields the satellite-receiver range.

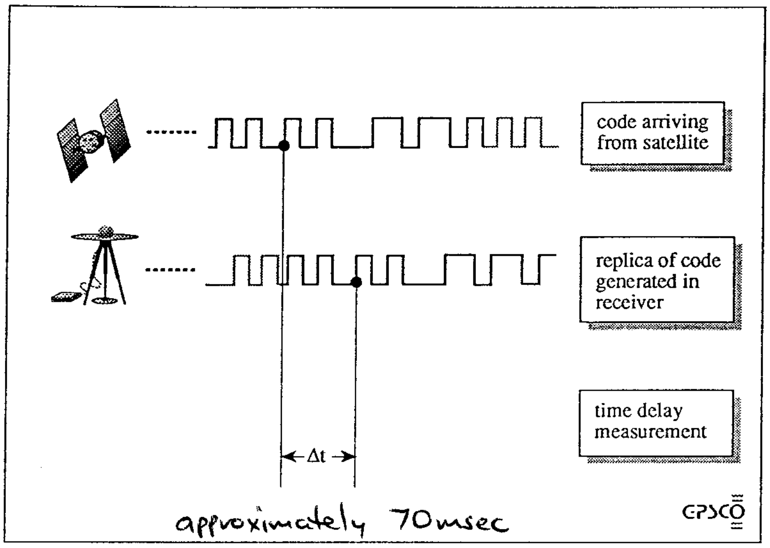


Figure 3.2-1. One-way ranging using PRN codes.

Measuring ranges simultaneously in this fashion to three satellites would fix the receiver's position at the intersection of three spheres of known radii (the satellite ranges), centred at each satellite whose coordinates can be calculated from the Navigation Message, as illustrated in Figure 1.4-3.

In reality the situation is more complex:

(1) Receivers are generally equipped with quartz crystal clocks that do not necessarily keep the same time as the more stable satellite clocks (these clocks can be approximately synchronised to GPST using the clock correction model transmitted in the Navigation Message -- §3.3). Consequently **each range is contaminated by the receiver clock error**. This is the reason this range measurement is referred to as a "**pseudo-range**". Hence, to determine position using pseudo-range data, a minimum of four satellites must be tracked and the position determination problem is therefore one requiring the solution of four equations (one per observation), each containing four unknowns: the three-dimensional position components and the receiver-clock offset

(from GPST). This is the basis of GPS (real-time) navigation as described in §1.4.

(2) Ranging (and hence receiver position determination) can be carried out using the C/A code or the P code. P code ranging can be performed on either of the two frequencies, or a linear combination of the L1 and L2 pseudo-ranges that largely eliminates the bias due to ionospheric refraction (§6.2). Furthermore, the C/A code resolution is "coarser", and hence the C/A derived ranges are subject to greater measurement "noise". The absence of a C/A code on L2 is intentional, as one of the accuracy limitations of the GPS system. Others are the ability under the policy of **Anti-Spoofing** to restrict access to the secret Y code to only "authorised" users (such as the military and those working in the "national interest" of the U.S. and its allies), and implementation of the policy of **Selective Availability**. (See §2.4 for further discussion.)

(3) This distinction between the ranging codes, and the associated policies for their use (in peacetime and in times of global emergencies), results in the provision of two GPS positioning services: the **Precise Positioning Service** based on P code (dual-frequency) ranging, and the **Standard Positioning Service** based on single frequency C/A code ranging (§2.4).

Recovery of PRN Ranging Codes from the Incoming Signals

The PRN codes are accurate time marks that permit the receiver's navigation computer to determine the time-of-transmission of any portion of the satellite signal. Before examining this in detail it is necessary to consider, in general terms, how the incoming satellite signal is processed within the GPS receiver. Within the electronics of a receiver tracking "channel" the L1 carrier modulated by the C/A code is mixed with a locally generated replica C/A code. The local C/A code is generated on a different time scale to that of the incoming C/A code (due to non-synchronisation of the receiver clock to GPST, and the travel time of the signal from the satellite to the receiving antenna). Alignment of the incoming signal with the receiver generated C/A code is carried out by the code-tracking loop, or the *"delay-lock loop" electronics.* As soon as the incoming signal and the receiver C/A code sequences are aligned within the receiver (by sliding the received code sequence against that internally generated sequence), the "0"s and "1"s of the two codes cancel, leaving the incoming carrier signal modulated only by the binary Navigation Message. This process is summarised in Figure 3.2-2.

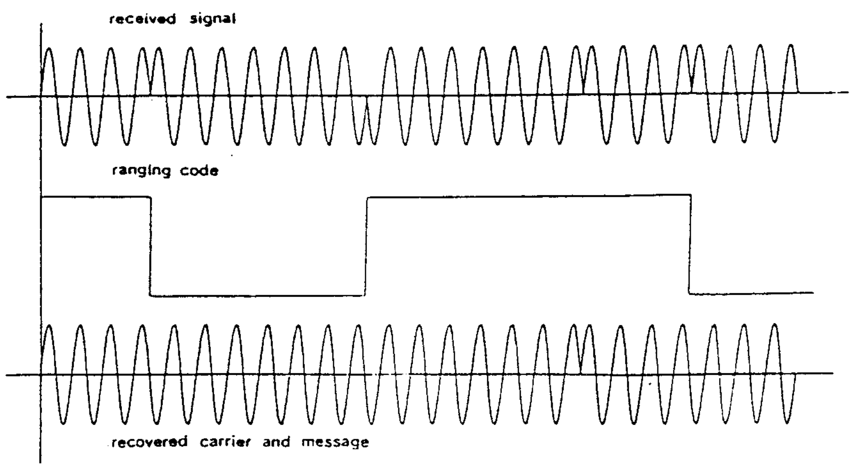


Figure 3.2-2. Recovery of ranging code.

Because of the complexity of the P code sequence (its length and higher chipping rate), a *sliding correlation* technique as described above for the C/A code cannot be used in practice without a very good estimate of GPST and receiver position. Typically a P code receiver must acquire lock on the C/A code first, then use a timing mark known as the "Handover Word", contained within the Navigation Message, to enable the correct portion of the P code to be generated within the receiver and thus initialise the P code delay-lock loop.

Extraction of the Pseudo-Ranges

As mentioned already, the extraction of the pseudo-range, or more precisely, the determination of the amount by which the receiver generated PRN code must be shifted to align it with the incoming signal, is carried out with the aid of a PRN code correlator in some delay-lock loop scheme (see, for example, TALBOT, 1987; LANGLEY, 1993; KAPLAN, 1996). *How accurate is this carried out?* The C/A code has a chip rate of 1.023Mbps, corresponding to a wavelength of about 300m (speed of light divided by the frequency). The P (or Y) code, on the other hand, has a chip rate of 10.23Mbps, and hence a wavelength of about 30m.

As a "rule-of-thumb": the alignment of the incoming and receiver generated codes is generally possible to within about 1-2% of the chipping rate, hence the measurement precision of C/A code ranging is of the order of 3-5m, and for P code ranging it is of the order of 0.3-0.5m. (Modern "narrow correlator" technology has demonstrated 10 times better correlation performance for the C/A code than that above.)

The main advantages gained by using the P code therefore are:

- ❑ Because of the higher chipping rate, and hence higher measurement precision, P code ranging translates into a more accurate position fix.
- ❑ The P code is modulated on both the L1 and L2 carriers, hence the ionospheric signal delay can be overcome.
- ❑ P code receivers are better suited to high dynamic environments and resist signal jamming better than C/A code receivers.

Both the P and C/A code ranges are susceptible to **multipath** (though the susceptibility is inversely proportional to the signal frequency). Multipath is caused by extraneous reflections from nearby metallic objects or water surfaces reaching the antenna and causing the signal measurement process to become noisy than normal. Some characteristics of multipath are (§6.2):

- Multipath can cause "jumps" in the signal measurement of the order of its (effective) wavelength. For pseudo-ranges this could mean tens or hundreds of metres, but of the order of only centimetres for carrier phase measurements.
- Multipath is receiver-satellite geometry dependent, and the causes of multipath tend to be permanent features (metallic fences, buildings, chimneys, superstructure, water surfaces, etc.), hence the multipath effect will generally repeat on a daily basis at the same receiver site -- see Figure 8.1-2.
- As the receiver-satellite geometry changes (and hence the angle of incidence and reflection of the signal with respect to the reflective surface changes), the multipath effect changes, and generally "averages out" over a period from several minutes to a quarter of an hour, or more. This makes static GPS positioning more accurate and reliable than in the case of positioning a moving GPS receiver (using either pseudo-range or carrier phase data).

3.2.3 CARRIER PHASE MEASUREMENTS

The wavelengths of the carrier waves are very short -- approximately 19cm for L1 and 24cm for L2 -- compared to the C/A and P code chip lengths. Assuming a measurement resolution of 1-2% of the wavelength, this means that **carrier phase can be measured to millimetre precision compared with a few metres for C/A code measurements (and several decimetres for P code measurements)**. Unfortunately, a phase measurement is "ambiguous" as it cannot discriminate one (either L1 or L2) wavelength from another. In other words, time-of-transmission information for the L-band signal cannot be imprinted onto the carrier wave as is done using PRN codes (this would be possible only if the PRN code frequency was the same as the carrier wave, rather than 154 or 120 times lower in the case of the P code, and 1540 or 1200 times lower for the C/A code). The basic phase measurement is therefore in the range 0° to 360° (Figure 3.2-3). It is nevertheless the basis for GPS surveying, and high precision kinematic positioning.

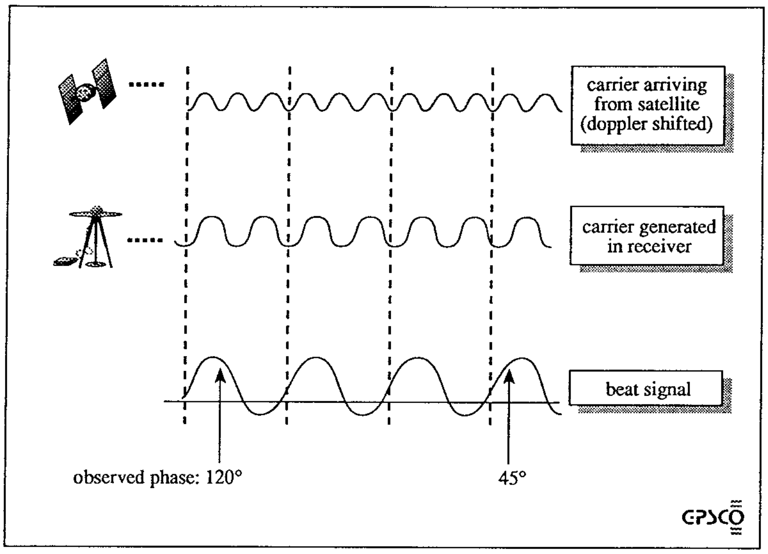


Figure 3.2-3. Carrier phase measurements.

There are essentially two means by which the carrier wave can be recovered from the incoming modulated signal:

- **Reconstruct** the carrier wave by removing the ranging code and broadcast message modulations.
- **Squaring**, or otherwise processing the received signal without using a knowledge of the ranging codes.

In the first technique the ranging codes (C/A and/or P code) must be known. The extraction of the Navigation Message can then be easily performed by reversing the process by which the bi-phase shift key modulation was carried out in the satellite. In the latter method no knowledge of the ranging codes is required (see WELLS et al, 1987; HOFMANN-WELLENHOF et al, 1994; for details). More complex signal processing is required to make carrier phase measurements on the L2 signal under conditions of Anti-Spoofing (see §4.2).

Extraction of Carrier Beat Phase: Reconstructing the Carrier Wave

This is the technique used within code-correlating receivers. When the spread spectrum signal (Figure 3.1-7) is received at the GPS antenna, the signal power is below the background noise (Figure 3.2-4). After the ranging code modulations are removed by the procedure described above, the satellite signal collapses into the original very narrow carrier frequency band and signal power is again boosted well above the background noise.

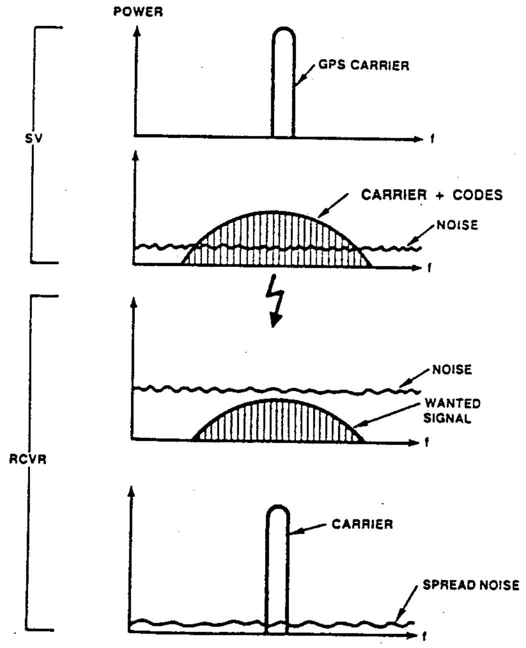


Figure 3.2-4. Spreading and de-spreading the spectrum of the carrier wave. *(NATO, 1991)*

By mixing a locally generated sine wave at the same frequency as the "reconstructed" received carrier (modulated only by the Navigation Message), the broadcast message can be extracted. The incoming and receiver-generated sine waves are continuously aligned within a "phase-lock loop" (§4.1). Periodic sampling of the phase of the local carrier provides the carrier beat phase observable (Figures 3.2-3 and 3.2-5), which although useful for some applications such as the "phase smoothing" of pseudo-ranges, is still not suitable for survey applications. A much more useful carrier phase observable can be constructed through the "integration" of carrier phase measurements (see below).

Measurement of carrier beat phase on L2 by this technique requires a knowledge of the P code generating algorithm. Under the policy of Anti-Spoofing, the Y code is secret and hence cannot be used in this code-correlating mode. The easiest option for GPS instrument manufacturers is to use the "squaring" technique (or some variation of it) to make L2 phase measurements. However, **the primary advantage of the code-correlating approach is that it results in a far better signal-to-noise ratio, and hence better quality measurements, than any other signal processing technique.**

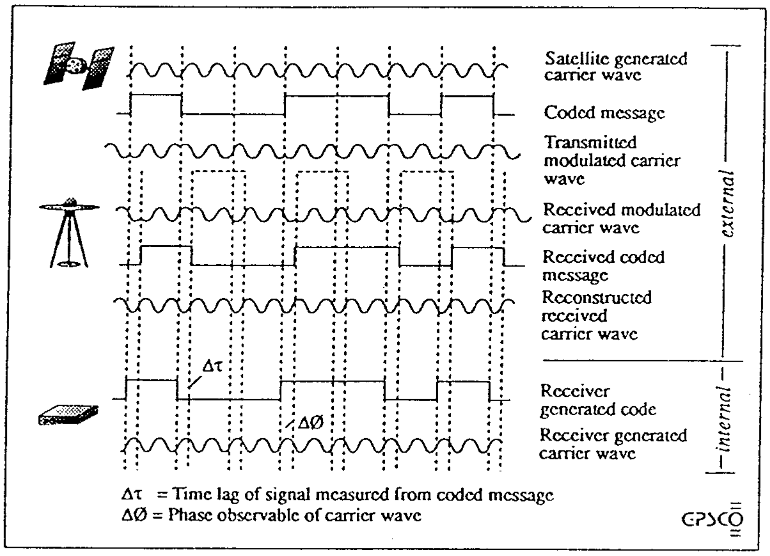


Figure 3.2-5. Reconstructing the carrier wave and extraction of pseudo-range data.

Extraction of Carrier Beat Phase: "Squaring" the Carrier Wave

In principle, the operation of a squaring receiver is very simple. The incoming signal is first converted to an intermediate frequency (IF) signal. The carrier, or rather the beat frequency carrier wave, is obtained simply by *squaring this signal.* Any phase inversions in the IF signal due to the PRN codes or message are removed. (This happens because a phase inversion is a change in the IF signal amplitude from "+1" to "-1", or from "-1" to "+1", and the instantaneous amplitude is either "+1" or "-1". Squaring the signal results in a signal with constant amplitude of unity, and hence the codes and message information are lost.) However, *squaring the signal also squares the noise.*

Aside from resulting in a noisier measurement, the squared carrier wave measurement is made on a carrier wave of double frequency. That is, the effective wavelength is of the order of 9.5cm on L1 and 12cm on L2. Figure 3.2-6 illustrates this measurement scheme.

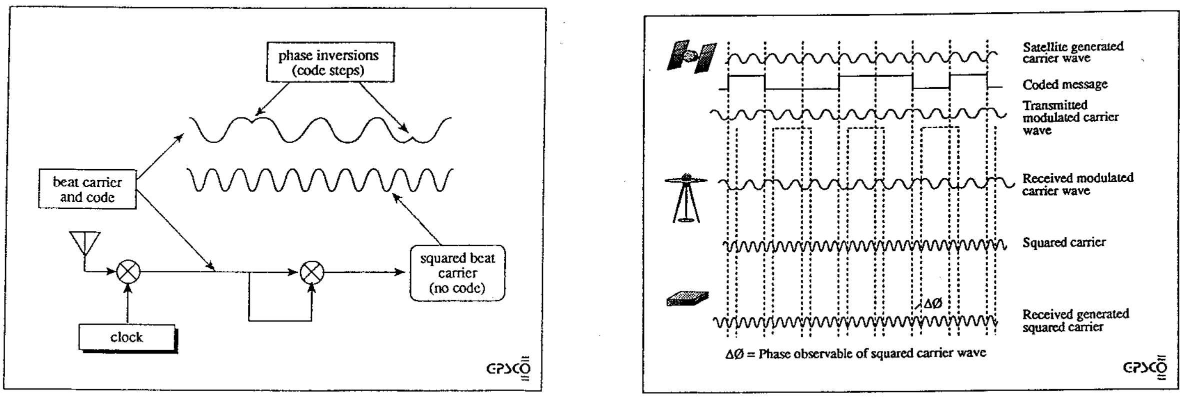


Figure 3.2-6. Extracting carrier phase from incoming GPS signals by carrier wave squaring.

Integrated Carrier Beat Phase

Raw carrier phase measurements are generally the by-product of all GPS receivers. These phase measurements cannot be used as "range" observations because they are *ambiguous*, and furthermore, the ambiguity changes continuously. The ambiguity is therefore a function of both the receiver channel tracking the satellite, and time. (This is analogous to making terrestrial distance measurements using only the "reader" portion of a steel band.) It is very difficult to resolve the continuously changing unknown ambiguity in a navigation solution (as can be done in the case of the receiver clock bias).

But all is not lost! If it were possible to keep track of the number of whole wavelengths of the carrier wave as it is sampled within, for example, a phase-lock loop, then the **integrated carrier phase** observation could be generated:

$$\phi_j^i(T_j) = \Delta\phi_j^i(T_j) + [\, C_R(T_j) + C_{Ro} \,] \qquad (3.2\text{-}1)$$

where $\Delta\phi_j^i$ is the *fractional phase* (measured as an angle in the range 0° to 360°, where 360° corresponds to about 19cm for the L1 phase and 24cm for the L2 phase, see Figure 3.2-3), and C_R is the current reading on a zero-crossing "counter", that only registers the number of whole cycles since lock-on when the counter had an initial value of C_{Ro} (usually zero). The term in square brackets is therefore an *integer*. The additional electronics to count the whole cycles since lock-on is the identifying characteristic of GPS "surveying" receivers.

The relationship between $\phi_j^i(T_j)$ and the range $\rho_j^i(T_j)$ is:

$$(\, f_0 / c \,)\, \rho_j^i(T_j) = \phi_j^i(T_j) + n_j^i + v(T_j) \qquad (3.2\text{-}2)$$

where n_j^i is the ambiguity term, and v contains all the biases and errors affecting this measurement (§6.2). (f_0 / c) scales range into units of cycles. Note that n_j^i is assumed to be constant over time, for a particular receiver-satellite combination, as illustrated in Figure 3.2-6. In order to convert this phase observation into range, the cycle ambiguity has to be determined. If the integer n_j^i can be correctly determined, then the resulting "phase-range" (or "carrier-range") will be a very precise range (at the level of a few millimetres).

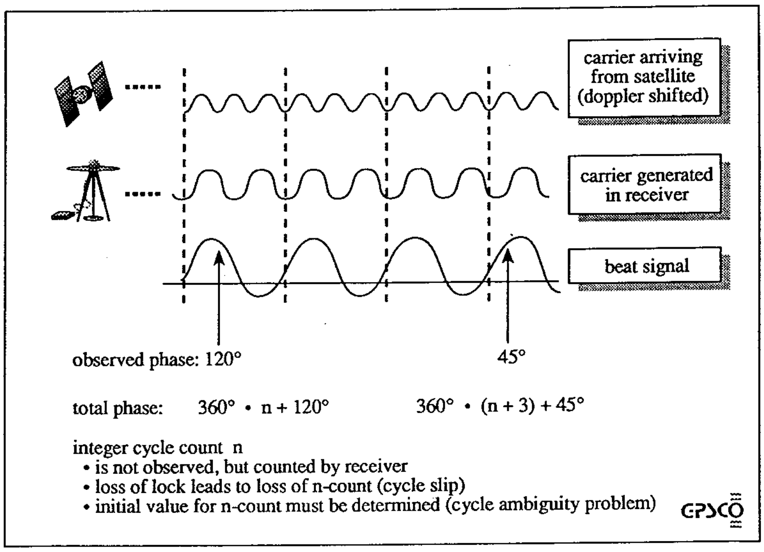


Figure 3.2-7. Integrated carrier phase and the ambiguity term.

3.3

THE GPS NAVIGATION MESSAGE

By mixing a locally generated sine wave at the same frequency as the received carrier *after* the ranging codes have been removed, the binary Navigation Message can be obtained -- see Figure 3.3-1. The extraction of the Navigation Message can therefore only be carried out in a code-correlating receiver, and not one that employs carrier wave "squaring" technique.

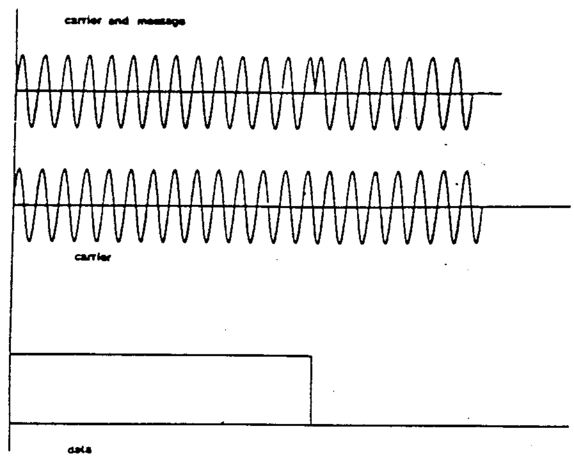


Figure 3.3-1. Recovery of the Navigation Message.

Message Characteristics and Content

The contents of the Navigation Message, and the manner in which the data "blocks" or "frames" are arranged is illustrated in Figure 3.3-2. Further information on the content of the Navigation Message can be found in VAN DIERENDONCK et al (1980), LANGLEY (1991b) and NAVSTAR (1993).

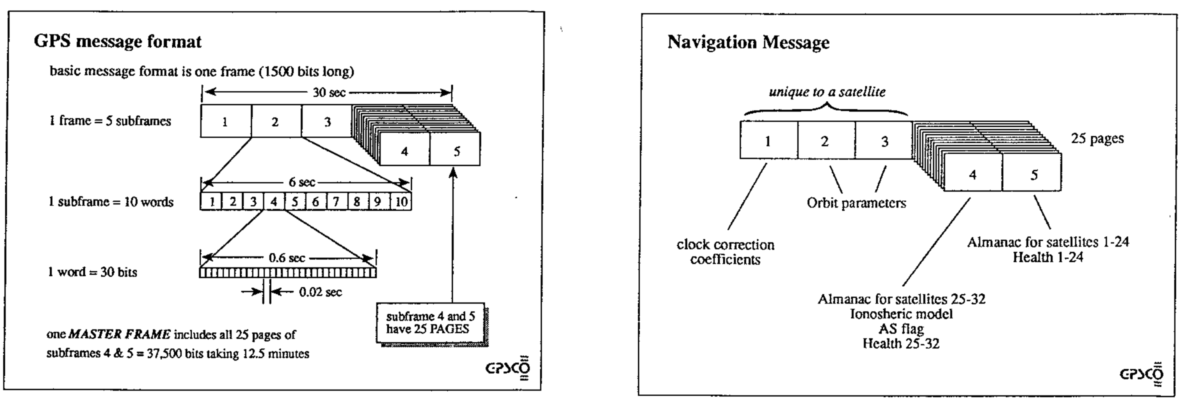


Figure 3.3-2. The GPS message format and content.

3.3.1 SATELLITE CLOCK ERROR DATA

The GPS clocks are free-running and their rate is monitored against GPS Time (GPST). GPST is itself kept synchronised to UTC as defined by the U.S. Naval Observatory.(after taking into account the leap second offsets -- §2.1), see LANGLEY (1991d). The clock behaviour so determined is used to occasionally reset the satellite clock so that it is kept within 1 millisecond of GPST, and is also made available to all GPS users via clock error coefficients in a polynomial form:

$$\varepsilon = a_0 + a_1 (t - t_{oc}) + a_2 (t - t_{oc})^2 \qquad (3.3\text{-}1)$$

where:

- a_0 is the clock bias term,
- a_1 is the clock drift term,
- a_2 is the clock drift-rate,
- t is the satellite clock time (seconds in the GPS week), and
- t_{oc} is the reference epoch for the coefficients (seconds in the GPS week).

The clock bias, drift and drift-rate are explicitly determined in the same procedure as the determination of the satellite ephemeris. What is therefore available to users is really a **prediction** of the clock behaviour some time into the future. As the random deviations of even cesium and rubidium oscillators are not predictable, such deterministic models of satellite clock error (that is, non-synchronisation with GPST) are accurate to about 20 nanoseconds, or approximately six metres in equivalent range. Typical values for these coefficients are:

SV PRN number	a_0 (μsec)	a_1 (μsec/day)	a_2 (μsec/day^2)
1	-73.208	-0.196	0.0
2	-58.831	-0.157	0.0
3 *	-21.739	-1.306	0.0
4	3.579	0.131	0.0
5	14.151	0.196	0.0
7	42.269	0.206	0.0
12 *	18.436	-8.418	0.0
14	3.454	0.020	0.0
15	45.545	0.216	0.0
16	-71.942	0.029	0.0
17	-38.796	-0.079	0.0
18	-3.796	-0.010	0.0
19	132.122	0.216	0.0
20	36.064	0.018	0.0
21	-14.313	-0.059	0.0
22	77.125	0.304	0.0
23	2.402	0.029	0.0

* Block I satellites

Note, under the policy of Selective Availability now implemented on all Block II/IIA satellites, the satellite clock error data are intentionally falsified and hence the residual satellite clock error is of the order of several dekametres (corresponding to about a hundred nanoseconds accuracy in the signal time delay).

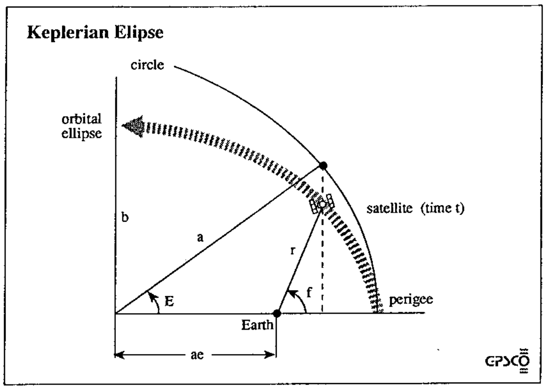
3.3.2 BROADCAST EPHEMERIS DATA

Forces of gravitational and non-gravitational origin perturb the motion of the GPS satellites, causing the orbits to deviate from a Keplerian ellipse in inertial space (Figure 3.3-3) -- defined by the six elements a (semi-major axis), e (eccentricity), i (inclination), Ω (longitude of the ascending node), ω (angle of perigee), and f (true anomaly) or E (eccentric anomaly) (LANGLEY, 1991b). The perturbations are characterised by periodic and secular components, and must be continually determined through the analysis of tracking data. In the case of the GPS broadcast ephemerides, this procedure is a three-step process:

(1) An off-line orbit determination is performed through the analysis of tracking to generate a *reference orbit for each satellite.* This is an initial estimate of the satellite trajectory computed from about one week's tracking data collected by the five Control Segment monitor stations (§2.2).

(2) An on-line daily updating of the reference orbit within a Kalman filter as new tracking data are added. This provides the *current estimates of the satellite trajectory which is used to predict the future orbit.*

(3) The ephemeris is derived by extrapolating the estimated orbit for 1 to 14 days into the future. To obtain the necessary broadcast information, *curve fits are made to 4 to 6 hour portions of the extrapolated ephemeris*, and hourly orbit parameters determined.



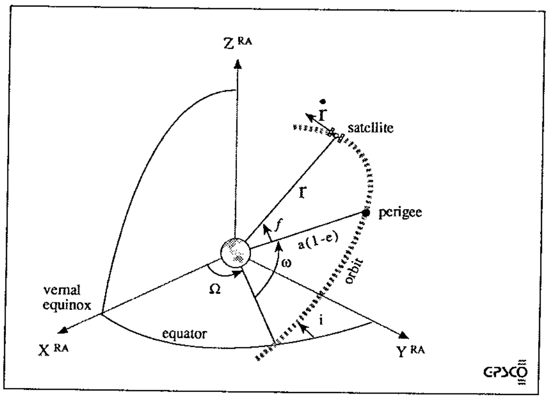


Figure 3.3-3. The Keplerian ellipse in space.

In order to adequately describe the GPS orbits during the interval of time for which the ephemeris information is transmitted (at least an hour), a representation based on *Keplerian elements plus perturbations* is used -- see Table 3.3-1 and Figure 3.3-4.

Note that the satellite ephemerides are broadcast two hours in advance of the epoch for which they were calculated (and valid). Generally there is a daily update, though sometimes more frequent. Therefore the portions of ephemeris data for the second through fourteenth day are not normally transmitted, except when upload is not possible. At the same time each satellite's clock state is estimated, then extrapolated into the future, and the information is formatted into the Navigation Message.

The parameters are given in terms of the ephemeris reference time t_{oe} -- nominally the centre of the transmission period. (GPS system time, as derived from the coded signals, and t_{oe} is measured in seconds from the start of the GPS week, at Sunday midnight.) Although the Keplerian representation has physical meaning, additional parameters are required to model the perturbations about the Keplerian orbit. The following parameters are introduced:

- Δn represents the secular change in the mean anomaly (or argument of perigee).
- $\dot{\Omega}$ and $\dot{i}$ describe the secular drift of the ascending node in the equatorial plane (mainly due to the "flattening" of the earth) and the change in inclination with time, respectively.
- C_{uc}, C_{us}, C_{rc}, C_{rs}, C_{ic}, C_{is} are the amplitudes of the cosine and sine harmonic (periodic) correction terms to the alongtrack, radial and angle of inclination (≈crosstrack) values at t_{oe}, respectively.

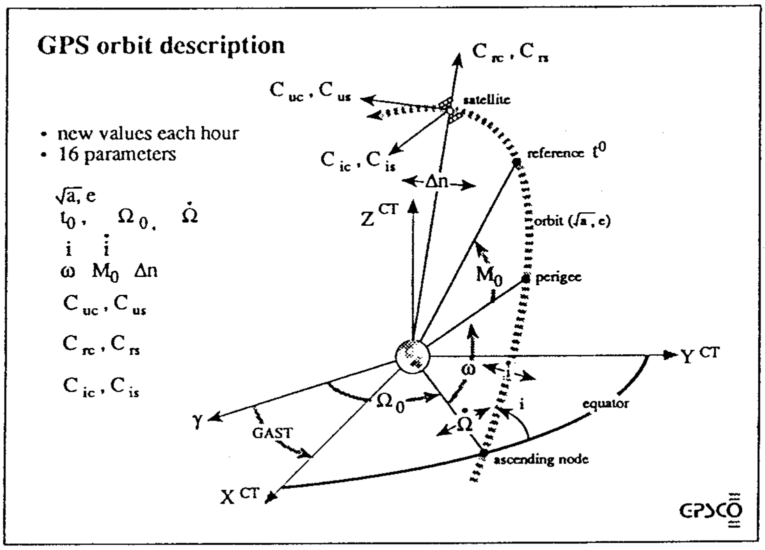


Figure 3.3-4. Broadcast Ephemeris orbital representation.

Table 3.3-1. Broadcast Ephemeris representation parameters.

M_0	Mean anomaly at reference time
Δn	Mean motion difference from computed value
e	Eccentricity
$\sqrt{a}$	Square root of the semi-major axis
Ω_0	Right Ascension at reference time
i_0	Inclination angle at reference time
$\dot{i}$	Rate of change of inclination angle
ω	Argument of Perigee
$\dot{\Omega}$	Rate of change of Right Ascension angle
C_{uc}, C_{us}	Amplitude of the cosine and sine harmonic correction terms to the argument of latitude
C_{rc}, C_{rs}	Amplitude of the cosine and sine harmonic correction terms to the orbit radius
C_{ic}, C_{is}	Amplitude of the cosine and sine harmonic correction terms to the angle of inclination
t_{oe}	Ephemeris reference time (seconds in the GPS week)

These parameters are obtained from a curve fit to the predicted satellite ephemeris over an interval of 4 to 6 hours. They are not true Keplerian elements as they only describe the ephemeris over the interval of applicability and not for the whole orbit. (Although only intended for use during the transmission period, they do, however, describe the orbit to the required accuracy over intervals of 1.5 to 5 or more hours.) While it is difficult to judge the accuracy of the broadcast ephemerides (the procedures are undergoing continual improvement), it is expected that it is on average better than 10m.

A sample of the broadcast Navigation Message for one satellite is presented in Table 3.3-2.

Table 3.3-2. Sample Broadcast Ephemeris message for satellite PRN26.

PRN	**Date/time of clock t_{oc}**	**a_0 (μsec)**	**a_1 (μsec/day)**	**a_2 (μsec/day^2)**
26	92 07 26 08 51 44.0	1.635868102312D-06	-3.410605131648D-13	0.000000000000D+00
	Age of ephemeris (sec)	**C_{rs} (m)**	**Δn (rad/sec)**	**M_o (rads)**
	2.170000000000D+02	1.128125000000D+01	4.738768932810D-09	5.863412966114D-02
	C_{uc} (rads)	**e**	**C_{us} (rads)**	**$\sqrt{a}$ ($m^{0.5}$)**
	4.898756742477D-07	8.141674217768D-03	9.039416909218D-06	5.153610269547D+03
	t_{oe} (secs in GPS wk)	**C_{ic} (rads)**	**Ω_o (rads)**	**C_{is} (rads)**
	3.190400000000D+04	-7.636845111847D-08	-1.371976967685D+00	-4.470348358154D-08
	i_o (rads)	**C_{rc} (m)**	**ω (rads)**	**$\dot{\Omega}$ (rad/sec)**
	9.603279283700D-01	2.067812500000D+02	-1.491259472097D+00	-8.059264366977D-09
	$\dot{i}$ (rad/sec)		**GPS week number**	
	-4.760912775126D-10	1.000000000000D+00	6.550000000000D+02	0.000000000000D+00

The procedure set out in Table 3.3-3 allows the user to derive the earth-centred / earth-fixed Cartesian coordinates from the broadcast orbital parameters in Table 3.3-1. The earth-fixed reference system is presently based on the WGS84 (prior to January 1987 the system used was WGS72). The algorithm in Table 3.3-3 is implemented within every code-correlating GPS receiver, be it the "navigation" or "surveying" variety. Further information on this can be found in VAN DIERENDONCK et al (1980), and NAVSTAR (1993).

3.3.3 IONOSPHERIC MODEL

To aid SPS single receiver real-time GPS navigation, the "Klobuchar model" is often used to compute the zenith time delay to the transmitted L1 signal (IBID, 1993):

$$t_{zion} = DC + A.\cos\left[\frac{2\pi.(t - t_0)}{P}\right] \qquad (3.3\text{-}2)$$

where:

t	is the local time at receiver (in seconds),
t_0	is the local time of maximum ionospheric correction (say 14:00hrs).
t_{zion}	is the time delay due to the ionosphere in the zenith direction (in seconds),
DC	is the base ionospheric time delay (taken as 5×10^{-9} s),
A	is the amplitude of the ionospheric delay function (in seconds), and
P	is the period of the ionospheric delay function (in seconds).

The quantities A and P are computed from the Navigation Message *alpha* and *beta* coefficients:

$$A = \sum_{i=0}^{3} alpha_{i-1} \cdot \phi_{om}{}^{i-1} \tag{3.3-3a}$$

$$P = \sum_{i=0}^{3} beta_{i-1} \cdot \phi_{om}{}^{i-1} \tag{3.3-3b}$$

where ϕ_{om} is the **geomagnetic latitude** of the ionospheric subpoint (expressed in semi-circles):

$$\phi_{om} = \phi_i + 0.064.\cos(\lambda_i - 1.617) \tag{3.3-4}$$

ϕ_i and λ_i are the user's latitude and longitude in semi-circles. (Multiply "semi-circles" by 2π to obtain the quantity in units of "radians".) The "alpha" terms are the coefficients of a cubic equation representing the *magnitude* of the vertical delay, and the "beta" terms are the coefficients of a cubic equation representing the *period* of the model.

The t_{zion} quantity must be scaled by the **mapping function** to determine the time delay along the slant direction to the satellite:

$$t_{ion} = t_{zion} \cdot SF \tag{3.3-5}$$

and

$$SF = \sec[\ \sin^{-1}[\ \frac{r.\cos(Elev)}{(r+h)}\ ]\ ] \tag{3.3-6}$$

where:

Elev	is the elevation angle to satellite (in radians),
r	is the mean earth radius, and
h	is the mean ionospheric height (say 350kms).

An even simpler approximation is often used within GPS receivers:

$$t_{ion} = [1 + 16.(0.53 - EL)^3] \cdot 5x10^{-9} \tag{3.3-7}$$

where EL is the elevation angle to satellite in units of semi-circles.

Table 3.3-3. Broadcast Ephemeris computational procedure.

CONSTANTS (WGS84):	
• earth's universal gravitational constant	$\mu = 3.986008 \times 10^{14} m^3/sec^2$
• mean earth rotation rate	$\Omega_e = 7.292115147 \times 10^{-5} rad/sec$
•	$\pi = 3.1415926535898$
COMPUTE TRUE ANOMALY:	
• time since reference epoch	$t_k = t - t_{oe}$
• corrected mean motion	$n = \sqrt{\frac{\mu}{a^3}} + \Delta n$
• mean anomaly at t_k	$M_k = M_o + n.t_k$
• solve iteratively for E_k	$M_k = E_k - e.\sin E_k$
• true anomaly υ_k	$\upsilon_k = \arctan[\frac{\sqrt{1-e^2}.\sin E_k}{\cos E_k - e}]$
ARGUMENT OF LATITUDE:	
• argument of latitude	$\phi_k = \upsilon_k + \omega$
• correction	$\delta u_k = C_{us}.\sin 2\phi_k + C_{uc}.\cos 2\phi_k$
• corrected argument of latitude	$u_k = \phi_k + \delta u_k$
CORRECTED ORBIT RADIUS:	
• radius	$r_k = a(1 - e.\cos E_k) + C_{rs}.\sin 2\phi_k + C_{rc}.\cos 2\phi_k$
• position in orbit	$x_k = r_k.\cos u_k$
• position in orbit	$y_k = r_k.\sin u_k$
CORRECTED INCLINATION:	
•	$i_k = i_o + C_{is}.\sin 2\phi_k + C_{ic}.\cos 2\phi_k + \dot{i}.t_k$
CORRECTED ASCENDING NODE:	
• longitude	$\Omega_k = \Omega_o + (\dot{\Omega} - \Omega_e).t_k - \Omega_e.t_{oe}$
WGS84 COORDINATES:	
• earth-fixed coordinates	$x_e = x_k.\cos\Omega_k - y_k.\sin\Omega_k.\cos i_k$
	$y_e = x_k.\sin\Omega_k + y_k.\cos\Omega_k.\cos i_k$
	$z_e = y_k.\sin i_k$

3.4

GPS ELECTRONIC INFORMATION SOURCES

Electronic Bulletin Board Services -- USA

Navigation Information Center (formerly the GPS Information Center)

- Operated by the U.S. Coast Guard
- Information on constellation status, scheduled outages, almanac data, downloadable files, etc.
- Also accessible via INTERNET
- Connect details:
  - Dial in: 300 - 14400 baud
  - Connect parameters: N-8-1 [a]
  - Connect number: `+1-703-3135910`
  - Further information: ph: `+1-703-3135900`; fax: `+1-703-3135920`

U.S. Naval Observatory Automated Data Service

- Information on constellation status, scheduled outages, user advisories, email, time-transfer performance, etc.
- Also accessible via INTERNET
- Connect details:
  - Dial in: over 1200 baud
  - Connect parameters: N-8-1 [a]
  - Connect number: `+1-202-6530155 / 6530068 / 6531079`
  - Further information: ph: `+1-202-6530487`; email: `res@tuttle.usno.navy.mil`
  - Call for password: ph: `+1-202-6531525`; email: `fmv@tycho.usno.navy.mil`

Holloman GPS BBS (formerly Yuma BBS)

- Operated by the U.S. Air Force at Holloman Air Force Base, New Mexico
- Information on constellation status, almanac data, email, downloadable files, etc.
- Connect details:
  - Dial in: will automatically adjust for protocols
  - Connect number: `+1-505-6791525`
  - Further information: ph: `+1-505-6791784`

Electronic Bulletin Board Services -- International

Grinel - Professional BBS (AFRICA)

- Operated by Grinel - Professional Services on behalf of the Southern Africa GPS User Group
- Status advisories, almanac data, etc.
- Connect details:
  - Dial in: 300 - 2400 baud
  - Connect parameters: N-8-1 [a]
  - Connect number: `+27-12-8038318`

AUSLIG Geodesy Electronic BBS (AUSTRALIA)
- Operated by the Australian Survey and Land Information Group, Canberra
- GPS information on recent & historical constellation status, almanac data, availability of differential services, downloadable files, sundry geodetic information such as solar/ionospheric data, datum transformations, availability of geoid height data for Australia, etc.
- Also accessible via INTERNET
- Connect details:
  - Dial in: 300 - 600 baud
  - Connect parameters: N-8-1 [a]
  - Connect number: `+61-6-2014375 / 2014378`
  - Further information: ph: `+61-6-2014347`; fax: `+61-6-2014366`

LIC BBS (AUSTRALIA)
- Operated by the Land Information Centre, Dept. of Land & Soil Conservation, Bathurst
- Downloadable RINEX files from LIC base station, geographical names, geodetic information on control stations and benchmarks, etc.
- Connect details:
  - Dial in: up to 14400 baud
  - Connect parameters: N-8-1 [a]
  - Connect number: Bathurst `+61-63-316799 / 318935 / 315826 / 316880 / 316885` and Sydney `+61-2-2641972 / 2642013 / 2642168 / 2642182 / 2642242 / 2642261`
  - Further information on subscription, etc.: ph: `+61-63-328444`; fax: `+61-63-318095`

Other Private & Government BBS (AUSTRALIA)
- There are several private and government operated BBS offering RINEX data files generated by permanent GPS base stations
- Downloadable RINEX data and navigation message files
- Connect details for Ballarat DGPS BBS:
  - Dial in: up to 9600 baud
  - Connect parameters: N-8-1 [a]
  - Connect number: `+61-53-333255`
  - Further information on subscription, etc.: ph: `+61-53-336202`; fax: `+61-53-336234`

Electronic BBS (DENMARK)
- Operated by Kort-og Matrikelstyrelsen, Copenhagen
- GPS status advisory notices, almanac data, historical data, etc.
- Connect details:
  - Dial in: up to 2,400 baud
  - Connect parameters: N-8-1 [a]
  - Connect number: `+45-3-1853541`
  - Further information: ph: `+45-3-875050`; fax: `+45-3-875052`

GIBS - GPS Informations und BeobachtungsSystem (GERMANY)
- Operated by the Institute for Applied Geodesy (IfAG), Frankfurt am Main
- GPS status advisory notices, almanac data, historical data, geoid height data, real-time integrity, datum transformations, availability of differential services, GLONASS information, etc.
- Also accessible via INTERNET
- Connect details:
  - Dial in: up to 19200 baud
  - Connect parameters: N-8-1 [a]

- Connect number: +49-69-6333379 / 6333418
- Further information: ph: +49-69-63331; fax: +49-69-6333425

Electronic BBS (NORWAY)

- Operated by the Norwegian Mapping Authority
- GPS status advisory notices, almanac data, historical data, etc.
- Connect details:
  - Dial in: 1200 and 2400 baud
  - Connect parameters: N-8-1[a]
  - Connect number: +47-32124045 / 32118369
  - Further information: ph: +47-67-18100; fax: +47-67-18101 / 26190

Electronic BBS (SWEDEN)

- Operated by the National Land Survey Agency in Gavle, Sweden
- GPS/DGPS information, almanac data, precise ephemeris, historical data, geoid height data, datum transformations, etc.
- Connect details:
  - Dial in: up to 19200 baud
  - Connect parameters: N-8-1[a]
  - Connect number: +46-26-153748
  - Further information: ph: +46-26-153000; fax: +46-26-106232

Electronic Videotext BBS (THE NETHERLANDS)

- Operated by the Survey Department, Dutch Ministry of Transport & Public Works
- GPS/DGPS information, almanac data, receiver concepts & features, historical data, equipment prices & options, GPS policy statements, etc.
- Connect details:
  - Dial in: 1200 and 2400 baud
  - Connect parameters: N-8-1[a]
  - Connect number: +31-15-561959
  - Further information: ph: +31-15-691111; fax: +31-15-618962

UKCSG Electronic BBS (UNITED KINGDOM)

- Operated by the Institute for Engineering Surveying & Space Geodesy, University of Nottingham
- Full access is available only to members of the UK Civil Satnav Group

INTERNET Access

AUSLIG Geodesy (AUSTRALIA)

- Operated by the Australian Survey and Land Information Group, Canberra
- Information as in BBS, plus other geodetic information from Australia
- Downloadable files:
  - Using Gopher[b]: URL[c] --> gopher.auslig.gov.au
  - Via WWW[d]: URL[c] --> http://www.auslig.gov.au/geodesy/geodesy.htm

Canadian Space Geodesy Forum

- Operated by the University of New Brunswick, Fredericton, New Brunswick
- Information on constellation status, ionospheric disturbances, NANUs[e], news and discussion of GPS and other space-based positioning systems through email
- LISTSERV[f] based discussion and information:
  - List address: CANSPACE@UNB.CA

  - To subscribe: send one-line message [sub CANSPACE your_name] to `listserv@UNB.CA`
  - Further information: Richard Langley at `LANG@UNB.CA`
- Downloadable files:
  - By "fingering"[g] `GPS@GEOMAC.SE.UNB.CA`
  - Using Gopher[b]: URL[c] --> `gopher://unbmvs1.csd.unb.ca:1570/1EXEC%3aCANSPACE`
  - FTP[h] from directory `PUB.CANSPACE` on `[unbmvs1.csd.unb.ca]` `(131.202.1.2`[i]`)`
  - Via WWW[d]: URL[c] --> `http://www.unb.ca/geodesy/CANSPACE.html`

Czech Technical University GPS/GLONASS Info Service

- Operated by the Department of Radio Engineering, Faculty of Electrical Engineering, Czech Technical University, Prague
- Information on historical constellation status, almanac data for both GPS and GLONASS, etc.
- Downloadable files:
  - Using Gopher[b]: URL[c] --> `gopher://gopher.feld.cvut.cz/11/satelit`

National Geodetic Survey Division (USA)

- Operated by the National Geodetic Survey, Rockwell, Maryland
- Downloadable software and RINEX data files
- Downloadable files:
  - Via WWW[d] for software: URL[c] --> `http://www.ngs.noaa.gov`
  - FTP[h] RINEX files from directory `dist/cors1/rinex` on `[cors.grdl.noaa.gov]`
  - FTP[h] software from `[ftp.ngs.noaa.gov]`

International GPS Service for Geodynamics

- Operated by the IGS Central Bureau at the Jet Propulsion Laboratory (JPL), Pasadena
- Information about IGS tracking stations, Data Centers, Analysis Centers, IGS products, NANUs[e], RINEX data files from IGS network, IAG information such as conferences, etc.
- Further information: `IGSCB@COBRA.JPL.NASA.GOV`
- Downloadable files:
  - Via WWW[d]: URL[c] --> `http://igscb.jpl.nasa.gov/`
  - Via anonymous FTP[h] from directory `/igscb`. Downloading the file TREE.TXT from this directory gives **124**ucture and available files.

GPS Digest

- Moderated[j] mailing list, manually maintained (that is, not on LISTSERV[f])
- Forum for discussion of topics related to GPS and other satellite-based positioning systems through email:
  - To post a message: `GPS@esseye.si.com`
  - To join or quit the list: `GPS-request@esseye.si.com`

GPStech

- Unmoderated[k] mailing list, manually maintained (that is, not on LISTSERV[f])
- Forum for discussion of topics related to high precision geodetic GPS through email:
  - To post a message: `GPStech@cotopaxi.stanford.edu`
  - To join or quit the list: `GPStech-request@cotopaxi.stanford.edu`

GPS Newsgroups

- Several USENET[1] newsgroups exist that have occasional postings related to GPS, e.g.:
  - `sci.geo.satellite-nav` *(appears to be the best at the moment)*
  - `comp.infosystems.gis`
  - `sci.engr.surveying`
  - `rec.aviation.misc`
  - `rec.aviation.products`
  - `rec.boats`
  - `sci.electronics`
  - `sci.space.news`

GPS WWW Sites

- WWW[d] hypertext documents on GPS related documents accessed via Netscape or other WorldWideWeb browser program.
- Examples of official academic, government and scientific GPS WWW sites:
  - **U.S. Coast Guard Navigation Information Center:** URL[c] --> `http://www.navcen.uscg.mil/gps/gps.htm`
  - **U.S. Naval Observatory Navstar GPS Operations:** URL[c] --> `http://tycho.usno.navy.mil/gps.html`
  - **National Geodetic Survey:** URL[c] --> `http://www.ngs.noaa.gov/index.html`
  - **AUSLIG:** URL[c] --> `http://www.auslig.gov.au/geodesy/geodesy.htm`
  - **Geodetic Survey of Canada:** URL[c] --> `http://www.geod.emr.ca/`
  - **German Institute of Applied Geodesy:** URL[c] --> `http://www.potsdam.ifag.de/english/info/ifag-geodaesie-e.html`
  - **GLONASS Updated Information Service:** URL[c] --> `http://www.nz.dlr.de`
  - **IGS Central Bureau:** URL[c] --> `http://igscb.jpl.nasa.gov/`
  - **International Earth Rotation Service:** URL[c] --> `ftp://mesiom.obspm.fr/iers/`
  - **OSU Center for Mapping:** URL[c] --> `http://www.cfm.ohio-state.edu`
  - **UNB Canspace:** URL[c] --> `http://www.unb.ca/geodesy/CANSPACE.html`
  - **UNAVCO:** URL[c] --> `http://www.unavco.ucar.edu/`
  - **Official GLONASS information service:** URL[c] --> `http://mx.iki.rssi.ru/sfcsic/sfcsic_main.htm`
- Examples of private and public information GPS WWW sites:
  - **GPS, GIS, remote sensing:** URL[c] --> `http://www.zilker.net/~hal/geoscience/gps.html`, `http://www.ggrweb.com/`, `http://www.truck.net/gps.html`
  - **GPS World Magazine:** URL[c] --> `http://www.gpsworld.com/`
  - **GPS applications & principles:** URL[c] --> `http://galaxy.einet.net/editors/john-beadles/introgps.htm` or `sum_dgp.htm`
  - **Reference links to GPS general information sources:** URL[c] --> `http://www.inmet.com/~pwt/gps_gen.html`
  - **NMEA0183 and GPS information directory:** URL[c] --> `ftp://sundae.triumf.ca/pub/peter/index.html`

- **Tutorial on GPS:**
  URL[c] --> `http://wwwhost.cc.utexas.edu/ftp/pub/grg/gcraft/notes/gps/gps.html`
- **U.S. Institute of Navigation:** URL[c] --> `http://www.ion.org/`
- **Intelligent Transport Systems (ITS) America:**
  URL[c] --> `http://www.itsa.org/`
- **Ashtech Inc.:** URL[c] --> `http://www.ashtech.com/`
- **Etak Inc.:** URL[c] --> `http://www.etak.com/`
- **GEOsurv Inc.:** URL[c] --> `http://infoweb.magi.com/~geosurv/`
- **Magellan Systems Corp.:**
  URL[c] --> `http://www.bahnware.com/bahnware/commerce-market/magellan/`
- **Navtech Seminars & GPS Store:**
  URL[c] --> `http://www.navtechgps.com/`
- **Rockwell Int. Corp.:** URL[c] --> `http://www.rockwell.com/`
- **Trimble Navigation Limited:** URL[c] --> `http://www.trimble.com/gps/index.htm,` or `solutions/index.htm,` etc.

• A search for specific GPS WWW sites can be made using search facilities, e.g. located at URL[c] --> `http://www.altavista.digital.com/`. Use keywords such as "GPS surveying", "GPS hardware", otherwise you will be swamped with information!

Glossary:

a N-8-1: No parity, 8 data bits, 1 stop bit.

b GOPHER is a user-friendly menu interface program to make accessing files over the INTERNET easier. Gopher software can be obtained for your PC (Mac or IBM) by anonymous FTP from a variety of sources.

c URL: Uniform Resource Locator. A form of addressing often used with the WWW.

d WWW: World Wide Web (or simply the "Web") is a sophisticated client/server system operating in the INTERNET consisting of three parts: servers, clients and the protocol that connects them. This protocol (HTTP) supports hypertext which allows links to be made which may point to another place in the document, to another document, to another type of electronic resource or an image anywhere in the INTERNET. It is best used via a graphical interface program such as Mosiac, Lynx, Explorer or Netscape.

e NANU: Notice Advisory to Navstar Users

f LISTSERV: a program that electronically manages mailing lists. You can join or leave mailing lists using this program. To subscribe send a one-line message, e.g. [sub CANSPACE your_name] where CANSPACE is the list name to LISTSERV@UNB.CA, and UNB.CA is where the list is located. To unsubscribe send message [unsub CANSPACE].

g To "finger": Use the FINGER program to get information about email addresses, logins.

h FTP: File Transfer Protocol. A protocol that allows transfer of files electronically from one computer to another. This is the most primitive of the techniques for accessing data files over the INTERNET.

i TCP/IP email addressing for the INTERNET. Now largely replaced by forms such as "J.DOE@UNSW.EDU.AU".

j Moderated mailing lists have a moderator to filter out messages that are boring, irrelevant, or otherwise unsuitable.

k Unmoderated mailing lists don't do this and so anything that is sent to the list is automatically added.

l USENET: a collection of electronic discussion groups (or Newsgroups) on the INTERNET. They operate as an open bulletin board (no need to subscribe) that allows discussion relevant to the topic of the group. Newsgroups are structured hierarchally by topic, subtopic, etc. For example, `sci.physics.fusion` is a science newsgroup that concentrates on the physics aspects of nuclear fusion. Major heirachies are *comp* (computer hardware, software & protocol discussions), *misc* (for topics that can't be easily pigeonholed), *news* (for topics related to Usernet operations), *rec* (for recreational subjects), *sci* (recognised sciences), *soc* (social topics, world culture), and *talk* (current affairs discussion). *These groups spring up and disappear very quickly and without warning!*

Chapter 4: GPS Instrumentation

4.1 GPS RECEIVER DESIGN

The following components of a *generic* GPS receiver can be identified (Figure 4.1-1):

☞ **Antenna and Preamplifier Electronics:** Antennas used for GPS receivers have broadbeam characteristics, thus *they do not have to be pointed to the signal source* as in the case of satellite TV receiving dishes. The antennas are rather compact and may be tripod or vehicle mounted. In the case of GPS surveying applications, the actual position determined is the electronic phase centre, which then has to be correctly related to the ground mark being coordinated.

☞ **Radio Frequency Section & Computer:** *The RF section contains the signal processing electronics.* There is, in addition, a powerful onboard microprocessor to not only carry out "computations" such as extracting the ephemerides, determining the elevation/azimuth of the satellites, etc., but also to control the tracking and measurement function within modern digital tracking circuits. Different receiver types may use different techniques to process the signal, but most navigation receivers are of the so-called "code-correlating" variety.

☞ **Control or Interface Unit:** *The interface unit enables the operator to control and query the functions of the microprocessor.* Its size and type varies greatly, ranging from a handheld keyboard unit to a series of soft keys surrounding an LCD screen fixed to the front of the receiver "box". There are also many types of electronic interfaces for specialised applications, including the digital map displays, computer generated speech output, and interfaces to other instrumentation.

☞ **Recording Device:** In the case of GPS surveying receivers, *the measured data must be stored on some suitable medium for subsequent baseline computation* ("post-processing"). A variety of data storage devices were utilised in the past, including cassette and tape recorders, floppy disks and computer tapes, etc. Nowadays almost all receivers utilise solid state (RAM) memory or removable "memory cards", but they can also be connected directly to a computer and data recorded directly on the harddisk.

☞ **Power Supply:** *Most transportable GPS receivers these days use low voltage DC power.* The trend towards more energy efficient instrumentation is a strong one and GPS surveying receivers are flexible enough to operate from a number of power sources, including internal NiCad or Lithium batteries, external batteries such as wet cell car batteries, or from mains power.

Although there is a great variety of GPS hardware, one basic classification system for hardware is based on the type of observable that is tracked: (a) civilian navigation receivers using the C/A code on the L1 frequency, (b) military navigation receivers using the P (Y) code on both L-band frequencies, (c) single frequency (L1) carrier phase tracking receivers, and (d) dual-frequency carrier phase tracking receivers. The latter two are discussed further in §4.2.

The antenna and RF technology components are discussed briefly below. For further details the reader is referred to WELLS et al (1987), LANGLEY (1991a) and KAPLAN (1996).

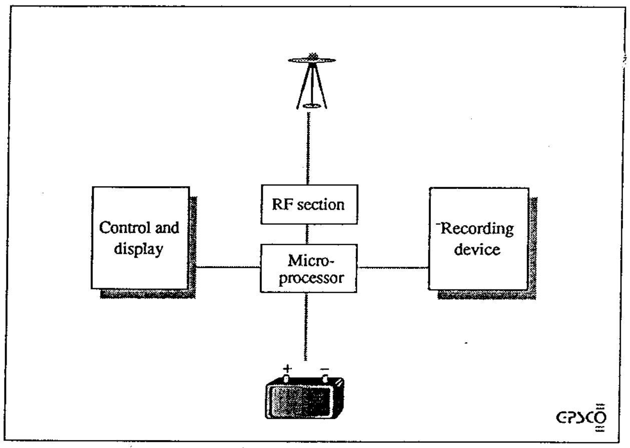


Figure 4.1-1. The generic GPS surveying receiver.
(WELLS et al, 1987)

4.1.1 ANTENNAS

The role of the antenna is to filter, amplify and down-convert the incoming signals into an electric current that can be processed by the receiver electronics within the RF section. There are a number of special considerations as far as GPS antenna design is concerned:

- The antenna must be able to pick and discriminate very weak signals (the signal strength is approximately the same as those from geostationary TV satellites).
- The antenna may need to operate at just the L1 frequency, or at both the L1 and L2 frequencies.
- As the signals are circularly polarised, the GPS antenna must also be circularly polarised.
- Antenna gain pattern that enhances the ability of the RF section to filter multipath and low elevation signals.
- An essential requirement is a stable electrical centre which is coincident with the geometric centre and insensitive to the rotation and inclination of the antenna.
- The construction of the antenna consists of the: (a) an omnidirectional antenna element, (b) antenna preamplifier electronics, and (c) a groundplane.

Several types of GPS antenna element-groundplane assemblies have been used:

- Monopole, or dipole, configurations.
- Quadrifilar helices.
- Spiral helices.

- Microstrip.
- Planar rings ("choke ring"), and other multipath-resistant designs.

Figure 4.1-2 illustrates these basic antenna types, and a sketch showing examples of specific manufacturers' antennas is included at the end of §5.3.

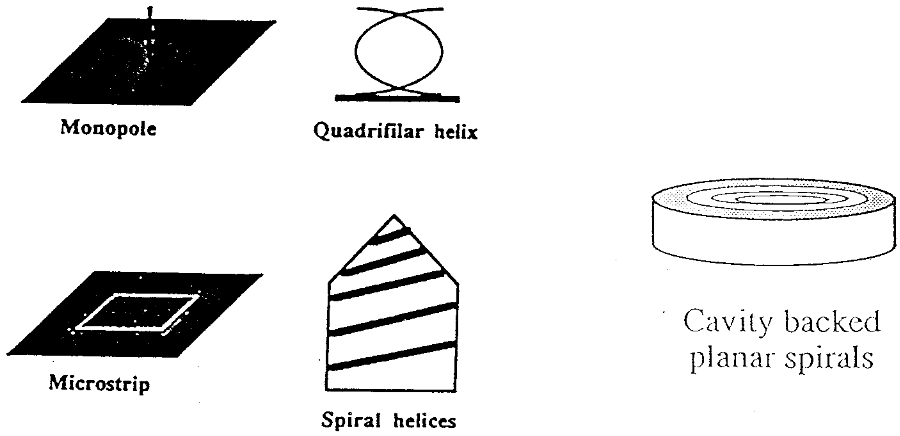


Figure 4.1-2. GPS antenna types. *(WELLS et al, 1987)*

The basic functions of the antenna preamplifier are (NATO, 1991):

(a) Increase incoming signal power level to overcome cable attenuation between antenna and receiver.
(b) Reject signal interference through appropriate filtering.
(c) Convert the L-band signals to an intermediate frequency to reduce cable signal losses.

GPS surveying antennas are required to be rugged, simple in construction, have stable electronic phase centres, be resistant to multipath, have good gain and pattern coverage characteristics. The following general comments can be made with regard to the antenna types identified above:

❑ **Quadrifilar Helix**:
- single frequency
- difficult to adjust for phasing and polarisation
- not azimuthally symmetric
- good gain pattern
- no ground plane required

❑ **Monopole**:
- single and dual-frequency (dipole construction)
- requires extensive ground plane
- very stable phase centre
- simple construction

❑ **Microstrip**:
- rugged and simple construction
- single or dual-frequency
- low profile (ideal for aircraft applications)
- low gain (offset by appropriate preamplifier)

- most common antenna available today
- much development work still going on

❑ **Spiral Helix**:
- dual-frequency operation
- good gain pattern
- high profile
- azimuthal asymmetry (requires careful ground orientation)

Present antennas are generally either *microstrip* or *quadrifilar* in construction. There is, however, intensive effort being invested in multipath-resistant antennas for both high precision geodetic use and the modern "rapid static", "stop & go" and "kinematic GPS" techniques (§5.5). There is also a trend to integrating the antenna assembly with the receiver electronics.

4.1.2 GPS SIGNAL PROCESSING TECHNIQUES

The antenna preamplifier of a GPS receiver generally converts the incoming signal (Figure 4.1-3) to a signal of a more useful frequency. This *intermediate frequency* is obtained by mixing the incoming signal with a pure sinusoidal signal generated by the local oscillator (usually a quartz "clock"). The frequency of this *beat frequency* is the difference between the original (doppler-shifted) received carrier frequency and the local oscillator. The **intermediate or beat frequency** is then processed by the signal tracking electronics.

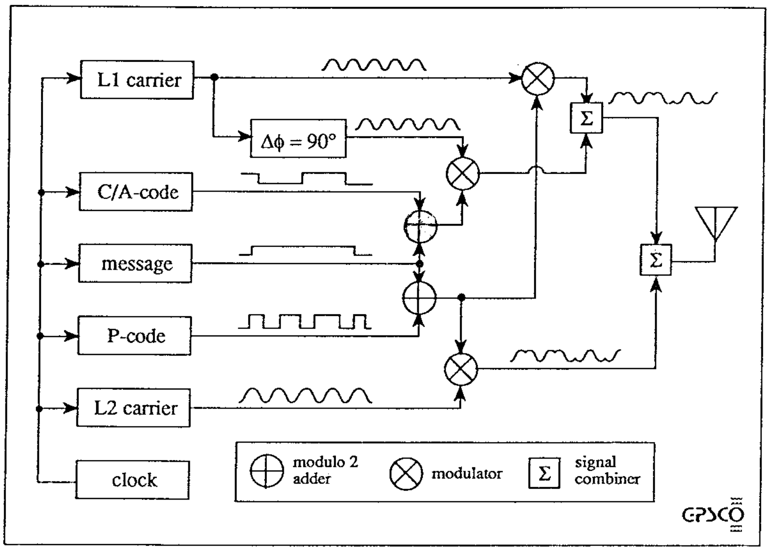


Figure 4.1-3. The incoming signal.

Frequency 1 = L1 carrier + P* code + C/A code + Navigation Message

Frequency 2 = L2 carrier + P* code + Navigation Message

(Changed to Y code under AS)*

There are several classes of signal processing techniques that can be employed to make pseudo-range or carrier phase observations, as well as several proprietary implementations of tracking *technologies*. However, it is beyond the scope of these notes to discuss in detail the electronic circuitry that is required. At a conceptual level it is only necessary to deal with:

(1) the two most commonly used signal processing techniques: "code-correlating" and "codeless" approaches, and some hybrid implementations such as "semi-codeless", and

(2) some details of the processing *architecture* within the receiver, and in particular the characteristics of the "tracking channel(s)".

Code-Correlating vs Codeless Receivers

Code-correlating receivers employ tracking loops to extract the necessary measurements and Navigation Message data from the beat signal. A typical GPS receiver contains two types of tracking loops:

☞ the *delay-lock*, or code-tracking, loop, and

☞ the *phase-lock*, or carrier-tracking, loop.

The **delay-lock loop** is used to align the PRN code sequence (C/A or P code) that is contained in the signal from a satellite with an identical PRN code generated within the receiver. A correlator in the delay-lock loop continuously cross-correlates the two code streams, time shifting the receiver generated stream until alignment is achieved. **The time shift is then the pseudo-range observation.** Once the code-tracking loop is aligned, the PRN code can be removed from the satellite signal (§3.2). The stripped signal then passes to the **phase-lock loop** where the satellite message is extracted. Once the local oscillator is locked onto the satellite signal it will continue to follow the variations in the phase of the carrier as the satellite-receiver distance changes. The *integrated carrier beat phase observable* is obtained by simply counting the *whole elapsed cycles* (by noting the "zero crossings" of the beat wave) and by measuring the *fractional phase* of the locked local oscillator signal (§3.2 and §6.1).

The carrier beat phase can be measured by another method besides this code-tracking/phase-tracking combined technique. The basic "codeless" or "squaring" technique takes the incoming signal and multiplies it by a copy of itself (§3.2). (Other variations also exist.) The reader is referred to VAN DIERENDONCK (1995) for further details on tracking channel terminology.

The organisation of a *code-correlating tracking channel* and a *squaring tracking channel* is shown schematically in Figures 4.1-4 and 4.1-5 respectively (and discussed in, for example, WELLS et al, 1987). There are several advantages and disadvantages to GPS code-correlating and codeless receivers:

- ❑ Only code-correlating receivers have access to the satellite Navigation Message. Either the P or the C/A code is required to despread the arriving signal without destroying the message, hence squaring receivers need an external satellite almanac in order to operate the tracking channels, and require externally-provided satellite ephemeris data for subsequent data processing.

- ❑ Codeless receivers must be synchronised with each other and UTC (or GPS Time) before observations begin. Code-correlating receivers can synchronise themselves with GPS Time at a point with unknown coordinates after acquiring signals from a minimum of four satellites, and carrying out a pseudo-range navigation solution.

- ❑ Code-correlating receivers enjoy a significant signal-to-noise advantage because of the inherent insensitivity of spread-spectrum signals to noise in a decoding receiver.

- Code-correlating receivers are capable of both pseudo-range and carrier phase measurements. Most non-military code-correlating receivers exclusively use the C/A code, which is only available on the L1 frequency. Dual-frequency ionospheric delay correction of either the pseudo-range or carrier phase data is only possible if measurements can be made on the second (L2) frequency.
- Codeless receivers can access both the L1 and L2 frequencies without knowledge of the PRN codes. This is particularly important under Anti-Spoofing conditions, unless special techniques are developed to overcome this handicap.
- Cycle slips in pure squaring receivers may sometimes be harder to repair as the wavelength of the phase data is only half that of the carrier frequency.
- Single frequency C/A code receivers are the cheapest GPS receivers on the market.

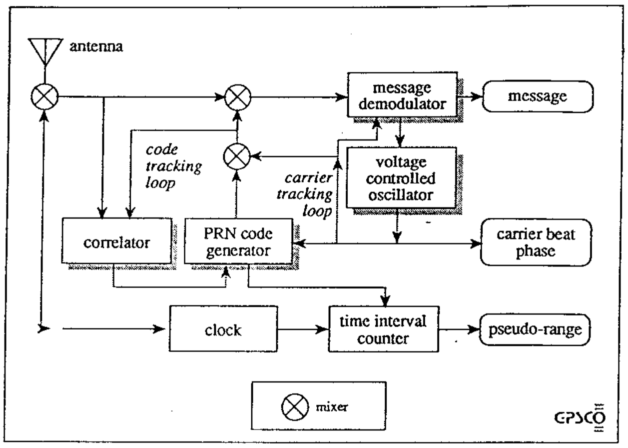


Figure 4.1-4. Code-correlating tracking channel.

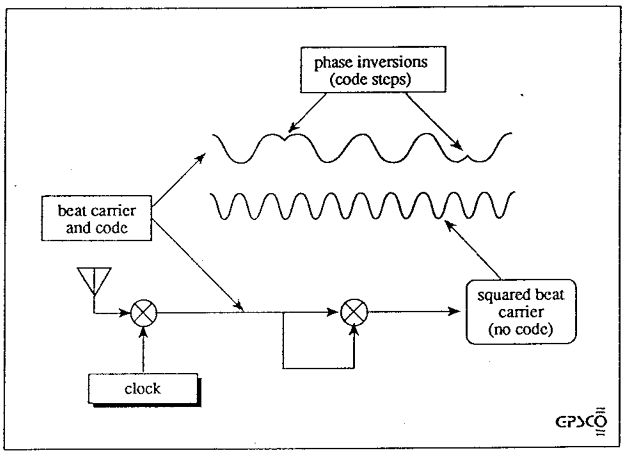


Figure 4.1-5. Squaring (codeless) tracking channel.

Nowadays the tracking channels are entirely digital in design, making GPS hardware smaller, lighter, and less expensive to build. Using analogue-to-digital conversion techniques a modern receiver usually consists of multiple digital tracking channels fed by a single receiver frontend (Figure 4.1-6). The characteristics of tracking channels, both the varieties used in earlier GPS receivers, and those used in presentday equipment are discussed below.

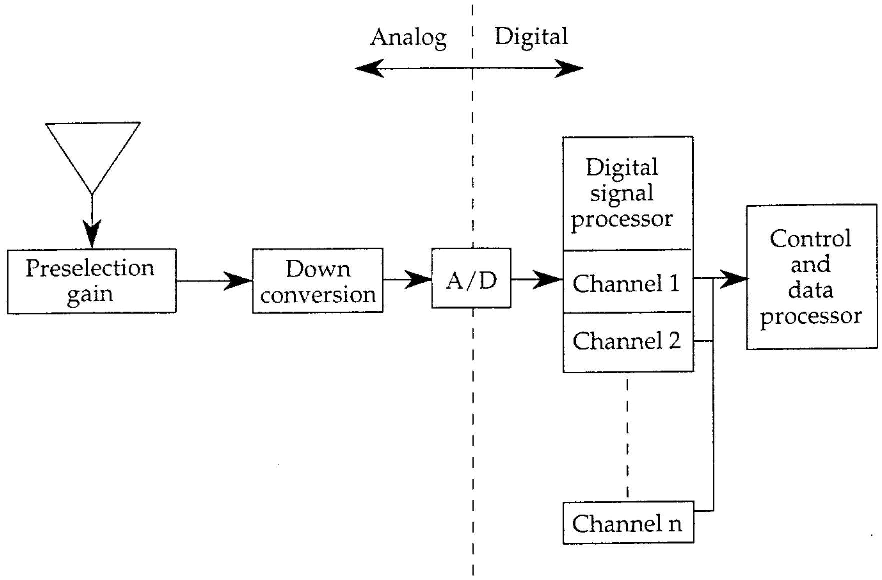


Figure 4.1-6. Digital GPS receiver design.
(CLARKE, 1994)

Hardware and Software Components of Tracking Channels

A GPS receiver may be described as "continuous" or "switching", depending on the type of channel(s) it uses.

A **continuous-tracking receiver** consists of dedicated *hardware* channels (Figure 4.1-7). Each channel tracks a single satellite and maintains continuous code and/or phase lock on the signal. Each channel is controlled and sampled by the receiver's microprocessor with input/output operations being performed fast enough so that tracking is not interrupted. Continuous-tracking receivers may enjoy a signal-to-noise advantage over switching receivers in that the satellite signal is continuously available and may be more frequently sampled. They are therefore particularly suited to high dynamic (kinematic) applications. A further advantage is a potential redundancy capability for should one of the hardware channels fail, it may still be possible to obtain sufficient data from the remaining channels. One disadvantage of a multi-channel receiver is that the differences in signal path delay in the channels, the so-called inter-channel biases, must be well calibrated. *This is the most common channel architecture used in survey-type GPS receivers.*

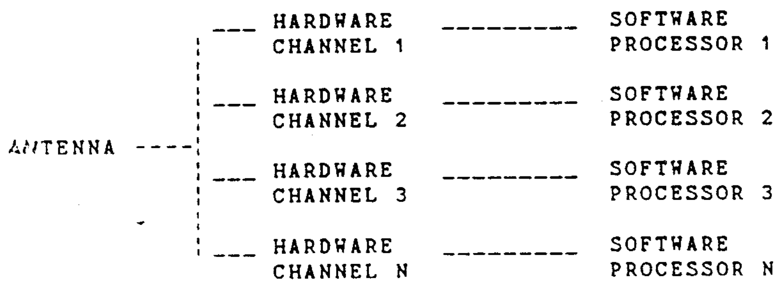


Figure 4.1-7. RF configuration with continuous tracking channels.

A **switching receiver** has one or more hardware channels. Each channel samples *sequentially* the incoming signal from more than one satellite (Figure 4.1-8). Code and/or carrier phase tracking for the individual signals is controlled by software (or, more correctly, the "firmware") within the microprocessor. As a result, greater demands are placed on the microprocessor in a switching receiver and its programming is necessarily more complex -- *in effect hardware complexity is exchanged for software complexity*. (The term *pseudo-channel* is sometimes used to identify software tracking channels.) Depending on the application and the environment, a switching receiver may be more susceptible to cycle slips than a continuous-tracking receiver. If the cost of hardware components is a significant factor in determining the sale price of a GPS receiver then, in principle, the cost of a switching receiver should be less than a continuous receiver -- this is certainly the case with the low-cost GPS navigation units.

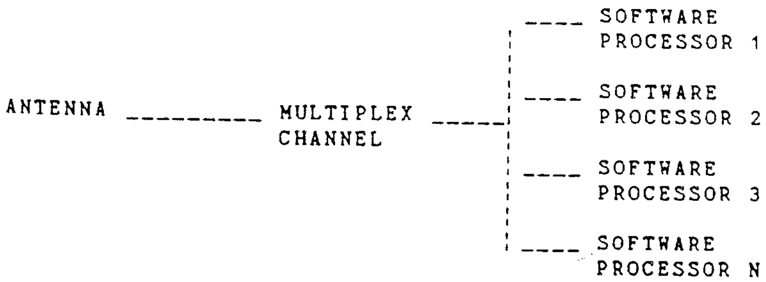


Figure 4.1-8. RF configuration with switching tracking channel(s).

There are three basic types of "switching" receivers, distinguished by the time required to *sequence* through the signals tracked by a particular channel.

A **multiplexing channel** is one for which the sequencing time to sample all satellites assigned to the channel is equal to 20 milliseconds, the period of one bit in the satellite Navigation Message (§3.3). The sampling is organised in such a way that no message bit boundary is spanned by any tracking interval. In this way, the messages from all satellites phase tracked by the channel can be read at the same time as the measurements are made. A multiplexing channel can be used to obtain both L1 and L2 data by alternating between the frequencies every 20 milliseconds. The early Texas Instruments TI-4100 GPS receiver had a single multiplexing hardware channel that was used to sample both L1 and L2 signals.

If a channel switches between signals at a rate which is asynchronous with the message bit rate, the channel is referred to as a **sequencing channel**. A *fast sequencing channel* is one which takes the order of fractions of a to perhaps several seconds to sequence through the signals. A *slow sequencing channel* may take many seconds. A single sequencing channel would lose bits in a particular satellite message during those intervals spent sampling the signals from other satellites. Consequently, sequencing receivers may have an extra hardware channel just for message decoding. Alternatively, the Navigation Message must be decoded before the receiver starts the tracking cycle for real-time positioning (remember that the message only changes once an hour).

Figure 4.1-9 illustrates the tracking coverage for different types of channels.

The configuration of channels is selected so that, in the case of a navigation receiver, a minimum of four satellites can be tracked at the same time. This may call for a single switching channel, or two switching (between two satellite signals) channels, or even five channels (one for calibrating the other four primary tracking channels, and decoding the Navigation Message).

In the case of GPS surveying receivers it is now recognised that it is very advantageous to track as many satellites simultaneously as possible (the so-called *"all-in-view"* tracking capability is preferred). To increase the number of simultaneously tracked satellites but still limit the number of channels (each requiring a "board" in the first generation of analogue GPS receivers), a combination of continuous and switching channels were sometimes used within the same receiver. An example of a *hybrid* geodetic receiver was the early Wild WM102, which was able to track six satellites on both frequencies. However, with the cost of electronic circuitry falling dramatically (due to the fabrication of digital tracking channels on VLSI chips), the latest generation of geodetic receivers have many continuous-tracking channels, some code-correlating and some employing a codeless tracking technology, or some variation of these. Configurations of 8, 12,16, 24 and even 48 channels are now used.

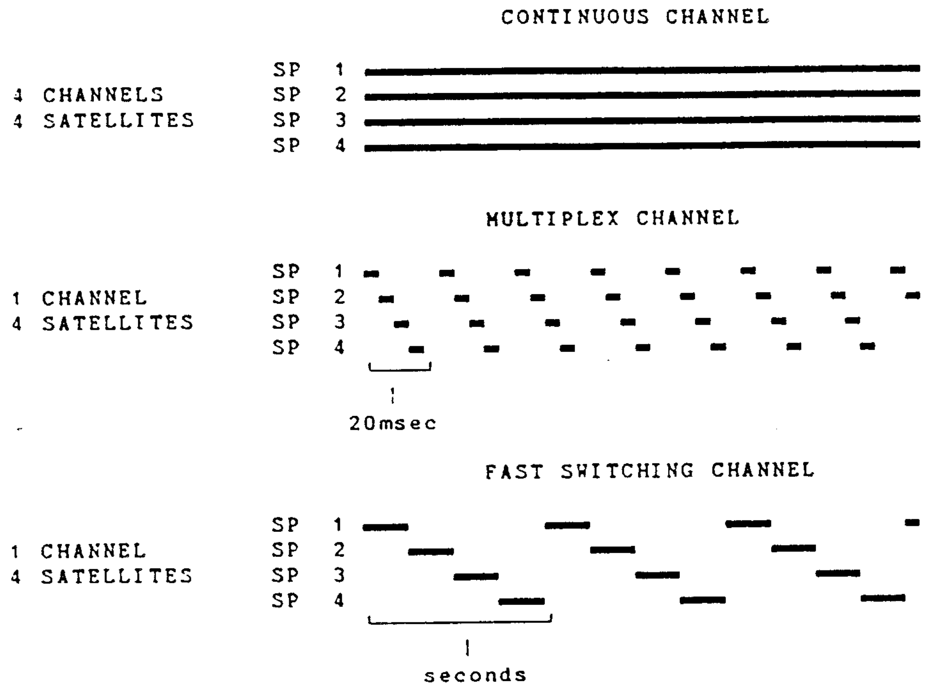


Figure 4.1-9. Tracking pattern for different types of channels.

4.1.3 TRENDS IN GPS INSTRUMENTATION

It is difficult to be precise about trends in GPS instrumentation in general. The task is no easier if attention is focussed only on a specific market, such as that for GPS surveying receivers. Nevertheless, speculation based on R&D activities being undertaken at present gives some clues:

- The third generation of GPS receivers presently available already exhibit significant gains in miniaturisation, reduction in power consumption, and portability, over the earlier models. This trend can be expected to continue. Surveying receivers are likely to decrease in size quickly to that of a car radio. However, there is also a trend to pack more electronics into the receiver box. Some navigation receivers are "handheld", and "credit card" sized units have made their appearance. The miniaturisation of much of the electronics (particularly the tracking channels) onto VLSI circuits means that less power is needed for the receiver to function. But the challenge, as with powering notebook computers, is to develop small, lightweight, longlife batteries.
- There have been many predictions of low-cost GPS receivers. Many handheld navigation

receivers are available for less than $1000, and GPS manufacturers have put the market on notice and announced units as cheap as a few hundred dollars. It is doubtful whether the cost of GPS *surveying* receivers will ever fall so low, for a number of reasons:
- The market (present and projected) volume is relatively small.
- The tendency to pack more electronics and sophisticated features into the receiver "package".
- There is a considerably greater software investment in a GPS surveying product than is the case for GPS navigation receivers.

- <u>There is a tendency towards product differentiation</u>, with many different configurations of tracking channels, data recording options, and, in particular, software options. Although some of these trends may be due to manufacturers wanting to give their product "an edge" in the marketplace, it is equally valid to suggest that this is in response to the different demands (some very specialist) of the market. The choice confronting GPS manufacturers is whether to maintain a broad-based development program, and market a product that is versatile enough to satisfy many applications (including the provision of interchangeable components such as antennas, RF units, memory, etc., to accomplish this), or to focus on specialist applications (military, navigation, differential navigation, geodesy, surveying, kinematic, etc.). To date most manufacturers are split between those with products for the surveying/geodesy market, and those focussing on the low-cost, high-volume navigation market. Only a few attempt to address all markets.

- <u>There is a tendency, particularly for surveying receivers, to incorporate more channels and strive for "all-in-view" continuous tracking</u>. Multiplexed or fast switching tracking channels are unlikely to be used in future surveying receivers. In the operational GPS constellation 10 satellites or more are visible above the horizon at any one time. Third (and subsequent) generation instrumentation will be able to track all of these, and optionally on both frequencies (requiring at least 20 channels!). There is only a slight trend to develop instruments that can also track the signals from the Russian GLONASS system -- there are only a small number of "top-of-the-line" navigation receivers intended for aviation use. *It remains to be seen whether GLONASS does have a future at all!*

- <u>There will be no strictly codeless surveying receivers</u>. Proprietary technology is, however, being used extensively to make L2 phase and pseudo-range measurements possible (§4.2), as the Y code generating algorithm will be secret.

- <u>There are many trends evident in GPS navigation receiver instrumentation and applications</u>. The core unit of a navigation instrument will be mass produced on one or more chips. These low-cost components will then have "value-added" features grafted to them. An exciting marriage is between satellite navigation and satellite communication, combining real-time positioning with instantaneous transmission of position. GPS receivers may be fitted to many different platforms (some unmanned, for example, railway rolling stock and ship cargo containers), and their locations remotely monitored at a central site via a satcom link. In addition, for land navigation these could include electronic map displays to aid the driver. *What is clear is that for many applications the navigation data provided by a GPS receiver will merely be the first link in a complex spatial information and control system.*

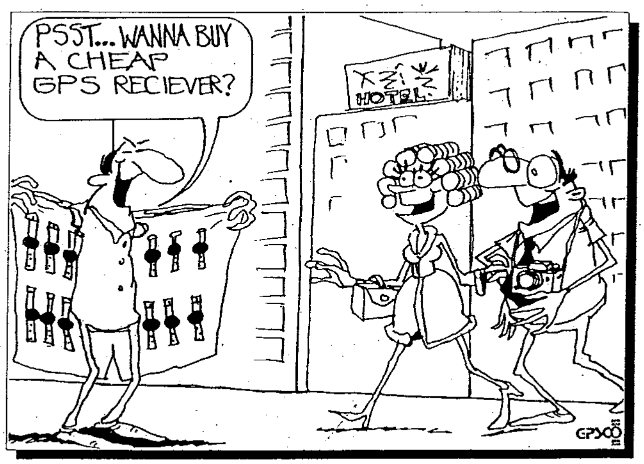
- <u>There is likely to be more self-checking (or "intelligence") capabilities built into surveying instruments</u>. In addition, much of the pre-processing of the data will be carried out within the receiver, including operations such as cycle slip editing, data compression, automatic and continuous operation of the receiver.

- <u>The trend towards kinematic positioning with survey accuracy is very strong</u> (§5.5). This mainly requires advances in the development of processing software, but also in antenna and tracking technology. Although an exciting prospect for airborne and marine applications, its potential to replace static GPS surveying on land will revolutionise further the practice of surveying and mapping. Instead of requiring half an hour or more of phase observations at two (or more) sites, "rapid static" and "stop & go" GPS surveying techniques represent a tremendous boost in productivity. *In addition, real-time relative position to this accuracy is already possible with the appropriate communication link and processing software.*

- Perhaps the most significant trends will be in software development. Many more applications may be addressed with the appropriate (and often specialised) software. Such software will take advantage of the advances in instrumentation and make, for example, real-time high precision positioning a simple and routine operation. There is already the possibility of receiver output compatible across different instrument types. Such a common output format has already been devised, the so-called RINEX format (GURTNER, 1994), and has been adopted (albeit sometimes reluctantly) by most surveying receiver manufacturers. The significance of this is that it is (almost) as easy to use a mixture of instrument types for a GPS survey, as a single brand of receiver hardware and software.

- The next generation of GPS surveying equipment will begin to incorporate recent advances in antenna technology in order to more fully exploit the accuracy of GPS. These trends include:
  - *Accurate calibration of presently used antennas:* Some of the present antenna designs are particularly unsuited to the GPS surveying application, while others have proven well suited. It is important that these antennas be accurately and correctly calibrated in order to assure suitable performance.
  - *Increased antenna flexibility:* Some antenna designs, while well suited to some applications, are poorly suited to others. It would be useful to have a variety of interchangeable antennas in order to meet specific application requirements.
  - *Receiver software to include phase centre:* It would be useful to include antenna phase centre calibration data within the receiver data processing software.
  - *New antenna design:* Antennas are required which are designed with GPS phase stability specifications in mind. Such antenna design should also incorporate amplitude pattern control for minimisation of multipath interference.
  - *Improved manufacturing:* Mechanical tolerances required for millimetre level phase centre stability are more stringent than those in effect for communication antenna fabrication.

- The choice of GPS instrumentation is unlikely to ever be a straightforward decision. Many factors will need to be taken into account, and not all of them of a technical nature. The hardware/software development cycles by different GPS manufacturers are generally out of synchrony, hence just as an instrument from one manufacturer is "maturing", new (but untried) systems will be released.



4.2

GPS HARDWARE FOR SURVEYING

In 1980 there was only one GPS receiver available on the market. A decade and a half later there are several hundred different makes and models available, most of them having appeared in the last 5 years. Different receivers have varied features and capabilities. For example, some are designed for military users; others for civilians; some receivers are intended for navigation, others for the most precise geodetic surveys; some receivers use the C/A code, others additionally use the P code; there are single and dual-frequency receivers; there are integrated GPS-GLONASS receivers; and there are handheld and rack-mounted configurations. However all receivers share a number of common features. A typical GPS receiver comprises the following **hardware** components (§4.1):

- antenna and preamplifier,
- radio frequency section,
- microprocessor,
- terminal, or control and display unit,
- data recording section, and
- power supply.

In this and the next chapter the focus is on the *total* GPS surveying "package", and therefore to the above list must be added the **software** components:

- survey planning software component,
- data logging, data downloading, pre-processing, editing, etc.,
- phase data processing module, and
- result presentation modules (network combination, transformation, etc.).

Below are highlighted some characteristics of the hardware component of GPS surveying instrumentation, with particular emphasis given to the most popular presentday instrumentation from the three major GPS survey receiver manufacturers: Ashtech, Leica and Trimble. The software component is discussed in §4.3.

4.2.1 HARDWARE CHARACTERISTICS

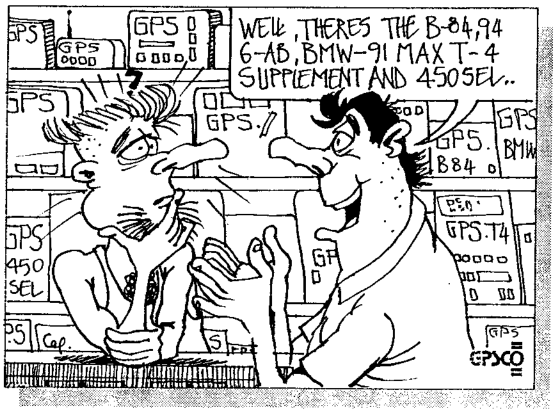
The following are some characteristics of GPS receiver hardware to be considered:

- P code, C/A code measurements, as well as carrier phase?
- Code-correlation, or squaring, or other proprietary phase tracking techniques?
- Single-frequency or dual-frequency capability?
- Continuous or switching, or hybrid tracking channel architecture? How many channels?
- External clock input?
- Strobe timing output?
- Type of data storage, and capacity?
- Power requirements and options?
- Antenna type, and antenna-receiver configuration?
- Nature of "packaging": backpack, handheld, tripod-mounted or rack-mounted?

- Availability and sophistication of software, reliance on external computer?

Some of these characteristics have already been discussed in §4.1. GPS receivers to be used for high precision surveying applications generally have the following features:

- ❑ Able to measure the integrated carrier phase on at least one of the two L-band carrier frequencies.
- ❑ Have the capability of tracking substantially more than the minimum four satellites simultaneously (preferably an "all-in-view" capability).
- ❑ Record the data for subsequent post-mission processing.
- ❑ Minimum data measuring rate of 15-60 seconds (preferably higher for kinematic applications).
- ❑ Antennas are chosen with stability and quality in mind so as to minimise phase centre variations and sensitivity to multipath, with the phase centre clearly marked on the antenna housing.
- ❑ Must be of rugged construction, light and portable.



The following are some hardware issues that need to be borne in mind when weighing up the advantages of one GPS surveying instrument over another, and whether the performance of an instrument is subject to factors beyond the operator's (or manufacturer's) control:

- CODE-CORRELATING receivers can measure both pseudo-range and carrier phase.
- Instrumentation has been developed to give decimetre precision pseudo-range data from C/A CODE RANGING, but there are still problems with multipath effects.
- Given a choice, CODE-CORRELATING receivers make a superior phase measurement.
- SQUARING permits phase measurements on L1 or L2, however CODE-CORRELATING permits only L1 measurements (phase or pseudo-range) with C/A code.

- CODE-CORRELATING with P code permits dual-frequency measurements to be made.
- ANTI-SPOOFING* affects CODE-CORRELATING with P code, but not C/A code.
- Proprietary techniques have been developed to extract L2 phase and pseudo-range data under AS conditions WITHOUT THE NEED FOR SIGNAL SQUARING (see below).
- There is no hardware "fix" for SELECTIVE AVAILABILITY** --> *only military receivers will have the "crypto-key" to overcome SA.*

* *AS is now fully implemented*
** *SA is fully implemented on Block II/IIA satellites though the decision is up for review*

Single or Dual-Frequency?

One of the most common questions that arises with regard to the selection of a GPS receiver to purchase (or use) is whether dual-frequency instrumentation is in itself *superior* to single frequency instrumentation. The following comments can be made:

- Dual-frequency instruments <u>eliminate</u> *ionospheric delays (a measurement bias).*
- Ionospheric delay is largely a problem for baselines greater than about 20-30km in length.
- Single frequency observations are adequate for <u>short baselines</u> *using conventional GPS surveying techniques.*
- Dual-frequency instruments are *more expensive* than single frequency receivers.
- *Different implementations of dual-frequency capability*, for example, code-correlating, squaring or proprietary techniques (see below).
- Dual-frequency instruments usually are either:
  - *hybrid* (code-correlating on L1, squaring or proprietary technology on L2), or
  - special L2 tracking design which allows for *switching* (for example, between P code-correlating and squaring/proprietary technology when AS turned on).
- Dual-frequency instrumentation is *essential for high productivity* "rapid static", "stop & go" and "kinematic" GPS surveying procedures.
- *Intensive R&D to improve tracking performance* on L2 carrier wave.

Selecting a GPS Receiver for Surveying Applications

The selection of a receiver that best suits the surveyor's needs should take the following factors into account:

- **ACCURACY:** What level of accuracy is sought? Over what baseline length? *Perhaps a dual-frequency receiver is recommended.*
- **RUGGEDNESS:** How portable must the instrumentation be? Will it be mainly used for fieldwork, or will it merely operate as a (semi-) permanent base station?
- **RELIABILITY:** What generation of instrument is it? Is it a revolutionary design newly released (with potential reliability problems)? Has the software been thoroughly debugged?

- **POWER REQUIREMENTS:** How do they compare with competitors's instruments? How many options for power input are there?
- **SERVICING:** What competitive advantage is there here? What is the local agent support like?
- **FINANCING:** How much GPS surveying work is available? Considered a consortium agreement whereby a pool of instruments is purchased by different partners? Lease or buy option?
- **FLEXIBILITY OF USE:** How many options for field operation are there? Are these options expensive? Do they involve the purchases of additional hardware/software? Will it be used for kinematic work? Used for real-time operation? etc.
- **STATE-OF-THE-ART:** Can it support the modern GPS surveying techniques?
- **STORAGE MEDIUM & CAPACITY:** Is it adequate? Is it appropriate?
- **SUITABILITY OF SOFTWARE:** Is it provided? At what cost? Is it up to the task or accuracy requirements? How reliable is it?
- **EASE OF USE:** How easy is the receiver to operate? How sophisticated (but easy to use) is the data post-processing software?

The actual hardware configuration selected for single frequency instrumentation is largely immaterial as all receivers are capable of measuring the raw L1 phase to a similar accuracy. The measurement of L2 phase is more problematic, and proprietary techniques are generally used. Each technique, however, has its advantages and disadvantages (though they may only be obvious after careful and thorough testing). Ultimately the quality of the results will probably be more influenced by the processing software and field procedures, than the phase measuring hardware.

P Code Availability

P code-correlation permits both the precise pseudo-range on the L1 and L2 frequencies, and the carrier phase on the L2 carrier wave to be measured. The C/A code-correlating receivers are used for only L1 pseudo-range and phase observations.

P code based L2 phase measurement:

- P code-correlating receivers are commonly specified for geodetic work.
- L2 phase measurement preserves full wavelength (24cm).
- An advantage for "rapid static" and "kinematic" surveying techniques based on dual-frequency data processing.

P code pseudo-range measurements, *on their own or in combination with phase data:*

- Have lower noise than C/A pseudo-ranges.
- Permit the use of new surveying techniques that mix pseudo-range and phase data.

However, there is no P code-correlating capability while AS is turned on!

L2 Measurements Under AS Without Knowledge of the Y Code

Without knowledge of the Y code, receivers have to apply codeless or quasi-codeless proprietary techniques to make measurements on the L2 carrier wave. The tracking techniques presently available are summarised below and in Table 4.2-1. (Much of the following information is taken from ASHJAEE, 1993; HOFMANN-WELLENHOF et al, 1994. The reader is also referred to VAN DIERENDONCK, 1995.)

The Squaring technique has already been discussed in §4.1. In essence, the received signal is multiplied with itself resulting in an unmodulated carrier with twice the frequency (half the wavelength). Its disadvantages are so great that it is no longer used in any GPS instrument:

- *All signal information is lost*, hence no pseudo-ranging on L2 is possible and the Navigation Message is lost from the L2 signal (though not a problem if the code-correlating tracking technology is applied on L1).
- As the L2 wavelength is only of the order of 12cm, *ambiguity resolution is much more difficult.*
- The *signal-to-noise-ratio is the worse of all the L2 tracking techniques*, resulting in frequent signal loss, poor data quality and many cycle slips (especially in moving antenna applications, or when the ionosphere is particularly active).

The Cross-Correlation technique was proposed 10 years ago, and is now implemented within the Rogue and Trimble GPS receivers. The technique makes use of the fact that the unknown Y code modulation is the same on both carrier waves. Due to the frequency dependence of the ionospheric delay, the Y code on the L2 signal is slightly slower than the Y code on the L1 signal. Hence the time delay of the L2 signal in relation to the L1 signal can be measured. The following comments can be made:

- *The L2 carrier measurement is obtained by adding the (L2 - L1) delay* (in units of cycles) to the L1 (C/A code-correlated) measurement.
- The L2 pseudo-range measurement is obtained in a similar fashion, by adding the (L2 - L1) delay (in units of metres) to the C/A code pseudo-range measurement.
- The L2 carrier measurement is of *full wavelength* (24cm).
- However, the *signal-to-noise ratio is only slightly improved* over using the pure squaring technique.

The Code-Correlation Plus Squaring technique is an improved squaring technique developed 5 years ago by the Magnavox corporation, and is used in its parent company's Leica System 200 GPS receivers. The technique requires the correlation of the received Y code on the L2 signal with a locally generated P code. This is possible because the Y code is formed by the modulo-2 addition of the P code and the encrypting W code. As a result of the W code frequency being about 20 times less than the P code, there will always exist Y Code portions which are identical to the original P code portions. Hence the P code correlation occurs between the internally generated P code and the *underlying* P code of the incoming signal, and the result can be low-pass filtered. Then the signal is squared to get rid of the code. A modified version of this technology has recently been implemented in Leica's System 300. The following comments can be made:

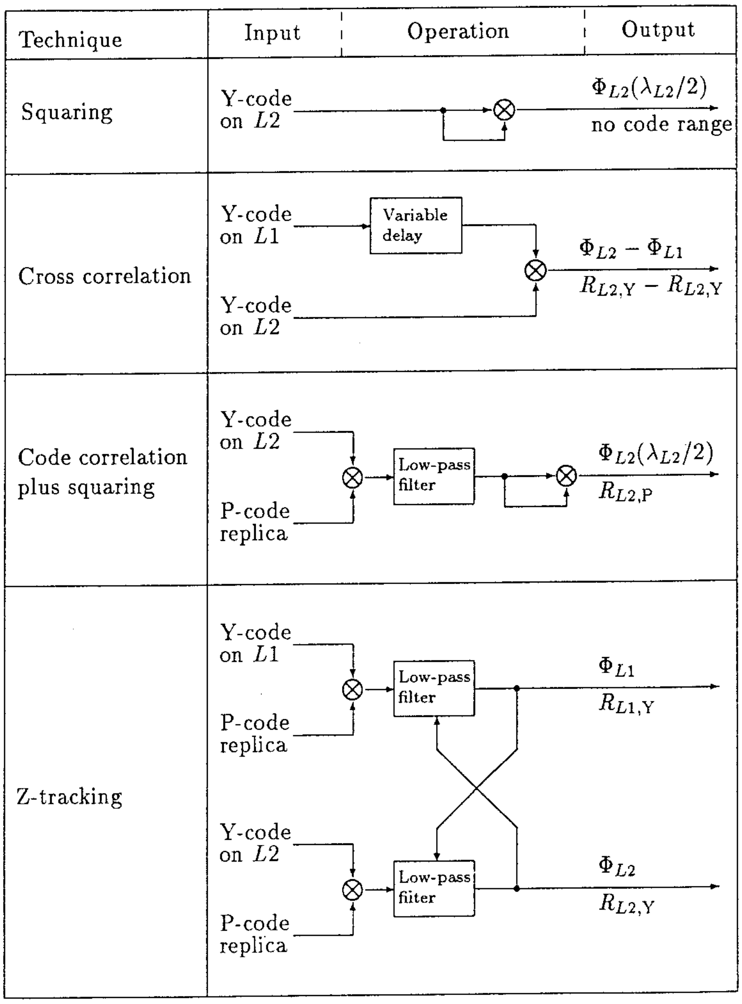
- The L1 and L2 pseudo-range measurement can be obtained.
- The L2 wavelength (System 200) is only of the order of 12cm, *hence ambiguity resolution is much more difficult.* This drawback is overcome in the new System 300.
- The correlation with the P code yields better jamming immunity and an improvement in multipath rejection performance.
- The *signal-to-noise ratio is significantly improved* over using the above two techniques.

The Z-Tracking technique is one of the recent quasi-codeless techniques, and is used in the Ashtech GPS receivers. The technique is based on the *removal* of the encrypting W code through a relative complex procedure described in ASHJAEE (1993). The following additional comments can be made:

- The L1 and L2 pseudo-range measurements can be obtained.
- The *L1 and L2 carrier phase measurements are obtained,* both with full wavelength.
- The *signal-to-noise ratio is significantly improved* over the Code-Correlation Plus Squaring technique.
- In reality the *Z-Tracking technique has not always delivered the best tracking performance.*

All techniques developed to overcome AS are sub-optimal, compared with using the P code-correlation technique!!

Table 4.2-1. Codeless techniques to track L2 under AS.
(HOFMANN-WELLENHOF et al, 1994)



4.2.2 SOME INSTRUMENTATION

LEICA GPS System 200

Manufacturer: **LEICA, Heerbrugg, Switzerland.**

Applications: **All survey & precise navigation applications.**

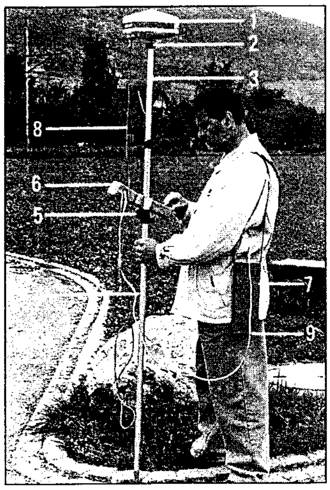
Components: **SR299 sensor with integrated microstrip antenna (optional SR299E configuration with external AT201 antenna), CR233 controller (optional CR244 for I/O of RTCM messages).**

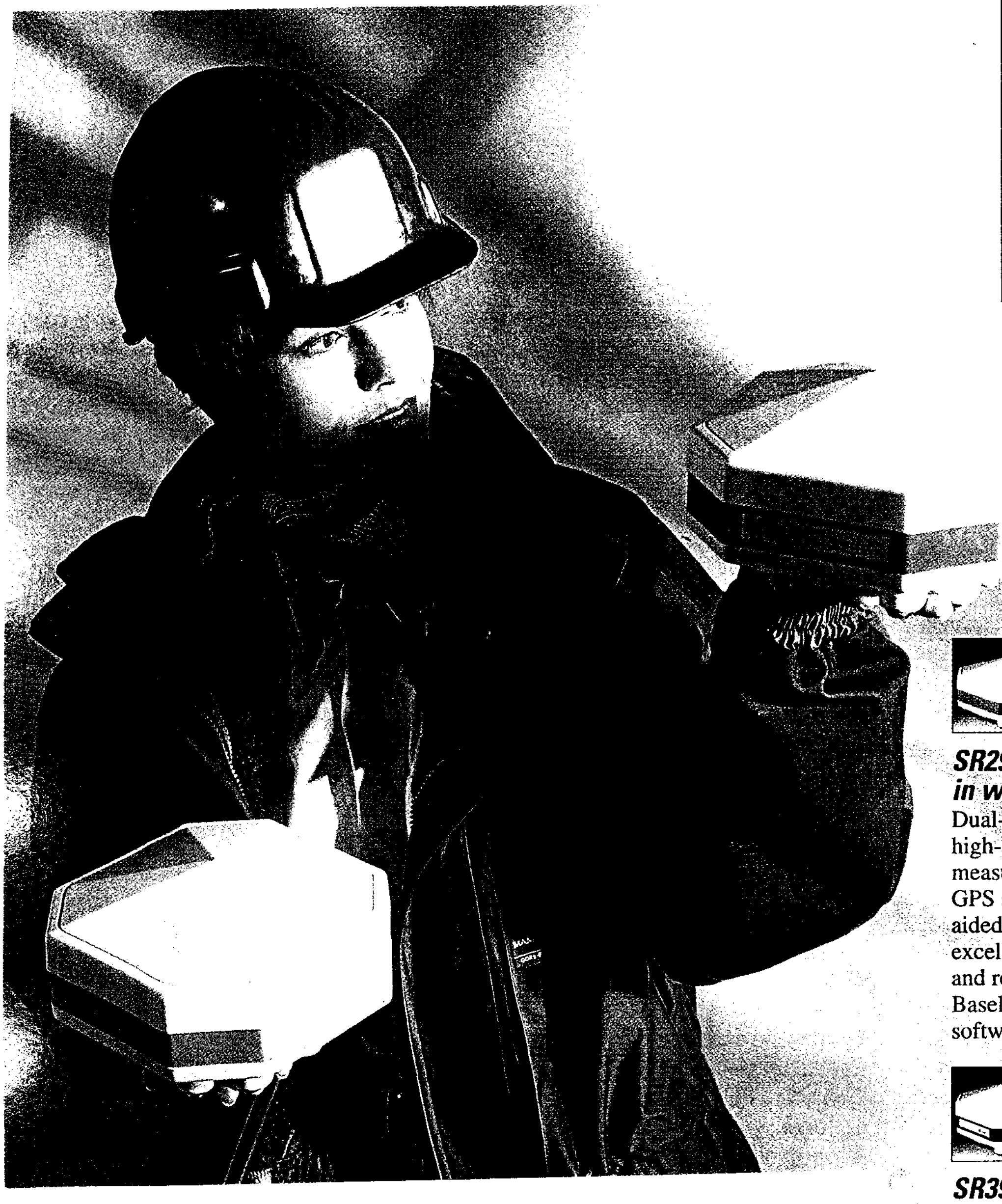
Features:

- *Precision:* **±(5mm + 1ppm) for static GPS, cm-dm accuracies for baselines <100km for kinematic (resolved ambiguities), submetre DGPS accuracy.**
- **Tracks 9 satellites on both L1 & L2 frequencies. C/A & P2 pseudo-ranges. L1 phase tracked by reconstructing carrier via C/A code. L2 phase tracked by: (a) P code-correlation (when AS off), or (b) Leica proprietary codeless technique (when AS on).**
- **CR233/244 is a water-resistant handheld computer, with major functions to control the sensor, monitor progress of survey, and to log data. Connected to sensor by cable.**
- **SR299 weighs 2.3kg, SR299E weighs 2.0kg, AT201 weighs 0.6kg, CR233/244 weighs 1.1kg.**
- *Power consumption:* **8.5W (for SR299/299E), max. 12W (total). External batteries for 12V DC supply, or other power source. Power can be supplied separately to SR299/299E and CR233/244, or together.**
  Example: **7 Amp.hr. battery will run SR299/299E and CR233/244 for about 5-6hrs.**
- **CR233/244 has 8x40 character LCD display (illuminated for night operations). Full alphanumeric keyboard for data and command entry.**
- **Data output 1-60 seconds.**
- **Log data to CR233/244 memory cards (memory card reader required for data download), optional internal memory or direct to PC. 512Kb, 1Mb & 2Mb cards.**
  Example: **512Kb cards hold about 9hrs of 5 satellite, 15 second rate data.**
- **Optional time-mark output.**
- **Optional RTCM I/O capability for real-time DGPS.**

Software: **SKI™ post-processing software: conventional static, as well as "rapid static", "stop & go", "reoccupation", "kinematic" survey modes, differential code solutions.**

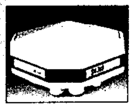
Options: **Real-time kinematic surveying, "on-the-fly" ambiguity resolution, DGPS navigation (RTCM I/O capable), PC base station software, aerial photogrammetry application.**





SR299 survey receiver: in widespread use

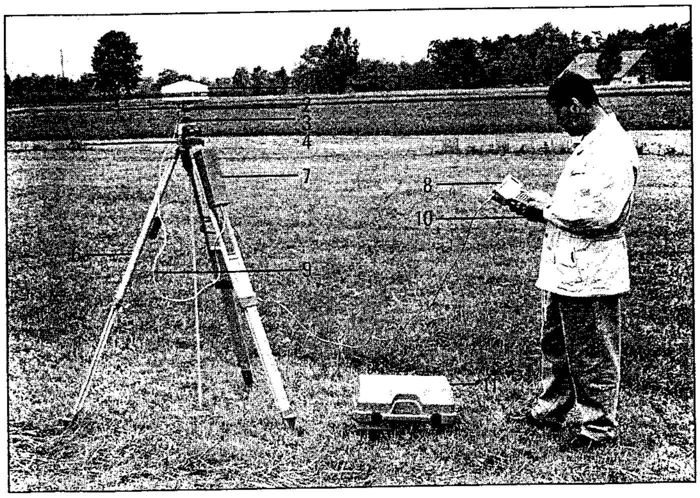
Dual-frequency receiver with high-accuracy carrier-phase measurements for all types of GPS surveying. Patented code-aided squaring ensures excellent signal-to-noise ratios and reliable tracking under AS. Baseline rms with SKI software: 5mm + 1 ppm.



SR399 geodetic receiver **new**

Additional observables, enhanced tracking:

- P-code on L1 and L2
- Full L1 and full L2 carrier-phase measurements, even under AS
- Half-meter differential-code positioning on L1 and L2, even under AS
- Best possible signal-to-noise ratios
- Baseline rms: 5mm + 1ppm



LEICA GPS System 200
(Single Frequency System)

Manufacturer: **LEICA, Heerbrugg, Switzerland.**

Applications: **Survey & precise navigation applications.**

Components: **SR261 sensor with external AT201 microstrip antenna, CR233 controller (optional CR244 for I/O of RTCM messages).**

Features:
- *Precision:* **±(5-10mm + 2ppm) for static, "rapid static" & "reoccupation" baseline solutions, ±(10-30mm + 2ppm) for "stop & go" & "kinematic" solutions, 1-3m DGPS accuracy.**
- **6 channel single frequency instrument tracking L1 phase & C/A pseudo-ranges (phase smoothed).**
- **CR233/244 is identical to that used with the dual-frequency System 200. Antenna and sensor are tripod-mounted.**
- **SR261 weighs 1.0kg, AT201 weighs 0.6kg.**
- *Power consumption:* **8.5W (for SR299/299E), max. 12W (total). External NiCad batteries, or any other 12V DC power source.**
  Example: **7 Amp.hr. battery will run SR261 and CR233/244 for 10-12hrs.**
- **Log data to CR233/244 memory cards.**
- **Optional RTCM input capability for real-time DGPS (CR244).**

Software: **SKI-L1 is the L1-only version of the SKI™ post-processing software supporting conventional static, as well as "rapid static", "stop & go", "reoccupation", "kinematic" survey modes, differential code solutions.**

Comment: **"Rapid static" using single frequency measurements requires longer observation times, 15-20 mins for 5km baseline, more than 30 mins for 10km baseline.**

Options: **Real-time DGPS navigation (RTCM capable), PC base station software.**

LEICA GPS System 300

Manufacturer: **LEICA, Heerbrugg, Switzerland.**

Applications: **All System 200 applications plus geodesy.**

Components: **SR399 sensor with integrated microstrip antenna (optional SR399E configuration with external AT201 antenna), CR333 controller (optional CR344 for I/O of RTCM messages).**

Features:

- ***Precision:*** **±(5mm + 1ppm) for static GPS, ±(5-10mm + 1ppm) for "rapid static" & "reoccupation" solutions, ±(10-20mm + 1ppm) for "stop & go" & "kinematic" solutions, submetre level DGPS accuracy.**
- **Tracks 9 satellites on both L1 & L2 frequencies. C/A (narrow-correlator technology), P1 & P2 pseudo-ranges. L1 phase tracked by reconstructing carrier via C/A code. L2 phase tracked by: (a) P code-correlation (when AS off), or (b) Leica proprietary codeless technique (when AS on) which is <u>different</u> from that used in the System 200.**
- **Physical characteristics, power consumption, memory options, etc., are identical to System 200.**
- **Most hardware and software options are identical to those offered for the System 200.**

Comment: **The System 300 overcomes the criticism levelled at it that the L2 phase was only a half-wavelength observable when AS is on.**

Software: **SKI™ post-processing software: conventional static, as well as modern "rapid static", "stop & go", "reoccupation", "kinematic" survey modes, differential code solutions.**

Options: **Real-time kinematic surveying, "on-the-fly" ambiguity resolution, DGPS navigation (RTCM I/O capable), PC base station software, aerial photogrammetry application.**

ASHTECH MS-XII

Manufacturer: **Ashtech Telesis Inc., Sunnyvale, California, USA.**

Applications: **Static surveying, DGPS navigation.**

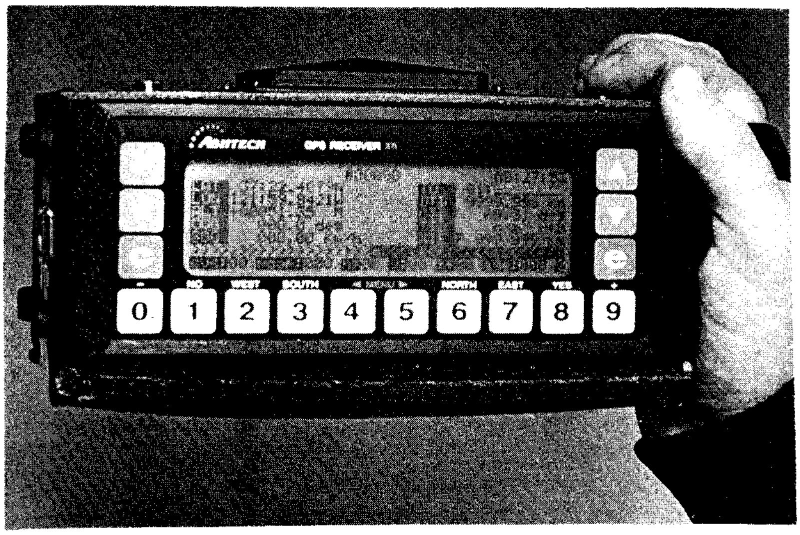
Components: **Single handheld unit with integrated antenna & receiver electronics.**

Features:

- ***Precision:*** **±(10mm + 2ppm) for L1 phase-based solutions, 1-3m DGPS accuracy.**
- **12 channel instrument, measuring L1 phase & C/A pseudo-ranges for up to 12 satellites.**
- ***Dimensions:*** **124x127x216mm, 3.3kg.**
- ***Power consumption:*** **4.1W, internal or external batteries (10-18V DC). Internal NiCad supports 8hrs operation.**
- **Optional RTCM or Ashtech format DGPS input capability.**
- **Two RS232 I/O ports for data communication.**
- **Data output from 1 per second.**
- **1Mb internal memory (expandable to 4Mb).**
- **8x40 character LCD display.**

Comment: **Single frequency instrument for static surveying and real-time DGPS applications.**

Software: **GPPS™ post-processing software (old) or PRISM™ suite (new), real-time differential, etc. Interface to electronic map and GIS.**



ASHTECH Dimension

Manufacturer: **Ashtech Telesis Inc., Sunnyvale, California, USA.**

Applications: **Static & kinematic surveying, GIS applications.**

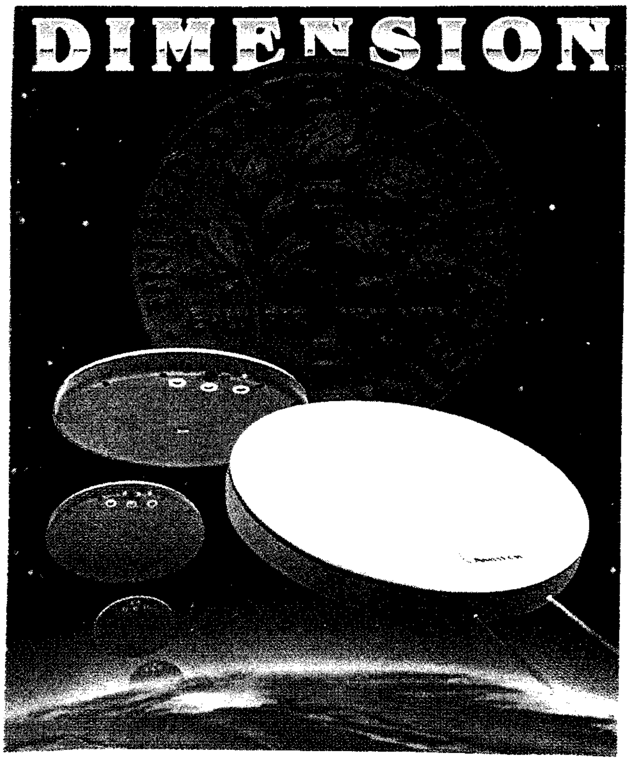
Components: **Single unit with integrated antenna & receiver electronics in "frisbee" shaped package.**

Features:

- ***Precision:*** **±(10mm + 2ppm) for surveying, 1-3m code-based kinematic GIS, dm accuracy for static GIS (with data averaging).**
- **Single frequency, L1 phase & C/A pseudo-range, tracking on up to 12 satellites.**
- **235mm dia., 43mm high, 1.6kg (without batteries & pole).**
- ***Power consumption:*** **4.1W, NiCad or alkaline battery operation.**
- **Optional RTCM or Ashtech format DGPS input capability.**
- **Two RS232 I/O ports for data communication.**
- **Data output from 1 per second.**
- **2Mb internal memory (expandable to 4Mb).**
- **HP-95 or CMT-CV5 datalogger interface.**

Comment: **Single frequency, convenient pole-mounted instrument specially designed for GIS applications.**

Software: **GPPS™ post-processing software (old) or PRISM™ suite (new), real-time DGPS, etc.**



ASHTECH Z-12

Manufacturer: Ashtech Telesis Inc., Sunnyvale, California, USA.

Applications: Geodesy, all survey & precise navigation applications.

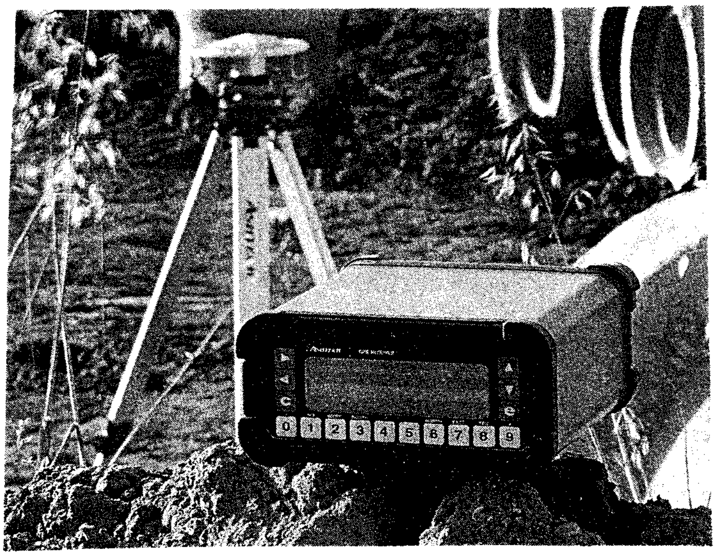
Components: Antenna, tripod-mounted.
Receiver box with LCD display and internal memory.

Features:

- *Precision:* ±(5mm + 1ppm) for static GPS, ±(10mm + 1ppm) for kinematic GPS, cm-dm accuracies for baselines <100km for kinematic (resolved ambiguities).
- Tracks up to 12 satellites, L1 & L2 phase, P1 & P2 pseudo-range measurements.
- Same size as M-XII models: 99x203x216mm, 3.9kg.
- Microstrip antenna (with groundplane 292mm dia.), 1.7kg.
- *Power consumption:* 18W, external batteries (10-32V DC), & other power sources.
- Four RS232 ports for data communication.
- Optional RTCM or Ashtech format DGPS I/O messages.
- 8x40 character LCD display.
- Internal RAM memory (expandable).

Comment: The Z-12 model is especially versatile even when AS is on (full P1, P2, L1, L2 observables using Ashtech proprietary codeless technology). Also available in configuration as Z-12 Field Surveyor™ in which some of the Z-12 features are optional.

Software: GPPS™ post-processing software (old) or PRISM™ suite (new), plus several software packages for kinematic survey (PNAV™), real-time DGPS, real-time surveying, etc.



TRIMBLE Land Surveyor II

Manufacturer: **Trimble Navigation, Sunnyvale, California, USA.**

Applications: **All survey & precise navigation applications.**

Components: **Separate pole-mounted dome antenna with backpack receiver & handheld data logger/controller.
Alternatively tripod-mounted receiver box with LCD display & internal memory, plug-on microstrip antenna.**

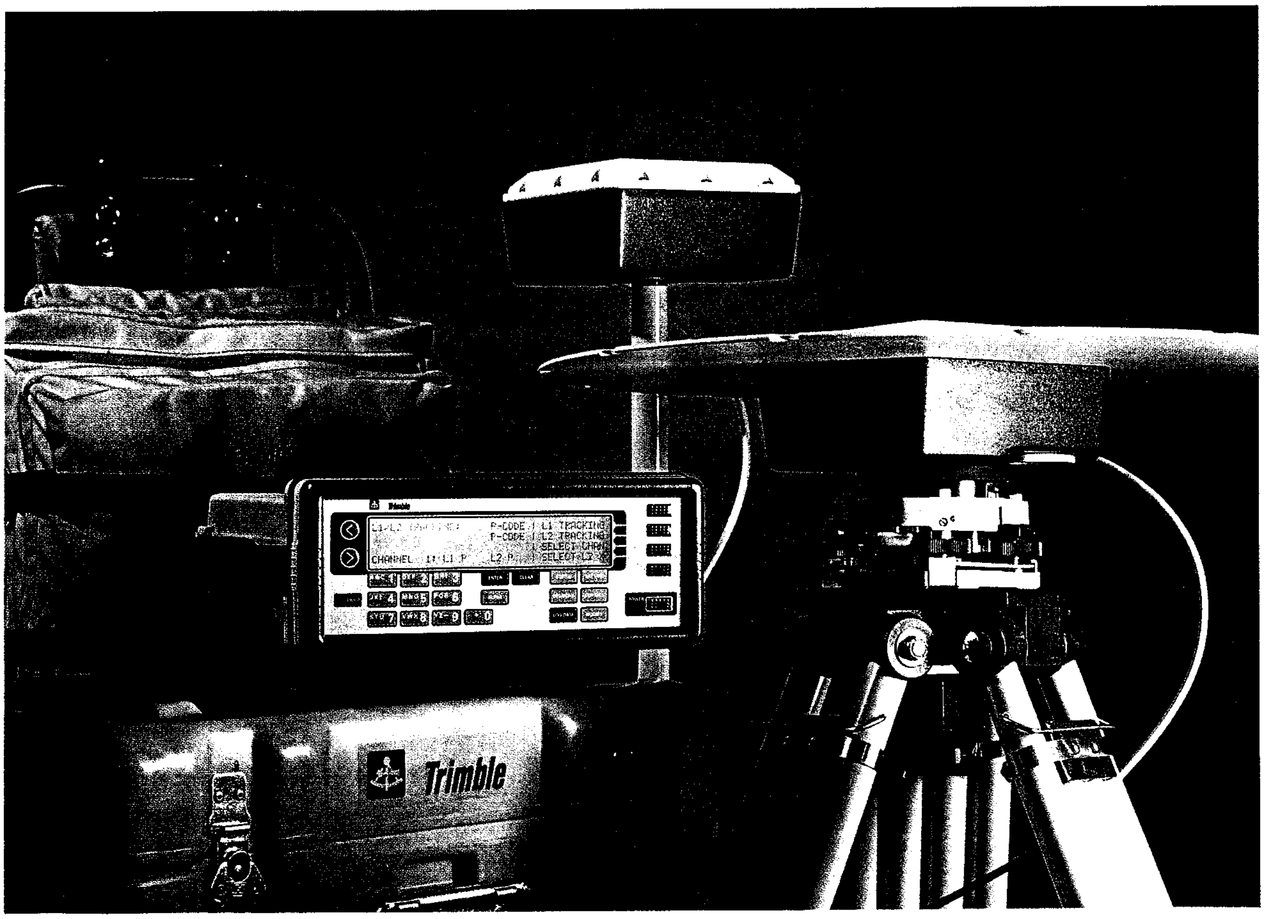
Features:

- *Precision:* **±(10-20mm + 2ppm) phase baseline solution, static or kinematic positioning mode.**
- **9 channel single frequency L1 phase and C/A code pseudo-range measuring instrument.**
- *Dimensions:* **102x248x280mm, 3.1kg (without batteries, etc.).**
- **Microstrip antenna plug-on receiver box (when tripod-mounted), or optional pole-mounted kinematic antenna.**
- *Power consumption:* **Max. 6W, external camcorder batteries (10.5-35V DC), & other power sources.**
- **Two RS232 ports for data communication.**
- **System Surveyor II™ model has RTCM capability, time-mark output, event mark input. RTCM input (standard) for real-time DGPS operation, RTCM output (optional).**
- **Data output from 2 per second.**
- **4x40 character backlit LCD display, softkey entry.**
- **Internal RAM memory (expandable from 1mB), or optional handheld data logger/controller.**
  Example: **1Mb RAM typically holds between 1 & 36hrs of data (depending upon rate data).**

Comments: **This is the "entry-level" single frequency instrument based on the most recent technology. Many options can be incorporated into this model giving rise to a confusing array of systems, e.g. System Surveyor II™, Site Surveyor II™, Trimble GPS Total Station™, etc.**

Software: **TRIMVEC-plus™ post-processing software (old), GPSurvey™ suite (new), TRIMMAP™ topographic mapping software, other specialised software for real-time operations, etc.**

Options: **Real-time (RTK) surveying receiver is the Site Surveyor SE™, fully optioned DGPS configuration is the System Surveyor II™, optional separate handheld data logger (memory cards or RAM) and controller.**


Trimble

TRIMBLE Geodetic Surveyor

Manufacturer: **Trimble Navigation, Sunnyvale, California, USA.**

Applications: **Geodesy, all survey & precise navigation applications.**

Components: **Separate pole-mounted kinematic antenna with backpack receiver, & optional handheld data logger/controller.
Alternatively receiver box with LCD display & internal memory, with tripod-mounted geodetic microstrip antenna.**

Features:

- *Precision:* **±(5-20mm + 1ppm) phase baseline solution, static or kinematic positioning mode.**
- **9 channels (optional 12 channels), tracking both L1 & L2. P1 & P2 pseudo-ranges (AS off), P1-P2 (AS on). L1 phase tracked by reconstructing carrier using C/A code. L2 phase tracked by: (a) reconstructing carrier via P code (when AS off), or (b) Trimble proprietary codeless technique (when AS on).**
- *Dimensions:* **As for Land Surveyor receiver box.**
- **Geodetic antenna 480mm dia. (groundplane), 90mm high, 3.2kg. Kinematic antenna 160x160x90mm, 1.6kg.**
- **Geodetic System Surveyor™ model has RTCM capability, time-mark output, event mark input. RTCM input (standard) for real-time DGPS operation, RTCM output (optional).**
- *Power consumption:* **Max. 9W, external camcorder batteries (10.5-35V DC), & other power sources.**
- **Optional external frequency input.**
- **Physical characteristics, memory options, etc., are identical to Land Surveyor models.**

Comments: **This is the first dual-frequency Trimble receiver developed to measure L2 phase when AS is on.**

Software: **TRIMVEC-plus™ post-processing software (old), GPSurvey™ suite (new), TRIMMAP™ topographic mapping software, other specialised software for real-time operations, etc. Supports "stop & go", "fast static", etc., modes.**

Options: **Real-time surveying receiver is the Site Surveyor SSE™, optional separate handheld data logger (memory cards or RAM) and controller, fully optioned real-time DGPS configuration is the Geodetic System Surveyor™.**

TRIMBLE Geodetic Surveyor SSi

Manufacturer: **Trimble Navigation, Sunnyvale, California, USA.**

Applications: **Geodesy, all survey & precise navigation applications.**

Components: **Separate pole-mounted kinematic antenna with backpack receiver, & optional handheld data logger/controller.
Alternatively receiver box with LCD display & internal memory, with tripod-mounted geodetic microstrip antenna.**

Features:

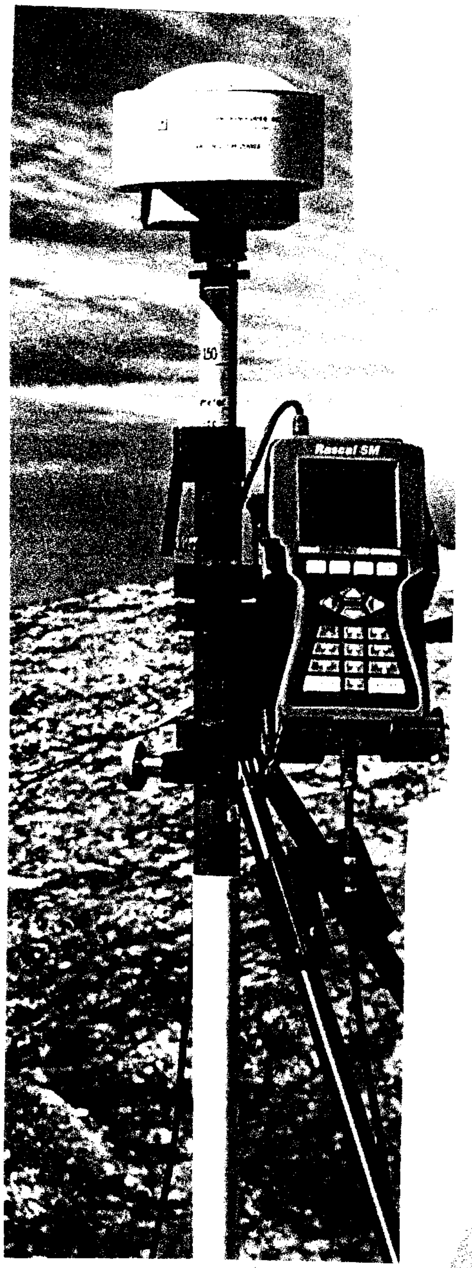
- *Precision:* **±(5-20mm + 1ppm) phase baseline solution, static or kinematic positioning mode.**
- **9 channels (optional 12 channels), tracking both L1 & L2. P1 & P2 pseudo-ranges (AS off), P1-P2 (AS on). L1 phase tracked by reconstructing carrier using C/A code. L2 phase tracked by: (a) reconstructing carrier via P code (when AS off), or (b) Trimble proprietary codeless technique (when AS on).**
- *Dimensions:* **As for Geodetic Surveyor receiver box & antennas.**
- **Geodetic System Surveyor SSi™ model has RTCM capability, time-mark output, event mark input. RTCM input (standard) for real-time DGPS operation, RTCM output (optional).**
- **Internal 2.5Mb RAM memory (expandable to 20Mb), or optional handheld data logger/controller.**
- *Power consumption:* **Max. 10.5W, external camcorder batteries (10.5-35V DC), & other power sources.**
- **Otherwise physical characteristics and options are identical to Geodetic Surveyor models.**

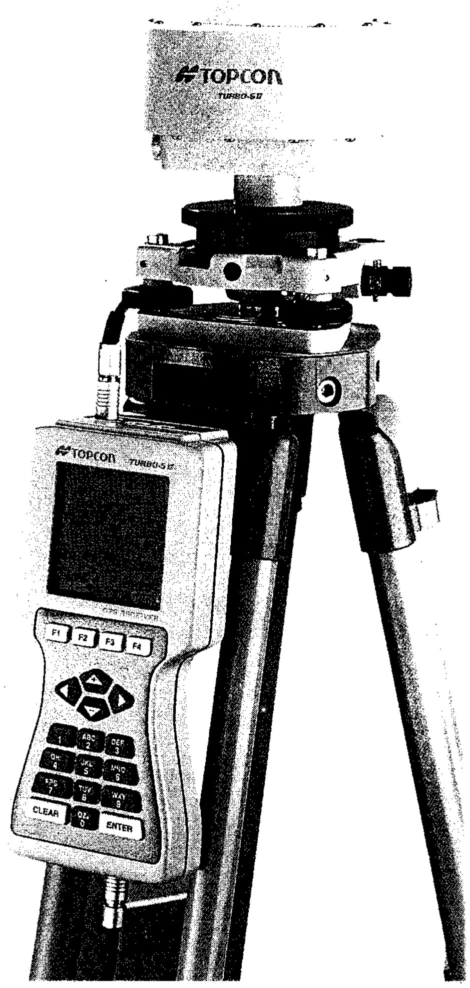
Comments: **This is an improved dual-frequency Trimble receiver able to measure L2 phase when AS is on, boasting super-trak™ signal processing technology to suppress multipath, and allow low noise tracking.**

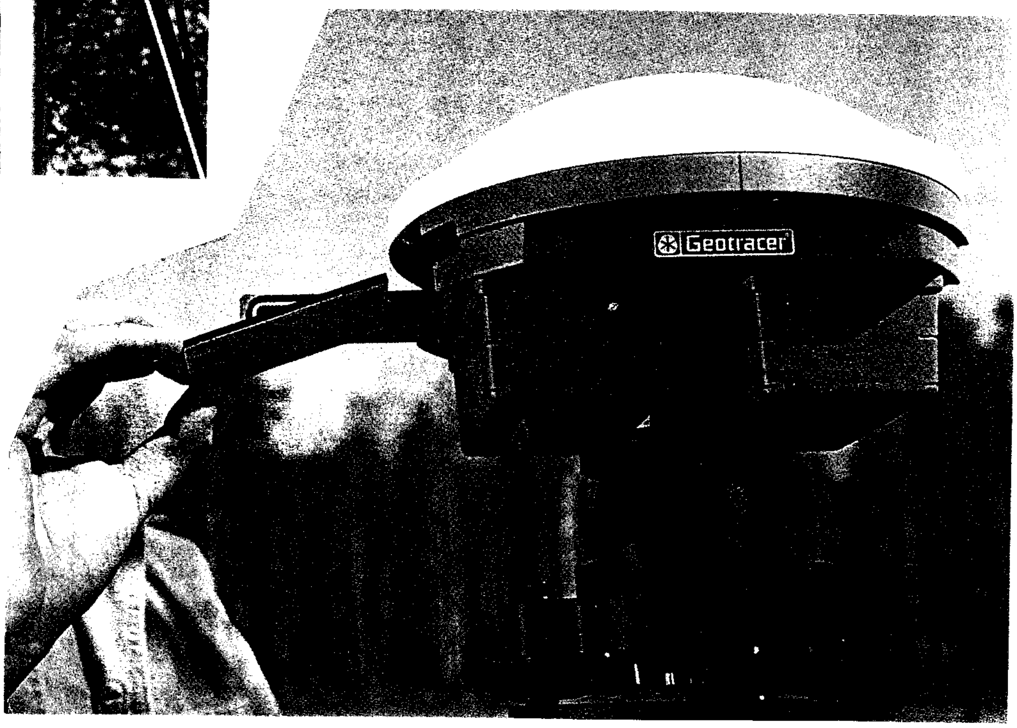
Software: **GPSurvey™ suite, TRIMNET-plus™ network adjustment software, TRIMMAP™ topographic mapping software, other specialised software for real-time operations, etc. Supports "stop & go", "fast static", etc., modes.**

Options: **Real-time surveying receiver is the Site Surveyor SSi™, optional separate handheld data logger (memory cards or RAM) and controller, fully optioned real-time DGPS configuration is the Geodetic System Surveyor SSi™.**

Other GPS Surveying Systems:







4.3
GPS SURVEYING SOFTWARE

A variety of software packages are available for GPS phase data processing. These packages have either been developed by universities and government departments (for research uses, for internal operational purposes, or for very high precision scientific applications), or by the receiver manufacturers. Although the detailed architecture of GPS software varies considerably from one package to another, there are a number of functions that most comprehensive GPS surveying software packages must carry out.

The major components of any package include software to aid pre-survey planning, decision making and reconnaissance (§5.2 and §5.3); to support field observations (including, increasingly, the high productivity survey methods such as "stop & go", "kinematic" and other such techniques, §5.5), data pre-processing and checking (§5.4 and §7.3); single-session data processing (§9.2); network adjustments and quality control (§9.3 and §10.3), and the transformation of the results into a local geodetic reference system (§11.1 and §11.2).

4.3.1 THE IDEAL SOFTWARE PACKAGE

Such a package does not exist! GPS manufacturers are continuously refining the software that they offer with their products. There are, however, a number of reasons why software refinement cannot be carried out to the degree that is perhaps possible, and one of the most important has to do with the dynamic nature of the GPS technology itself. Valuable development resources are generally directed to the creation of new, or significantly revised software, rather than enhancing existing software. Nevertheless, it is worth considering what makes a "good" software system so that purchasers of GPS technology can carefully evaluate the software component of the total GPS surveying "package", as well as the hardware component.

Various criteria can be used to evaluate software. The following is a list of some of the characteristics that may be appropriate.

Primary Criteria:

- Accuracy.
- Reliability.
- Support high productivity surveying techniques.
- Efficiency of data flow.
- Ease of operation.

Secondary Criteria:

- Level of automation.
- Processing speed.
- Processing flexibility.
- Data analysis and quality control capabilities.
- Software portability and ease of maintenance.
- Data file storage and I/O requirements.
- Computer hardware requirements.

Accuracy and **reliability** are dependent not only on the data processing algorithms that are used, but also on the planning and field procedures, as well as the hardware. Hence, if the field data collection procedures lead to poor quality data, or an inadequate quantity of data, then no matter how sophisticated the software, the accuracy and reliability of the baseline results will suffer. However, they do directly influence the selection of the mathematical models and the processing strategies to be employed for any given survey application.

Although efficient and fast processing is desirable, this will obviously be dependent upon the computer implementation, the algorithms used, the amount of data to be processed, and the accuracy required. High accuracy applications generally requires the flexibility to intervene at various stages to control the processing. **For *scientific* or geodetic applications, commercial software is not adequate.** "Part per million" accuracy (standard surveying) applications place a higher priority on automation or ease-of-use, with minimal operator intervention.

Nowadays, **an important marketing factor for a GPS "package" is the ability to support high productivity survey techniques** (§5.5). The field procedures are very different to those of the *conventional* static GPS surveying method. However, it is the innovative data processing algorithms that make possible the determination of high accuracy results in a fraction of the time required by the conventional GPS baseline processing technique.

Efficient data flow implies a natural flow of data from raw to finished form in some *logical and efficient sequence of operations*, and consequently has a direct impact on the software structure. Once a software architecture has been specified, it then becomes possible to address the related secondary criteria of automation, data storage, and computer hardware requirements, etc. (The integration of a Data Management System is central to ensuring this desired outcome -- it is in this area that many of the early commercial software packages were weakest.)

Many of the above secondary criteria will be determined to a large part by the observation models and processing methodologies used, and hence will have different optimal realisations depending upon the application. It is also important that the software structure allow for flexibility in processing options, and be adaptable to changes as new models and methodologies are developed. Finally, the software should be portable to allow easy implementation in different computing environments. This may be difficult to achieve if the user interface is device dependent and highly automated. Hence the programming language, operating system, and DMS structure selected will require a similar trade-off having to be made between efficiency, speed, and automation on the one hand, and software *portability and maintainability* on the other.

For the following discussions, GPS data processing can be considered as consisting of essentially four distinct tasks:

- **Planning**
- **Pre-processing**
- **Phase data adjustment**
- **GPS survey result analysis**

Each of these is discussed below.

Comments on Planning Software

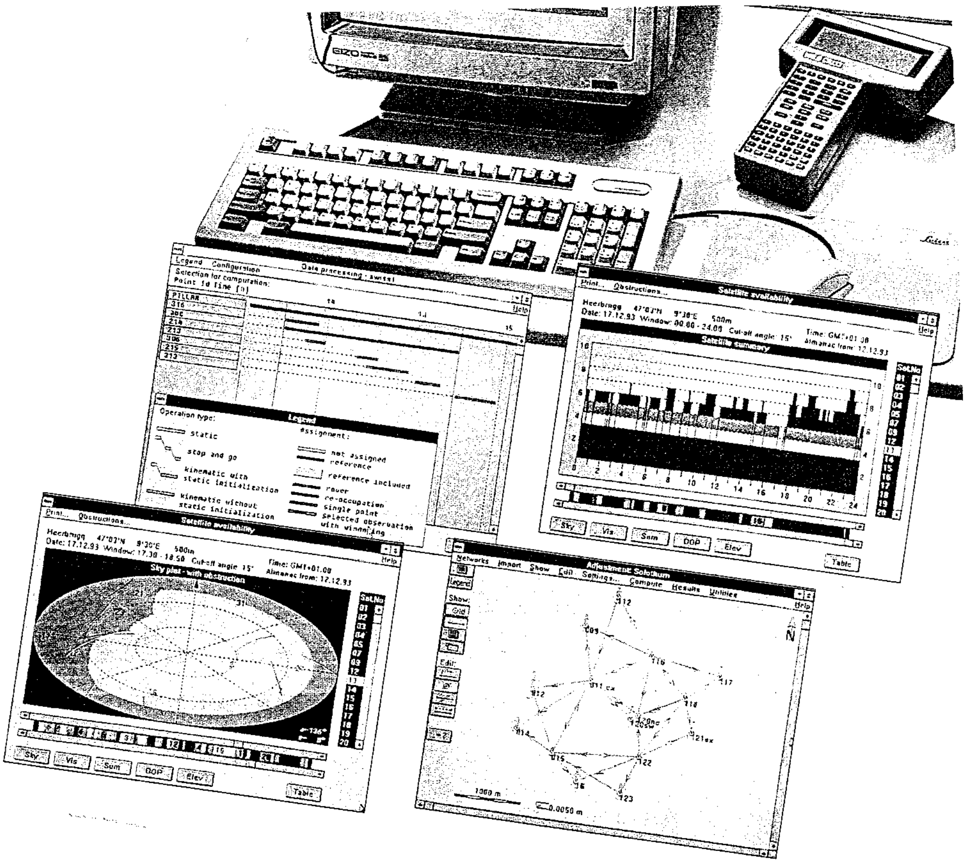
It is generally this software component that most impresses, as it makes full use of colourful "menus" and dazzling graphics. It is necessary to distinguish between the comparatively simple satellite availability-visibility software, and the software intended to support pre-analysis of a GPS observation session, or of an entire campaign.

The following are some comments on mission planning software:

- The software can generate some, or all, of the following:
  - graphs of satellite availability-visibility at a site,
  - graphs of rise-set times of GPS satellites,
  - graphs of GDOP (and other DOP factors),
  - graphs of azimuth-elevations of GPS satellites,
  - satellite groundtrack plots, and
  - skyplots.

  A flexible user interface to permit rapid recomputation when the scenario is varied (time of day, satellites to be excluded, site location, etc.). All output is directed, in the first instance, to the screen, with hardcopy output always possible.
- This data is generated from stored satellite ephemeris data, hence there are a number of options for inputting this information. The simplest is to install the planning software on the same computer as the data processing software. An automatic procedure would ensure that whenever real data files (phase data, Navigation Message, satellite health and almanac information, etc.) are decoded, a check is made and, if necessary, the computer-resident ephemeris-health-availability information is updated. In general, commercial software requires proprietary file format input for the satellite ephemerides. Some software permit alternative input modes, such as the NASA bulletin data, approximate orbital elements, RINEX files, Yuma almanac files, or other bulletin board formats.
- The satellite ephemerides need not be very accurate. For high altitude satellites such as GPS the orbits are very predictable, and hence almanacs up to several months old can still be used. (However one should always guard against satellites being withdrawn from service or moved into other orbital positions.)
- Planning data, such as that listed above, is specific to a site. However, changes to the coordinates of the site of quite large magnitude (>100km) do not noticeably affect a skyplot, or a satellite rise-set graph. Hence the planning software need only be executed for one centrally located site within the network of points to be surveyed.
- Because the GPS orbits are synchronous with sidereal time, the information is valid day after day, but for an approximate time shift four minutes earlier each day.
- Pre-analysis software is analogous to network optimisation software used for planning conventional control surveys. In the first instance, an observation scenario is tested to estimate the internal precision ("noise-only") of a GPS survey on a single-session basis. Then the quality of the propagated GPS survey can be tested by considering the manner in which the independent and redundant baselines combine into a network. *Such software is generally not provided by GPS manufacturers.*
- Pre-analysis software to test the external accuracy of a GPS survey is a subsequent step, which makes use of the coordinate information (and its quality) of selected GPS stations from other sources. Generally the geodetic coordinates of some stations are known in the local datum, and as they are often held fixed when GPS results are transformed into the local datum, they can be considered external constraint information. The (simulated) distortion of the GPS-only network to the external control can then be studied. *Such software is generally not provided by GPS manufacturers.*

- ❑ Pre-analysis software that takes into account systematic errors in the residual biases (for example, orbits, atmospheric refraction) is not generally available.



Comments on Pre-Processing Software

The pre-processing component of the software package generally:

(1) Downloads the recorded data from internal memory or removable memory cards, to a computer.

(2) Prepares the files containing the raw observations and Navigation Message data. If a Data Management System is integrated within the software, the appropriate file names are assigned and catalogued for later use. In addition, the station record file is interrogated and an appropriate entry made in a log file. This file contains such items as user-entered station name, receiver serial number, antenna height, etc.

(3) Reformats the data files if necessary. For example, if the data is to be archived, or transmitted to a processing centre where different phase reduction software is to be used, a "standard" (receiver-independent) format may be used. The RINEX (Receiver INdependent EXchange) format is the preferred format at present (GURTNER, 1994), see §7.3.

(4) Computes preliminary station positions using pseudo-range data. As a by-product, the receiver clock offsets from GPS time can be calculated.

(5) Repairs cycle slips in the phase data (§7.3). This is often undertaken after a triple-difference phase solution (see below).

The outcome of the pre-processing step is a set of "clean" data files, and ancillary information, to support subsequent phase data processing. The following additional comments need to be made:

❑ The pre-processing software is invariably written by the GPS instrument manufacturers, and is specific to the operation of their receivers. Some pre-processing modules (particularly the ones that carry out operations (3), (4) and (5) above) are an integral part of third-party GPS software as well.

❑ Step (4) may not be necessary if the code-correlating GPS receiver automatically resets its own clock to GPS Time during data tracking (as, for example, in the case of the Leica GPS receiver -- see §6.3).

❑ **Cycle slip repair** is generally a time consuming and laborious task if carried out *manually*. This is no longer necessary for standard short baseline GPS surveying, even in the "kinematic" mode. These days "standard" pre-processing for cycle slip detection and repair is carried out automatically (receivers may correct them in the field as data is being recorded). Only in the case of precise, scientific, long baseline applications is some level of manual screening of the data required.

❑ Some pre-processing software *creates differenced data files* (for example, between-receiver differences, or double-differences), perhaps requiring the analyst to make a decision on the processing strategy at the pre-processing step.

❑ Some pre-processing steps (data downloading, reformatting, receiver point positioning) *can occur on a single site basis*, but other steps require that the data from all receivers operating simultaneously be first brought together.

❑ The RINEX format has received strong endorsement as the "official" GPS data format. Many GPS receivers output data either directly in this format, or via a utility program to translate the (downloaded instrument-specific) internal format into the RINEX data format (IBID, 1994).

❑ There is a trend to include pre-processing software within the receiver itself, and hence making it largely "transparent" to the user. *This is particularly true for real-time GPS surveying!*

Comments on Phase Data Reduction Software

This component is at the "heart" of the GPS software package. Phase data reduction software tends to fall into three broad classes:

☞ Software developed by the instrument manufacturer, and offered as a "package" with the GPS receiver hardware to address standard land surveying applications. *The so-called "commercial" software.*

☞ Software developed by third-parties, generally government departments or academic institutions *intended for very high precision "scientific" (or geodetic) applications.*

☞ Software *intended for "specialist" (or unusual) applications*, such as to support airborne and marine operations, GIS data capture, multi-antenna attitude determination systems, GPS systems integrated with other sensors (including aerial cameras, inertial systems, etc.).

There are several distinctions to be made between these classes of software:

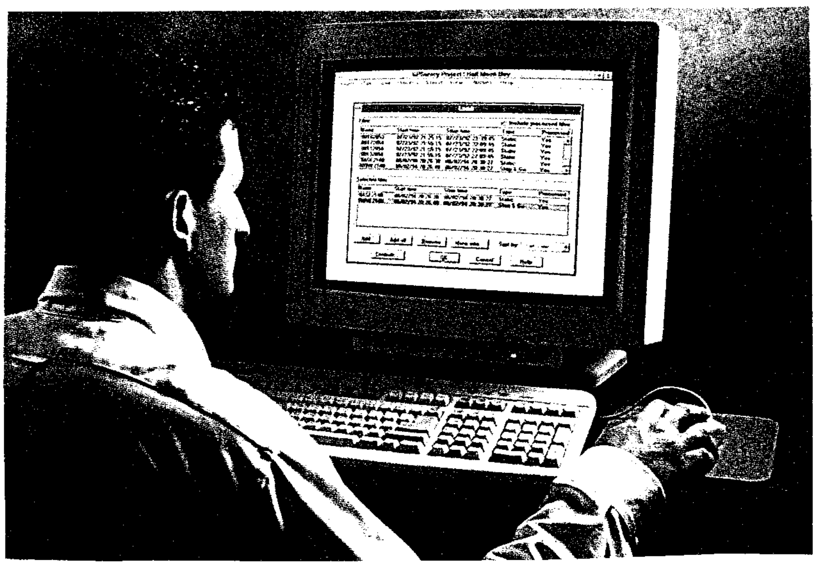
- ❑ The *commercial* and *specialist* software is invariably written to handle data from one instrument type. The *scientific* software generally is instrument independent, accepting data in the RINEX format.
- ❑ The *commercial* software tends to be "user-friendly", requiring a minimum of analyst input and runs on PC computers. The *scientific* software tends to have been developed for research purposes, offering many options and requiring more analyst skill to use. In addition, because such software has many more features and supports more complex data modelling, the computer requirements are generally more severe.
- ❑ The *commercial* software is optimised for GPS surveying accuracies (a few parts per million relative positioning accuracy), whereas *scientific* software generally addresses higher accuracy applications. The *scientific* software has more sophisticated modelling and processing strategies, such as the ability to adjust orbital parameters, estimate tropospheric scale factors, process more than one observation session simultaneously, etc.
- ❑ The *commercial* software tends to use sub-optimal algorithms, generally processing data on a single baseline mode (even if more than two receivers were operating simultaneously), whereas the *scientific* software generally has multi-baseline and multi-session capability.
- ❑ The range of *specialist* software is growing rapidly. Some *specialist* software may be quite "polished" products, others may have rather crude interfaces.

Although these above remarks are generalisations, there is nevertheless a difficult *compromise* to be made between insisting that the operation of the manufacturers' commercial software be largely transparent, and: (a) having the data processed in as mathematically rigorous a manner as possible, or (b) the software having the necessary versatility so as to address many of the unusual applications as well. Such a tension is not evident with scientific software developed specifically for high accuracy applications, or software developed for specialist applications.

It is not the intention of these notes to address topics in "GPS geodesy", or to comment on many of the unusual applications that GPS is increasingly being called to address. The focus will instead be on studying the elements of GPS phase data reduction software which are commonly found in the commercial software provided by the instrument manufacturers as part of their GPS "package". Some of these elements are:

- There is generally a sequence of processing steps from the least rigorous to the "optimal" solution:
  - **triple-difference solution**: moderate accuracy, but robust (insensitive to cycle slips) and hence ideal for preliminary station coordinate solutions,
  - **double-difference solution with free ambiguities**: for long baselines this may be the "best" solution obtainable,
  - **double-difference solution with fixed ambiguities**: for short baselines when the estimated ambiguities can be "resolved" to their nearest correct integer value, and the solution repeated using "phase-range" or "carrier-range" data.
- The comments above are valid for conventional static GPS baseline solutions. *Modern "rapid static", "stop & go" and "kinematic" survey techniques require resolved ambiguity solutions only* (§5.5 and §8.3).

- Software for dual-frequency instrumentation allows for several options (depending on the baseline length), some of which lead to ambiguity-fixed solutions, others provide ambiguity-free solutions (§8.4).
- Most commercial software packages can process only single baselines, even when more than two receivers had been operating simultaneously in the field (PoPS™ and TRIMMBP™ are exceptions -- however both packages have been superseded with later single-baseline versions from the manufacturers). (Rigorous mathematical processing requires a multi-baseline approach in which the between-station correlations are taken into account.)
- Processing options are generally fixed so that in an operational data processing environment phase data reduction is executed according to a "recipe".
- Pseudo-range data is generally not used during phase data reduction, certainly not for conventional static GPS surveys.
- Data from more than one observation session cannot normally be rigorously processed in one step (PoPS™ was the lone exception in commercial software -- now superseded by the SKI™ single-baseline program).
- If the observation residuals indicate an unresolved cycle slip, then the data has to be re-scanned, and the phase data eduction process repeated. This is not usually done in commercial software running in "automatic" mode.
- The phase data reduction step (be it in the single-baseline or multi-baseline mode) results in a minimally constrained "mini-network".
- The output of the phase reduction process (be it single-baseline or multi-baseline) is then input to a network adjustment program.
- The output variance-covariance information is generally optimistic and does not truly reflect the accuracy of the GPS adjustment (the accuracy may have to be artificially "deflated" in the network adjustment).
- The output coordinate results are given in an approximate geocentric WGS84 coordinate system (utility programs are available to transform these results into geodetic or map projection coordinates), and refer to the groundmark over which the GPS antenna was plumb.



Comments on Network Analysis Software

The processing of GPS phase data may be carried out for the smallest unit possible: a single baseline, or in a large simultaneous adjustment of all data collected in a survey, or in any combination in between. In any case, there are several common features of all phase data reduction:

☞ They are *minimally constrained* solutions based on the coordinates of one station in the network (or baseline) being held fixed.

☞ The results are in the form of 3-D coordinates (either in the Cartesian system X,Y,Z, or as geodetic coordinates ϕ,λ,h referred to some ellipsoid).

☞ The coordinate results are referred to the WGS84 reference system (at the level of accuracy that this is defined by the fixed datum station).

If the phase data reductions are performed piecemeal, that is **single-baseline processing** or (the preferable) **single-session processing** mode, the separate results must be combined together in a subsequent **network adjustment**. Hence the secondary network computation software must be able to:

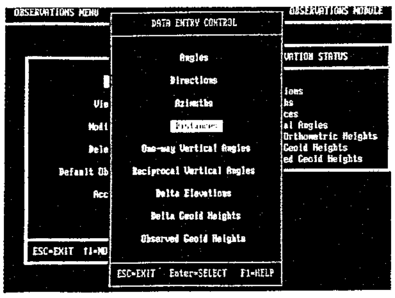
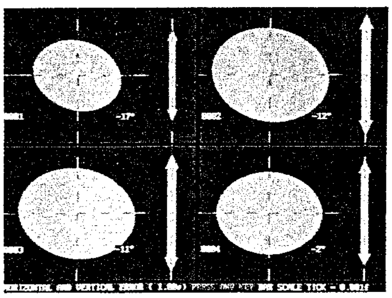
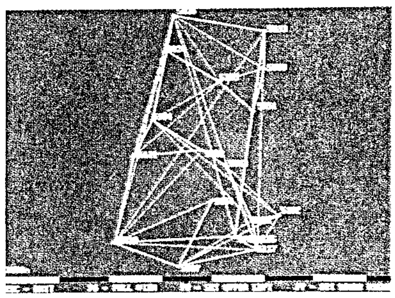
- ❑ Combine individual baseline (or multi-baseline) solutions into a network. As there are invariably redundancies in the GPS survey (for example, multiple GPS station occupancies, repeated baseline observations, etc.), the best way of obtaining an optimal coordinate solution is to *input* the GPS results (and associated variance-covariance information) into a network analysis program. The combined solution could then be left as minimally constrained -- the coordinates of only one station are held fixed.
- ❑ Include external station control coordinates within the GPS network solution to provide both a means of relating the GPS-only solution (in the quasi-WGS84 system) to the local geodetic datum, and in order to constrain the GPS-only solution to fit the local control. The former requires the determination of the transformation parameters between the GPS and local geodetic system.
- ❑ Convert the GPS ellipsoidal heights into the more useful orthometric heights through the input of geoid height information for some or all of the surveyed stations.
- ❑ Report writing and presentation software able to support survey report generation with coordinate lists, network diagrams, table formatting and spreadsheet capabilities.

The above tasks can be performed by *network adjustment software* which may be written explicitly for this purpose, or by conventional geodetic network adjustment software which has been modified to handle the GPS output (from the phase reduction software), as another type of geodetic "observable". This is in addition to terrestrial survey data such as horizontal directions, distances, zenith distances, etc. Some of the characteristics of network adjustment software used in GPS survey computations are (see also chapter 9):

- The software may be developed by the GPS instrument manufacturers, and offered as an additional module within the overall GPS software "package". However, in the past, the preferred software packages have been written by geodesists experienced in conventional geodetic adjustments and are offered as a separate package.
- The major modifications to conventional geodetic adjustment software necessary to accommodate GPS baseline solutions are: (a) to permit a 3-D adjustment capability (most conventional network adjustment programs can only handle 2-D data), (b) to partition coordinate datasets into several different datums, related to each other through the

appropriate selection of transformation model (see §11.1), and (c) permit full VCV matrices to be input.

- The most popular software packages have "toolkits" that can read the output files of most commercial phase data reduction software, thus obviating the need to manually type in the "data" (GPS coordinates and variance-covariance information).
- It is often necessary to "rescale" the VCV information provided by the phase reduction software in order obtain more realistic estimates of GPS accuracy (it is invariably over optimistic).
- The network adjustment software is a useful means of monitoring the overall quality of the GPS survey as it progresses. The results of a baseline, or session, adjustment can be incrementally fed into the network program, and the solution checked. If an outlier is detected (for example, a bad baseline) then this could require a modification to the survey plan (perhaps a reobservation of the suspect baseline, or the survey of additional stations).
- The output of such a program is a set of station coordinates (on the WGS84 datum or transformed into a local geodetic datum) and the VCV information, generally presented in the form of absolute or relative error ellipses (or ellipsoids).
- Some adjustment programs have a simulation mode. GPS survey plans can be tested (or "optimised" in the parlance of geodetic adjustment) by specifying which stations will be occupied so as to fulfil some predefined accuracy and reliability criteria.
- Some network adjustment programs may not have a geoid computation capability built-in, hence a separate package may be required (see §11.3). Geoid computation programs are relatively uncommon, though there are geoid height datasets. For example, there is a program available for interpolating from a precomputed gridded Canadian geoid, as well as one for Australia. In addition, programs have been developed that can compute a geoid height value using a geopotential model as input (though these are nowadays being increasingly incorporated within the network adjustment software). A very small number of programs are able to compute geoid height values from surface gravity data.
- In Australia, the commonly used network adjustment packages are NEWGAN™ (developed by J.S. Allman, a former staff member of the School of Geomatic Engineering, UNSW), COMPUNET™ (developed by F. Smith) and GEOLAB™ (a Canadian product). These packages can process conventional geodetic observations, either alone or together with GPS results. There are many other "inhouse" programs that can handle GPS-only solutions (with and without transformation parameter determination capability).
- Trimble, Ashtech and Leica all offer network adjustment software as modules within their commercial software packages, for example TRIMNET™, FILLNET™ and SKI™.
- There is a trend to append *result presentation software* to manufacturers GPS packages to aid client report generation, by being able to plot station networks, prepare coordinate lists, etc.

Summary of Capabilities and Features of GPS Surveying Software

In summary, the ideal software package for the processing of GPS data *should* have the following capabilities and features:

- The ability to process data from all available receiver types quickly and efficiently using standard methods of pre-processing for data of similar types, apart from receiver-specific decoding software requirements.
- The ability to process one-way phase data as well as single-, double-, and triple-differences, formed by various differencing bases.
- The ability to process dual-frequency observations, either L1 and L2 data separately, or in various combinations.
- The ability to handle all field scenarios, such as "rapid static", "stop & go", "kinematic", etc.
- The ability to invoke automatic processing involving minimal operator intervention after an initial selection of options for "standard" accuracy applications.
- The ability to estimate tropospheric parameters if required for high precision vertical surveying, or for long baseline applications.
- The ability to combine data from many receivers operating simultaneously, and to process the observations in a multi-baseline solution, if required.
- The ability to efficiently update previously processed solutions with additional information gained from new data, to permit incremental construction of a network solution.
- The ability to exploit the inherent integer nature of the carrier phase ambiguities when processing double-differenced data collected by the receiver(s) in either static or kinematic mode, through the implementation of an ambiguity resolution capability "on-the-fly".
- The ability to include precise ephemerides.
- The ability to assign realistic apriori weights to the observations, and ground station coordinates.
- The ability to incorporate fixed control stations and derive transformation parameters that relate GPS coordinates to local geodetic coordinates.
- The software should be implementable on a microcomputer system under MS-DOS or Windows™ for field processing applications.

There is as yet no such software package! ***Though each year we move ever closer to the ideal.***

4.4

TESTING GPS SURVEYING SYSTEMS

The dilemma facing individuals and government Certifying Agencies is deciding on just what aspect of the GPS technology to actually test, and then to decide on how to go about doing it. Consider the following possibilities:

- ❑ Testing of the GPS "system" as a whole -- *primarily the Space and Control Segments* (§2.2).
- ❑ Testing a class or brand of GPS instrument, involving the combination of hardware and software -- *the User Segment* (§2.2).
- ❑ Testing a particular GPS instrument unit -- *on a regular basis, or in response to a suspected problem.*
- ❑ Testing the skill of personnel -- *in field and data processing procedures.*

In this section several testing strategies are mentioned. The most common type of test is based on the results of the baseline solution, a product of the processing of datasets explicitly collected for the purpose of testing specified hardware and software. The results are compared against "ground truth" data, either provided by previous GPS surveys (preferably using a method which is of a higher accuracy than that being tested) or conventional survey techniques. *It is less common to test the raw data quality.*

However, before proceeding to consider some testing strategies, certain difficulties shared by some or all of the testing procedures should be mentioned:

- GPS errors/biases are a function of time -- *how can a test then be truly "conclusive" if it is carried out only once?*
- Some GPS errors/biases are a function of geographic location -- *how can a test be considered "conclusive" if it is carried out in only one location?*
- The propagation of most GPS errors/biases into the baseline solution is a complex combination of factors, such as time, location, baseline length and satellite geometry -- *should all baseline lengths be sampled?*
- The quality of GPS baselines is not just a function of errors and biases, but also the length of observation session and the type of carrier phase solution (§8.1) -- *what operational procedures for data collection and data processing should be insisted upon?*
- The quality of GPS baselines is influenced by the data editing and pre-processing procedures which are used -- *should automatic processing procedures be insisted upon?*
- What should be the outcome of the testing -- *a lifetime "stamp-of-approval", or an annual "certification"?*

4.4.1 CALIBRATION TESTS

Characteristics of these tests are:

- Some tests strive to determine the overall characteristics of GPS performance by testing the total system, others try to isolate as many of the "external" errors/biases as possible.
- Some tests are conducted once, either in a laboratory (of the manufacturer, or an independent organisation) or "in the field", others are conducted on a continuous basis.
- Some tests are conducted as a form of certification, others are conducted by individuals as they see the need to do so.
- Some tests involve specific GPS instrument units, others are tests applied to an entire class or brand of instrumentation.

Three test strategies can be identified under this category:

(1) GPS system "Integrity Monitoring".
(2) Zero baseline tests.
(3) Laboratory testing.

Integrity Monitoring

The following comments are made regarding this test procedure:

- The impetus has come from the civil aviation community. As GPS is being adopted as the primary navigation aid for the Future Air Navigation System (FANS), concerns have been raised about the reliability, availability, integrity and accuracy of the system for GPS navigation users.
- One strategy for monitoring "integrity" is to use a ground network of permanent GPS receivers, located on points of known coordinates, which track all visible satellites, monitor certain performance parameters, and issue a warning when the performance of the GPS system (Space and Control Segments) degrades to such a point that GPS navigation performance is adversely affected.
- A record of the "integrity" of the GPS system from such an Integrity Monitoring (IM) network can be referred to by surveyors when evidence is sought for possible periods of degraded GPS performance.
- In Australia, IM for the civil aviation community (as well as for others) is the responsibility of the Federal Government's Australian Survey and Land Information Group (AUSLIG). AUSLIG has established for this (as well as for other geodetic purposes -- see §12.1) the Australian Regional GPS Network, a permanent network of up to 15 GPS tracking stations (eight of which are located on the Australian continent) -- see Figure 12.1-1. All tracking data is telemetered back to Canberra where real-time algorithms process the data and generate navigation warning messages, which are transmitted to users via a dedicated channel on the OPTUS communications satellite.

It is expected that the GPS survey community will in the near future have easy access to GPS Integrity Monitoring data. Although there may not always be a correlation between poor GPS navigation performance and poor GPS survey baseline results (as GPS integrity is generally very good), any periods of poor system performance as detected by the IM network must be considered with suspicion.

Zero Baseline Testing

A "zero baseline" test can be used to study the precision of the receiver measurements (and hence its correct operation), as well as the data processing software. The following comments are made regarding this test procedure:

- The experimental setup, as the name implies, involves connecting two GPS receivers to the same antenna. The antenna "splitter" can be purchased from specialist electronics shops. Obviously this test cannot be applied to integrated antenna/receiver systems such as some of the Leica GPS instruments (§4.2).
- When two receivers share the same antenna, biases such as those which are *satellite* (clock and ephemeris) and *atmospheric path* (troposphere and ionosphere) dependent, as well as errors such as *multipath* CANCEL during data processing. The quality of the resulting "zero baseline" is therefore a function of random observation error (or noise), and the propagation of any receiver biases that do not cancel in data differencing (§6.3).
- The impact of residual bias effects on the baseline solutions which are a function of baseline length (such as satellite ephemeris bias, the handling of observation time-tags and atmospheric delay) cannot be evaluated.
- An important advantage of this test is that it is comparatively simple to administer -- no specialised software or "ground truth" data is required, and the location of the antenna is immaterial.
- Some GPS surveying receiver manufacturers use this procedure to perform final product testing of all receiver units before they leave the factory.
- No significant time-dependency to the quality of the zero baseline results should be evident, apart from a small effect that is due to the daily variation in receiver-satellite geometry.

Laboratory Calibration Tests

GPS receiver manufacturers perform a variety of component tests; such as on the oscillator, onboard data memory (or memory cards), the receiver firmware, the antenna, batteries, etc. Furthermore, such tests may be carried out over a large range of temperatures (generally from -30° to +50° C), as well as under conditions of artificially induced signal jamming.

A commercial receiver testing service is now on offer by the NAVSTAR GPS Joint Program Office (§2.1). Unfortunately, not a lot is known of this service at the present time, but it is intended primarily for GPS receiver manufacturers, not for the general user community.

CREDIBILITY

\,kred-ə-'bil-ət-ē\

1 : the quality or power of inspiring belief
2 : the quality or power GPS equipment specifications gain with this seal

Commercial Receiver Test Program

GPS EQUIPMENT TESTING FOR COMMERCIAL AND NON-DoD GOVERNMENT CUSTOMERS

NAVSTAR GPS Joint Program Office
internet: gpshouse@nosc.mil *Fax: (714) 458-9174*

4.4.2 CERTIFICATION TESTS

The characteristics of such tests include:

- An attempt to test the performance of the GPS instrumentation (hardware -- §4.2, and software -- §4.3) under realistic conditions.
- The implicit assumption is that the contribution of all GPS system biases to baseline results is largely a function of baseline length and not a function of time or geographic location.
- Some tests are conducted once, for example in the case of newly developed GPS surveying instrumentation, while other tests may be conducted as deemed necessary.
- Some tests are conducted as a form of certification, or validating the claimed performance parameters of GPS survey instrumentation.
- Some tests involve specific GPS instrument units, others are tests applied to an entire class or brand of instrumentation.

Three types of tests in this category shall be discussed:

(1) FGCS GPS survey system testing.
(2) "Legal Traceability" tests.
(3) Pre-mission testing.

FGCS Tests

In January 1983, the U.S. Federal Geodetic Control Sub-committee (FGCS) conducted the first in a series of tests of GPS satellite surveying systems and associated commercial software. The first system tested was the Macrometer™ V1000, a codeless carrier phase survey system. Many other GPS survey systems have been tested since late 1985, including: Texas Instruments TI4100, Trimble 4000S (4-channel), ISTAC 2002, Trimble 4000SX (5 channel), Wild-Magnavox WM101, Motorola Eagle, Sercel NR52, Ashtech XII, Ashtech LD-XII (24 channel), Trimble 4000SST (16 channel), and several recent models of the Trimble, Ashtech, Leica, as well as other companies (§4.2).

The following additional comments can be made:

- The FGCS have tested all GPS surveying instrumentation as it has been released onto the market, and continues to do so. These receivers have included single frequency and dual-frequency instruments. No other country has systematically tested GPS surveying receivers in this way. It therefore functions as a de facto Certifying Agency.
- All tests are performed over a network of up to ten stations located in the vicinity of Washington D.C. (Figure 4.4-1). The baselines for the network vary in length from 186m to about 105km. The standard for evaluating the GPS baseline results are coordinate differences compared with those determined using 1-2 part per million (ppm) terrestrial survey techniques, and a data base of accumulated baseline measurements from all the FGCS test surveys.
- The data collection is carried out by survey personnel provided by the GPS manufacturer, and is reduced by the same personnel using the manufacturer's software. The raw GPS observations are immediately processed using the broadcast ephemerides, under campaign conditions as close to those that would normally be encountered as possible. The tests are only carried out once.

- Such a network was initially used only to test conventional static GPS surveying techniques, however modern high productivity surveying techniques (§5.5) have been recently tested as well, but only for the short baselines. Some real-time GPS surveying tests have been conducted.
- The analysis of the results of the short baselines (< 2km), estimates the **base uncertainties**.
- The analysis of the results of the medium length baselines (ranging in length from 8 to 105km), estimates the **line-length dependent uncertainties**.
- GPS results are evaluated in terms of: Cartesian coordinate difference (dX, dY, dZ), baseline lengths, ellipsoidal height differences, and azimuth.
- Computations are analysed according to: repeat baseline measurements, loop miscloses, minimally constrained 3-D Least Squares adjustments, comparisons with the terrestrial standard, and comparisons with past FGCS GPS test survey results.

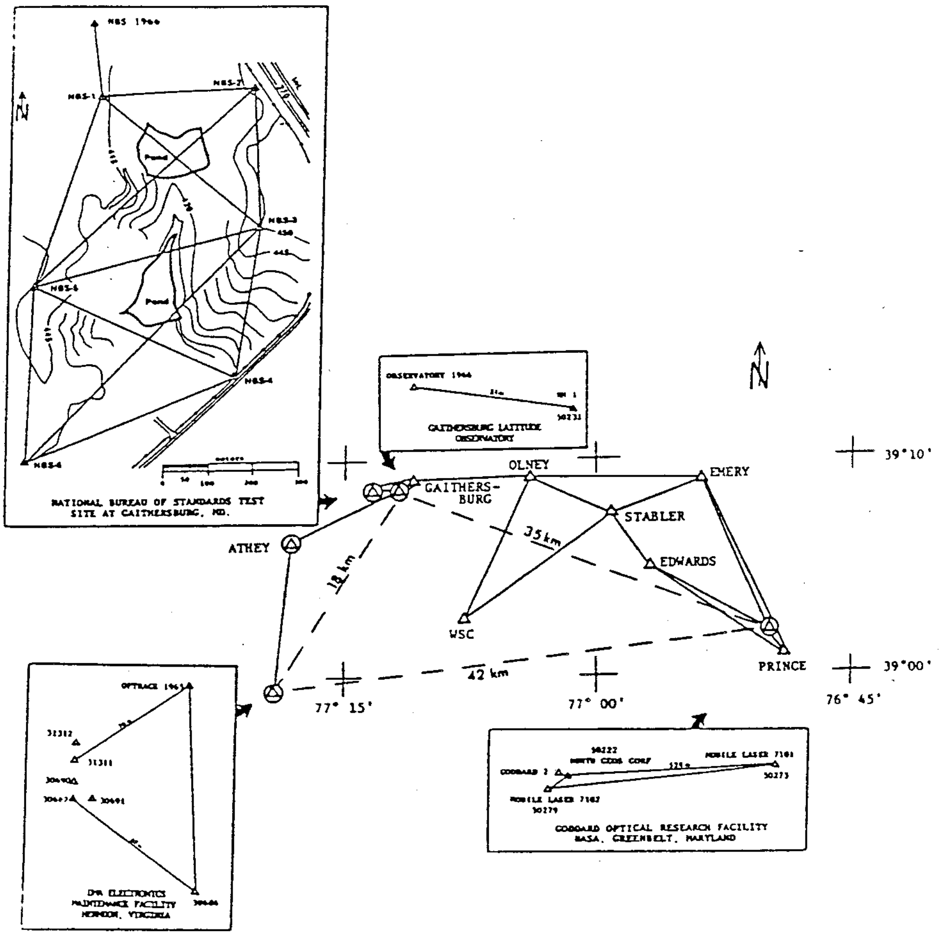


Figure 4.4-1. FGCS test network within the metropolitan area of Washington D.C.

Summary of data analyses and results:

- Fixed orbital coordinate solution using the broadcast ephemerides (assumed to have introduced an uncertainty no larger than 1-2 ppm).
- Ionospheric refraction correction accounted for in dual-frequency systems by using the L3 "ionosphere-free" combination for long baselines (§6.4), and ignored for short baselines. *Other systems which used L1 phase data only could not ensure the same accuracy results for long baselines.*
- Internal consistency (or repeatability) of the baseline results were characterised by the following estimated uncertainties in each component (at the 1-sigma, 67% confidence level):
  - "Base" error: 0.3 to 1cm
  - "Line-length dependent" error: 1 to 2 ppm
  - **Combined uncertainty therefore is (cm):**

$$\pm\sqrt{(0.3 \text{ to } 1\text{cm})^2 + [0.1 \ . \ (1 \text{ to } 2\text{ppm}) \ . \ B]^2}$$

where B is the baseline length in kilometres.

The Issue of "Legal Traceability"

There has recently been a move to establish procedures in Australia for the regular testing of GPS survey instrumentation and staff. This initiative derives from the need to define "legal traceability" of GPS survey results for cadastral surveys (BOEY & HILL, 1995), in a similar manner to the calibration of EDM on official baselines. It has been recognised that there may need to be several levels of testing:

(1) An "integrity monitoring" network that continuously tests the health of the GPS system.

(2) The definition of a set of test procedures which are the responsibility of State Lands and Survey Departments for administering. Such a network could, for example, be a number of high precision State GPS survey control points, in a configuration that may be similar to that of the FGCS network.

(3) Network and procedures may have to be defined for both the conventional static GPS survey technique, and the modern "stop & go" and "rapid static" GPS techniques.

(4) An alternative approach to the above could be based on the principles of "Total Quality Management", by making the surveyor responsible for ensuring that his/her GPS instrumentation is in satisfactory compliance with the operational performance standards defined by the State Certifying Authority. *In this case, a user-pays service for rigorous testing of GPS receivers may be provided.*

The following additional comments can be made:

- <u>Each State would develop its own guidelines for testing</u>. For example, such tests maybe specifically concerned with validating GPS for cadastral surveys. Alternatively, a general certification may be given.
- It is important to define not just the network of test points to be surveyed, but also the <u>field and processing procedures to be used</u>.
- If modern high productivity GPS procedures, such as "stop & go" and "rapid static", are to be tested, a *typical* survey environment must be ensured. Hence the <u>inclusion of trees and other obstructions is important</u> in order to realistically test ambiguity reinitialisation procedures.

- There should be a variety of baseline lengths within the test network. In the case of networks to test conventional static GPS, the baselines could be up to 50km in length (or greater). However, "rapid static" and "stop & go" GPS surveys are generally carried out only over baselines less than 20km in length.
- The problem with such *campaign* style GPS testing is that they provide merely a "snapshot" of GPS performance. The challenge is how to investigate the impact of varying observation conditions:
  - changes in receiver-satellite geometry (as experienced through the day),
  - different numbers of available satellites (as a result of signal obstructions),
  - day and night observations (to induce differences in atmospheric conditions),
  - varying temperature conditions (seasonal effects?), and
  - varying site conditions (to induce a variety of multipath and signal jamming scenarios).

See IBID (1995) for a discussion of the options being considered for the state of Victoria, Australia.

Pre-Mission Testing

If there is any doubt concerning the proper functioning of the GPS equipment, or there is unfamiliarity by the field staff with the equipment to be used, and to verify that the hardware and software is functioning to the required accuracy, then some localised testing could be undertaken. Such a test could be carried out over:

- a "zero baseline", or
- a micro-network, to test instrument functioning, antenna connections, battery and power requirements, etc., or
- a "standard" validation test network, as described above (under "Legal Traceability"), to verify the average performance of GPS.

Such testing is strictly the prerogative of the surveyor, hence he/she could make the testing as sophisticated and comprehensive as they wished. Furthermore, regular testing could be considered prudent if the surveyor is operating under the principles of "Total Quality Management". At present there are no test networks established for these purposes in Australia. The Australian "Standards & Practices for Control Surveys" (ICSM, 1994 -- see §10.2) sets out some "test procedures" for GPS equipment, however these are just suggested practices and do not have anything to do with "certification".

4.4.3 INVESTIGATIVE TESTS

The defining characteristics of these tests is the desire to *understand* a certain phenomenon that impacts on GPS performance through a carefully designed experiment. Invariably such testing is carried out by academic institutions. Because there are so many factors impacting on baseline accuracy, it is necessary to constrain or standardise as many of the factors as possible. These could be done in the following way:

- Do all the tests on a single baseline, but varying the other factors such as time of day, length of observation session, etc.

- Test several different GPS instruments *simultaneously* and, if possible, over the same baseline.
- Use the same GPS datasets, but vary the processing options, and even the software, in order to gauge the *impact of different modelling and processing strategies.*
- In a controlled manner *induce* a disturbing influence in order to determine realistic operational accuracies across all plausible conditions.

Tests include those that study:

(1) the cycle of receiver-satellite geometry effects on baseline results,
(2) the impact of using different data pre-processing strategies,
(3) the quality of the raw tracking data, under different site conditions,
(4) the use of different styles of GPS antenna,
(5) the impact of using different data processing software, or
(6) the impact of using different observation models.

One institution undertaking considerable GPS receiver testing is the UNAVCO organisation (University NAVstar COnsortium), in particular into the effects of multipath. Their reports can be accessed via Internet (see §3.4). Many other multipath investigations have also been reported in the literature.

Another interesting experiment is reported by HUSTI et al (1994). The test involved collecting phase data with a variety of GPS surveying receivers, on a number of baselines of varying length. The data was converted into the RINEX (Receiver INdependent EXchange) format (GURTNER, 1994). The quality of the data collected by each receiver was judged to be of similar high quality because when the data was processed using receiver-independent software (the "Bernese software") the results that were obtained were very similar. However, when each of the instrument datasets was processed using the proprietary commercial software, differences in the baseline results were noticed. Several tests such as these have been carried out, and the basic conclusion is that the mixing of GPS receivers and software is still a risky proposition.

The Netherlands Geodetic Commission has also compared the performance of different codeless GPS receivers under A/S conditions. SLUITER & HAAGMANS (1995) report the results of investigations into the susceptibility of the receivers to intentional and unintentional jamming. (Unfortunately, they do not name the best performing receiver!)

Another type of test seeks to investigate the *variability* of the baseline results. GPS positioning accuracy is influenced by such factors as (§2.4):

- The observation type -- *and hence the measurement precision.*
- The signal disturbances present -- *and are responsible for systematic biases.*
- The algorithms used -- *appropriate for the level of biases and the accuracy sought, but tempered by operational issues.*
- Operational issues such as satellite geometry, length of observation session, static or moving antenna, baseline length separating a pair of antennas, etc.

Assuming a certain hardware configuration, and acknowledging the presence of an acceptable level of *residual* biases (those that remain after the application of processing algorithms based on double-differencing simultaneously observed tracking data -- §6.3), it is the operational issues that influence accuracy the most. The operational issues listed above will impact the quality of the positioning results in several ways, for example:

(a) the magnitude of the residual biases is most affected by baseline length (the longer the baseline, the larger the residual biases, and hence there is a commensurate reduction in accuracy), and

(b) the length of the observation session affects the sensitivity of the solution to residual biases, geometry, etc. (all other things being equal, solutions based on short observation sessions are generally less reliable than those from longer sessions).

Hence, there are two options:

- Test a single baseline by collecting 24 hours of GPS data, *and partitioning the dataset into different session lengths (15, 30, 45, 60 ... minutes) before processing the data sessions.*
- Test a variety of baseline lengths by collecting 24 hours of GPS data, *and partitioning all the datasets into the same observation session length before processing the data sessions.*

In both cases there are many results that then have to be systematically analysed in order to ascertain the overall **performance characteristics** of GPS (and especially how *variable* the baseline accuracy is) as a function of parameters that can be influenced such as: baseline length and observation session length; and as a function of parameters over which there is little control such as: satellite geometry, multipath and signal disturbances. An example is illustrated in Figure 4.4-2, where a 24 hour dataset collected on an approximately 4km baseline has been partitioned into consecutive 5 minute (Figure 4.4-2a) and 30 minute (Figure 4.4-2b) observation sessions, and the same data processing algorithm applied to the phase data (double-differenced ambiguity-free solutions).

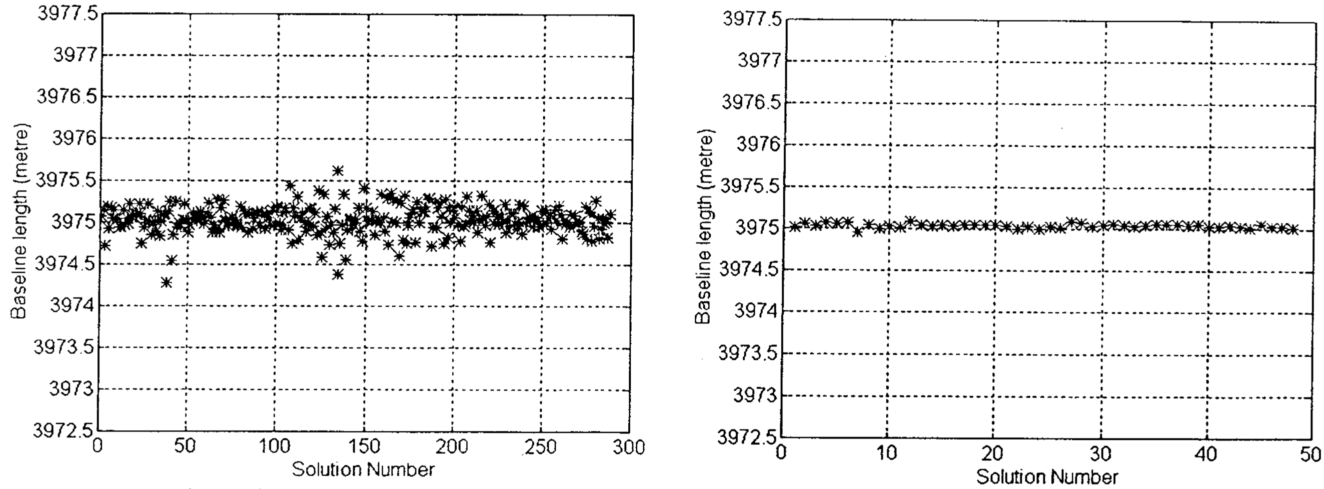


Figure 4.4-2. Double-difference phase solution (ambiguity-free) for 4km baseline: (a) 5 minute observation sessions, (b) 30 minute observation sessions.

Chapter 5: GPS Satellite Surveying

5.1 GPS SATELLITE SURVEYING: SOME CONSIDERATIONS

In this section several important matters concerning the GPS satellite surveying methodology will be discussed:

☞ The fundamental differences between the *GPS Navigation* and *GPS Surveying* modes of positioning.

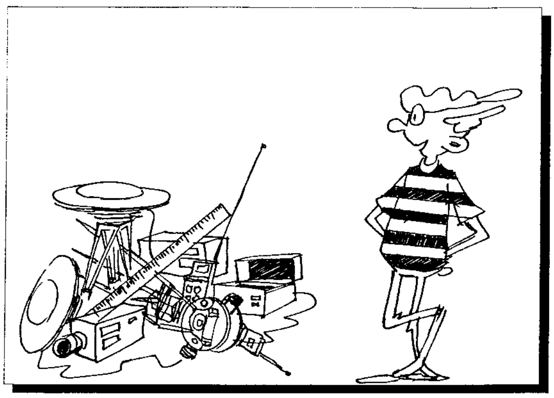
☞ The factors which have encouraged the adoption of GPS, in place of competing positioning technologies, for a range of possible land surveying users.

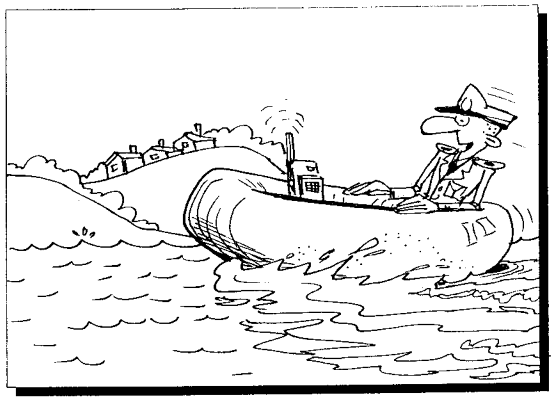
5.1.1 GPS SURVEYING versus GPS NAVIGATION

The distinction between *GPS Surveying* and *GPS Navigation* can be made according to a variety of criteria, for example:

- ❑ According to "when", "where" and "how" the GPS technology is applied. This focuses on the applications, and the following simplistic distinction is therefore made: GPS Navigation supports the *safe passage of a vessel or aircraft*, from the port of departure, while underway and to its point of arrival; while GPS Surveying is mostly associated with the traditional functions of establishing geodetic control, supporting engineering construction, cadastral surveys and map making.
- ❑ According to operational aspects, such as the real-time, absolute positioning aspects of Navigation, as opposed to the post-processed, "unhurried", relative positioning characteristics of GPS Surveying.
- ❑ According to the type of measurement made and the GPS instrumentation used. GPS Navigation-type receivers are comparatively low-cost, code-correlating instruments that only measure pseudo-range, whereas GPS Surveying receivers are expensive, phase measuring instruments that include many special features and complex software in order to support their function.
- ❑ According to the mathematical models used. For example, because the primary measurement in GPS Navigation is the pseudo-range, the biases are dealt with in a more "casual" fashion (with the exception of the clock errors, they are all ignored!). In contrast, GPS Surveying requires a more careful treatment of the biases during the data processing.

The particular characteristics of GPS (land) surveying, the modes of positioning and a review of the treatment of GPS biases (and the resulting operational requirements) are summarised below.





Characteristics of GPS Surveying and GPS Navigation

Some characteristics of *GPS Satellite Surveying* are:

- The points being coordinated are stationary.
- GPS data are collected over some "observation session".
- Relative positioning modes of operation, and hence high accuracies.
- The measurements are made on the L-band carrier wave, hence requiring special instrumentation and software.
- Mostly associated with the traditional surveying and mapping functions.

Some characteristics of *GPS Satellite Navigation* are:

- The points being coordinated are generally in motion.
- GPS is collected for an "instant", and the solution is obtained in real-time.
- Absolute and relative positioning modes of operation, of comparatively low accuracy.
- The measurements are typically made on the PRN codes, and requires the processing of pseudo-range data.
- Mostly associated with defining safe passage of ships and aircraft.

The main positioning modes for GPS surveying and navigation are (§2.4):

- ❑ ABSOLUTE or POINT positioning: coordinates are in relation to a well-defined global reference system.
- ❑ DIFFERENTIAL or RELATIVE positioning: coordinates are in relation to some other fixed point. *In GPS surveying this is referred to as baseline determination.*
- ❑ STATIC positioning: coordination of stationary points, either in absolute or relative mode. *This is generally synonymous with the SURVEYING mode of positioning, based on the analysis of carrier phase observations.*
- ❑ KINEMATIC positioning: coordination of moving points, either in absolute or relative mode. *This is generally the NAVIGATION mode of positioning, based on*

pseudo-range observations.

There are some distinctions to be made in the way the data is processed in order to minimise the effect of biases in the measurements.

As all GPS observations are plagued with biases, hence for both navigation and surveying applications an appropriate combination of measurement and processing strategies must be used to minimise their effect on the positioning results.

Pseudo-range data is relatively "noisy", and the significant biases are accounted for in the following way:

- In the point positioning mode, satellite clock error is ignored, as it is assumed (after applying the broadcast clock error model -- §3.3) to be smaller than the measurement noise,
- Receiver clock error is estimated in real-time through redundant measurements, because all data is contaminated by the same bias.
- In the relative positioning mode, all satellite and propagation biases are significantly reduced.

This is the NAVIGATION mode of positioning, as results are obtained in real-time (when four or more pseudo-ranges are processed simultaneously). Relative navigation is of higher accuracy as the primary biases due to orbit error, atmospheric refraction and SA are minimised.

Integrated carrier beat phase data is very precise, hence any contamination by systematic errors is of greater concern than in the case of pseudo-range measurements. Appropriate processing techniques must therefore be used. However, the primary drawback of this data type is its range "ambiguity". In GPS surveying the major biases are accounted for in the following ways:

- Differencing data collected simultaneously from two or more GPS receivers, to several GPS satellites, *between satellites and between receivers.* This eliminates, or significantly reduces, most of the biases. All position results are therefore expressed relative to (fixed) datum stations.
- The "ambiguity" bias is often estimated, though a weaker solution can be obtained from the appropriate triple-differenced observable (§6.3).

This is the SURVEYING mode. The fact that the receivers are stationary, and that data is collected over some observation period, permits the ambiguities to be reliably estimated and a strong solution obtained. There are alternative means of estimating ambiguities that permit real-time kinematic baseline determination to be carried out as well.

The GPS Satellite Surveying Methodology: Some Comments

Comments to the *operational aspects* of GPS Surveying:

- ❑ Survey planning considerations are derived from:
  - The nature and aim of the survey project --> *as for conventional surveys.*
  - The unique characteristics of GPS, and in particular no requirement for station intervisibility --> *a simplification in survey design.*
  - The number of points to be surveyed, the resources at the surveyor's disposal, and the strategy to be used for propagating the survey --> *a logistical problem.*
  - Prudent survey practice, requiring redundancy and check measurements to be incorporated into the network design.

- ❑ The requirement for training in the operation of GPS survey receiver hardware, and post-processing software, as well as being aware of calibration and test procedures.

- ❑ Field operations are characterised by requirements for:
  - Setup of antennas over predefined ground marks.
  - Simultaneous operation of two or more GPS receivers.
  - Coordinate data gathering operation so that data collected has the same time-tags, involves the same satellites, etc.
  - Common data collection over some observation session.
  - Coordinated demount of GPS antennas and transport to new stations.

- ❑ Field validation of data collected, in order to:
  - Verify sufficient common data collected at all sites operating simultaneously.
  - Verify quality of data to ensure that acceptable results will be obtained.
  - Where data dropout is high or a station has not collected sufficient data, reoccupation may be necessary.

- ❑ Office calculations:
  - To obtain GPS solutions for single sessions or baselines.
  - To combine the results of single sessions into a network solution.
  - To incorporate external information (for example, local control station coordinates), and hence modify the GPS-only network solution.
  - To transform the GPS results to the local geodetic datum, and to derive orthometric heights.
  - To verify the accuracy and reliability of the GPS survey.

Comments to the *GPS Survey Solution:*

- ❑ The fundamental unit of a GPS solution is a 3-D baseline vector joining the antennas of two GPS receivers that have been tracking simultaneously the same satellites. GPS software to carry out the solution task is usually provided by the instrument manufacturer.

- ❑ One end of the baseline is held "fixed" (its coordinates are assumed known), and the other station's coordinates are determined relative to it (in effect, the baseline components are estimated).

- ❑ Solutions may be obtained from triple-differenced, or double-differenced data solutions, with different resultant accuracies and reliabilities.

- ❑ All results are obtained in the quasi-WGS84 reference system, but relative to a fixed station (the WGS84 coordinates of one end of the baseline are assumed known).

- ❑ All results refer to the antenna phase centres, and the height of antenna and any offsets must be applied in order to reduce the coordinates to the ground marks.

- The quality of the baseline vector solution is dependent on, amongst other things:
  - length of the (common) observing session,
  - the number of satellites tracked by the receivers,
  - the quality of the data (multipath and cycle slips, single or dual-frequency data, presence of noise and other biases),
  - the type of baseline solution: triple-difference, double-difference, etc., and
  - the software used to reduce the data.
- If during an observation session more then two receivers were deployed, independent baselines need to be processed, either in a single combined solution or separately.
- If a network needs to be surveyed over a number of sessions (because the number of points is greater than the number of GPS receivers), a combination of separate baseline solutions is needed in a subsequent "network adjustment" step.
- This network solution may then be constrained and transformed into the local geodetic datum if sufficient geodetic control stations (with coordinates in the local datum) are also surveyed as well.
- GPS survey results should be "quality controlled" at all stages of survey and data processing.

5.1.2 FACTORS INFLUENCING THE ADOPTION OF GPS FOR LAND SURVEY APPLICATIONS

GPS does not need to be only competitive against conventional terrestrial techniques of surveying, but also other extraterrestrial techniques and new technologies. GPS relative positioning technology can, in principle, be employed for a wide range of activities. As a starting point in discussions concerning the special advantages (and disadvantages) of GPS for land surveying applications, applications are classified into three general categories (§2.3):

Class A (Scientific):	better than 1 ppm
Class B (Geodetic):	1 to 10 ppm
Class C (General Surveying):	lower than 10 ppm.

Several criteria for judging the utility of GPS over other competing technologies can be identified for each of these categories:

(1) **Cost benefit**: *issues such as the capital cost of equipment, ongoing operational costs, data processing costs, development, training and maintenance costs.* This can best be measured according to productivity. The direct cost of a GPS survey (not including equipment and training costs) can be estimated during the planning phase (§5.2). It needs to be established whether the competing technologies offer lower costs.

(2) **Ease of Use**: *issues such as servicing, timeliness of results, expertise of users.* Experience indicates that the primary factors affecting servicing are those of distance to the servicing agents, their technical expertise, and their customer service. To ensure quality results in a reasonable time (real-time operations may not be required) it is important that all personnel (field and office) be well trained. The complexity of the processing software must also be factored in.

(3) **Accuracy**: *obviously related to the class of user.* The level of accuracy sought will directly influence many other factors such as: type of instrumentation, expertise of personnel, sophistication of software, "reasonable" cost of survey, field operations,

etc. The primary (non-hardware dependent) accuracy limitations are due to errors in the satellite ephemerides and atmospheric uncertainties; both requiring special software processing for very high accuracy surveys.

(4) **External factors**: *such as availability of satellite ephemerides and other performance constraints such as superior GPS networks to connect into, base station operation, etc.*

GPS vs Conventional Terrestrial Surveying -- ADVANTAGES:

- ☞ **Intervisibility not required**
- ☞ **Operations are weather independent**
- ☞ **Network independent site selection, hence sites placed where needed**
- ☞ **Around-the-clock operation**
- ☞ **Economic advantages from greater efficiency and speed of survey**
- ☞ **Geodetic accuracies easily achieved**
- ☞ **3-D coordinates are obtained**

GPS vs Conventional Terrestrial Surveying -- DISADVANTAGES:

- ☞ **High productivity places greater demand on survey planning and logistical considerations**
- ☞ **No sky obstructions can be tolerated, therefore cannot be used underground, under foliage or structures**
- ☞ **GPS surveying is generally "targeted" to satisfy a specific survey need**
- ☞ **No azimuth control for subsequent non-GPS surveys**
- ☞ **Horizontal and vertical coordinates from GPS must be transformed if they are to be useful for conventional survey applications**
- ☞ **GPS accuracies are generally higher than the surrounding existing control**
- ☞ **High capital cost of GPS instrumentation**
- ☞ **New skills needed**

Geophysical, Engineering, Large Scale Cadastral and Mapping Surveys

The overwhelming majority of civilian GPS surveys will be carried out in support of geophysical prospecting activities, engineering projects, large scale point coordination for Geographic Information System (GIS) data collection, land parcel surveys, and for map control. Indeed, GPS is ideally suited for surveys associated with the establishment of a coordinated cadastre (apart from the constraints of sky visibility in highly built-up areas). The relative accuracy requirements are likely to be of the order of 1 part in 10^4 to perhaps better than 1 part in 10^5.

The majority of these prospective users will have had little or no previous experience with satellite surveying, therefore the introduction of GPS into surveying practices will be mainly influenced by cost-benefit considerations. Reservations concerning GPS technology will only be overcome if the surveyor is convinced that GPS could perform the positioning function, to the required accuracy, in a shorter time and with greater efficiency (and hence less cost) than any other technique. This means that the surveyor would have to satisfy himself/herself that the GPS positioning technology is superior to the conventional EDM and theodolite procedures with which he/she is already very familiar.

Although a definitive statement on the relative competitiveness of GPS and EDM-theodolite technologies is not possible, some criteria for evaluating competitiveness can be identified:

(a) Receiver cost.
(b) Ease of operation.
(c) Productivity.

Receiver Cost

The average cost of an electronic tacheometer (or "total station") is between $10000 and $20000. As a result of small production runs, and the fact that such instruments are a combination of mechanical, optical and electronic components, the costs of these surveying instruments have not undergone the dramatic price reductions experienced in the handheld calculator or personal computer market. Aside from inflation, the prices of traditional surveying instruments are not expected to vary significantly through the 1990's. Prices of over one hundred thousand dollars were the norm for first generation GPS receivers. However, instrument prices have been falling and third generation phase measuring instrumentation is already of the order of $10000 to $15000US for the basic models and perhaps double this for top-of-the-line "geodetic" receivers. There is a greater interest in "enhanced" GPS navigation receivers. These, when operated in differential mode, can deliver accuracies from a few metres to sub-metre over baselines up to hundreds of kilometres, and generally cost below $10000US. *(A pair of receivers would be needed for relative positioning, hence doubling the cost.)*

Ease of Operation

GPS will be more readily adopted if it can be shown to be at least as easy to use as medium or long-range EDM. That is, GPS observation and reduction procedures should not require a high degree of operator training. To allow in-the-field positioning (or "real-time" processing), the processing software should be resident in the receiver, or on a portable field computer, and include communication links between receivers and easy-to-use software that requires a minimum of operator intervention. Differential GPS operation using pseudo-range data can deliver low accuracy results in a largely "automatic" mode and is hence ideally suited for GIS mapping applications.

Productivity

As surveyors, it is usual to compare GPS efficiency with that of EDM-theodolite procedures in terms of time spent in the field acquiring data. The fallacies in such comparisons are illustrated by the following three survey scenarios:

1. *The coordination of distinct and unrelated points or structures, such as control points for aerial mapping or the positioning of oil production platforms, etc.* For these applications, relative positioning with respect to a distant control station is the most efficient technique. In contrast, EDM-theodolite traversing "carries" the coordinate information to its destination by measuring azimuth and distance between mutually intervisible instrument stations.
2. *The establishment of a network of coordinated points to support a geophysical survey, the monitoring of construction activity, data capture for GIS, or for surveying rural land boundaries.* In these applications the location and relative disposition of the control points does not depend on considerations such as network shape, intervisibility or maximum separation of stations, but rather on optimum layout for carrying out the intent of the survey (for example, determining the shape of a land parcel). With conventional techniques the control points would be established by traversing and intermediate instrument setups would only be used if the line-of-sight is obstructed or the distances between control points is too great. With GPS only the minimum complement of stations need be surveyed.
3. *The establishment of a network of coordinated points for control densification or for subsequently carrying out detail surveys on engineering construction sites, or for urban mapping control.* Here the natural advantage of GPS in not requiring intermediate instrument setups is eroded by the need to have many control points close together, and for the points to be intervisible (to support conventional surveying). Modern "kinematic", "rapid static" or "stop & go" GPS techniques have natural advantages for such scenarios (§5.5). *This class of GPS instrumentation is even more expensive than standard GPS surveying receivers.*

When GPS is forced to operate in a pseudo-traversing mode, EDM-theodolite techniques maintain an edge in competitiveness over the standard static (or even "rapid static") GPS baseline survey mode. However, there is probably a critical station-separation above which GPS would be the more efficient technique. Initially, interstation distances of ten or more kilometres (less in rugged terrain) are likely to be most efficiently bridged using GPS. Shorter baselines can be economically surveyed using modern GPS surveying techniques in which station occupancy times have fallen to just a few minutes or less. An additional factor to be considered is that GPS gives height information as well, although this height is not orthometric elevation and must be "corrected" by subtracting the geoid height quantity.

Geodetic Surveys

GPS has already replaced TRANSIT for establishing, maintaining and densifying geodetic networks. Over distances of a few hundred kilometres, multi-station TRANSIT gave relative positioning accuracies of a few decimetres (this figure is largely independent of station spacing). Unlike TRANSIT, GPS can give accuracies of 1ppm even over distances as short as a few kilometres. Furthermore, GPS is much faster, with measurement times as short as half an hour (in standard mode), or just minutes (for "stop & go" techniques), being required to measure a baseline. GPS can also compete with conventional methods which are slow, labour intensive, and suffer from unfavourable error propagation that degrades accuracies over long distances.

Unlike triangulation, traversing and trilateration, control networks established with GPS are not bound by constraints such as station intervisibility and network shape. With GPS the spacing between control stations in unsurveyed regions can be increased over traditional techniques. The increased flexibility of GPS also permits control stations to be established at easily accessible sites rather than being confined to hilltops as has hitherto been the case.

According to the "rule-of-thumb" in §6.2, the orbital information necessary to support 1ppm surveying should be accurate to 20 metres or better. At present the Broadcast Ephemerides appears to be adequate for surveying applications at the part per million level. Since immediate positioning results are not needed for the applications considered here, post-processed ephemerides could also be used. These could be obtained from an "ephemeris service" such as the International GPS Service for Geodynamics (IGS), as discussed in §6.2 and §12.2.

Frequently, the relative positioning results obtained with GPS have a higher internal precision than that of the geodetic network to which they must be tied. Discrepancies between GPS and published coordinates will cause headaches for surveyors carrying out high precision GPS surveys for many years to come. *Should the GPS-derived values be distorted by performing an adjustment in which the local control is held fixed?* Alternatively, should only the minimally constrained solution (holding only one station fixed) be insisted upon? These are issues that will be raised in later chapters. **The entire geodetic network may ultimately need to be strengthened and redefined using GPS as is happening in many countries, including Australia** (chapter 12).

Scientific Surveys

The measurement of crustal deformation is central to our understanding of earthquake processes, plate motion, rifting, mountain building mechanisms and the near-surface behaviour of volcanoes. The scale of deformation associated with these processes varies from tens to thousands of kilometres, but much of the deformation of interest is found near plate boundaries that are typically 30 to 300km wide. The requirement for crustal motion studies is distinctly different from routine surveying and mapping requirements in that the objective is not one of establishing absolute coordinates, but of measuring *changes* in position, displacement or strain with time. *Hence one seeks to repeat the measurements under as nearly an identical set of circumstances and to as high an accuracy as possible.* The relative positioning accuracies that are obtained for GPS crustal movement surveys are typically between 10^7 and 10^8 (that is, a few centimetres in 1000km!).

Until the 1970's, terrestrial techniques provided the only geodetic observations of presentday crustal movements. Over the last two decades measurements using Very Long Baseline Interferometry (VLBI) and Satellite Laser Ranging (SLR) technologies have resulted in accuracies of several centimetres over distances of many thousands of kilometres. Unfortunately, the equipment required and the personnel costs have discouraged the use of such systems in the density necessary to study crustal deformation mechanisms at the shorter baseline lengths that are of interest in most seismic zones. Therefore, although VLBI and SLR are very valuable methods for the determination of global plate movements and large scale deformation within the plates, their future general use for baselines less than roughly 300km in length can no longer be justified from an economic point of view.

The cost of making differential measurements using GPS receivers is low enough to consider their use for large numbers of crustal movement measurements with baseline lengths ranging from a few kilometres to thousands of kilometres. The GPS receivers are inexpensive compared to SLR and VLBI equipment and the personnel requirements can be reduced to one person per receiver (or none at all when operated in the automatic base station mode). In fact, GPS provides a much needed technique to bridge the "spatial gap" between conventional trilateration-triangulation for short baselines and VLBI-SLR for long baselines. GPS will therefore be applied to many of the geophysical problems presently studied using ground techniques over 5 to 30km baselines. There is also a trend to using GPS for baselines up to several thousand kilometres long, in order to densify the global network of fundamental stations beyond those established by SLR or VLBI techniques.

This high accuracy mode is achieved using "top-of-the-line" receivers and very sophisticated "scientific" software. In addition to being used by specialists, the trend is now to use these techniques to strengthen the basic datum network so that "standard" GPS (geodetic) surveys at the ppm level can be easily incorporated within such a **"zero-order" network** without the distortion that presently has to be experienced when GPS results are combined with earlier conventional survey results.

The vertical accuracy, while it is less than that obtained for the horizontal components, will have a major impact on geodynamic studies. The cost is much less than for precise levelling for distances greater than a few kilometres. Elevation differences are obtainable with much greater ease and may be undertaken by a small group of investigators. This requires relatively little expenditure of time, compared to the effort and large team required for first-order levelling. Furthermore, ellipsoidal height differences are obtained directly, not orthometric height differences, which must be corrected for any geoid height changes to give the true geometrical (ellipsoidal) height change.

Ionospheric refraction effects on the GPS measurements can be quite large over the distances considered here. The effects at either end of the baseline would not necessarily be the same and thus would not be expected to cancel in long baseline measurements. Dual-frequency receivers capable of measuring and recording phase on the L1 and L2 frequencies are therefore essential for high precision geodetic applications. Another significant error source is in modelling the tropospheric refraction error. Although the dry part of the atmosphere can be adequately modelled, the wet part, due to the tropospheric water vapour, is more difficult to determine.

A major limitation in GPS baseline measurements is the accuracy of the satellite ephemerides used in the data reductions. For survey accuracies of 1 part in 10^7 or better, ephemerides accurate to a few metres, or better, are required. Only carefully computed post-processed ephemerides would satisfy these accuracy requirements. Post-processed ephemerides also contain reference system information which must be reconciled with survey datums. *The IGS was established with this dual role very much in mind: a source of high precision satellite orbit information, and a means of defining and maintaining a global terrestrial reference datum.*

Summary Remarks

Just as GPS *complements* the other space techniques for high precision applications, it will also complement the traditional EDM-theodolite techniques for routine surveying activities. Indeed the traditional techniques are likely to continue playing the dominant role for some time to come -- for traverses less than a few kilometres in rugged terrain or less than a ten or so kilometres in flat terrain. For longer distances GPS is a potential competitor. However this is more likely to occur if the cost of modern GPS receivers capable of high speed (even real-time) surveying falls to a level comparable with "total station" instrumentation.

Figure 5.1-1 illustrates the accuracy and spatial ranges of the various geodetic methods of positioning, and reinforces the remarks made above concerning the "niche" that GPS occupies between EDM-theodolite traversing at the short range end of the survey spectrum and SLR/VLBI techniques at the long range end. (Note that the plot is double logarithmic!)

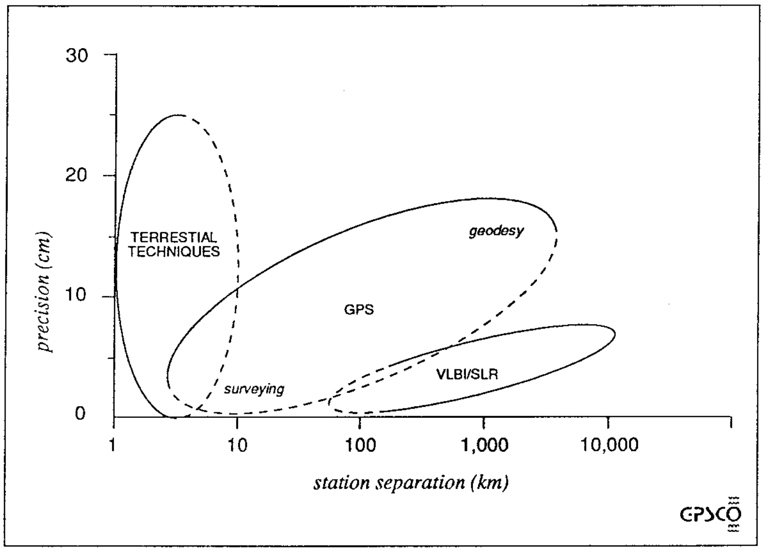


Figure 5.1-1. Accuracy and ranges of geodetic methods of positioning.

As already mentioned, GPS surveying has several unique characteristics related to:

- instrumentation (chapter 4)
- positioning strategies (§5.2)
- field operation (§5.4 and §5.5)
- data processing (chapters 7 and 8)
- integration of results with prior surveys (chapter 11 and 12)

Elements of the GPS survey task:

☞ **Definition of the task: how many points? accuracy required? horizontal & vertical? connect to datum? distribution of points? resources available? etc.**

☞ **Planning: logistical and structural considerations, connection to control, standards & specs for GPS surveys, number of receivers/parties, site selection, observation schedule, etc.**

☞ **Reconnaissance: satellite visibility & availability, site conditions & access, station marking, etc.**

☞ **Field procedures: equipment checklist, on-site procedures.**

☞ **Post-processing & result presentation: baseline processing, minimally constrained solutions, fitting GPS network results to geodetic control, QC, heights, etc.**

5.2
GPS SURVEY PLANNING

Project planning is one of the most important aspects of GPS surveying, as careful planning maximises the chances of the survey achieving the desired accuracy, within a reasonable time and to budget.

Before commencing the planning of a GPS survey it must first be established (a partial list only!):

- What is the purpose of the survey?
- What are the accuracy and reliability requirements?
- What resources are available?
- What previous surveys have been carried out?
- Are there any special (or unusual) characteristics of the project?
- Is the surveyor suitably equipped (in the broadest sense of the word) to carry out the GPS survey for the client?

Elements of the GPS Survey Planning Process

Unlike conventional surveying technologies there has generally not been sufficient time for the average surveyor to have amassed the "conventional wisdom" needed to reliably execute GPS surveys. Careful planning is therefore still a critical issue and consists of the following elements:

(1) **PROJECT DESIGN:** involves project layout and network design, and is driven primarily by accuracy and station location / density requirements (defined by the client), productivity / economic considerations (of concern to the GPS surveyor), and standards and specifications (promoted by the geodetic control authority).

(2) **OBSERVATION SCHEDULE:** giving consideration to such factors as the number of GPS receivers, occupation time per site, number of sites per day, requirement for multiple station occupancy, etc.

(3) **INSTRUMENTATION & PERSONNEL:** instrumentation appropriate for the task, mainly driven by what may be available inhouse or what could be hired from outside the organisation. Also, adequately trained personnel are needed in order to carry out the survey and process the data.

(4) **LOGISTICAL CONSIDERATIONS:** issues such as transportation (appropriate to ensure observation schedule can be adhered to), special site requirements (for example, power, intervisibility, etc.), and factors related to network design and observation scheduling such as the receiver deployment pattern, etc.

(5) **RECONNAISSANCE:** which may or may not be necessary, depending upon how "critical" the GPS stations are to be to the overall project, whether permanent marks will be established, etc.

5.2.1 PROJECT DESIGN

Designing the project layout is one of the most important planning tasks of the GPS surveyor. **The final network / project design is usually a *compromise* between technical requirements and economics, worked out within the framework of explicit recommended practices for GPS surveys (or at the very least, prudent practices that ensure the job gets done to the appropriate standards of accuracy and reliability).** The surveyor must take the following factors into account:

☞ Definition of the network: size and shape of the overall network, the number of stations, station spacing, any intervisibility requirements, new and existing (known) stations.

☞ Spacing of the existing (known) stations: for what purpose are they intended? as a quality check, for densification of existing control, for the determination of transformation parameters, etc.

☞ Accuracy requirements (defined by client) and standards (defined by the geodetic control authorities), for both horizontal and vertical surveys.

If the number of stations (known and new) to be surveyed is greater than the number of instruments, the GPS survey will have to be carried out over a number of observation sessions. Hence "project design" and "observation scheduling" are inextricably linked together.

Propagation of the GPS Survey

One of the significant advantages of the GPS survey technique over conventional surveying techniques is that sites may be placed *where they are required*, irrespective of whether intervisibility between stations is preserved. Generally, the GPS stations would be "clustered" around the project *focus*, for example a road, dam, powerline corridor, etc. This is in contrast to traditional geodetic control that was generally evenly spaced and the stations located in prominent locations such as at the tops of hills, to ensure that they were visible from afar. In addition, extra survey stations that "carry in" the control from the nearest geodetic control stations to the project area are not usually necessary for GPS work. **Hence even spacing of stations and selection of stations on the basis of terrain are no longer important considerations** (Figure 5.2-1).

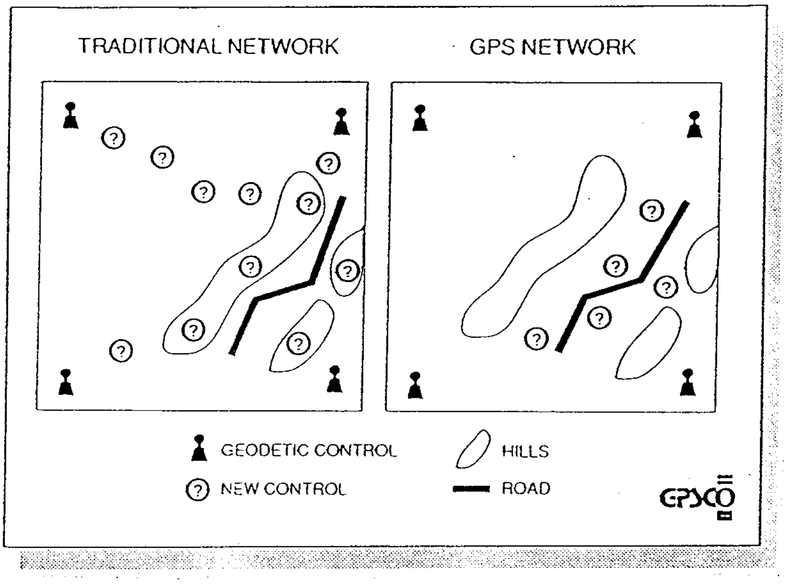


Figure 5.2-1. Terrain need not influence site selection.

Once the number of GPS stations has been decided upon, and their approximate locations have been determined, other considerations may influence where additional stations may need to be located, or where refinements to the network design could be made:

- *Some intervisible stations may need to be included to define starting azimuths for subsequent conventional surveys.* This may be best satisfied by the provision of some additional azimuth marks surveyed by GPS, rather than altering the original network design to incorporate station intervisibility. The reference marks may be set up a couple of hundred metres away (and, depending on the terrain, etc., perhaps up to a kilometre or more distant).

- *There may be several existing stations in the area which could be included in the GPS survey.* There are many reasons for occupying already established stations (apart from the obvious one: the client requests it!). Generally the reasons range from simple expediency (save on establishing new monumentation), to necessary ties to previous work, or to ensure datum definition or the calibration of GPS heights (Figure 5.2-2).

- *There is generally no need to establish GPS stations to "connect" the main net to surrounding control unless distances are large.* Although relative error is a function of interstation distance ("parts per million" is the ratio of baseline vector error to baseline length, Figure 2.4-8), it is not a linear relationship. For distances up to 20-30km an "ambiguity-fixed" solution is generally possible (§8.1), and the accuracy of the position solution is of the order of 1 to 5 ppm. When, at longer distances, ambiguity resolution is not possible, the "ambiguity-free" solution is generally weaker by a factor of 2 to 3 compared to the "ambiguity-fixed" solution. It remains at this new (lower) relative accuracy for distances up to 100km. The optimal solution for baselines longer than 100km (if using commercial GPS software) is one based on triple-differenced phase observables.

- Certain baselines may be designated as primary ones, for example, the baseline(s) defining the axis of a tunnel or engineering structure. These are critical to the final network and must be at both a higher accuracy and higher reliability. Hence they may need to be measured several times, or special instrumentation and procedures used. Network design may therefore not simply involve points, but baselines (or figures) as well.

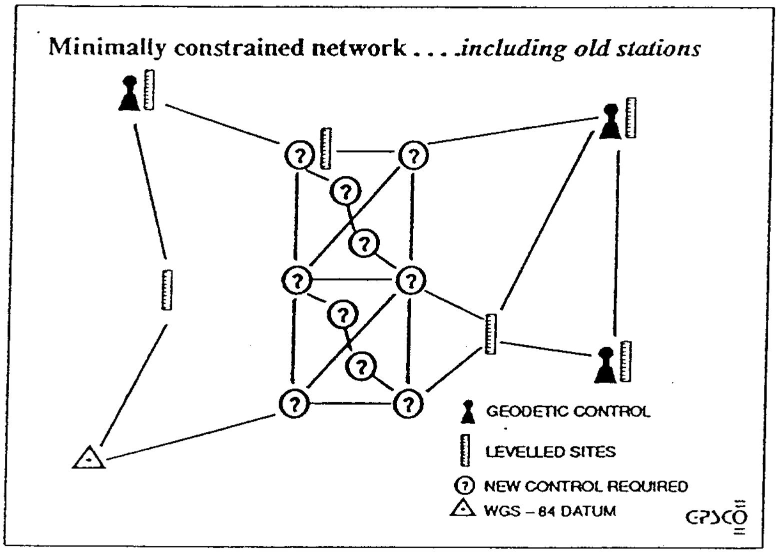


Figure 5.2-2. New, old and datum GPS stations.

Why New AND Known Stations?

> **A GPS survey typically requires the occupation of new stations and stations whose coordinates (2-D, 3-D, or height) are already known, in either the GPS datum or the local geodetic datum.**

Some reasons for surveying both existing (known) and new stations:

☞ Required by the relevant GPS survey standards & specifications.

☞ For the determination of local transformation parameters between the GPS datum and the local geodetic datum.

☞ For quality control (QC) purposes.

☞ In order to determine the geoid-spheroid separation.

☞ In order to connect new GPS points into surrounding geodetic control.

☞ However, a minimum of one known station must be used as the datum station in the GPS survey -- its coordinates must be known in the WGS84 system.

The number and distribution of known stations, and the accuracy with which the coordinates of the known stations are required is strongly dependent on the use to which the known stations will be put. Guidelines can usually be found in the relevant GPS "standards & specifications" (§10.2). **For most purposes a minimum of three or four stations**

around the perimeter of the survey project area is sufficient.

These known stations may be occupied in the course of the GPS survey, or they may be continuously operating "base" or "fiducial" stations. For example, there are several IGS and other continuously operating base stations in Australia which may be incorporated into the survey network if they are in the vicinity of the survey area (see §12.1). In future it is likely that there will be many more of these permanently operated base stations, maintained by government departments, universities, or private companies, to support a range of GPS activities (navigation, surveying and geodesy).

Accuracy Issues

Accuracy of New Stations:

- Required accuracy is defined by the project specifications, but the accuracy *classification* may be defined by GPS survey standards & specifications.
- Latitude is typically better determined than longitude. Height is worse than both. (The actual accuracy is dependent upon the instrumentation used, observations made and software used for the phase data reduction.)
- Typically, the horizontal coordinates are required in a local geodetic datum, and the vertical results in the orthometric height system. GPS derived 3-D coordinates are nominally in the WGS84 system.
- The accuracy of the GPS derived 3-D coordinates depend on the accuracy of the datum station(s) incorporated into the GPS solution (GPS phase solutions are relative, hence something has to be held fixed in WGS84 -- see §7.1).
- The final coordinate accuracy (in relation to the local geodetic datum) depends on the precision of the transformation between the GPS derived coordinates and the coordinates of the known stations in the local network.
- However, after transformation, the internal consistency (or precision) of the GPS results is unchanged.
- The final height accuracy depends on the precision of the height transformation: the geoid-spheroid separation, which can be determined in a number of ways (Chapter 11).

Accuracy of Known Stations:

Both GPS derived coordinates and the coordinates of geodetic control are essentially RELATIVE. However, GPS derived coordinate accuracy *relative to the local geodetic datum origin* depends on:

- the datum station(s),
- the transformation process, and
- any secondary distortion of the GPS network resulting from fitting to the datum defined locally by the known geodetic stations.

The following should therefore be noted:

- Datum station accuracy is generally defined by GPS survey standards & specifications. This influences the accuracy and precision of the GPS results.
- Final transformed GPS coordinate accuracy depends on the quality of the transformation parameters (themselves dependent on the accuracy of the known stations used to derive the transformation parameters).
- If the GPS transformed results are also to be fitted into the surrounding control, the final transformed GPS coordinate accuracy and precision depends on the accuracy and precision of the known stations in the local geodetic system.

Accuracy as a Classification Criteria:

Different geodetic control authorities may have adopted different terminologies for the classes of GPS survey, and have varying numerical accuracy limits used to define the categories, but all distinguish these categories by some relative accuracy measure. For example, in Australia and the U.S., relative error is defined for the various categories of survey by the specification of the maximum allowable "base error" (a) and "line-length error" (b), at the 95% confidence level, for the relative error ellipse (or ellipsoid):

$$e = a + b.L \qquad \text{(for Australia)} \qquad (5.2\text{-}1a)$$

$$e = \sqrt{a^2 + (b.L)^2} \qquad \text{(for the U.S.)} \qquad (5.2\text{-}1b)$$

where L is the interstation distance in kilometres, the quantities e and a are in millimetres, and b is expressed in "parts per million" (ppm).

There are essentially two classifications for accuracy:

☞ An internal one based on the minimally constrained adjustment of the GPS-only survey.

☞ An external one based on a constrained adjustment, where existing geodetic stations of the GPS network are held fixed to the published values of the terrestrial geodetic datum.

In Australia, the former corresponds to a survey's CLASS, while the latter defines its ORDER (see §10.2). For details concerning the Australian classification system see ICSM (1994), and for the U.S. see FGCC (1988). Table 5.2-1 is an extract from the Australian Standards and Practices for Control Surveys. Table 5.2-2 is the U.S. equivalent. Note the significant difference in definition of base error and line-length error between the Australian and U.S. standards. The Australian standards address all types of surveys, not just GPS.

Table 5.2-1. Australian horizontal (2-D) survey accuracy classifications.

CLASS*	Minimum geometric accuracy standard (95% confidence level)	
	a -- Base error (mm)	**b** -- Line-length error (ppm)
3A	0.2	2
2A	0.6	8
A	1.5	18
B	3	35
C	6	75
D	10	125
E	20	250

*CLASS is a function of field and reduction procedures.
ORDER is a function of both CLASS and fit to existing control.

Table 5.2-2. U.S. three-dimensional GPS survey accuracy classifications.

ORDER-CLASS	Minimum geometric accuracy standard (95% confidence level) a -- Base error (mm)	b -- Line-length error (ppm)
AA	3	0.01
A	5	0.1
B	8	1
1	10	10
2-I	20	20
2-II	30	50
3-I	50	100

The product of the Network Design Process is generally a **project sketch** depicting such information as: geographic graticule, symbol for station type (known and new), non-trivial baselines to be measured, repeat baselines, number of independent station occupancies, azimuth reference stations, and any other information that may help in the logistical design. See Figure 5.2-2.

Network Shape

- As in the case of conventional surveys, there is an impact arising from "structural" considerations.
- Some networks are superior to others with regard to "strength".
- Only independent baselines contribute to network strength.

GPS networks may have different shapes (Figure 5.2-3), as well as different "strengths" arising from the number of independent baselines observed over a number of sessions (Figure 5.2-4).

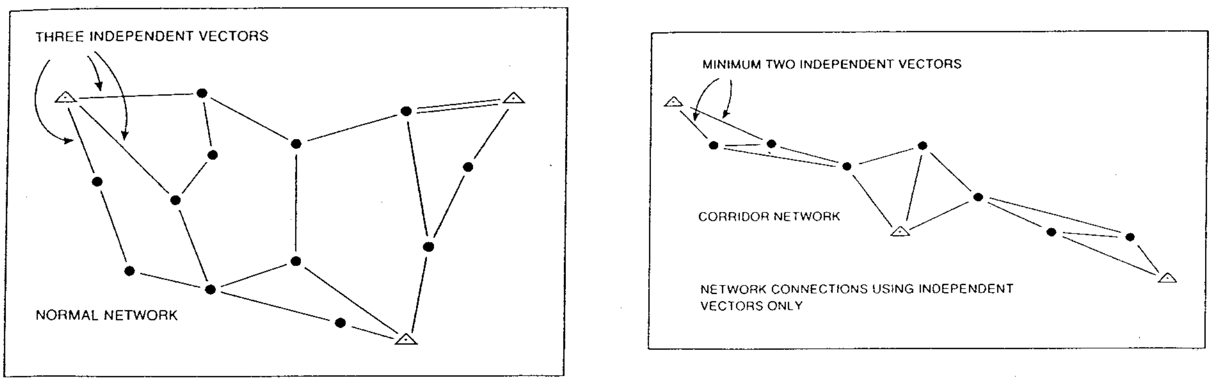


Figure 5.2-3. Network shapes: "wide" and "narrow".

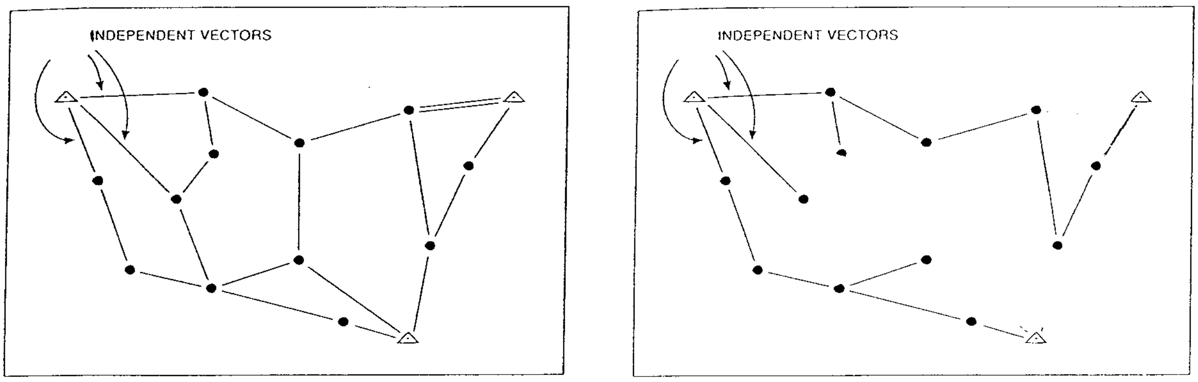


Figure 5.2-4. Network strength:
a function of the number and location of independent baselines.

5.2.2 OBSERVATION SCHEDULING

There are three considerations:

☞ Those that relate to the satellites themselves: how many to observe, for how long, etc.

☞ Those that relate to satellite-receiver geometry.

☞ Those that relate to logistical design: number of observation sessions per day, number of multiple site occupancies, etc.

Satellite Considerations

To prepare an observation schedule for a GPS survey it is necessary to first define the satellite constellation to be tracked, including such information as:

- Rise and set times of satellites above the observing horizon of a site. (Due to atmospheric refraction modelling difficulties GPS satellites are not normally tracked at elevations less than 15° to 20° above the horizon.)
- Health of the satellites. Each satellite broadcasts a health status indicator within its Navigation Message. This should be monitored during tracking, but a history of health problems may indicate a satellite which should be avoided if at all possible. (Note that although a satellite may be "unhealthy" for GPS navigation this does not mean that it cannot be used for surveying.)
- Satellite-site geometry: azimuth and elevation of satellites as a function of time.
- External Bulletin Board information on the status of satellites (§3.4), for example, planned orbit manoeuvres, shutdowns, testing, new satellite launches, etc.

This information is largely provided (in various graphical and tabular forms) by the "planning" software provided by GPS instrument manufacturers. Obviously the optimum "window" of satellite availability is required, but it may not always be possible to have such a window coincide with the daylight hours of the "working day". Hence, observation scheduling may be considered an exercise in *avoiding the worst observing windows!* An important criteria is the azimuth and elevation of the satellites over the day of interest, presented in the form of a "sky plot" as in Figure 5.3-1. (Azimuth and elevation can be computed using the formulae given in §1.2.)

Two tasks within the observation window scheduling process can be identified:

(1) Determine the minimum length of observation period per station. This is dependent on, amongst other factors, the following:
   - the satellite constellation to be observed,
   - the length of the baseline,
   - whether a "conventional" static GPS survey is being carried out, or "kinematic" and "rapid static" survey techniques will be used (see §5.5),
   - the observation period, as some periods may have better "geometric strength" than periods of the same length at other times. *(This is not an important issue for "conventional" GPS surveys, but it is important for modern "kinematic" techniques.)*

(2) Determine the satellites to be observed by all simultaneously operating GPS receivers. If instrumentation can only track less than the number visible satellites, then satellite selection becomes an issue. Modern instrumentation however tends to have "all-in-view" tracking capability.

The most critical is the **observation session length**. Unfortunately estimating the appropriate length of an observing session is very difficult, as it is a function of baseline length and environmental factors, as well as being dependent on the satellite constellation. Once the sites, instrumentation to be used, accuracy sought and time of day is fixed, then the constellation that can be tracked is known, and what remains to be determined is the session length. The following factors should be borne in mind:

- For GPS surveying, the minimum number of satellites that must be observed simultaneously by GPS receivers is two (to form one double-difference observable).
- The more satellites observed simultaneously, the shorter the observation period can be.
- The shorter the observation session, the more care is needed to guard against multipath and other biases.
- The "total" satellite geometry (all satellites tracked over a session) has an important bearing on the quality of the results. That is, sessions in which the constellation changes (satellites set and new ones rise) should be avoided, hence continuous tracking is preferred.
- The preferred solution is an "ambiguity-fixed" solution, hence any tracking beyond that required to successfully resolve the integer cycle ambiguities is useless. In other words, the optimum period of tracking is the one that permits ambiguity resolution to occur in the minimum period of time.
- An "ambiguity-free" solution, on the other hand, tends to improve with longer observation sessions.
- An "ambiguity-fixed" solution is unlikely to be possible for baselines longer than 20-30km using commercial GPS software packages.
- Experience indicates that to adequately resolve ambiguities, it is necessary to observe at least four satellites simultaneously for a minimum of 30 minutes. However, it has also been shown that in the presence of multipath, it can take two hours or more to properly resolve the ambiguities, if at all.

- In the case of baseline lengths greater than 20-30km, longer session lengths are recommended (Figure 5.2-5).
- It is not yet possible to predict in advance, with absolute certainty, how long an observing period should be to maximise the chances of an "ambiguity-fixed" solution being obtained.

There is unfortunately no GPS planning aid that takes all these factors into account: type of solution sought, baseline configuration, data sampling rate, available satellite constellation; and indicates the appropriate observation session length. Some attempts have been made to develop simple indicators of "good" and "bad" satellite geometry for GPS surveying, but even with these there is still great uncertainty in the expected baseline results due to unmodelled external factors influencing GPS adjustments, such as orbit errors, multipath and atmospheric refraction.

For conventional static GPS surveys it is generally recommended that the observation session lengths should be 0.5 - 2 hours in length, though this is dependent on the number of satellites to be tracked, baseline length, and to some extent the instrumentation and software being used (see, for example, Figure 5.2-5).

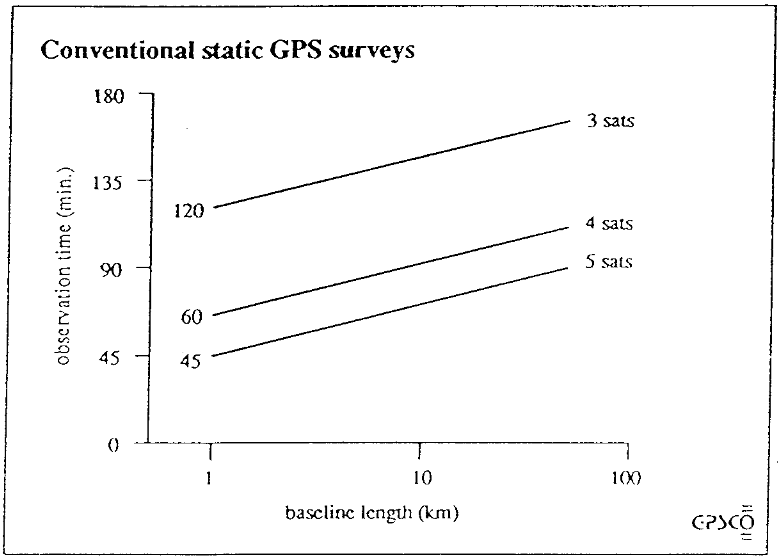


Figure 5.2-5. Suggested observation session lengths for "conventional" static GPS surveys.

Observation session lengths for "rapid static" GPS techniques are recommended as being of the order of 5-15 minutes. These are guidelines only, adequate for short baselines (<20km) and conditions of low PDOP (§1.4). Observation sessions lengths are a non-issue for "kinematic" and "stop & go" techniques as station occupancies are only of the order of seconds (§5.5).

Measures of Satellite Geometry

The accuracy of GPS-derived coordinates is generally a function of:

(1) the measurement precision,

(2) the systematic errors present,

(3) the processing strategy used, and

(4) the receiver-satellite geometry during the observing session.

Of these, items (2) and (4) are the most variable. The receiver-satellite geometry is highly predictable and is manifest in the Normal Equations for the Least Squares solution for the site coordinates. In fact the elements of the design matrix (the unit vectors for the receiver-satellite in question) can be computed beforehand and "optimised", similar to the planning procedures used for conventional terrestrial surveys. Hence, given a satellite constellation, the approximate coordinates of the receivers, the time of day and length of session, the Normal Equation System can be determined, inverted and the formal errors of the estimable parameters obtained. This process is described in, for example, MERMINOD et al (1990).

Unfortunately GPS receivers do not output this information. Instead, the "indicators of precision" (or more correctly the influence of satellite geometry on the solution) are borrowed from GPS navigation. In conventional surveying, the often used **indicators of precision** are the components of the error ellipse (in the case of 2-D Least Squares solutions) or error ellipsoid (for 3-D Least Squares solutions). The error ellipse associated with a computed point is described by three parameters: the lengths of the axes, and the azimuth of the major axis. Similarly, the error ellipsoid is described by six parameters: the lengths of the three axes of the ellipsoid, and the three orientation angles. However, the use of six parameters to describe the instantaneous precision of a GPS navigation "fix" is rather awkward, and hence GPS navigation has tended to use *simplified* precision indicators.

The error ellipsoid is approximated by an error sphere whose radius is equal to the root of the sum of the squares (RSS) of the ellipsoid axes. The **Position Dilution Of Precision** (PDOP) is defined as the radius of the RSS sphere, assuming the standard deviation of the pseudo-range measurements is unity (§1.4). Hence the six precision indicators associated with an error ellipsoid are compressed into one PDOP quantity which depends only on the relative geometry of the four or more satellites, and the approximate location of the point whose coordinates are to be determined. Other DOP factors can also be defined, in particular HDOP and VDOP, the horizontal and vertical DOP respectively, and GDOP. It is these DOPs (in particular PDOP and GDOP) that are output by GPS receivers, and have tended to be used (perhaps incorrectly) for accuracy studies. The inappropriateness of using PDOP for conventional GPS survey planning can be readily seen if a GPS survey adjustment is compared with a GPS navigation solution:

(a) The navigation solution is based on the *instantaneous* satellite-receiver geometry, whereas the GPS carrier phase adjustment is governed by the continually changing geometry of an observing session which may last up to several hours.

(b) PDOP gives the geometric strength of a *point position*, whereas the GPS carrier phase adjustment provides the *relative position* of two or more receivers.

(c) The pseudo-ranges are considered to be biased only by the errors of the receiver clock (which changes from epoch to epoch). In a GPS carrier phase adjustment, the double-differenced observation is predominantly biased by the integer cycle ambiguities (the satellite and receiver clock biases are eliminated). The geometric strength of double-differenced *ambiguous range* measurements (integrated carrier phase) is quite different to that of *pseudo-range*.

A simple example of the inappropriateness of GDOP for GPS carrier phase adjustment is the fact that GDOP is undefined for three satellites (it would involve the estimation of four parameters from three observations) and yet three satellites can be used in a baseline solution using integrated carrier beat phase. Alternative DOPs have been introduced, such as BDOP, RDOP and DGDOP (see IBID, 1990). Their usefulness, however, has been restricted by the unpredictable nature of the systematic errors, and in some cases by the lack of distinction being made between an "ambiguity-free" solution and an "ambiguity-fixed" one. As a consequence, mere "rules-of-thumb" can be given, such as: the best time to carry out a GPS survey is when the PDOP (or GDOP) is changing rapidly (rather than when it is low). But with the deployment of the full satellite constellation it unlikely that there will be periods when the conditions for a "good" geometry for GPS surveying are significantly better than at other times.

Modern GPS surveying techniques are in fact closer to GPS navigation solutions, hence PDOP does have a role to play in GPS planning.

So what constitutes a good tracking session?

As a rule-of-thumb, a good session is one which has four or more visible satellites above the 15° - 20° elevation angle at the start of the tracking period, and ends before the fourth satellite sets below this cutoff elevation angle.

The length of session then is a function of the baseline length, and whether it is a conventional static survey (say 30-120 mins in duration), or a "rapid static" survey (5-15 mins) or "stop & go" (1 min).

Observation Scheduling: Some Logistical Factors

Observation scheduling within a network relates to:

☞ The number of observation sessions in a day (dependent on length of the work day, and minimum session length).

☞ Total non-productive times: travel time between stations, station setup/takedown time, data downloading, etc.

☞ Number of occupations of each station (new and unknown).

The main consideration, however, relates to multiple occupancy of sites. These are required for a number of purposes:

- To provide a "link" in the datum between each session. The datum station's WGS84 coordinates can then be propagated through an entire network (Figure 5.2-6). How many common stations between sessions? The issue is simply that:
  - one is a bare minimum,
  - two provides one redundant link, and
  - three or more is even better.

- A quality control measure, as multiple occupancies provide different "pathways" through a network for checking on blunders, wrongly identified stations, errors in antenna height measurement, etc. However, the larger the number of links between sessions, the slower the progress and hence the more expensive the GPS survey.

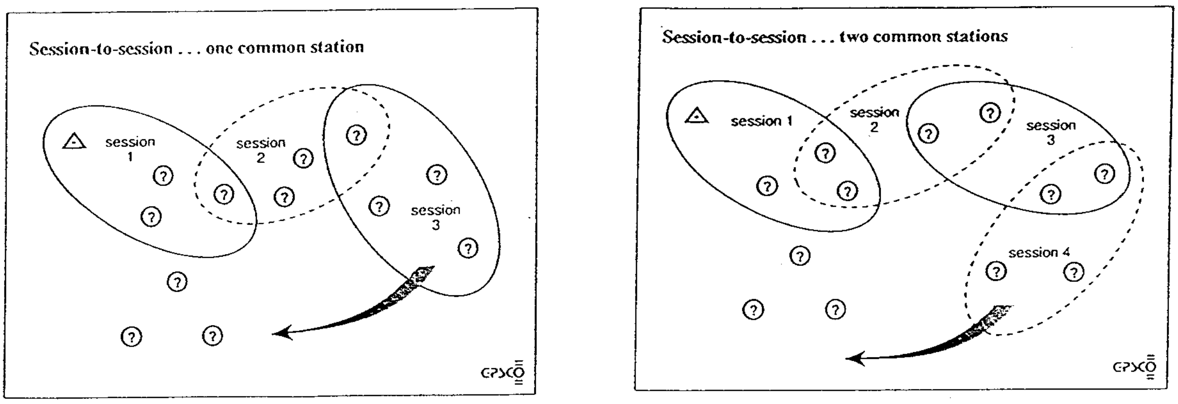


Figure 5.2-6a & b. Moving from session to session: multiple station occupancy.

The minimum number n of sessions in a network with s stations, using r receivers is (HOFMANN-WELLENHOF et al, 1994):

$$n = \frac{(s - o)}{(r - o)} \tag{5.2-2}$$

where o denotes the number of "link" or "pivot" stations between sessions, $o \geq 1$, $r \geq 2$, $r > o$ and n is an integer. An alternative relation that is simply based on the notion that *all* stations should have multiple occupancies (as is recommended by certain GPS "standards & specifications" -- §10.2) is (IBID, 1994):

$$n = \frac{m.s}{r} \tag{5.2-3}$$

where m is the number of times a station must be occupied, and n must be rounded up to the next higher integer. With this basic information a session-by-session plan is made that has as one of its major aims the survey of a network with relatively *homogeneous* accuracy and reliability. The outcome of this observation scheduling exercise is the preparation of an observation plan that is to be followed by the field parties. Table 5.2-3 is an example of such a plan, for the session-by-session survey appropriate for the project plan in Figure 5.2-6b.

Often the common link stations will be occupied by the same receiver and field party during two successive sessions. However, this does not constitute an independent set-up! Nevertheless, a link between sessions is established even if the field party is setting-up over the wrong groundmark! If stations linking successive sessions are occupied independently, and if the antenna is not set-up over the same mark, or the antenna height is incorrectly measured, the link between sessions is lost. Having a suitable degree of network redundancy (multiple setups, preferably independent) may permit such errors to be identified and remedied. Alternatively such common errors as setting-up over the wrong mark, or not measuring antenna height correctly, can be minimised if field surveyors are properly trained in GPS field procedures.

Guidelines on multiple occupancy of GPS stations, and the percentage of independent occupations of common stations, can usually be found in the prescribed GPS survey "standards

& specifications" (see "Logistics Design Principles" and §10.2).

5.2.3 INSTRUMENTATION & PERSONNEL CONSIDERATIONS

Instrumentation considerations as they relate to the project planning process include:

- Number of available GPS receivers: the larger the number of receivers in a session, the larger the number of directly connected stations, and hence a better network, faster progress and a less expensive survey. However, there may be a restriction on the availability of receivers and field parties, and, in addition, the logistical problems quickly multiply. The optimum number of receivers appears to be of the order of four to six.
- Receiver type: all geodetic GPS receivers produce, in principle, similar datasets and hence similar final accuracies. The mixing of different brands of receivers however can cause problems. For example, most data processing software is instrument-specific. This is particularly true for modern GPS surveying techniques (§5.5).
- Single or dual-frequency receivers: dual-frequency instruments permit compensation for the ionospheric delays on the GPS signals, hence they are essential for high accuracy applications. They are usually of little benefit for baselines <30km. Dual-frequency instrumentation is generally necessary if modern "rapid static" survey techniques are being used (§5.5). Guidelines may be found in the prescribed GPS survey "standards & specifications" (§10.2). Mixing of single and dual-frequency receivers, as in the situation of mixing receiver types, is not recommended.

Table 5.2-3. Sample field deployment plan.

Session	Time	Field parties			
		A	B	C	D
295-1	1000 1100	7352	20882	62755	15007
295-2	1110 1210	7352	20882	20888	70936
295-3	1515 1615	15001	20877	20888	70936
295-4	1800 1900	15001	20877	70418	62500

As noted earlier, many errors in field procedure can be detected if multiple station occupancies are incorporated into the GPS survey plan. Errors may also be minimised through adequate training of field personnel. If the field staff are unfamiliar with the GPS instrumentation to be used for a survey it is even more important to insist on pre-mission testing as part of the survey planning procedure (§4.4). This testing could be:

- over a micro-network of several metres in extent, for example, a square or other regular figure, for which the "ground-truth" can be directly obtained by precise distance and/or angle measurement, or
- over a special permanent test network, perhaps previously surveyed by GPS, or
- an effective simulation of the planned survey, in which lines of similar length are surveyed using procedures close to those to be applied during the actual survey. This is the most expensive option.

Such testing will establish:

- the proper functioning of antennas, receivers, data storage, cables, power supplies,
- health and availability of satellites,
- the competence of field parties,
- the proper functioning of field and office software and procedures, and
- the competence of the computing staff.

5.2.4 LOGISTICAL CONSIDERATIONS

The logistical considerations of a GPS project increase enormously with:

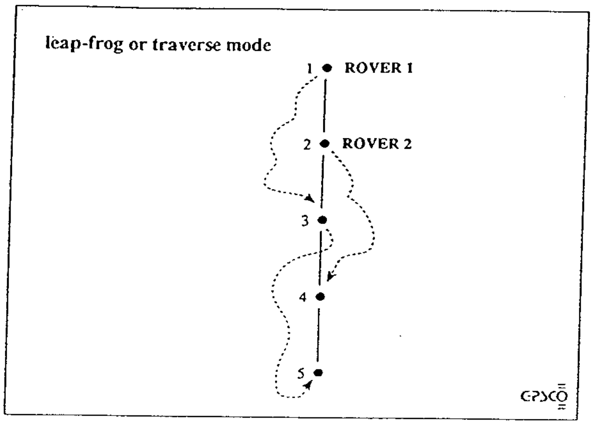
- an increase in the number of stations to be surveyed,
- an increase in the number of receivers deployed,
- an increase in the number of common stations occupied between sessions,
- an increase in the number of fixed stations to be occupied, and
- an increase in the number of sessions per day.

They are also complicated by such factors as:

- Terrain and the nature of the transport links? 4WD vehicles? helicopter?
- Requirements for contingency plans, in the event of mishaps, etc.
- Requirement for quality control checking and partial processing in the field office each day, hence the need for data from all receivers to be transferred to the field office.
- The amount of reconnaissance already carried out.
- Economic pressures, for example, to complete the project in the minimum time.
- Available accommodation to minimise travel time, etc., is it convenient to the project?
- Instrumental factors: do receivers need special equipment for data download?

The basic principle is to KEEP IT SIMPLE, hence it is important to keep baseline short (and hence minimise travel times), use say one hour observation sessions, and employ a suitable receiver deployment strategy.

There are a number of possible receiver deployment schemes that can be used. Each has its advantages and disadvantages with respect to logistical considerations such as cost, time, manpower, etc. Generally, some station or stations are occupied for all or some of the campaign, and the other receivers move between sessions. A combination of the "base station" (or "radiation") mode, and the "leap-frog" (or "traverse") mode, as shown in Figure 5.2-7, is usually used.



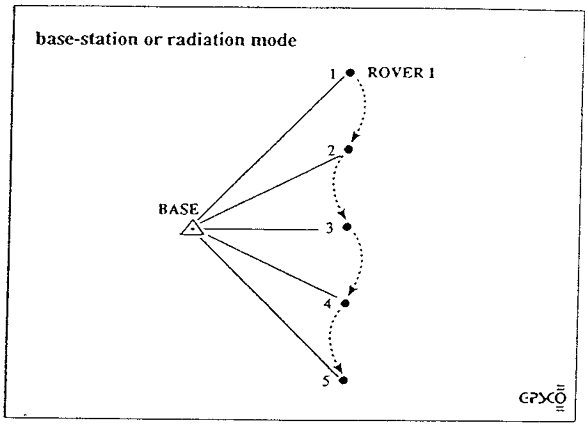


Figure 5.2-7. Receiver deployment modes.

Logistics Design Principles

With experience, the organisational design of a session-by-session observation schedule is a straightforward matter based on a few simple "rules-of-thumb":

- Receiver deployment schemes based on:
  - leap-frog mode,
  - base station operation,
  - or some combination of these.
- The "shape" of a GPS session, or final network (multi-session), plays little part in the final accuracy, unlike the situation with conventional surveys. Length of lines in a session do influence the final accuracy, both through the "ppm" relationship, as well as a change in the "ppm" value at distances above which ambiguity resolution can be carried out.
- Adjacent stations should be connected directly -- that is, keep baselines short.
- Stations on the perimeter of project area should be directly connected.
- Connect all stations in network through "pivot" or "link" sites common to two or more sessions, so that a minimally constrained GPS network is established. The percentage of multiple occupancies is directly related to the accuracy classification of the survey, and is usually defined in the prescribed "standards & specifications" (§10.2).
- Maximise geometrical redundancy through having multiple occupancy of sites (though only within reason). There are two types of redundancy:
  - Reoccupation of several sites at the same time to define repeat baselines (to allow for checks on the internal consistency of GPS surveys).
  - Form loops of stations occupied during different sessions, permitting checks on loop closure statistics.
- Where possible sites should be revisited by different field parties, hence ensuring independent setups, in order to minimise the chance of misidentifying the station mark.
- Avoid "no check" baselines (Figure 5.2-8), where one station of the baseline has only been visited once.

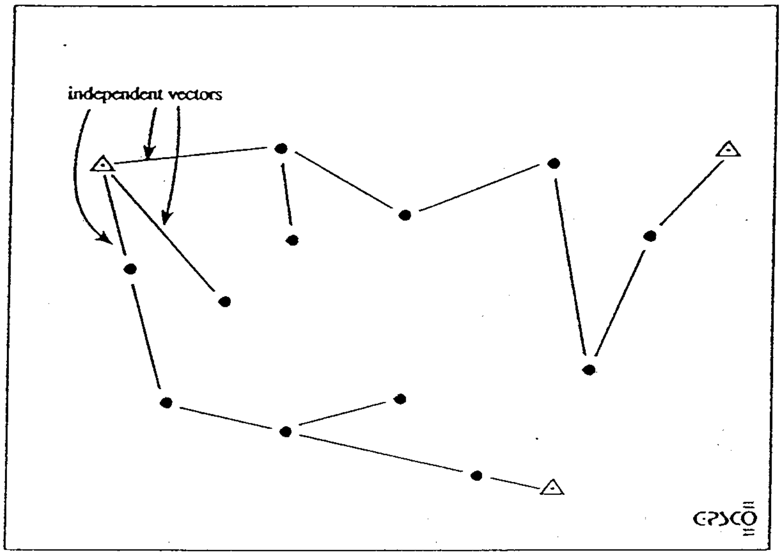


Figure 5.2-8. Factoring in redundancy to prevent "no check" baselines.

There is a danger in the *overuse* of the "base station" or "radiation" mode of survey (be it with GPS, or conventional EDM/theodolite) as, unless all the sites are visited more than once, all *radiated* baselines are in reality "no check" baselines. It is tempting to consider a two-base station configuration as having a natural *builtin* redundancy (there are two baselines coming into each site, one from each of the base stations -- one being the primary base station, and the other the secondary base station). However, although the site is no longer fixed by a single "no check" baseline radiation, the additional baseline from the secondary base station is not independent! **Hence, observation scheduling must take into account not only the requirement for multiple occupations to ensure redundancy, but also the degree to which these additional baselines must be truly independent.**

The reader is referred to SNAY (1986) and UNGUENDOLI (1990) for a discussion of organisational design as it applies to surveying a given network with a certain number of GPS receivers. However, it is unlikely that such "academic" planning exercises are ultimately useful given the extraordinary variety of GPS networks that surveyors may be called upon to observe. It is preferable that the surveyor develop a "sixth sense" when it comes to planning surveys, such that due consideration is always given to ensuring there are builtin redundancy and quality control measures, rather than incorporating them into a design as an afterthought.

The issue of GPS network design is discussed again in §5.5, in relation to the modern GPS surveying techniques such as "rapid static" and "stop & go".

Factors influencing GPS accuracy: Baseline results

The "shape" of the GPS network is irrelevant as far as the baseline solution quality is concerned - orientation of baseline does not greatly influence its quality. The following factors do influence baseline quality:

- Length of baseline.
- Single or dual-frequency instrumentation.
- Length of observation session.
- Number of observed satellites.
- Observable being processed.
- Processing software.
- Quality of ancillary information (orbits, fixed sites, etc.).

Typical horizontal baseline accuracy is expressed as:

$$e = \sqrt{a^2 + (b.L)^2}$$

where L baseline length in km,
a = 0.2 - 1 cm (centring error),
b = 1 - 5 ppm

Height is the weakest component (2-3x worse)

Factors influencing GPS accuracy: Overall network quality

- Baseline quality.
- Homogeneity of GPS survey (single and dual-frequency results, etc.).
- Number and distribution of independent baselines.
- Other geodetic observations (distance, angle, etc.).

The overall network accuracy is a function of the number and distribution of <u>repeat</u> baselines and <u>redundant</u> station occupations.

5.3
GPS SURVEY RECONNAISSANCE

Reconnaissance issues for GPS surveying include:

☞ Satellite availability: satellite selection, satellite health, observation window, etc.

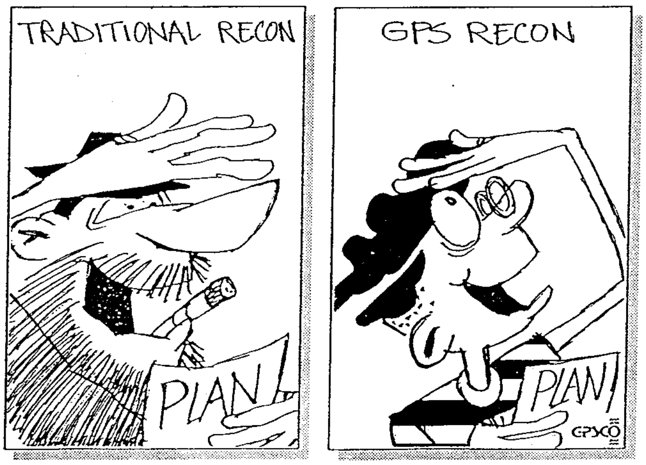
☞ Satellite visibility: checking onsite obstructions.

☞ Clearly identifying the groundmark over which the GPS antenna is to be set up. *Particularly for night observations.*

☞ Identifying, if necessary, eccentric stations to be occupied if the primary groundmark cannot be used, and other azimuth stations if required.

☞ Station access: critical for minimising non-productive travel times and unscheduled delays in getting on-site. *Particularly important if night travel is involved.*

☞ Site conditions: on-site power? multipath environment?



Satellite Availability and Visibility

To schedule a GPS survey the following factors need to be taken into account:

- Satellites are not normally tracked below an elevation of 15° to 20° due to large atmospheric refraction errors at low elevation angles.
- There is a 24 hour observation window for GPS.
- At least 30 minutes of four (or more) satellite coverage is the accepted norm for standard static GPS surveys. *There are periods when many more satellites are visible.*
- The satellites' positions in the sky are predictable. They can be computed and output in a convenient graphical form, and taken out into the field during reconnaissance.

- A popular representation of satellite availability is the **skyplot**, which is a plot of satellite tracks on a zenithal projection centred at the GPS ground station. The satellite azimuth and elevation is shown as a function of time. Figure 5.3-1 is a schematic of a skyplot with several groundtracks plotted.

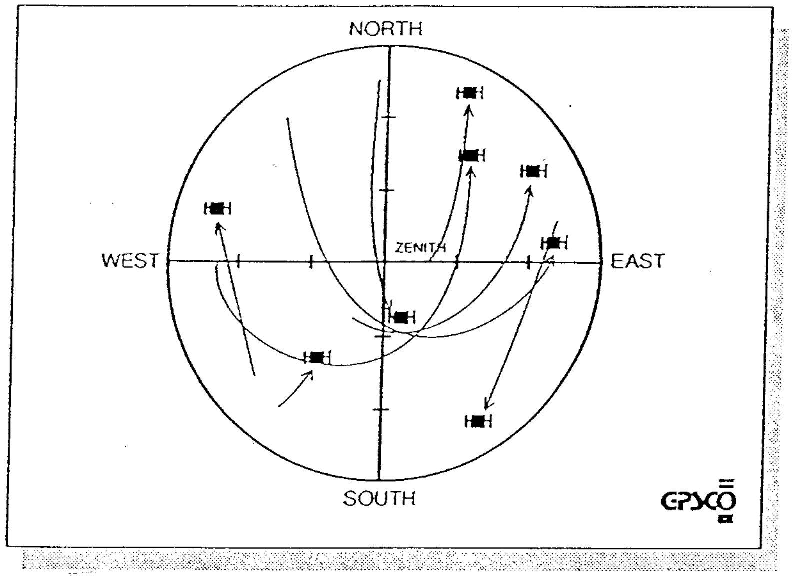


Figure 5.3-1. A GPS skyplot.

Skyplots can be used during reconnaissance to ensure that the satellite signals from the selected satellites can be acquired by the receiver. The obstructions at a site can be plotted on a zenith plot independently of a skyplot as in Figure 5.3-2 (or a photograph of the zenith could be taken with a fish-eye lens!), and during the survey planning stage a decision can be made on which satellites to be tracked at the sites by comparing Figures 5.3-1 and 5.3-2 (remembering that only common satellites between sites are useful!). Skyplots annotated with possible obstructions are also useful during actual data tracking because if the data quality is poor (as indicated by the signal-to-noise ratio displayed by the receiver during operation), the possible source of interference can be identified and an appropriate note made to assist subsequent data processing.

How to best use skyplots during reconnaissance? While GPS manufacturers software can generate the plots, it is not convenient to go into the field with a computer and printer! A compromise is therefore to use a manual method based on a series of skyplots (generated for a representative site such as the centroid of the network), for each hour. At each site the obstructions can then be mapped, and if there are a significant number of potential signal obstructions, they may be visually compared with the various one hour skyplots and the best tracking period selected (that with the minimum obstructions).

How may the obstructions be mapped during reconnaissance? The basic tools are a compass and a clinometer, and every field party should be equipped with them. However, if the site location has not yet been fixed with certainty, as may be the case with sites in urban environments, the reconnaissance assumes greater importance. As there may be many candidate locations for each site, some shortcut techniques should be considered. For example, keep several graphs handy that relate distance to and height of a potential obstruction, to the elevation angle that it subtends (see Figure 5.3-3). Efficient reconnaissance techniques can be developed if the heights of buildings and trees are known, and distances are easily measured using, for example, a handheld laser rangefinder.

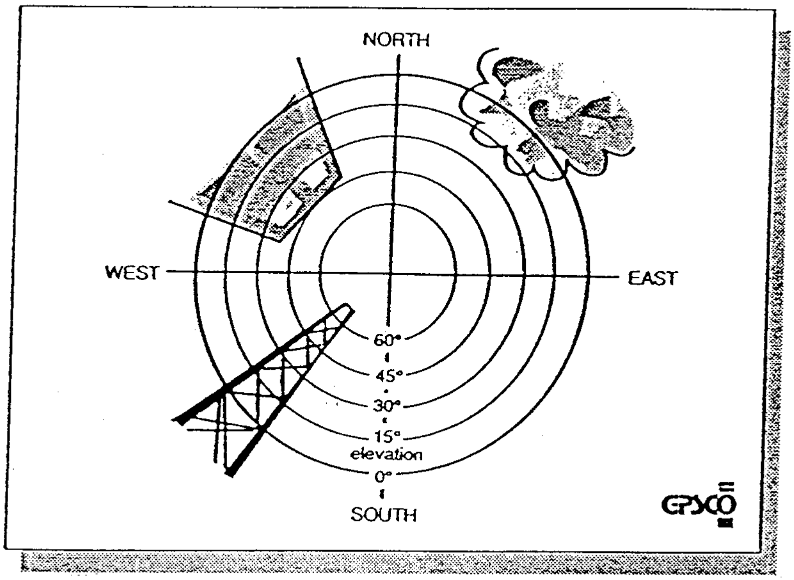


Figure 5.3-2. Plotting signal obstructions on a skyplot format.

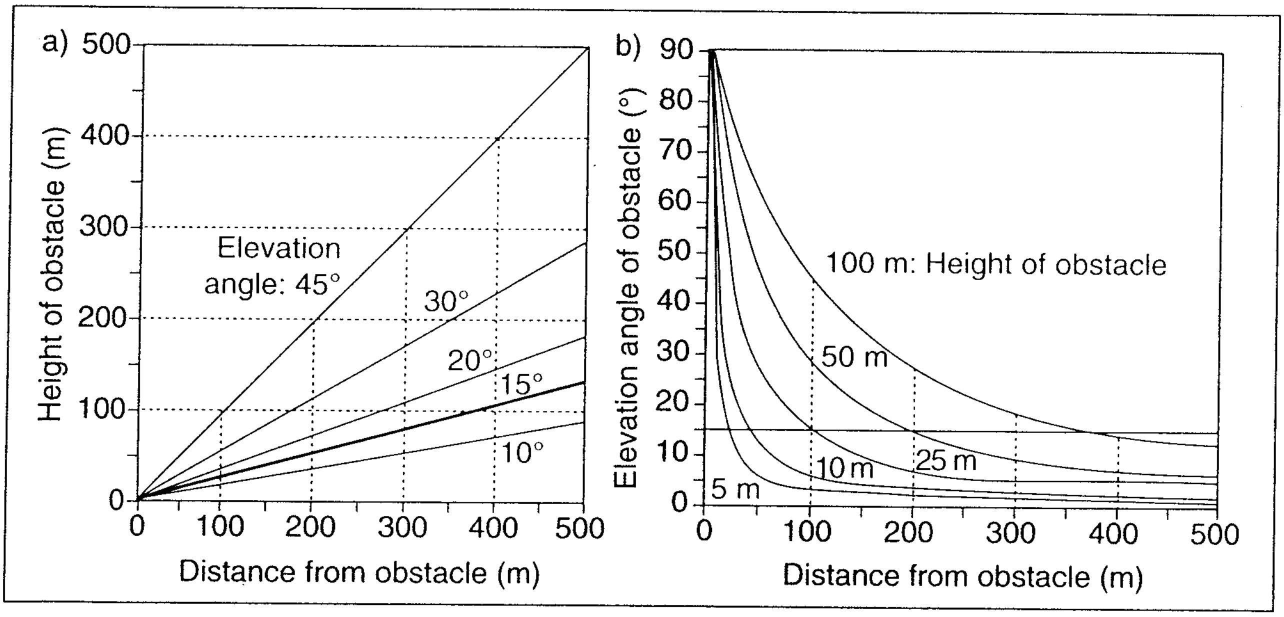


Figure 5.3-3. Graph a) shows height of an obstruction as a function of its distance and elevation angle, and
Graph b) shows the elevation angle of an obstruction as a function of its distance and height.
(SANTERRE & BOULIANNE, 1995)

The pattern of signal obstructions at a site is very sensitive to the height of the antenna. By increasing the height of the antenna in urban areas using some form of telescopic pole, the extent of signal obstruction may be greatly reduced, as dramatically indicated in Figure 5.3-4.

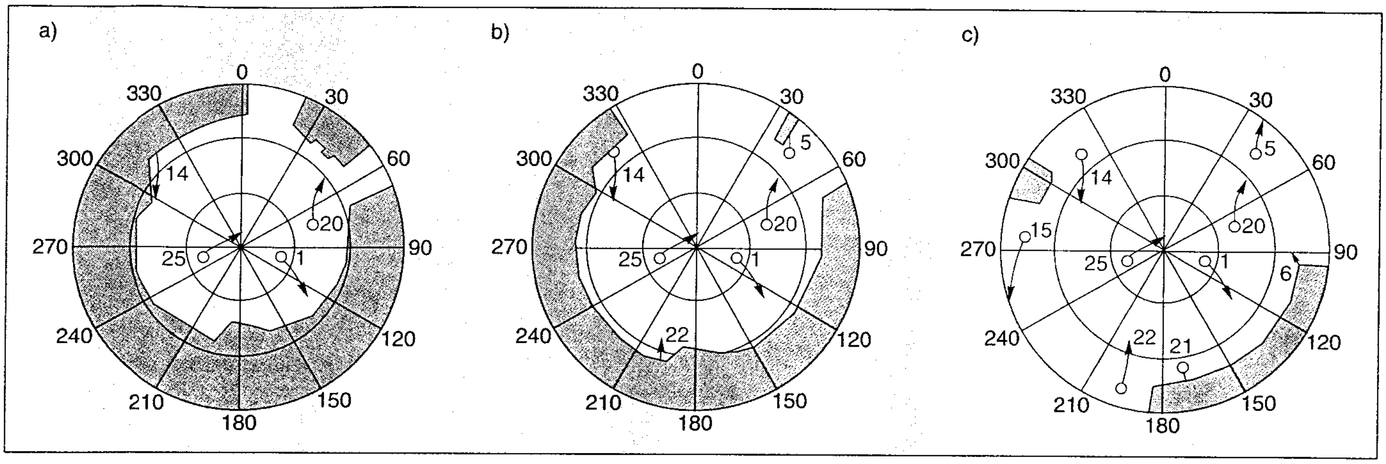


Figure 5.3-4. A sequence of skyplots showing obstruction masks and groundtracks of visible satellites for antenna heights of 1.5, 5 and 10m respectively.
Note the number of visible satellites increases from four (at an antenna height of 1.5m) to nine (at an antenna height of 110m).
(SANTERRE & BOULIANNE, 1995)

When preparing skyplots the following factors should be kept in mind:

- Only approximate coordinates of the intended GPS station are required. Only one such plot need be produced if the GPS network has sides of the order of 100km.
- One skyplot can be used for an entire observing campaign, as the satellite tracks across the sky repeat daily, except that the satellite passes over the same ground point four minutes earlier each day.
- The preparation of skyplots requires approximate satellite ephemerides. Different skyplot generating software may require different satellite data. The most convenient is to use an old Navigation Message (collected during a previous GPS survey or during pre-mission testing). *This will contain the necessary GPS Almanac data.*
- The Navigation Message or Almanac need not be the most up-to-date. For planning purposes even messages a month old (or longer) may still be used. (Check that satellites have not been moved, or new ones launched in the meantime!) Up-to-date Almanac files can generally be downloaded from Bulletin Board Services (§3.4).

Point Recovery and Station Access

This information should be clearly stated in words as well as described in some graphical form. This is critical for minimising down-time due to difficulties in finding stations, or if access involves caretakers (for example, to visit the roofs of buildings), etc. An example of very detailed information sheets designed for this purpose are attached at the end of this section.

Station Selection and Marking

As with details of the station access and point description, the area around the site should be studied carefully. Depending on the aim of the project, and its accuracy requirements, this task may be very elaborate and include, for example:

- Investigating the provision of on-site power.
- Testing soil stability and defining the appropriate antenna mount (tripod, pillar, etc.).
- Noting the presence of any potential multipath causing structures.
- Noting any UHF/TV radio or radar transmitters (they could affect a receiver's operation).
- Establishing permanent monumentation -- using previous marks helps avoid this.
- Establishing nearby azimuth marks.
- Clearing the area of possible obstructions caused by trees or shrubbery.
- Taking photographs of the surrounding area, including any tree cover.

> ***What is a good site?***
>
> **(1) No signal obstructions above 20°.**
>
> **(2) No multipath causing surfaces, such as metallic fences, structures and water surfaces.**
>
> **(3) No nearby electrical installations, such as high tension cables, radio/radar/TV transmitters.**

In some countries, account may have to be taken of the season. For example, tree cover may be much thicker during summer than at other times of the year (Figure 5.3-5). Care should therefore be taken to ensure that site selection (during reconnaissance) is not carried out during the time of year when density of foliage is less than what will be experienced during the survey.

Figure 5.3-5. In spring, the branches of trees may still be bare (left), while in summer the trees are covered with thick trees.
(LACHAPELLE & HENRICKSEN, 1995)

NASA JPL

GPS Site Information Report

Date__________

This report should be completed in detail to provide uniformity in data essential to station occupation and GPS experiment planning. It also aids in station recovery and reoccupation. Please provide the requested information where applicable.

Site Name:__________ 4-Char. ID:__________ Station ID #:__________
Location:__________ City:__________ Country:__________
Site Address (if applicable):__________

Reported By:__________ Agency:__________ Telephone:__________

Approximate Geodetic Coordinates

Latitude:__________ Datum:__________ Current Magnetic Declination:__________
Longitude:__________ Source of Position:__________ Rate and Direction of Change:__________
Height:__________ Source of Declination:__________

Topographic Map Showing Location

Name__________ Scale__________
Agency__________ Date__________

Time offset from UTC

Standard______Hr. Daylight______Hr.
☐ No Seasonal Time Changes are Observed Locally

Information Included in this Report (Check all that Apply)

☐ How to Reach the Station and Verbal Site Description
☐ Site Vicinity Sketch
☐ Maps Showing Site Location
☐ Ancillary Information Form
☐ Horizon Mask Diagram
☐ Site Survey Information/Results
☐ Photos of all Marks (Label all Photos)
☐ Photos of Station From Cardinal Points (50m dist)
☐ Photos of Landmarks From Station
☐ Other (Specify)__________

Site Ownership/Permission

Who owns the site?__________

Primary Site Contact:__________
Title:__________
Address:__________

City, State, Zip:__________
Country:__________
Tel.:__________ Fax:__________

Discuss procedure required to obtain permission to occupy this site.

Attach Additional Sheet if Necessary

Existing Geodetic Ties

Is this station part of an existing geodetic control network?............ ☐ Yes.................. ☐ No.................. ☐ Don't Know
If yes, what type of network is it?.. ☐ Horizontal........ ☐ Vertical............ ☐ 3-D
What agency is respossible for the network?__________
What is the accuracy classification of this network?__________
What other space geodetic systems use this site?............................ ☐ Fixed VLBI...... ☐ Mobile VLBI.... ☐ SLR
☐ None................ ☐ Don't Know
What agency is responsible for the system's operation?__________
Has a survey tie been made to the GPS station?.............................. ☐ Yes.................. ☐ No.................. ☐ Same Station
If yes, include relevant information in the site survey section of this report.
Comments:__________

Jet Propulsion Laboratory M/S 238-600, 4800 Oak Grove Dr., Pasadena, CA 91109 USA, Tel. 818-354-8330/393-6051, Fax 818-393-4965

Summary of On-Site Geodetic Markers

Agency	Inscription	Approx. Location From Sta. Distance (m)	Azimuth	Mark Type	Cond-ition	Mark Setting	Soil Type
		N/A	N/A				

See Code Below — Comment Where Necessary (Mark Type, Condition, Mark Setting, Soil Type)

Mark Type: (Include all that Apply)
J = JPL Rod and Plaque
R = Steel Rod Only
B = Brass Plaque
P = Other Plaque (Specify)
M = Magnetic Material in Monument
O = Other (Specify)

Condition: (Include all that Apply)
S = Mark Solid in Setting
L = Mark is Loose in Setting
M = Mark is Marred
U = Mark is Un-Marred
C = Center Point is Unreadable
N = Center Point Not Well Defined (Specify)
O = Other (Specify)

Mark Setting: (Include all that Apply)
P = In Poured Concrete Form or Post
D = In Concrete in Depression in Rock
C = Cemented into Drill Hole in Rock
M = Mounted on Mast or Pier (Specify)
S = In Concrete Structure, ie Foundation (Specify)
O = Other (Specify)
F = Set Flush with Ground Surface
B = Set Below Ground Surface (Specify)
A = Set Above Ground Surface (Specify)

Soil Type/Foundation:
A = Alluvium or Soil
B = Boulder
C = Compact Sediments
R = Rock Outcrop
U = Bedrock Under Soil
O = Other (Specify)

Comments Where necessary comment on type, condition, setting, and soil type. Comment on any factors that may affect the stability of the marks (ie. land slides or erosion). If constructing monument describe what is underground. Use sketches and attach additional sheets if necessary.

Site Survey

Has a survey been performed to determine the relative,
3-D locations of the station and reference marks?........................ ☐ Yes.......... ☐ No.......... ☐ Don't Know
If yes, what type of survey was it?.. ☐ GPS.......... ☐ Conventional
What is the expected/achieved accuracy of the survey?__________ Date of Survey :__________
What agency performed the survey?__________
Who has the original field notes or data?__________
Who performed the data reduction and adjustment?__________
Comments on the site survey:__________

Attach all available survey information to this report.

Site Access What type of transportation is needed to access the site (boat, 2WD, 4WD, inaccessable by vehicle, etc.)? Comment on the effort required (long hike, help needed to pack gear, mules. etc.). Indicate availibility of assistance, make recommendations.

Attach Additional Sheet if Necessary

On-Site Facilities Discuss any on-site facilities available to support GPS operations (Can the receiver be placed in a building? How far is the building from the observing mark? Is office or storage space available on-site? etc.).

Attach Additional Sheet if Necessary

Power

Is AC power available on-site?.. ☐ Yes................ ☐ No

Voltage___________V. Frequency__________Hz. Distance from power source to receiver:___________ Meters

Comment on Stability and Reliability:

Discuss any special equipment that may be required, describe outlets (sketch if necessary) and state costs if required.

Where can batteries be purchased (try to include several options, comment on type and cost)?

Where can batteries be charged (try to include several options, indicate best option)?

Is a generator or solar panel needed at this site?... ☐ Yes............... ☐ No

Equipment Recommendations List ancillary equipment that may be useful at this site (antenna cable length, tent, etc.)

Attach Additional Sheet if Necessary

Communications What modes of communication are available at or near the site (telephone, fax, telex, radio, mail, indicate locations)? State the best mode of communication. Make recommendations and give instructions. Supply all relavent numbers or radio frequencies. Assess quality, operating hours, frequency of delivery etc.

Attach Additional Sheet if Necessary

Adverse Conditions Discuss any conditions that could adversely affect operations. Include things like potential sources of multipath, radio frequency interference, cultural or political sensitivities, adverse weather conditions etc.

Attach Additional Sheet if Necessary

Other Site Contacts List all persons providing assistance at this site. Include all those visiting the site with you, other local contacts providing assistance, US embassy contacts etc. Indicate roles and include institutions, mailing addresses, telephone and fax numbers etc.

Attach Additional Sheet if Necessary

NASA JPL

GPS Site Description Form

Date:____________

Site Name:______________________________ 4-Char. ID:__________ Station ID #:__________

Location:__________________________ City:________________ Country:________________

Reported By:_______________________ Agency:________________ Telephone:________________

Verbal Site Description Describe how to reach the site from an identifiable location such as a major intersection or landmark. Include street names, distances and directions of travel. Verbally describe the site. Indicate the relative locations o all monuments and the monument locations relative to land marks. Describe each monument (what it is set in, what it is, inscription and condition).

Site Sketch Sketch the site vicinity. Show all relevant features, landmarks and the relative locations of all markers. Include a North arrow and approximate scale. If not to scale indicate so.

NASA JPL

Ancillary Site Information

Attach Additional Sheets as Necessary

Date__________

Site Name:__________ 4-Char. ID:__________ Station ID #:__________

Location:__________ City:__________ Country:__________

Reported By:__________ Agency:__________ Telephone:__________

Customs and Shipping Detail customs procedure if applicable. Indicate limitations or exemptions and requirements for duty-free or temporary importation. Include applicable forms if available. If possible include a shipping address and contact.

Entry Requirements Discuss country entry requirements. Discuss passport and visa requirements and costs. State any required immunizations.

Air Transportation Indicate nearest commercial airport. Discuss frequency of flights and connecting information.

Local Transportation How do you get from the airport to the site or hotel? Supply information on car rental agencies and other local transportation. Where can 4-wheel drive vehicles be rented? Include phone numbers and addresses. Indicate costs.

Food and Lodging Make recommendations on available food and lodging. Note proximity to site, acceptability and price. Supply applicable addresses and telephone numbers. Where can you get fresh water?

Local Facilities Describe the nearest community (size, availibility of standard services, stores etc.).Discuss any repair/maintenence facilities at or near the site (hardware, electronic, construction materials etc.). Include locations, telephone numbers, proximity to site etc.

Personal Recommendations Discuss items that site personnel should know for future operations at this site. Make recommendations for safer, easier, and more enjoyable operations. Include all pertinent and useful information. <u>Clearly define any "must know" items.</u> The following is a partial listing of items you may want to discuss: immunizations and medications; precautions for local sanitation, drinking water, diseases and wild animals; ground environment; seasonal variations in climate; clothing and personal equipment; local language and customs; monetary currency used and exchange rate; location of bank and check cashing capability; drivers license requirements; etc.

Emergencies Discuss procedures and facilities available in case of emergencies. Discuss evacuation procedures, location of the nearest US embassy or consulate and who to contact to keep abreast of developing situations. Include the location telephone number of the nearest hospital and pharmacy. Is an english speaking doctor available?

5.4

GPS STATIC SURVEYING: FIELD PROCEDURES

In general, a GPS survey involves a minimum of two field parties (each responsible for one GPS receiver unit) operating *independently* of each other, but in a *coordinated* manner.

5.4.1 EQUIPMENT LISTS

Equipment necessary to carry out a GPS survey can be categorised according to whether it is:

☞ **Equipment per site for field work.**

☞ **Equipment at a base station or the field office.**

Equipment for Instrument Station

The following list may be taken as a guide:

- GPS receiver, antenna and associated cabling.
- External batteries (including spares), battery charger.
- Data storage consumerables, for example memory cards, diskettes, and possibly a P.C. computer for logging or downloading data.
- Antenna tripod, tribrachs or adaptors for mounting antenna on pillars.
- Compass and clinometer for determination of the azimuth and elevation of possible obstructions to satellite signals -- reconnaissance may be performed "on-the-run" during a survey!
- Pocket tape, 30m tape, plumbobs, umbrella and supports, etc.
- Theodolite for eccentric station survey, or sun/star azimuth observations.
- Thermometer, psychrometer and barometer for met observations (if insisted upon).
- Fieldbook(s), maps, access details, observation schedule, skyplots, instructions, etc.
- Useful ancillary equipment: camera, watch, communications equipment.
- Surveyors toolkit: spanners for trig stations, construction of masts (to raise antenna above trees), clearing undergrowth, etc.
- Transport vehicle.

GPS receivers are complex and expensive instruments -- *has everyone read the manual?* Keep the antenna, receiver and cabling together in a "kit". It is vital that the cables be looked after and the connections kept clean *as they must carry the signal to the receiver.*

Equipment at Base Station or Field Office

The following may be taken as a guide:

- Portable computer with suite of software for the downloading, checking, pre-processing, and perhaps baseline processing, of GPS data collected by individual field parties.
- List of station coordinates, topographic maps, observation schedules, recovery/access diagrams, client instructions, useful contact addresses and telephone numbers.
- Data storage consumerables, for example diskettes, for the archiving and storage of tracking data.
- Cables and ancillary equipment for downloading data from GPS receivers.
- Computer modem for transmitting data to head office.
- Communication equipment, for example radios, or at the very least telephone procedures to ensure contact between field parties and head office.
- Spare GPS receiver(s), cables, batteries, and other field equipment that may not be needed by field parties every day such as, for example, theodolites, EDM, etc.
- Transportation.

5.4.2 PARTY ORGANISATION & LOGISTICS

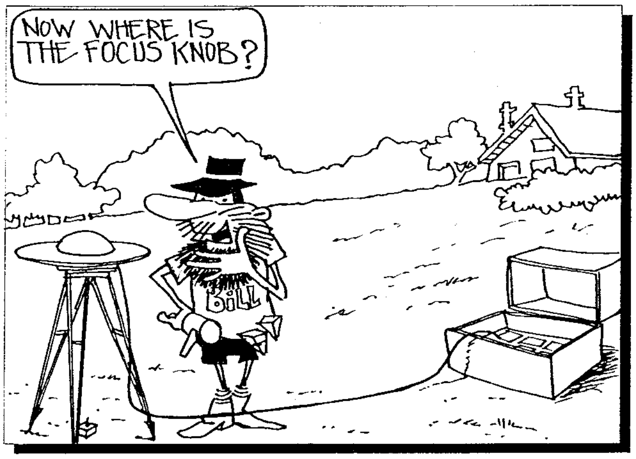
Depending upon the number of field parties, and the complexity of the GPS survey, there may be quite an elaborate field organisation employed. Each **field party** generally comprises of one or two members trained to operate the GPS and ancillary equipment, and well briefed on the observation plan (both in detail for each session: which satellites to track, when to start/stop tracking, etc.; as well as the general schedule of stations to occupy: where other field parties are, the contingency plans, etc.). **The Plan must be flexible, in order to cope with changing and unforeseen circumstances during fieldwork.**

The **campaign manager**, on the other hand, is responsible for carrying out (or at the very least supervising):

- The survey planning, the technical as well as the administrative tasks such as contacting the relevant authorities to gain station access permission; booking accommodation and transport; investigation of facilities such as communications, power supplies (for computers and receivers), etc.
- The pre-mission tasks such as logistical planning; testing and validating equipment; reconnaissance; site preparation; briefing of staff; etc.
- The day-to-day management of the field parties, including preparation of contingency plans in the event of instrument failures, etc.; and implementing changes to the observing schedule.
- The supervision of the field processing of GPS data to check data validity; perform a preliminary network adjustment as the survey progresses; recommend alterations to the schedule in order to reobserve sites, observe new stations, etc.

5.4.3 ON-SITE PROCEDURES

All procedures for the operation of GPS receivers should strictly follow the manufacturer's instructions in the operator's manual. For example, warmup times for the receiver oscillator, operations as "cold" and "warm" starts, minimum power operation, data storage capacity, etc.



Antenna Setup and Height Measurement

The following are some procedures that should be adhered to in this regard:

- The antenna normally bears a *direction indicator* that should be oriented in the same direction at all sites using a compass. This ensures that any antenna centre offset (as measured from the mechanical centre to the electrical phase centre) will propagate into the baseline solution (groundmark to groundmark) in a systematic manner.
- The same antenna, receiver and cabling should be maintained together in a "kit".
- Because of the high precision of GPS surveys, the *centring of the antennas is important.* If centring is poor, the accuracy of the overall survey will suffer hence plumbobs should be avoided. Tribrachs with built-in optical plummets should be regularly calibrated.
- The antenna assembly should therefore be mounted on a standard survey tribrach with an optical plummet, on a good quality survey tripod.
- Setting up on a *pillar* is, of course, reasonably effortless and to be preferred.
- If the receiver is to remain on-site for two or more observing sessions, the antenna should be re-positioned each time.
- Care must be taken with *antenna height measurement.*

As the latter is probably the most critical of all antenna setup operations some further comments must be made. *The height of the antenna above the station marker, measured from the standard reference point on the antenna housing should be measured to the nearest millimetre, and should be done at the beginning and at the end of each session.* As this is a common source of error, the measurement should be checked, for example by independent measurement by another person, or by measurement in imperial units, and cross-checked.

Different antenna types have different recommendations for height measurement (some of the common antenna types are indicated on the attached sample booking sheet at the end of this section). Figure 5.4-1 illustrates two common antenna height measurement procedures for a tripod-mounted antenna. All antenna height measurements must be carefully noted, preferably with a diagram, as indicated on the sample field sheets.

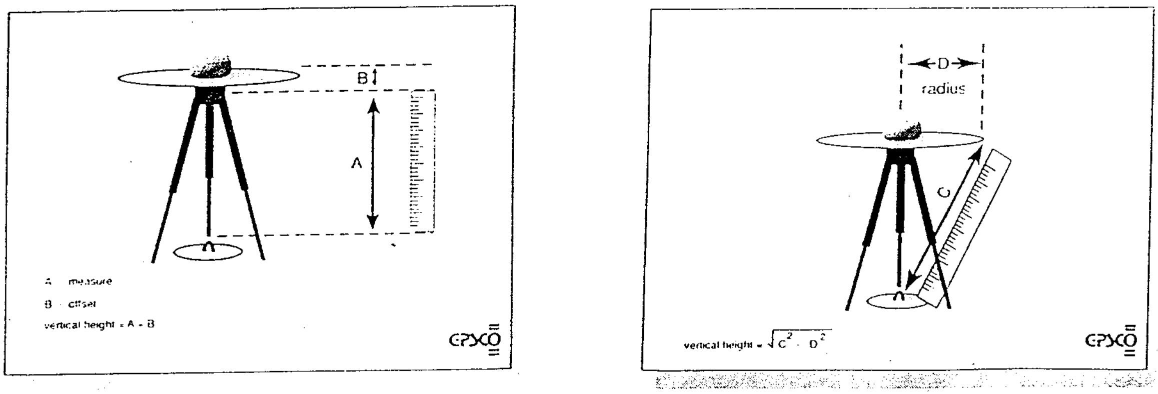


Figure 5.4-1. Measuring the antenna height.

Synchronising Observation Sessions

- ❑ All receivers must track during the same time period (or session) -- *good planning & reconnaissance will help.*
- ❑ Receivers must track the same constellation of satellites -- *use "all-in-view tracking" if possible.*
- ❑ All receivers must record data for the same epochs to within a *few microseconds*, hence:
  - they should start at a well defined instant in time
  - they should track at the same data rate
- ❑ For example, start on the minute and make observations every 15 seconds: 0, 15, 30, 45, 0, ...

> ***How to coordinate data collection?***
>
> **Modern GPS receivers will independently synchronise to GPS Time, and as long as they are programmed for the same data observation rate, then the receivers will be automatically synchronised to the required accuracy.**

Field Log Sheets

A field log should be maintained, in which pertinent information concerning the site being occupied and the data collection process itself is entered. Such a sheet would typically contain the following information:

- Date and time, field crew details, etc.
- Station name and number (including aliases, site codes, etc.).
- Session number, or other campaign indicator.

- Serial numbers of receiver, antenna, data logger, memory card, etc.
- Start and end time of observations (actual and planned).
- Satellites observed during session (actual and planned).
- Antenna height (several measurements), and eccentric station offsets (if used).
- Weather (general remarks), and meteorological observations if requested (such as temperature, pressure and relative humidity).
- Receiver operation parameters such as data recording rate, type of observations being made, elevation mask angle imposed, data format used, etc.
- Any receiver, battery, operator or tracking problems that were noticed.
- Sketch of the site showing all marks, possible obstructions, etc.

Examples of GPS field log sheets are attached. One is a rather elaborate example used for high precision geodetic GPS surveys. The other is an example of a simple one page, general-purpose booking sheet format.

On-Site Procedures: A Checklist

The following is a list of some GPS on-site field procedures:

- GPS receiver initialisation procedures.
- Set-up and orientation of antenna.
- Correct cable connection of antenna to receiver, receiver to battery, etc.
- Double (and triple) checking of centring and antenna height measurement.
- Receiver startup procedure, for example entry of site number, height of antenna, etc.
- Start of tracking.
- Survey of eccentric station.
- Temperature, pressure and humidity measurements (if required).
- Monitoring receiver operation and data recording.
- Field log entries.
- Photographs of point occupancy.
- Procedures at completion of session, for example communication, data transfer to P.C.
- Instructions in event of receiver problems, contingency plans, etc.

GPS receivers display a lot of information during a data measurement session. Devise a field routine to periodically check some of the important indicators ...

Don't forget to note in the field sheets if anything unusual is detected!!

Receiver Operation Monitoring Checklist:

- Battery status
- Memory capacity left
- Satellites being tracked
- Real-time navigation position solution
- Satellite health (also useful for post-processing)
- Date and time (UTC or local)
- Elevation & azimuth of satellites (compare with predictions or skyplot)
- "Signal-to-noise" ratios
- Antenna connection indicator
- Tracking channel status
- Amount of data being logged

What about collecting met data?

FORGET IT!!

Why?

- Commercial software does not accept meteorological data for tropospheric refraction correction.
- Met data cannot be measured to the required precision (particularly that part critical for the wet component of the troposphere).
- Field staff have other more important things to do while data are being collected.

Eccentric Station Survey

Unlike conventional surveys, GPS surveying requires a comparatively clear view of the sky above the elevation cutoff angle (15° to 20°). Sometimes the groundmark that is to be surveyed does not satisfy this condition, perhaps because it is a previously monumented mark at an already existing geodetic control station. In such circumstances an **eccentric station** may be occupied by the GPS antenna. However, *to reduce the baseline components to the required groundmark it is necessary to make certain site measurements.*

Three mutually intervisible marks are identified: the groundmark A, the antenna site B and a third mark C -- *mark C would typically be a "witness mark".* Some typical scenarios are:

- (a) Set up a theodolite at A, and orient it by observing a distant trig station; (b) observe the azimuth A to B (and C); (c) measure the distances A to B (and A to C) with a tape; (d) measure the difference in height A to B (and A to C). (Mark C provides a means of checking these operations.)
- (a) If orientation is not possible (no distant trig stations are visible), then set up a theodolite at A and observe the astronomic azimuth A to C (and A to B); (b) measure the distances A to B (and A to C) with a tape; (c) measure the difference in height A to B (and A to C).

- (a) If orientation is not possible (no distant trig stations are visible), then set up a second GPS receiver/antenna at C and observe the baseline B to C; (b) measure the distances A to B (and A to C) with a tape; (c) measure the difference in height A to B (and A to C) using a theodolite.

5.4.4 FIELD OFFICE PROCEDURES

It cannot be overemphasised that data should be processed as soon as possible after the observation session in order to assure the quality of the survey at an early stage. As a prerequisite therefore, all data should be systematically catalogued and archived between observation sessions (if there is time), or at the end of the working day at the very latest. Many problems can be identified at this stage.

The following are some typical field office procedures:

- Data handling tasks -- *transfer of data from receiver to P.C.*
- Data verification, backup and archiving in field office -- *transmit raw data to head office?*
- Preliminary computation of baselines in field office.
- Preliminary quality control procedures, such as the inspection of repeated baselines, loop closures, and evaluation of (incomplete) minimally constrained network.
- If appropriate software available, minimally constrained network for entire campaign can be built up one session at a time.
- Command and control of survey parties -- *develop contingency plans for repeated observation sessions.*
- Oversee calibration and testing of field equipment.
- Preparation of campaign report, and maintain ensure reporting to head office and/or the client.

Without data safely downloaded from the GPS receiver, the survey work should never be considered complete.

A few hints:

- Download data "asap".
- Follow procedures in the operator's manual.
- Most GPS receivers have many hours of internal memory, so daily download is a reasonable routine.
- Delete files from receiver memory when data download procedure has been verified.
- Download to P.C. harddisk, then to floppy disks, then make backup copies.
- Store backup disks separately.
- Label and write-protect floppy diskettes.
- Be ruthlessly systematic with diskette labelling convention.
- Cross-reference booking sheets to data files.
- Verify data download, for example check number and size of files, and process as quickly as possible.

It is often more economical to reobserve a session than to spend a lot of effort in the office trying to track down problems. This option can only be exercised if the field parties are still in the area and have not moved on to a different area, or have returned to their head office. *This is therefore another good reason for insisting on some field office processing.*

A common problem that may be identified at an early stage is <u>incorrectly measured antenna heights</u>. Confusion often results when antenna heights are extracted from field log sheets, and compared with heights entered into the receiver message file by the field party. (Further confusion can result with proprietary GPS software which have built-in offsets that are automatically applied for their own antennas, and the field and office staff do not fully understand or are aware of them.)

Guarding against the "ultimate fieldwork sins":

- ❑ **<u>Power loss</u> is the most common cause of GPS equipment failure.**

  ☞ *always have back-up power supplies!*

- ❑ **<u>Cable problems</u> are the next most common sources of failure.**

  ☞ *keep them in good condition!*

- ❑ **Incorrect <u>operation</u> of receiver.**

  ☞ *field staff must be trained!*

- ❑ **<u>Antenna height reading error</u> is probably the most common field error affecting GPS survey quality.**

  ☞ *know the antenna phase centre, check and recheck height reading!*

- ❑ **Are you on the <u>correct station</u>?**

  ☞ *good reconnaissance helps, get evidence of occupation!*

- ❑ **Data collection must be coordinated, only <u>common</u> data from a minimum of two sites can be processed.**

  ☞ *good teamwork, well designed observation schedule and well trained field staff!*

- ❑ **<u>Loss of data</u> after survey session ends.**

  ☞ *use ruthlessly systematic data management procedures!*

NASA JPL

GPS Station Occupation Report

Fill Out for Each Site Occupation or Receiver Installation

Site Name:______________ 4-Char ID:__________ Station ID #:__________
Location:______________ City:__________ Country:__________
Observing Monument Inscription:______________ Project:__________

Approximate Geodetic Coordinates

Latitude:__________ Datum:__________ Current Magnetic Declination:__________
Longitude:__________ Source of Position:__________ Rate and Direction of Change:__________
Height:__________ __________ Source of Declination:__________

Site Personnel

Name	Agency

Power Supply

☐ 110 V AC
☐ 220 V AC and Power Converter
☐ Generator
Type__________Watts__________
☐ Batteries
Type__________Number__________

Ancillary Data Collected

☐ Water Vapor Radiometer
☐ Solar Hygrometer
☐ Meteorological Data
(Attach Appropriate Log Forms)

Observing Sessions Summary (Start Dates)

	Local	UTC	UTC Day
Today's Date			
First Observing Day			
Last Observing Day			

Total Number of Observing Days at this Station __________
Local Time Offset from UTC __________ Hours

Equipment Used

	Type	Model	Serial Number
Receiver			
Antenna			
Tribrach			
Tripod			Cable Length

Receiver Software __________ Version __________
Equipment Owned By __________
☐ External Reference Frequency Used
Type ☐ Rubiduim ☐ Hydrogen Maser ☐ Other

Attach Photo or Rubbing of

Observing Monument Here

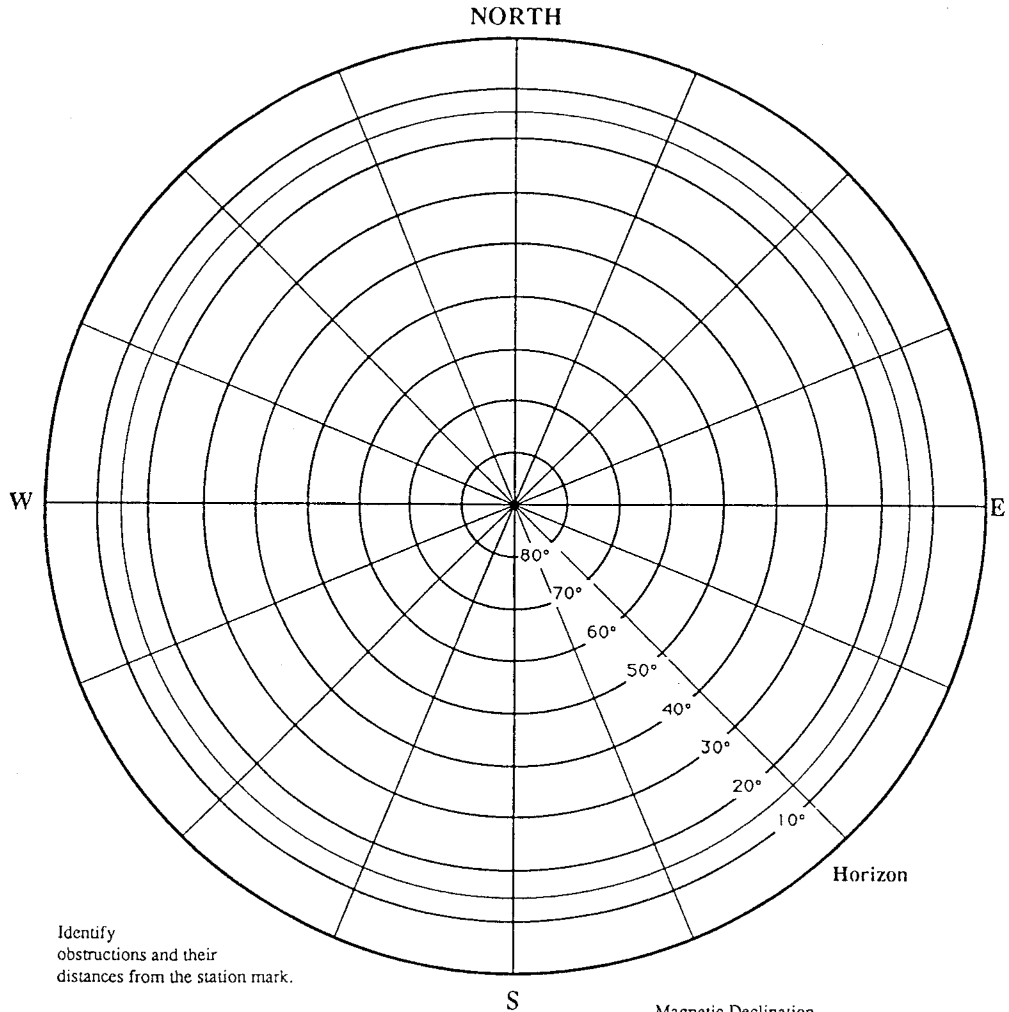
Indicate Exact Point Instrument was Centered Over

Include a North Arrow

Jet Propulsion Laboratory M/S 238-600, 4800 Oak Grove Dr., Pasadena, CA 91109 USA, Tel. 818-354-8330/393-6051, Fax 818-393-4965

NASA JPL

GPS Station Horizon Mask



Identify
obstructions and their
distances from the station mark.

Height Above Marker that Horizon
was Mapped From:__________

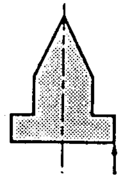
Magnetic Declination__________
Declination Applied to this Figure? ☐ Yes ☐ No

Site Name: __________ Abbreviation: __________ Station ID #: __________
Location: __________ City: __________ Country: __________
Observing Monument Inscription: __________ Your Name: __________

Jet Propulsion Laboratory M/S 238-600, 4800 Oak Grove Dr., Pasadena, CA 91109 USA, Tel. 818-354-8330/393-6051, Fax 818-393-4965

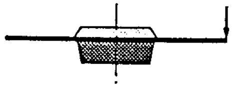
Circle Appropriate Antenna Type Include a sketch if the height was measured differently than shown or if nother antenna type was used. Indicate antenna dimensions and phase center location on sketch if possible.

TI-4100



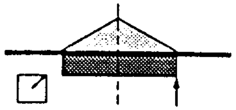
Measure slant height to bottom edge of baseplate.

TRIMBLE 4000 (SDT, STD & SST)



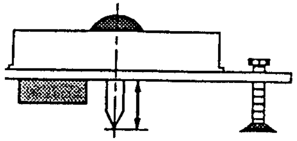
Measure slant height to top of backplane at outer edge or use Trimble measuring rod.

TRIMBLE 4000 (SLD)



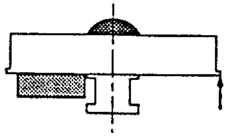
Measure slant height to bottom corner of pre-amp housing.

ROGUE (Spike Mounted)



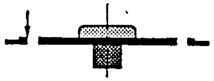
Vertical height to bottom of baseplate is stampped on spike.

ROGUE (Tripod Mounted)



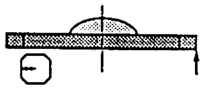
Measure slant height to bottom edge of chokering.

ASHTECH XII



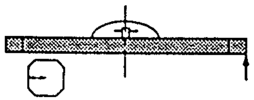
Measure slant height by placing Ashtech measuring rod through hole.

MINI-MAC (Portable)



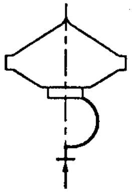
Measure slant height to bottom edge of baseplate at center.

MINI-MAC (Standard)



Measure slant height to bottom edge of baseplate at center.

WM 102



Measure vertical height to bottom of adapter.

Sketch

NASA JPL

GPS Daily Observation Log

Fill Out for Each Observation Session and After Power Failures

Site Name:________ 4-Char ID:________ Station ID #:________
Location:________ City:________ Country:________
Observing Monument Inscription:________ Project:________

	Type	Model	S/N
Receiver			
Antenna			

Receiver Software________ Version________
Collection Rate________ Solution Rate________

Primary Operator________
Agency________

Download Information

☐ Tapes ☐ Disks ☐ Backups Made

	Disk/Tape Label	File Name
1		
2		
3		

Antenna Height Above Mark

	☐ Slant Height Before	☐ Vertical Height After
1		
2		
3		
Average		Meters

☐ Antenna Aligned with North

Comments on Discrepancies________

Timing

	Local Time	Local Date	UTC Time	UTC Date	UTC Day
Scheduled Start Time					
Scheduled End Time					
Actual Start Time					
Actual End Time					

☐ Power Failure - Started Over with New Disks/Tapes and Log Form

Receiver Solution (Record Near End of Session)

UTC Time________
Latitude________
Longitude________
Height________ Meters

Satellites Tracked

(Circle Appropriate PRN Numbers)

Block I SV's 3 6 9 11 12 13
Block II SV's 2 14 15 16 17 18 19 20 21 22

☐ TI-4100 Scenario on Back

Did anything abnormal or unusual occur? ☐ *Yes* ☐ *No* ☐ *Comments on Back of Sheet*

Meteorological Data

☐ Relative Humidity Reported Instead of Wet and Dry Temp. ☐ Continued on Back

☐ Automated Meteorological Package
Serial No.________
UTC Start Time________ End Time________

☐ Conventional Meteorolgical Gear
Psychrometer Model and Serial No.________
Barometer Model and Serial No.________

UTC Time								
Temp. °C (Dry)								
Temp. °C (Wet)								
Pressure (Units______)								
Comments/Visual								

Station________ Project________ UTC Start Date________ UTC Start Day________

Jet Propulsion Laboratory M/S 238-600, 4800 Oak Grove Dr., Pasadena, CA 91109 USA, Tel. 818-354-8330/393-6051, Fax 818-393-4965

FIELD SHEETS

LIÇ BATHURST

JULIAN DAY	SESSION

OPERATOR ____________

RECEIVER No. ____________

ANTENNA No. ____________

SITE CODE ☐ ☐ ☐ ☐

SITE NAME ____________

STAND POINT ____________

HEIGHT OF ANTENNA

A = Heights from tops of marks to top of black adaptor
B = Antanna constants

Check Heights

Measured A: | A:
Constant B: 0.069 | B: 27
HEIGHT: ______ metres | Inches

	TIME	DATE
Start:		/ /
Obs Time:		
Finish:		/ /

LOCAL TIME ZONE:......................
(eg. UT + 10)

CHECKLIST!

Battery - Start ☐☐☐☐☐ Finish ☐☐☐☐☐
Standpoint number physically confirmed ☐
Centring checked ☐
Antenna oriented to North ☐
Antenna connection confirmed ☐
Data File name checked ☐
On site tracking confirmed ☐
Auto timer: time zone confirmed ☐
Auto timer: start time confirmed ☐
Auto timer: session length checked ☐
Obstruction diagram (on reverse) ☐

ATMOSPHERICS

Time	Dry	Wet	Baro

EVENTS DURING SESSION ...

ELEVATION MASK ☐
MIN # OF SVs: ☐
MEAS SYNC TIME: ☐

DATA FORMAT:
Compacted Standard
L1 + L2 L1 only

Time	Event

1/4 = 0.2 1/2 = 0.25 2/6 = 0.37 1/2 = 0.5 1/4 = 0.52 1/2 = 0.75 1/4 = 0.88

5.5

MODERN GPS SURVEYING: FIELD PROCEDURES

5.5.1 INTRODUCTION

Conventional GPS surveying has the following characteristics (§2.3):

☞ The points being coordinated are not moving.

☞ GPS data are collected over some "observation session", typically ranging in length from thirty minutes to several hours, or perhaps days for very precise applications.

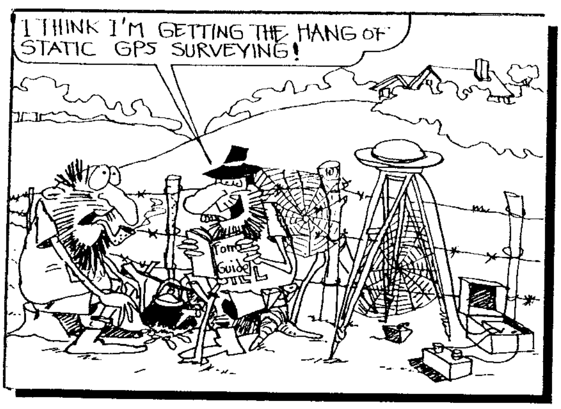
☞ Relative positioning mode of operation is the only mode employed.

☞ The measurements used for data reduction are those made on the transmitted L-band carrier wave, requiring specialised hardware and software.

☞ A variety of processing algorithms can be employed, including triple-difference, double-difference ambiguity-free and ambiguity-fixed solutions.

☞ Mostly associated with the traditional surveying and mapping functions.

During the late 1980's considerable attention was paid to the first two points, as they were considered to be unnecessarily restrictive for precise GPS technology. **That is, if antennas could be moving during a GPS survey, then *new applications for the GPS technology could be addressed.* If the length of time required to collect phase data for a reliable solution could be shortened, then *GPS survey productivity would improve and the technology would be attractive for many more surveying applications.*** It is possible now to distinguish between the two basic modes of GPS surveying: *static* GPS positioning and *kinematic* GPS surveying.



New GPS surveying methods have been developed with the two liberating characteristics of: (a) static antenna setups no longer having to be insisted upon, and (b) long observation sessions no longer essential in order to achieve survey level accuracies. These **modern GPS surveying** techniques are given a variety of names by the different instrument manufacturers, but the following generic terminology will be used in these notes:

- **Rapid static** positioning techniques.
- **Reoccupation** techniques.
- **Stop & go** techniques.
- **Kinematic** positioning techniques.

All require the use of specialised hardware and software, as well as new field procedures. GPS receivers capable of executing these types of surveys can also be used for conventional static GPS surveying. Although the field procedures are different from conventional GPS surveying, the principles of planning, quality control and network processing are the same for both modern and conventional GPS surveying practices.

5.5.2 MODERN GPS SURVEYING TECHNIQUES

Each of the techniques represents a technological solution to the problem of obtaining <u>high productivity</u> (measure as many baselines in as short a period of time as possible) and/or <u>versatility</u> (for example, the ability to obtain results even while the receiver is in motion) without sacrificing very much in terms of *accuracy* and *reliability*. **None of these techniques is as accurate or reliable as conventional static GPS surveying, and each of these techniques has its special strengths and weaknesses.** They represent the state-of-the-art in precision GPS positioning, and are a direct outcome of considerable innovation by instrument manufacturers addressing survey applications. In many cases the most significant advances are in the software, but nevertheless the receiver hardware is of the top-of-the-line variety.

Rapid Static GPS Surveying

> ***Static positioning with short observation times of 5-20 minutes (vs 1-2 hours) ... giving centimetre accuracies!***

Also referred to as fast-static or quick-static. The following characteristics distinguish rapid static techniques from other methods of GPS surveying:

- ❑ **Observation time requirements**: *These are significantly shorter than for conventional GPS surveying, and are a function of baseline length, whether dual-frequency instruments are used, number of satellites being tracked and satellite geometry.* Typically the receivers need only occupy a baseline for a period of 5-20 minutes (the lower value corresponding to baselines <5km and tracking six or more satellites; the upper value for longer baselines up to 20km, and/or where tracking is to only four satellites).

- **Hardware requirements**: Different GPS "products" have different hardware requirements. In some systems only dual-frequency phase measurement is sufficient, in system configurations dual-frequency pseudo-range measurements are also required. To date there is no "mixing" of receivers and software as in the case of conventional GPS surveying. The exact configuration is of course dependent on the software being used. For example, if the software for rapid static is quite sophisticated then there is less reliance on specialised top-of-the-line hardware. (For example, rapid static results have been obtained with single frequency phase data.) However, for other rapid static "products", full four observable instrumentation (L1 & L2 phase, P1 and P2 pseudo-range; or L1 & L2 phase, C/A pseudo-range and P1-P2 observable) is a prerequisite.
- **Specialised software**: The basis of this technique is the ability of the software to *resolve* the ambiguities (determine their integer values) using a very short observation period. There is a variety of software, with different characteristics and levels of sophistication, but the fundamental requirement is a *fast ambiguity resolution* capability.

The field procedures are much like those for conventional static GPS surveying except that the occupation times are shorter, the baselines should be comparatively short, the satellite geometry favourable and signal disturbances such as multipath should be a minimum. Since rapid static surveying is a relatively new technique it is not possible to define exactly how much data needs to be collected in order to produce quality baselines every time, and with high reliability. Instrument manuals typically give guidelines. Some receivers give an audio and/or visual indication when enough data has been collected in the field (but this cannot be confirmed until the data is downloaded and processing is completed). If the real-time mode is employed (the base station transmits tracking data to the mobile unit where it is processed immediately) then the "data quantity gamble" for rapid static GPS surveying can be overcome.

Such a technique is well suited for short range applications such as control densification and engineering surveys, or any job where many points need to be surveyed (Figure 5.5-1). Unlike the "kinematic" and "stop & go" techniques there is no need to maintain lock on the satellites when moving from one station setup to another.

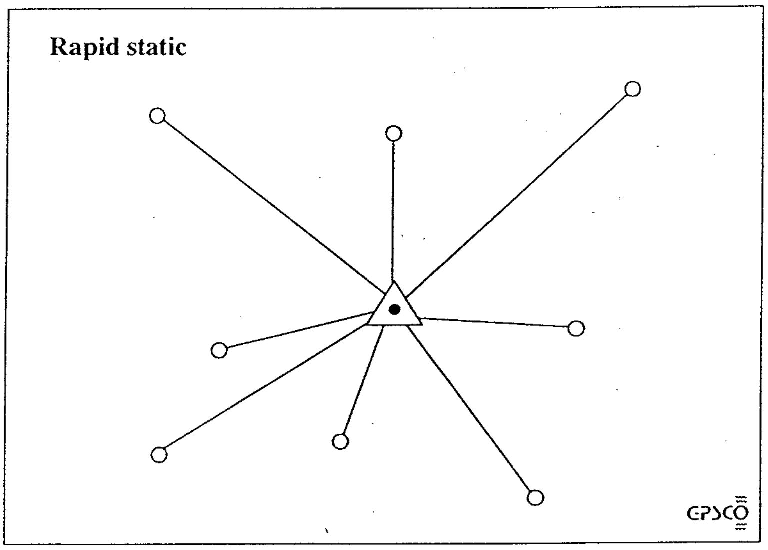


Figure 5.5-1. Field procedure for the "rapid static" surveying technique.

Reoccupation GPS Surveying Techniques

> ***Centimetre positioning accuracy with two occupations per site, each for a short static observation period (few minutes) ...***

Also known variously as pseudo-kinematic and pseudo-static. This technique exploits changes in satellite geometry across conventional observation sessions. Phase data from two short sessions of just a few minutes in length (perhaps up to 10 minutes), collected about one hour apart is sufficient to ensure a good quality ambiguity-free solution.

The field procedure is otherwise similar to the "rapid static" or conventional static techniques. One receiver is located at a known point (the "reference" receiver) while the second "roving" receiver moves from point to point. For example, the roving receiver stops at a site, where it is static for a short period, and then moves on to the next point. *The roving receiver must revisit the same point one or more hours later* (Figure 5.5-2). The second occupation is the same as the first: the receiver stops on the point, is static for 10 or so minutes, and then is off again. (To increase redundancy, the points may be revisited more than twice.) The receiver need not be tracking satellites between the sessions (it can in fact be switched off), however continuous tracking (no cycle slips) should be maintained during the on-site observation period. Furthermore, the satellite geometry should be *favourable*. **Two separate sets of ambiguities must be estimated, one for the first session, the other for the second session**. It is not necessary for the same constellation of satellites to be observed for both sessions.

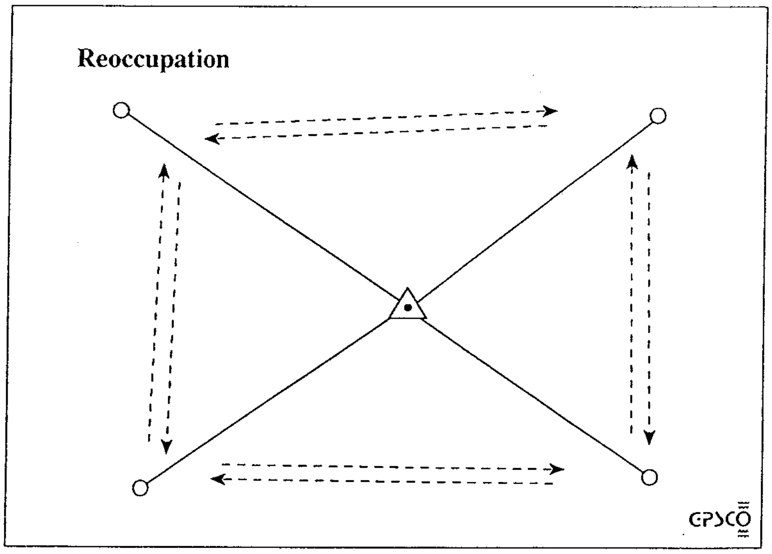


Figure 5.5-2. Field procedure for the "reoccupation" surveying technique.

The following are some of the characteristics of the technique:

- It is the technique that fits somewhere between conventional static and "kinematic" techniques in terms of productivity.

- It is faster than conventional static, but it is not as accurate, if only an ambiguity-free solution is obtained. An ambiguity-fixed solution, if obtained, is more accurate.
- It is an alternative to the "rapid static" technique, no faster but also not as accurate unless ambiguities are resolved (in which case it is identical to the "rapid static" technique).
- It is more flexible than the "stop & go" or "kinematic" techniques as it does not require that satellites be tracked while the receiver is being moved from site to site (Figure 5.5-4).

No special tracking hardware is required other than that used for conventional GPS surveys. However, the receiver must be able to record the data files in a manner that permits the processing software to sort out the pairs of observation sessions for a single site. Appropriate software is necessary for the subsequent data processing.

"Stop & Go" GPS Surveying Techniques

Centimetre accuracy positioning during very short static observation periods (<1minute) ... receiver moves carefully from point to point ...

This is a true kinematic technique because the receiver continues to track satellites while it is in motion. It is known as the "stop & go" (or semi-kinematic) technique because the coordinates of the receiver are only of interest when it is stationary (the "stop" part), but the receiver continues to function while it is being moved (the "go" part) from one stationary setup to the next. There are in fact three stages to the operation:

(1) **The initial ambiguity resolution**: This is carried out (generally in static mode) before the "stop & go" survey commences. The determination of the ambiguities can be carried out using any method, but in general it is one of the following:
   - A conventional static (or "rapid static") GPS survey determines the baseline from a fixed receiver to the first of the uncoordinated sites occupied by the second "roving" receiver. An ambiguity-fixed solution provides the integer values of the ambiguities.
   - Setup both receivers over a known baseline, possibly surveyed previously by GPS.
   - Employ a procedure known as "antenna swap". Two tripods are setup a few metres apart, each with an antenna on them (the exact baseline length need not be known). Data is collected by each receiver for a few minutes (tracking the same satellites). The antennas are then carefully lifted from the tripods and swapped, that is, the receiver 1 antenna is placed where the receiver 2 antenna had been, and visa versa (Figure 5.5-3). After a few more minutes the antennas are swapped again.
   - The most versatile, and most recent, technique is to resolve the ambiguities "on-the-fly" (that is, while the receiver is turned on but the receiver/antenna is moving).

(2) **The receiver in motion**: Once the ambiguities have been determined the survey can begin. The roving receiver is moved from site to site, collecting just a few minutes of phase data. *It is very important that the antenna continues to track the satellites.* In this way the resolved ambiguities are valid for all future phase observations, in effect converting the ambiguous carrier phase data to unambiguous "carrier-range" or "phase-range" data (by applying the integer ambiguities as data corrections). As soon as the signals are disrupted (causing a cycle slip) then the ambiguities have to be reinitialised (or recomputed). This can most easily be done by bringing the receiver back to the last

surveyed point, and redetermining the ambiguities by the "known baseline" method.

(3) **The stationary receiver**: The "carrier-range" data is then processed in the double-differenced mode to determine the coordinates of the roving receiver relative to the static reference receiver. *The trajectory of the antenna is not of interest, only the stationary points which are visited by the receiver.*

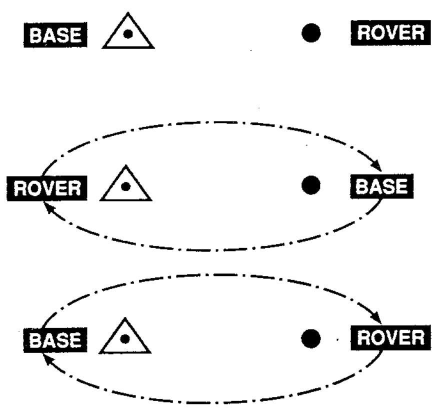


Figure 5.5-3. The antenna swap procedure for initialising ambiguities.

The technique is well suited when many points close together have to be surveyed, and the terrain poses no significant problems in terms of signal disruption (usually an audible signal is emitted by the receiver when it has lost lock on the satellites). The survey is carried out in the manner illustrated in Figure 5.5-4, and the ambiguities reinitialised using any of the techniques shown in Figure 5.5-6.

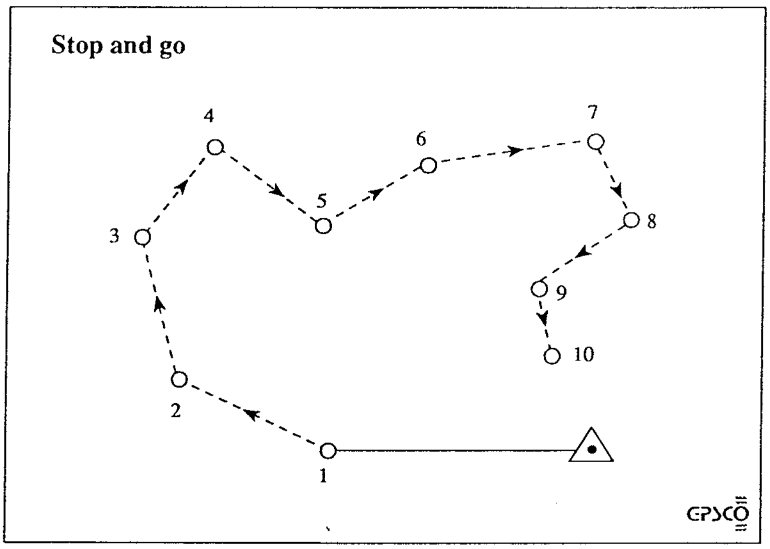


Figure 5.5-4. Field procedure for the "stop & go" surveying technique.

One particularly *negative* characteristics of this technique is the requirement that <u>phase lock must be maintained</u> by the roving receiver as it moves from site to site. This requires special hardware mounts on vehicles if the survey is carried out over a large area.

An additional requirement is that the stationary reference receiver must continue to track all the satellites being tracked by the roving receiver. The accuracy attainable is about the same as for the "rapid static" technique. As with the "reoccupation" technique, the receiver must have the ability to handle data files from several different sites. The software then has to sort out the recorded data for the different sites, and to differentiate the "kinematic" or "go" data (not of interest) from the "static" or "stop" data (of interest). It can be implemented in real-time if a communications link is provided to transmit the "carrier-range" data from the reference receiver to the roving receiver(s).

Kinematic GPS Surveying Techniques

Centimetre positioning accuracy of moving antenna ...

This is a generalisation of the "stop & go" technique. Instead of coordinating stationary points and disregarding the trajectory of the roving antenna as it moves from site to site, the intention of "kinematic" surveying is to determine the position of the antenna only while it is in motion (Figure 5.5-5). In many other respects the technique is similar to "stop & go". **That is, the ambiguities must be resolved before starting the survey, and the ambiguities must be reinitialised during the survey when a cycle slip occurs.** However, for many applications, such as the positioning of an aircraft (for example, for photogrammetric applications) or a ship (for example, a dredging operation), it is impractical to reinitialise the ambiguities if the "roving" antenna has to return to a stationary control point (as in Figure 5.5-6). **Hence much R&D effort has been invested in initially determining (and redetermining after a cycle slip) the ambiguities "on-the-fly".** Today the "kinematic" GPS surveying technique is undergoing tremendous improvement and "on-the-fly" ambiguity resolution is a routine procedure (though not yet by any means an entirely foolproof one!), making kinematic surveying techniques ideal for road centreline surveys, hydrographic surveys, airborne applications, etc.

There are a number of trends in "kinematic" surveying that are worth noting:

- There is a blurring of the distinction between "kinematic GPS surveying" and "kinematic GPS navigation". The former is carrier phase based (actually "carrier-range" data), whereas the latter has usually been taken to refer to pseudo-range based positioning. However, nowadays more navigation instruments are using "carrier phase smoothed pseudo-ranges" (§6.4). However, it is still valid to distinguish these techniques as "decimetre accuracy positioning" on the one hand, and "submetre accuracy positioning" on the other hand.
- There is trend to using a combination of both phase and pseudo-range data within the positioning algorithm itself, precise C/A code ranges as well as P code pseudo-ranges.
- There is an increasing sophistication of the algorithms, for example, incorporating Kalman filters.
- There are techniques based on single frequency data, as well as those top-of-the-line procedures requiring dual-frequency data.
- "On-the-fly" ambiguity resolution techniques will probably supersede all other "kinematic" techniques (and possibly the "rapid static" and "reoccupation" techniques as

well).

- Real-time operation is still a challenge, but offers considerable advantages in that the results are available immediately, in the field.

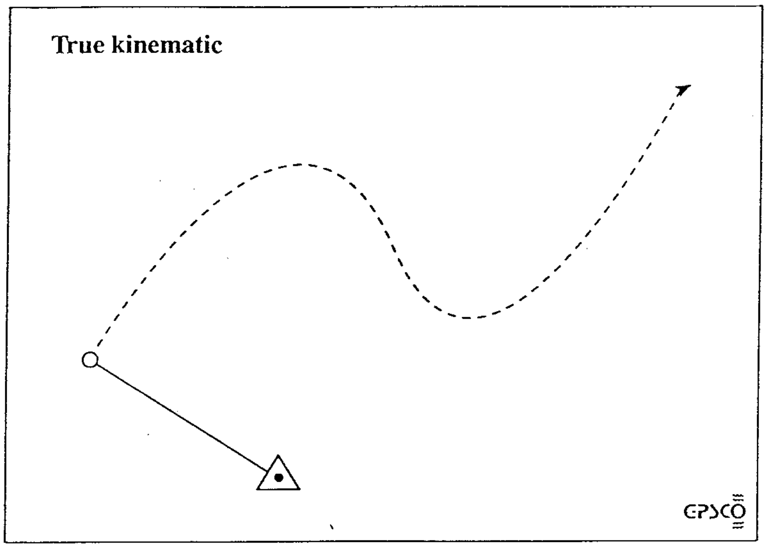


Figure 5.5-5. The "kinematic" GPS surveying technique.

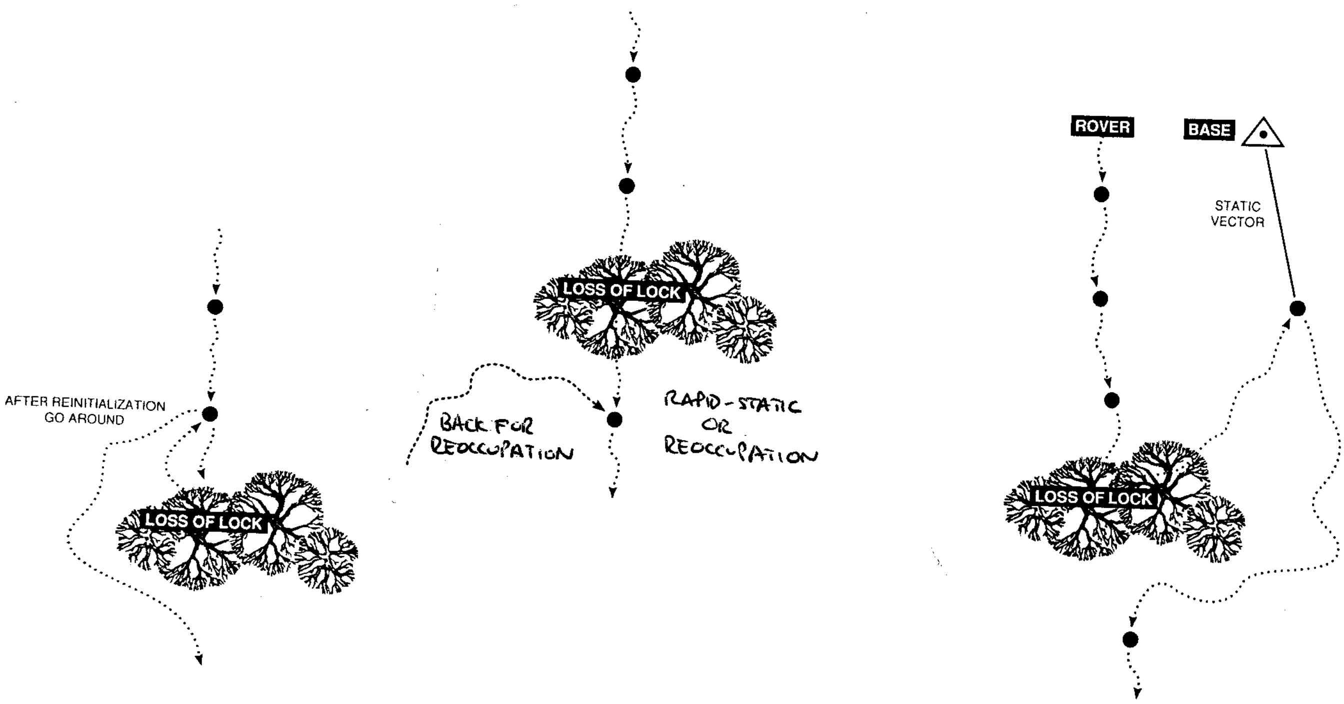


Figure 5.5-6. A variety of reinitialisation techniques for "stop & go" or "kinematic" surveys.

5.5.3 ISSUES RELATING TO MODERN GPS SURVEYING

Each of these modern GPS surveying techniques has its strengths and weaknesses (see below), however all are less accurate than the conventional GPS surveying technique. This should not be too great a drawback as it is often not necessary that relative accuracies of a few parts per million be insisted upon. Often a combination of conventional static and the other techniques makes for an ideal solution to a surveying problem.

Conventional versus Modern GPS Surveying Techniques

CONVENTIONAL (STATIC) GPS SURVEYING:

Advantages	*Disadvantages*
• Highest accuracy • Robust technique • Ambiguity resolution not critical • Minor impact of orbit error and multipath • Undemanding of hardware and software	• Long observation sessions • Inappropriate for engineering and cadastral applications

MODERN GPS SURVEYING:

Advantages	*Disadvantages*
• Higher accuracy than pseudo-range solutions • Appropriate for many survey applications • High productivity • Similar procedures to modern terrestrial surveying	• Special hardware and software • Susceptible to orbit, atmospheric and multipath disturbances • Higher capital costs • Ambiguity-fixed or continuous lock required

Two negative characteristics of these modern GPS surveying techniques are:

- They are susceptible to multipath disturbance to an even greater extent (affecting the receiver signals during *both* the kinematic and static stages of the tracking) than the conventional static technique. Multipath during the ambiguity resolution period is especially dangerous, as wrong ambiguities may result.
- The results from short observation sessions are more sensitive to bad satellite geometry (large GDOP) than the conventional static technique.

An example of the possible combination of conventional static GPS surveying and modern GPS techniques is illustrated in Figure 5.5-7. In this case conventional GPS surveying provides the control for lower order densification or topographic mapping surveys.

Applications and Productivity

There is no doubt that the main attraction of "rapid static", "reoccupation" and "stop & go" techniques is the *decrease in field time required to collect the data.* Hence they are well suited for surveys where 10 parts per million (or lower) accuracy is adequate, and the speed of survey makes GPS an attractive alternative to EDM/theodolite, and other such techniques. The increased productivity of modern techniques compared with the conventional GPS surveying technique is illustrated by an example in Table 5.5-1.

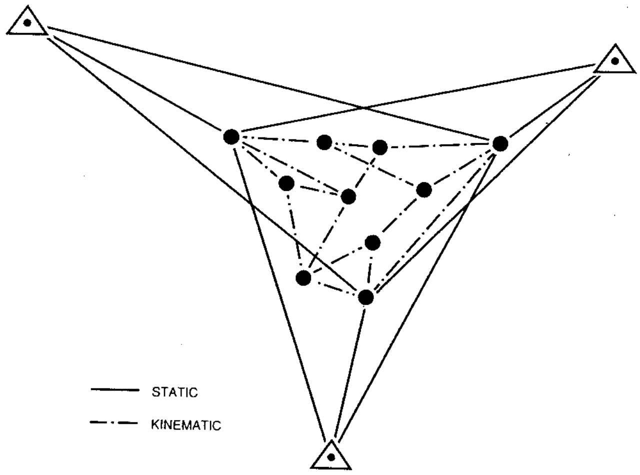
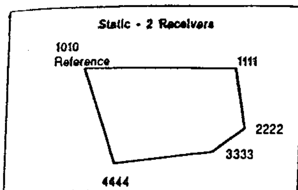


Figure 5.5-7. Combining conventional GPS and modern GPS survey techniques.

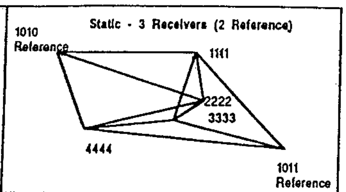
Table 5.5-1. A comparison of survey productivity between conventional static GPS surveying and the "stop & go" technique.

Static - 2 Receivers



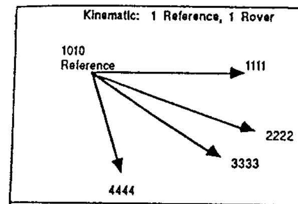
Observation 1010 - 1111	45 min
Move from 1010 to 2222 (setup)	13 min
Observation 1111 - 2222	45 min
Move from 1111 to 3333 (setup)	16 min
Observation 2222 to 3333	45 min
Move from 2222 to 4444 (setup)	14 min
Observation 3333 - 4444	45 min
Move from 3333 to 1010 (setup)	16 min
Observation 1010 - 4444	45 min
	284 min
Total time: 4 h 44 minutes	

Static - 3 Receivers (2 Reference)



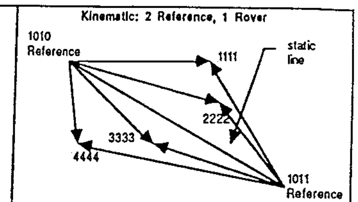
Observation 1010, 1111, 1011	45 min
Move time and setup	23 min
Observation 1111, 2222, 3333	45 min
Move time and setup	16 min
Observation 4444, 3333, 1011	45 min
Move time and setup	22 min
Observation 1010, 4444, 2222	45 min
	241 min
Total time: 4 h 01 minutes	

Kinematic: 1 Reference, 1 Rover



Initialize survey (antenna swap @ station 1010: 4 satellites)	13 min
Move to 1111 (setup)	11 min
Observation 1010 - 1111	1 min
Move to 2222 (setup)	12 min
Observation 1010 - 2222	1 min
Move to 3333 (setup)	11 min
Observation 1010 - 3333	1 min
Move to 4444 (setup)	9 min
Observation 1010 - 4444	1 min
Move to 1111 (setup)	18 min
Reobserve 1111 (as a check)	1 min
	79 min

Kinematic: 2 Reference, 1 Rover



This method takes less time than the first configuration:

79 minutes - 19 minutes (last reobservation of station 1111) =

68 minutes

Checks: Repeatability. Allows for Free Least Square Adjustment

There are a variety of combinations of several modern GPS surveying techniques which are possible. The greatest challenge likely to be faced by the GPS surveyor is to use the best combination of techniques for the terrain and logistical constraints that he/she faces. These are illustrated in Figure 5.5-8. In reality, the constraints are likely to include those in Figure 5.5-6 as well, that is, many trees which can "break" the signal reception.

The development of "on-the-fly" ambiguity resolution algorithms is a dramatic step forward because static ambiguity reinitialisation (as in Figure 5.5-6) is no longer necessary. The ambiguities will be resolved *while the antenna is moving* to the next stationary survey point. If a point X has been surveyed (that is, a few minutes of "carrier-range" tracking data has been collected), and as the antenna is moved from point X to point Y, an obstruction blocks the signals and causes cycle slips to occur, then the antenna does not have to go back to point X (nor need any of the procedures illustrated in Figure 5.5-6 be used). New ambiguities can be resolved "on-the-fly" as the antenna moves from X to Y. However, there must be a sufficient period of uninterrupted tracking for this to take place. Although this varies from receiver to receiver (and is influenced by the baseline length, satellite geometry, and several other factors), it may be of the order of several minutes. In an extremely unfavourable scenario, there may be so many signal obstructions that there is insufficient time for the O-T-F algorithm to work properly during the very short periods of uninterrupted tracking, and hence the survey is not possible using the "stop & go" technique, even when aided by an O-T-F capability.

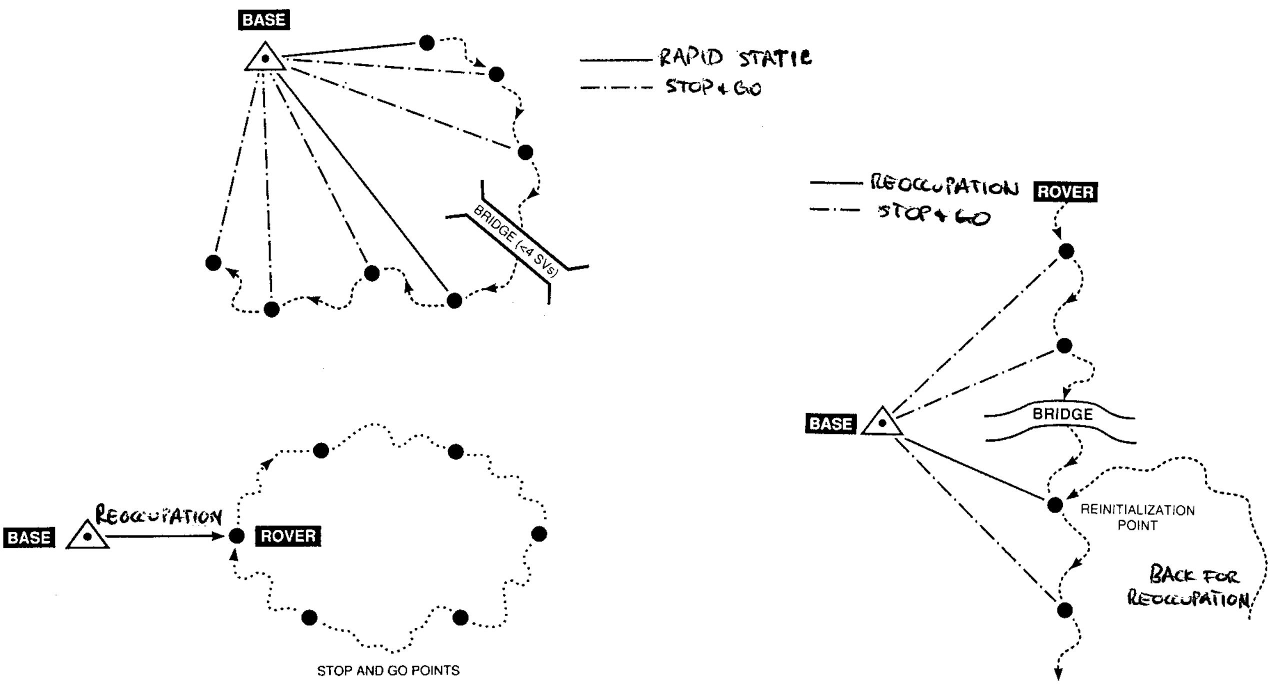


Figure 5.5-8. Examples of combinations of several GPS surveying techniques.

Planning and QC Considerations

Being able to determine baselines faster than using conventional GPS surveying does not, in its self, mean that the network planning and design guidelines discussed in §5.2 must be changed. However the following issues should nevertheless be addressed:

- As the accuracies attainable are lower than for conventional GPS surveying, the GPS survey "standards & specifications" may be relaxed.
- The high speed of survey would suggest that the most appropriate mode of receiver deployment is the "base station" or "radiation" mode (Figure 5.2-7). However, this provides no redundancy because every point is fixed by a "no check" vector.

- It is possible to ensure redundancy by deploying two base stations (Figure 5.5-9). Each roving receiver point is connected by two vectors (useful if one base receiver malfunctions!). But because the rover site is occupied only once (except when the "reoccupation" technique is being used), then the vectors are still of the "no check" variety as there is no way of knowing if the height of antenna has been measured correctly, or even if the correct station was occupied!
- Productivity improves as more GPS receivers are deployed, but the logistics also become more complex. An example of a "hybrid" scenario involving two base receivers, and two roving receivers is illustrated in Figure 5.5-9.
- It has been found that even though the resolved ambiguities (for example, using OTF techniques) are NOT correct, the relative positions between the surveyed points may be *correct* (though the position relative to the base receiver is INCORRECT).
- Because the modern GPS surveying techniques are likely to be used for land applications which were not addressed by conventional static GPS surveying (for example, cadastral surveys), new recommended specifications may have to be developed for fixed control placement, redundancy, ties to control, calibration, heighting procedures, and other network considerations.

REMEMBER: The network adjustment issue is the same whether the baselines were observed with 30 second data sessions, 3 minute data sessions, or 60 minute data sessions.

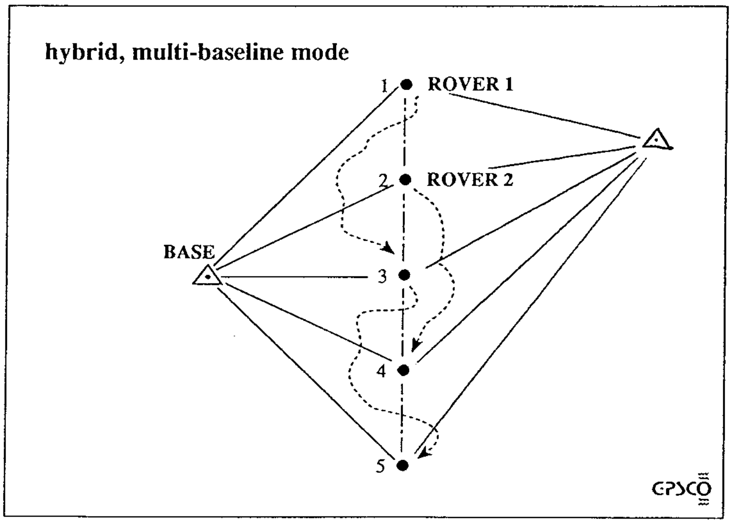


Figure 5.5-9. A multi-receiver deployment scenario for modern GPS surveying.

Standards & Specifications

This is an unenviable task. The speed with which GPS technology is developing is such that these new GPS surveying techniques were not accessible to the average GPS surveyor when Australia and the U.S. were developing their "standards & specifications" for GPS surveys. New Australian specifications have been developed (see §10.2), but some states have already produced "guidelines" for cadastral surveys using "high productivity" GPS techniques.

Future Trends

The following as some likely trends relevant to modern GPS surveying practice (§12.3):

- Increasing use of combined pseudo-range and phase data in a sophisticated Kalman filter, and other similar algorithms, able to perform "on-the-fly" ambiguity resolution -- *a "super" kinematic technique.*
- Replacement of "rapid static" and "reoccupation" techniques by the "super" kinematic technique referred to above, able to operate in both "static" and "kinematic" positioning mode without the need for clumsy static initialisation procedures.
- Significant receiver development to overcome the effects of AS (see §4.2).
- Increasing importance of real-time positioning -- *taking the "gamble" out of O-T-F ambiguity resolution.*
- Increasing applications for GPS in cadastral, engineering, topographical and detail surveying.

Chapter 6: Modelling GPS Observations

6.1 GPS OBSERVATION EQUATIONS

6.1.1 CARRIER BEAT PHASE

The GPS satellites transmit on two L-band frequencies: L1 at 1575.42MHz and L2 at 1227.60MHz. The observation equation for the carrier beat phase is developed below, and is valid for measurements made on either the L1 or L2 frequency. The following development is principally taken from GRANT et al (1990). The reader is also referred to TIBERIUS & DE JONGE (1995) for a condensed treatment of the derivation of the GPS carrier phase observation equation.

Clock Phase Error

The phase of an oscillator at time T can be represented by $\theta(T)$. The behaviour of phase with time in general obeys the relation:

$$\theta(T) = \theta(T_0) + \int_{T_0}^{T} f(T)\, dT \qquad (6.1\text{-}1)$$

where $f(T)$ is the time dependent frequency of the oscillator and T_0 is the time at some arbitrary reference epoch. In developing the observation equations for carrier beat phase it is useful to assume that every clock, or oscillator, can be compared directly with a "perfect" oscillator having a known and constant frequency f_0 in the reference time scale (for example, GPS Time). Because the frequency of the "perfect" oscillator is constant at f_0 the relation becomes:

$$\theta(T) = \theta(T_0) + f_0 \,.\, (T - T_0) \qquad (6.1\text{-}2)$$

The oscillators in the GPS satellites and receivers are used to generate timed signals such as the P code, the C/A code and the observation time-tags. It is common therefore, to think of them as clocks and to consider the errors caused by the variation of frequency as *clock errors*. This use of the term "clock error" can cause difficulties when it is necessary to distinguish between the carrier phase error caused by variations in the receiver oscillator frequency, and errors in the observation time-tag which are also affected by these frequency variations. It is preferable to use the terms **clock phase error** (or clock range error in the case of pseudo-range measurements) for the former and **time-tag error** for the latter (RIZOS & GRANT, 1990). Although the carrier phase and observation time-tags are generated by the same oscillator they are generally treated as if they were independent in GPS processing. For instance, the time-tags are often corrected according to the clock error estimates obtained from a pseudo-range point position solution. These corrections are not generally accompanied by an equivalent correction to the observed carrier phase.

Using the symbol $\phi(T)$ to represent the phase of a real (though imperfect) GPS satellite or receiver oscillator at time T the clock phase error $\varepsilon(T)$ can be defined in terms of the phase error $(\phi(T) - \theta(T))$:

$$\varepsilon(T) = \frac{1}{f_0}.(\phi(T) - \theta(T)) \tag{6.1-3}$$

or

$$\phi(T) = \theta(T) + f_0.\varepsilon(T) \tag{6.1-4}$$

Of course, in using eqn (6.1-3) or eqn (6.1-4) account must be taken of both the *integral* and *fractional* components of the phase observation. This point is important later when considering how phase is actually measured and modelled.

Signal Transit Time

Consider the transmission of a signal from a satellite i to a receiver j. The **time-of-reception** of a signal at receiver j is T_j. The **time-of-transmission** of that signal from satellite i is T_j^i. The **transit time** of the signal from i to j is $\tau_j^i(T_j)$ and is defined by:

$$\tau_j^i(T_j) = T_j - T_j^i \tag{6.1-5}$$

The beat phase formed as an observation in a GPS receiver is the difference between the phase of the local receiver oscillator and the phase of the received signal. *(This is the sense that is now adopted for the beat observation, earlier instrumentation had this difference reversed -- see, for example, GRANT et al (1990).)* The phase of the received signal at the time-of-reception T_j is equal to the phase of the transmitted signal at the time-of-transmission T_j^i. Therefore:

$$\begin{aligned}\phi_{bj}^i(T_j) &= \phi_{loj}(T_j) - \phi_{rj}^i(T_j)\\ &= \phi_{loj}(T_j) - \phi^{ti}(T_j^i)\end{aligned} \tag{6.1-6}$$

where :
- $\phi_{bj}^i(T_j)$ is the carrier beat phase for receiver j, satellite i, at reception time T_j,
- $\phi_{rj}^i(T_j)$ is the received signal phase from satellite i at receiver j at time T_j,
- $\phi_{loj}(T_j)$ is the local oscillator phase of receiver j at time T_j, and
- $\phi^{ti}(T_j^i)$ is the transmitted signal phase from satellite i at transmission time T_j^i.

From eqn (6.1-4) the expression relating phase and clock phase error of the local oscillator can be written as:

$$\phi_{loj}(T_j) = \theta(T_j) + f_0.\varepsilon_{rcj}(T_j) \tag{6.1-7}$$

where $\varepsilon_{rcj}(T_j)$ is the clock phase error of GPS receiver j. Using eqns (6.1-2), (6.1-4) and (6.1-5), an expression relating transmitted phase and satellite clock phase error, at reception time, can be formed:

$$\phi^{ti}(T_j^i) = \theta(T_j^i) + f_0.\varepsilon^{sci}(T_j^i)$$

$$= \theta(T_j) + f_0.\varepsilon^{sci}(T_j^i) - f_0.\tau_j^i(T_j) \qquad (6.1\text{-}8)$$

where $\varepsilon^{sci}(T)$ is the clock phase error attributable to the oscillator of GPS satellite i. Combining eqns (6.1-6), (6.1-7) and (6.1-8), leads to an expression for **carrier beat phase**:

$$\phi_{bj}{}^i(T_j) = f_0.[\ \tau_j^i(T_j) - \varepsilon^{sci}(T_j^i) + \varepsilon_{rcj}(T_j)\ ] \qquad (6.1\text{-}9)$$

The transit time component in eqn (6.1-9) is made up of two parts. The main part is derived from the **geometric range** $\rho_j{}^i$ between the satellite at the time-of-transmission and the receiver at the time-of-reception. This can be determined from the position vectors of the satellite and the receiver, when expressed in the same reference system. Given the speed of the signals (c, the speed of electromagnetic radiation in a vacuum) it is possible to calculate the time taken for the signal to travel this distance ($\tau_j{}^i$). The value used for c is the speed in a vacuum and hence this is the primary means of defining metric scale in GPS. It is the satellite and receiver position information contained in the geometric range that is important for GPS positioning.

The second part of the transit time term accounts for the extra time taken for the signal to travel through the earth's **atmosphere**. This is caused by a change in speed but can also be modelled as a time delay, a phase delay or as an increase in range. This can be incorporated as a *phase correction term* ϕ_{atmos}. (This may be further broken down into components for the different parts of the atmospheric delay; ϕ_{ion} for the ionosphere, ϕ_{trop} for the troposphere -- see §6.2.) Hence:

$$f_0.\tau_j^i(T_j) = (\ f_0 / c\ ).\rho_j^i(T_j) + \phi_{atmos} \qquad (6.1\text{-}10)$$

Combining eqns (6.1-9) and (6.1-10) a model for carrier beat phase in units of cycles can be developed:

$$\phi_{bj}{}^i(T_j) = (\ f_0 / c\ ).\rho_j^i(T_j) + f_0.[\ \varepsilon_{rcj}(T_j) - \varepsilon^{sci}(T_j^i)\ ] + \phi_{atmos} \qquad (6.1\text{-}11a)$$

and in metric units:

$$\Phi_{bj}{}^i(T_j) = \rho_j^i(T_j) + c.[\ \varepsilon_{rcj}(T_j) - \varepsilon^{sci}(T_j^i)\ ] + \lambda.\phi_{atmos} \qquad (6.1\text{-}11b)$$

where λ is the wavelength of the carrier wave ($\lambda = c\ /\ f_0$). The range term and its relationship to the two clock errors (receiver and satellite clocks) is illustrated in Figure 6.1-1 (note: the notation used is: $dT = \varepsilon_{rcj}$, $dt = \varepsilon^{sci}$, and $p = \Phi_{bj}{}^i$).

Measured Carrier Beat Phase

The measured carrier beat phase differs from the above modelled carrier beat phase in a number of important respects. *Firstly*, the carrier beat phase measurements will have random noise (and signal interference such as multipath) associated with the measurement process. This is contained in an additional term ϕ_{noise}.

Secondly, the definition of clock phase error in eqn (6.1-4) depends on the integral and fractional part of the carrier phase. When the carrier beat phase is measured in the receiver the measurement is ambiguous with regards to the number of integer cycles. This integer is set in a "counter" to some arbitrary value when the satellite signal is first acquired (for example, zero). There is therefore an unknown integer number of cycles n_j^i difference between the measured carrier beat phase and the model beat phase in eqn (6.1-11). This will be unique to a particular satellite-receiver pair. Furthermore, it will be a constant for as long as the receiver continues to track and count the integer number of cycles from the time the satellite signal is first acquired (see Figure 6.1-2). At any epoch other than the initial measurement epoch, the instrument measures the fractional phase $\phi_{fj}^i(T_j)$ and, in addition, takes a reading $C_R(T_j)$ on the counter. This combined fractional and integer phase observation is therefore referred to as **integrated carrier beat phase**, and n_j^i is the **cycle ambiguity**.

Thirdly, if the receiver at any time fails to track the signal correctly, and the ambiguity changes between observation epochs, then a **cycle slip** $S(T_j)$ has occurred.

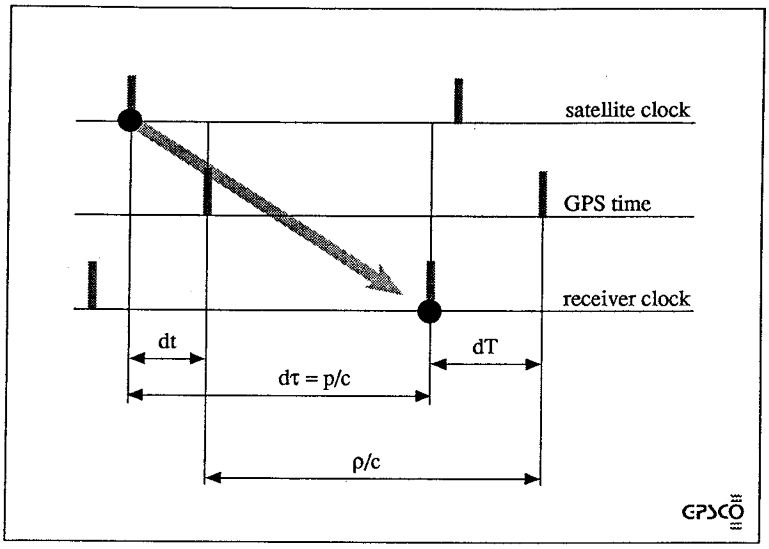


Figure 6.1-1. Geometric range and clock errors.

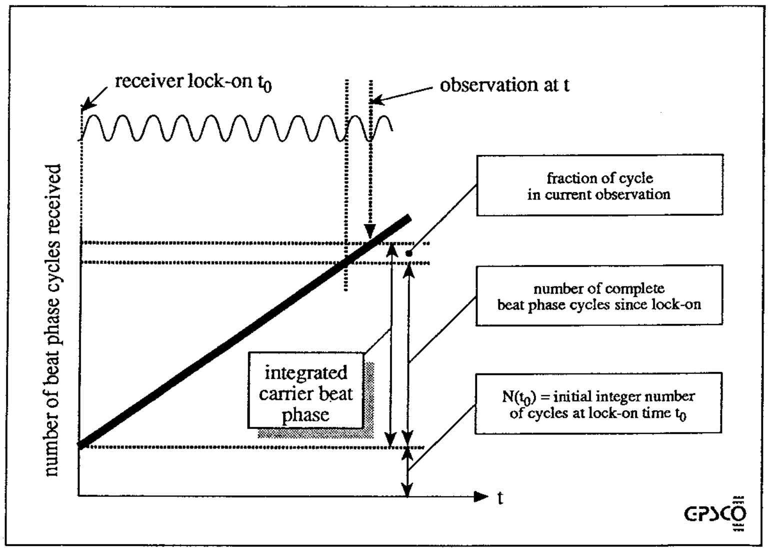


Figure 6.1-2. Integrated carrier beat phase measurement.

The model for the measured carrier beat phase therefore includes these three additional terms:

$$\begin{aligned}\phi_{bj}{}^{i}(T_j) &= \phi_{fj}{}^{i}(T_j) + C_R(T_j) + S(T_j) \\ &= (f_0 / c).\rho_j{}^{i}(T_j) + f_0.[\varepsilon_{rcj}(T_j) - \varepsilon^{sci}(T_j{}^{i})] \\ &\qquad + n_j{}^{i} + \phi_{atmos} + \phi_{noise}\end{aligned} \tag{6.1-12}$$

This is the basic observation equation for integrated carrier beat phase. The dependence of ϕ_{atmos} and ϕ_{noise} on the receiver, satellite or epoch, are not explicitly shown, however these terms will vary slightly for every observation. The change in observed carrier beat phase with time is therefore equal to the change in any one of the following quantities:

- the satellite-receiver geometric range,
- the satellite-receiver clock phase error difference,
- the number of cycle slips,
- the tropospheric delay,
- the ionospheric delay, or
- the measurement noise and signal disturbance.

Mathematical Model for Integrated Carrier Beat Phase

Eqn (6.1-12) is an example of a *physical* model of the integrated carrier beat phase observation. In order to develop an appropriate parameter model for carrier phase data processing, a *mathematical* model needs to be developed. The mathematical model incorporates only those terms of the physical observation model that will be parameterised for the adjustment, but without explicitly defining the functional models for these terms. Eqn (6.1-12) can therefore be written in the standard form:

$$\phi_{bj}{}^{i}(T_j) = (f_0 / c).\rho_j{}^{i}(T_j) + f_0.[\varepsilon_{rcj}(T_j) - \varepsilon^{sci}(T_j{}^{i})] + n_j{}^{i} + \phi_{atmos} \tag{6.1-13a}$$

or in terms of metric units:

$$\Phi_{bj}{}^{i}(T_j) = \rho_j{}^{i}(T_j) + c.[\varepsilon_{rcj}(T_j) - \varepsilon^{sci}(T_j{}^{i})] + \lambda.n_j{}^{i} + \lambda.\phi_{atmos} \tag{6.1-13b}$$

Note that the fractional and integer part of the observation are now combined into a single term $\phi_{bj}{}^{i}(T_j)$, and that the noise and cycle slip terms have been dropped. (Cycle slips are assumed to be independently determined during a "pre-processing" step, and hence are absent from the phase data as modelled above -- see §7.3.)

Eqn (6.1-13) is valid for either L1 or L2 carrier phase observations, however, there are several frequency dependent biases: the ionospheric delay and the ambiguity term (§6.2). All other terms are identical.

The mathematical model for pseudo-range, or "code-phase" as it is sometimes referred to, is also eqn (6.1-13), but without the cycle ambiguity term. However the sign of the ionospheric delay is reversed (see §6.4).

The observation model for Doppler data can be derived from eqn (6.1-13) by taking the time derivative of all the quantities. In that case, the constant ambiguity term is eliminated and other biases such as those due to the atmosphere can also be considered as being relatively insignificant. The rate of change of clock error, however, may still be significant enough to warrant additional parameterisation in the observation equation.

6.1.2 THE NATURE OF GPS OBSERVATION MODEL BIASES

The basic GPS observables of carrier beat phase and pseudo-range are essentially **biased ranges** (§1.3). The ideal quantity required for GPS data processing is the true range ρ_j^i. The true range contains all the geometric information necessary to determine the receiver coordinates and/or the satellite position. One of the challenges in GPS processing is therefore to develop data analysis strategies that best handle the measurement biases. These biases arise from a number of sources (§6.2):

- clock errors in both the receiver clock (ε_{rcj}) and satellite clock (ε^{sci}),
- the atmospheric refraction delay (ϕ_{atmos}), and
- cycle ambiguities (n_j^i) in the case of carrier phase observations.

Upon adopting eqn (6.1-13) as the basic measurement model, and proceeding with the normal linearisation process required for a Least Squares adjustment, a further class of biases is introduced into the residual quantity: observed range (phase or code) minus modelled or calculated (biased) range. This "O-C" quantity now also includes errors in the apriori information contained in the modelled range, primarily the errors in the assumed known **receiver coordinates** and **satellite ephemeris** information (used to derive the "calculated" value for the quantity ρ_j^i). There are two options, these additional biases can be:

(a) explicitly included in eqn (6.1-13) in the form of corrections to the apriori information (receiver and/or satellite coordinates) in the event of these parameters being adjusted, or

(b) neglected (for example when no receiver or satellite coordinate adjustment is performed), in which case these biases are now simply unmodelled or residual biases.

A significant challenge in GPS processing is therefore to develop strategies that either reliably estimate the additional model biases introduced into a Least Squares adjustment, or at the very least, minimise the effects of neglected residual systematic biases and random errors.

Accounting for Biases

Dealing with the various biases in GPS solutions based on the mathematical model according to eqn (6.1-13) is a considerable problem. In fact much of the development effort in GPS software has gone into investigating ways of best reducing the computational burden imposed by having to account for the biases present in GPS range/phase measurements. Several options are available (§2.4):

☞ They can be estimated as explicit (additional) parameters.

☞ Those biases linearly correlated across different datasets can be eliminated by differencing.

☞ Those biases which are a function of frequency can be eliminated by constructing linear combinations of dual-frequency data.

☞ The biases can be directly measured, for example using Water Vapour Radiometer (WVR) observations in the case of the tropospheric delay.

☞ The biases can be considered known, being adequately modelled, for example as is sometimes attempted for the tropospheric delay.

☞ The biases can be simply ignored.

These strategies may completely account for some of the biases, or reasonably handle other biases (the unaccounted for part being subsumed into "residual biases" which are usually then lumped together with the measurement "errors"). Table 6.3-1 summarises the options for handling GPS biases, affecting carrier phase as well as pseudo-range data.

The basic observable in GPS data processing contains the following information:

(1) geodetic parameters of interest, such as the receiver or satellite coordinates being estimated,

(2) explicit biases in the measurements that can be parameterised in eqn (6.1-13),

(3) the known quantities in the Least Squares functional model, such as the satellite orbits, fixed station coordinates, etc., which permit, in combination with the measurements, the "O-C" quantity to be computed, and

(4) residual errors (or biases) that may contaminate the "O-C" quantity and which are not parameterised in the observation model, arising from measurement noise, cycle slips, or faulty apriori knowledge of the biased range parameters such as atmospheric delays, orbit errors, etc.

In particular, attention will be focussed on the parameters that are explicitly modelled in GPS surveying applications: the geodetic parameters, and the remaining "nuisance" parameters (cycle ambiguity, clock errors, atmosphere). Within this latter group of nuisance parameters, the bias subset comprising the cycle ambiguity and clock phase error terms are often referred to collectively as the "**clock biases**".

Dealing with Clock Biases and the *Fundamental Differencing Theorem*

The two most frequently used options applicable in the case of handling the "clock biases" associated with GPS measurements: (1) explicit parameter estimation, and (2) elimination by observation differencing. The questions that then immediately spring to mind are:

- Does the extra (nuisance) parameter approach (option (1)) and the data differencing approach (option (2)) both lead to the same result?
- Which is the "best" approach?

The **Fundamental Differencing Theorem** is useful in this regard:

> **"Linear biases can be accounted for either by reducing the number of observations so that the biases cancel, or by adding an equal number of unknowns to model the biases. Both approaches give identical results."** *(LINDLOHR & WELLS, 1985)*

Thus, the effect of the clock phase errors (ε_{rcj} and ε^{sci}) and cycle ambiguity (n_j^i) may be eliminated by differencing the basic eqn (6.1-13) between satellites and between receiver sites, observed simultaneously, and perhaps between observation epochs as well; or by estimating them explicitly, and the result would be equivalent.

However, although they may be *mathematically equivalent*, there are significant differences in how the algorithms are implemented for GPS phase reduction. On the one hand, differencing leads to a smaller parameter set having to be estimated, but on the other hand there is more data management, or "book-keeping", required if satellites rise and set during an observation session. The undifferenced observation model option is more "intuitive", being more like a range observable, however the need to estimate many more parameters weighs against it heavily. Consequently only a few software packages employ the principles of undifferenced phase reduction, and these are invariably "scientific" software for high precision geodetic applications. *All commercial GPS software use the double- (and triple-differenced) approach.*

The observation modelled by eqn (6.1-13) is sometimes referred to as "one-way phase" or "undifferenced phase" to distinguish it from the more commonly used single-, double- or triple-differenced phase observations described in §6.3. The term "undifferenced phase" is, strictly speaking, a misnomer because eqn (6.1-6) is actually formed by differencing the incoming and local (receiver) oscillator phases. Nevertheless it is useful to refer to the two approaches to dealing with the clock biases as being the "undifferenced" or "nuisance parameter" approach on the one hand, and the "differenced" approach on the other.

"Undifferenced" & "Differenced" Approaches to Handling Clock Biases

However, a further comment on the equivalence of the "undifferenced" and the "differenced" approach is worth making. The double-differencing approach (difference between satellites to eliminate ε_{rcj}, and between receivers to eliminate ε^{sci}) does introduce mathematical correlations in the resultant double-difference observables (§7.2). Only if this correlation is reflected in the variance-covariance (VCV) matrix of the double-differenced observables will mathematical equivalence between the "differenced" and the "undifferenced" approach be assured. (The same holds true for triple-differenced data.) As a consequence, to ensure this equivalence, it is necessary to explicitly construct a new VCV matrix at each observation epoch (of course in practice, advantage is taken of the fact that the weight matrix would only change with a change in the satellite/receiver combinations used to construct the double- or triple-differences, and that happens only if satellites rise or set during the session).

In the "undifferenced" approach, there is more *flexibility in the clock modelling options*. For example, the clock phase errors in eqn (6.1-13) are a function of time. If the satellite clock, or receiver clock, or both, had a stable, predictable behaviour (as when external atomic frequency standards such as cesium or rubidium oscillators are used), a bias parameter model based on a clock polynomial could be used, and the coefficients estimated, together with the cycle ambiguity, as *session parameters*. In general however, the clock phase errors are assumed to possess the characteristics of "white noise" and are explicitly modelled by independent bias terms on an *epoch-by-epoch basis*. The total number of clock bias parameters is $(R+S)T+RS$, where R is the number of receivers, S the number of satellites and T is the number of epochs in a session of data. There is therefore a potentially large number of these bias parameters to be estimated. In the case of four receivers, four satellites and 60 (one minute) measurement epochs, the total number of estimable clock bias parameters would be 496!

6.2

MEASUREMENT BIASES & ERRORS

All GPS pseudo-range and carrier phase measurements are affected by a variety of **biases** and **errors** (Figure 6.2-1). To appreciate the impact of measurement biases on GPS data quality (and ultimately on the positioning accuracy and reliability that is obtained), the level of measurement *noise* should be kept in mind.

> **"Rule-of-Thumb":**
>
> **Measurement resolution is possible at the level of 1%, or better, of the wavelength. *This is therefore the level of measurement noise.***

For the two basic GPS range-like measurements, the implications are:

Pseudo-range "noise":

- C/A code "wavelength" is approximately 300m, hence 3m is the range resolution or range measurement noise.
  (There is, however, a trend to instrumentation with submetre C/A code resolution.)
- P code "wavelength" is approximately 30m, hence 0.3m is the range resolution, or range measurement noise.

Carrier phase "noise":

- L1 carrier wavelength is approximately 0.19m, implying millimetre resolution, or carrier measurement noise.
- L2 carrier wavelength is approximately 0.24m, implying millimetre resolution, or carrier measurement noise.

> **Measurements are also <u>biased</u> by:**
>
> - atmospheric refraction
> - clock errors
> - site and instrumental effects
> - Selective Availability effect
> - *phase is biased by unknown ambiguity*

Biases may be defined as being those effects on the measurements that cause the <u>true range</u> to be different from the <u>measured range</u> by a *systematic amount*, and which must be accounted for in the measurement model used for data processing. As well as entering through incorrect or incomplete observation modelling, biases can also enter through imperfect knowledge of constants (for example, any "fixed" parameters such as the satellite orbit data, station coordinates, speed of light, etc.). *Hence under the heading of "errors" are assembled all unaccounted for measurement effects, as well as any unmodelled or residual biases* (see §2.4).

Different GPS applications require different levels of GPS accuracy. There is therefore a different partitioning of "biases" and "errors". At one extreme, in the case of GPS pseudo-range point-positioning, all effects with the exception of the receiver and satellite clock "uncertainty" are treated as errors (§1.2 and §1.4). At the other extreme, GPS baseline determination to accuracies of 1 part in 10^8 for geodesy applications demand that all measurement biases are explicitly accounted for in any solution scheme. In the case of GPS surveying (falling as it does between these two extremes) the following categorisation can be adopted:

❑ **BIASES**

☞ **Satellite dependent:**
- Ephemeris uncertainties
- Satellite clock uncertainties
- Selective Availability effects

☞ **Receiver dependent:**
- Receiver clock uncertainties
- Reference station coordinate uncertainties

☞ **Receiver-Satellite (or Observation) dependent:**
- Ionospheric delay
- Tropospheric delay
- Carrier phase ambiguity

❑ **ERRORS**
- Unmodelled, residual biases
- Carrier phase cycle slips
- Multipath disturbance
- Antenna phase centre offset
- Random observation error (noise)

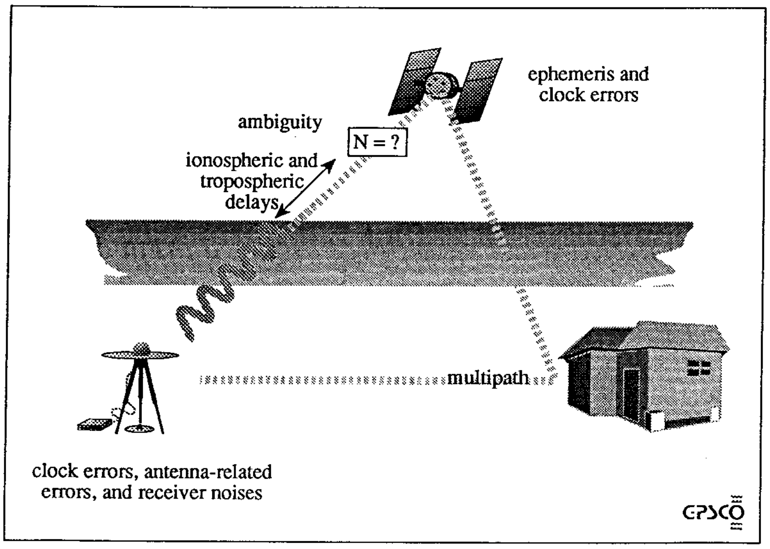


Figure 6.2-1. GPS biases and errors.

6.2.1 SATELLITE DEPENDENT BIASES

Satellite Clock Uncertainties

The GPS satellite clock bias, drift and drift-rate are explicitly determined in the same procedure as the estimation of the satellite ephemeris. The behaviour of each GPS satellite clock is monitored with respect to GPS Time, as maintained by an ensemble of atomic clocks at the GPS Master Control Station in Colorado Springs. The offset, drift and drift-rate of the satellite clocks are available to all GPS users as clock error coefficients broadcast in the Navigation Message (§3.3).

What is available to users is actually a *prediction* of the clock behaviour for some time into the future (24 hours or more ahead). As the random deviations of even cesium and rubidium oscillators are not predictable (see §1.3), such deterministic models of satellite clock error are accurate to about 20 nanoseconds, or six metres in equivalent range, depending upon the time since the last Navigation Message update. Selective Availability is a further artificial *dithering* of the satellite clocks causing several dekametres error in the range (or phase-range equivalent).

SATELLITE CLOCK BIAS

MAGNITUDE:

The residual satellite clock error (after correcting for the broadcast error model) cannot be neglected for any GPS surveying application as its magnitude is at the several dekametre level (under SA conditions). Clock error and corresponding range error:

1 nanosecond	==>	***0.3m***
1 microsecond	==>	***300m***

OPTIONS:

☞ **Construct a range-like observable from which the satellite clock error has been eliminated -- DIFFERENCE between-receivers.**

☞ **Model the satellite clock error as a "random process" -- ESTIMATE a constantly changing parameter.**

In the context of GPS phase data processing, the former involves observation differencing (§6.3), while the latter requires explicit estimation of the clock error on an epoch-by-epoch basis (§1.4). In general, the latter option usually requires the clock error to be modelled as "white noise". (This is in fact not a model in the stochastic sense, as no attempt is made to relate the clock error from epoch to epoch.)

STRATEGY:

Requires that two or more GPS receivers simultaneously track the same satellite. Both options are very effective at COMPLETELY eliminating the effect of the satellite clock bias.

Satellite Ephemeris Bias

The satellite ephemeris bias is the discrepancy between the *true* position (and velocity) of a satellite and its *known* value. This discrepancy can be parameterised in a number of ways, but a common way is via the three orbit components: alongtrack, crosstrack and radial (Figure 6.2-2). In the case of GPS satellites the alongtrack component is the one with the largest error.

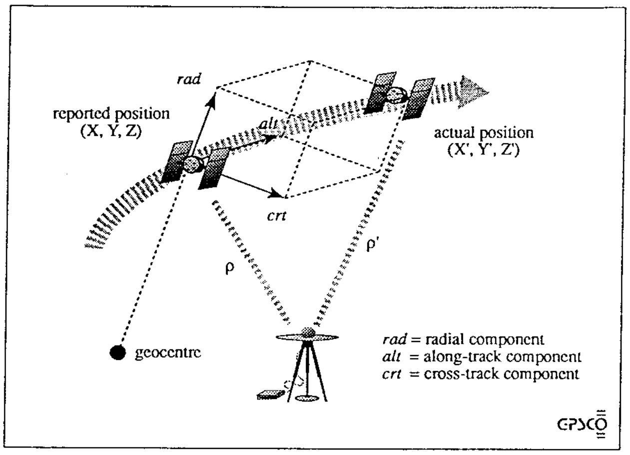


Figure 6.2-2. Satellite ephemeris bias.

What is the effect of the satellite orbit bias on GPS positioning? Expressed another way, if there was no other option but to assume that the available ephemeris data is correct, what is the effect of a satellite orbit error on GPS positioning? The following comments can be made:

- Height is a relatively weakly determined component, mainly because there are no satellites below the horizon.
- The East-West (longitude) component is slightly weaker than the North-South (latitude) component because of the motion of satellites (particularly in equatorial regions -- see Figure 2.2-8b).
- Effect on point positioning:

  Position error = PDOP . Orbit error

  Example: *If PDOP = 2 (§1.4), orbit error = 20m (or 1ppm)*
  Position error = 40m

  Complicated further because orbit error varies on different satellites.

- Effect on relative positioning, the following "rule-of-thumb" can be used (BESER & PARKINSON, 1982):

  $$\textbf{Baseline error} = \frac{d}{20000} \cdot \textbf{Orbit error}$$

  d = baseline length in kms

  Example: *If d = 10km, orbit error = 20m (or 1ppm)*
  Baseline error = 1cm (or 1ppm)

  Hence, the longer the baseline, the larger the effect of orbital error (this is illustrated in Figure 6.2-3). *Complicated further because orbit error varies on different satellites.*

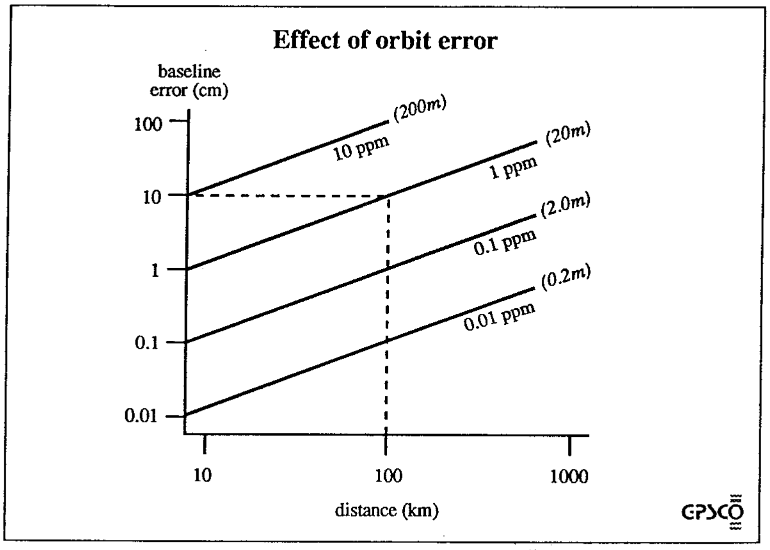


Figure 6.2-3. Approximate relationship between baseline length, accuracy and GPS satellite orbit error.

There are two basic classes of satellite orbit information:

- Ephemerides that are predicted from past tracking information, and are available to GPS users at the time of observation, and
- Post-processed ephemerides, which are orbit representations valid only for the time interval covered by the tracking data. Obviously this information is not available real-time as there is a delay between collection of the data, transmission of the data to the computer centre, the orbit determination process and the subsequent distribution to GPS users.

The former is available via the GPS Navigation Message (§3.3). The ephemeris computation takes place at the Master Control Station using tracking data acquired from the five monitor stations of the GPS Control Segment (§2.2). With regards to the accuracy of the broadcast ephemerides, there are (in principle) several distinct effects:

(1) There is the effect arising from the accuracy of the orbit *computation* procedure itself. The data used is P code pseudo-ranges, and although the tracking geometry is not strong (most of the tracking stations are in the equatorial belt), accuracies better than 10 metres are achievable.

(2) There are errors resulting from unpredictable orbital motion during the period since upload. These are essentially the *prediction* errors. Their magnitude can vary from a few decimetres (close to the time of Navigation Message update) to several tens of metres (for example, if using ephemeris data several hours beyond the reference epoch).

(3) There is the effect due to *Selective Availability*, which involves the deliberate degradation of the broadcast ephemeris parameters within the Navigation Message. The resulting orbit error is highly variable (the degradation algorithm is classified!), but it may be assumed that it could be as high as 100 metres or more. *(There is little evidence that such orbit data degradation is in fact at present taking place.)*

The accepted evidence suggests that the accuracy of the broadcast ephemerides is *below 10m for a single Navigation Message update per day, and better than 5m when three daily updates are performed* (HOFMANN-WELLENHOF et al, 1994). Many GPS surveying systems can output files containing the broadcast ephemerides in the RINEX format (see §7.3 for a sample).

Post-processed ephemerides are, in general, more accurate than predicted ephemerides, *with demonstrated accuracies well below the metre level.* Over the last few years there has been considerable activity in this area (see §12.2). The main requirement is a network of tracking stations and an orbit processing facility. Since the mid 1980's there have been tracking networks organised on regional, continental and global bases. There have been scientific, private and/or government initiatives, as well as military networks. Some networks have operated intermittently, for specific geodetic applications, others were organised on a semi-permanent basis. Several of these networks were the first examples of international civilian cooperation in the field of GPS ground infrastructure (§12.2). The following is a "potted history" of independent tracking and orbit computation initiatives:

- The U.S. Defense Mapping Agency (DMA) has generated "precise ephemerides" for many years, computed using tracking data from a global network of 8 or so tracking stations (comprising the monitor stations, supplemented by several DMA sites). *These ephemerides are now mostly used for internal DMA and Control Segment purposes.*
- Private initiatives such as that by Aeroservices Inc. have pioneered the non-military generation of GPS orbits. They originally established a number of tracking stations in the continental U.S. to collect data for satellite orbit computation. The orbits were then distributed to a small group of Macrometer™ users. *(This network was then subsumed into CIGNET.)*
- The U.S. Geological Survey contracted the Center for Space Research, the University of Texas at Austin, to generate post-computed ephemerides, beginning in late 1986, based on data from continental U.S. tracking stations. *Gradually data was incorporated from all CIGNET stations.*
- CIGNET (Cooperative International GPS Network) was a collection of tracking stations spread throughout the world, that operated on a continuous basis from the late 1980's until 1992 when the International GPS Service for Geodynamics (IGS) was established. The U.S. National Geodetic Survey (NGS) collected and archived the tracking data, and generated ephemerides, that were then made available to any user. *This was the first global tracking network initiative.*
- During the 1980's and early 1990's several regional and continental scale tracking networks were deployed in many parts of the world for national ephemeris computation (as well as GPS Integrity Monitoring -- §4.4), as well as in support of high precision GPS crustal motion surveys. *These networks were generally not established on a permanent basis.*
- In June 1992, a substantial "core" tracking network for the IGS was established under the auspices of the International Association of Geodesy. As well as organising the network (Figure 6.2-4), the IGS has several computing centres generating "products", including precise satellite ephemerides with decimetre level accuracies. A number of test campaigns and pilot services have been organised (see ZUMBERGE et al, 1995, for a review). *The full service was launched in 1994 and nowadays is the sole source of precise GPS ephemerides for the civilian user community.*

At present the precise IGS orbits are available via Internet (see §3.4), in the so-called SP3 ephemeris format, a format first developed by NGS for the CIGNET products. A sample of the SP3 ephemeris format is given below (Table 6.2-1). The SP3 file format is explained in a document that can be FTP'ed from the IGS Central Bureau by Internet.

Table 6.2-1. A sample of an IGS GPS satellite ephemerides "SP3 file".

Header portion (22 lines):

LINE 1

Definition:

col 1	symbol	#	col 2	version id	a
col 3	P/V mode flag	V	col 4- 7	year start	1993
col 9-10	month start	_1	col 12-13	day of month start	29
col 15-16	hour start	_0	col 18-19	minute start	_0
col 21-31	second start	_0.00000000			
col 33-39	number of epochs	_____96			
col 41-45	data used	____d	col 47-51	coordinate system	ITR91
col 53-55	orbit type	FIT	col 57-60	agency name	_JPL

Sample:

```
#aV1993  1 29  0  0  0.00000000      96     d ITR91 FIT  JPL
```

LINE 2

Definition:

col 1- 2	symbols	##	col 4- 7	GPS week	_681
col 9-23	seconds of week	432000.00000000			
col 25-38	epoch interval	__900.00000000			
col 40-44	mod. julian day start	49016			
col 46-60	fractional day	0.0000000000000			

Sample:

```
##  681 432000.00000000   900.00000000 49016 0.0000000000000
```

LINE 3

Definition:

col 1- 2	symbols	+_	col 5- 6	number of satellites	19
col 10-12	sat #1 id	__1	col 13-15	sat #2 id	__2
col 16-18	19-21 .... etc	until	col 58-60	sat #17 id	_26

Sample:

```
+   19    1  2  3 12 13 14 15 16 17 18 19 20 21 23 24 25 26
```

LINES 4-7 repeated with remaining sat IDs or zeros

Sample:

```
+        27 28  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
+         0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
+         0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
+         0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

LINES 8-22 contain various unused fields and comments

Epoch records:

LINE 23 (epoch header record), generally every 15 minutes

Definition:

col 1- 2	symbols	*_	col 4- 7	year start	1993
col 9-10	month start	_1	col 12-13	day of month start	29
col 15-16	hour start	_0	col 18-19	minute start	_0
col 21-31	second start	_0.00000000			

Sample:

```
* 1993  1 29  0  0  0.00000000
```

LINE 24 (position and clock record), repeated for each satellite

Definition:

col 1	symbol	P	col 2- 4	satellite id	__1
col 5-18	x-coordinate (km)	__14722.638510			
col 19-32	y-coordinate (km)	___6464.319150			
col 33-46	z-coordinate (km)	_-21020.844810			
col 47-60	clock (microsecond)	______-8.059218			

LINE 25 (velocity and clock record), repeated for each satellite

Definition:

col 1	symbol	V	col 2- 4	satellite id	__1
col 5-18	x-dot (decim/sec)	__-1196.628800			
col 19-32	y-dot (decim/sec)	__26950.022500			
col 33-46	z-dot (decim/sec)	___7502.277100			
col 47-60	cl rate (10e-4 msec/sec)	______0.000000			

Sample:

```
    P  1  14722.638510   6464.319150 -21020.844810      -8.059218
    V  1  -1196.628800  26950.022500   7502.277100       0.000000
    P  2 -24023.155300 -11843.563990  -1675.647210     -10.813964
    V  2   -769.152700  -3247.508000  31255.023300       0.000000
    P  3   2074.555420  19025.998840  17928.366120    -430.859048
    V  3  -6873.932300  22421.664200 -23147.529600       0.000000
    P 12  -6236.325580  13153.271260 -21964.100040    -108.945737
    V 12 -27384.917100   6805.632800  12337.728800       0.000000
    P 13 -13306.857100   4790.062110 -22360.523490     628.240251
    V 13   9739.738600 -26864.612400 -11662.222500       0.000000
              .              .              .              .
              .              .              .              .
    P 27 -19350.820260  -4003.111190  17582.690790      14.651464
    V 27  19491.879100 -11990.042400  18156.904400       0.000000
    P 28  13316.378500 -13959.644490  18317.660940      52.520005
    V 28    258.246400  23316.420800  17208.928500       0.000000
```

LINE 22 + number_epochs*(number_sats+1) + 1 (last line)

Definition: `col 1- 3    symbols    EOF`

SATELLITE EPHEMERIS BIAS

MAGNITUDE:

Orbit error is a residual bias, arising from mismodelling of the satellite trajectory, or accepting as "true" an ephemeris that has errors. In the case of the Broadcast Ephemerides within the GPS Navigation Message, these errors can range from (usually) less then 10m to (very rarely) up to 100m.

OPTIONS:

☞ **IGNORE the problem, and use the GPS Broadcast Ephemerides -- *assume the available orbit data are error-free.***

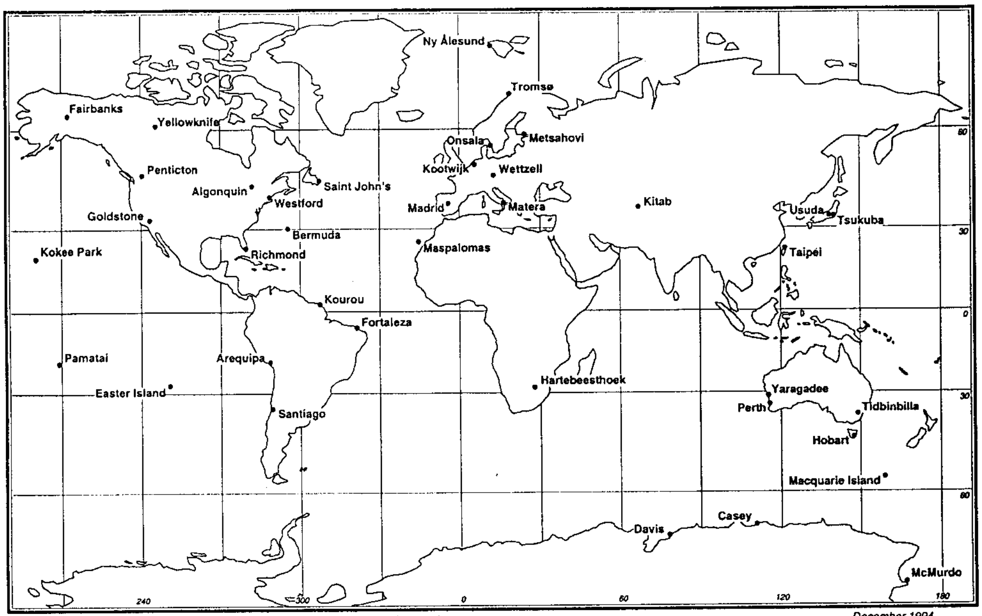
☞ **USE MORE PRECISE post-processed ephemeris -- for example the IGS orbits.**

☞ **ADJUST ORBIT as additional parameters -- *not an option with commercial GPS software.***

☞ **DIFFERENCE data between sites, and the effect of orbit error is minimised due to its high correlation over baselines.**

☞ **LENGTHEN observation session -- "average" out effect of baseline errors.**

☞ **SHORTEN baseline length -- reduces effect on baseline components when expressed in units of centimetres.**



† Processed by either 1) two or more IGS Analysis Centers on another continent or 2) a majority of Analysis Centers.

December 1994

Figure 6.2-4. The IGS core tracking network.

Selective Availability

As mentioned in §2.4, due to the surprisingly good Standard Positioning Service (SPS) accuracy for point position determination, the policy of Selective Availability (SA) was endorsed in order *to artificially widen the gap between the civilian and U.S. military positioning services* (GEORGIADOU & DOUCET, 1990). SA is an intentional degradation of the accuracy of GPS horizontal positioning to 100m (at the 95% confidence level), and height determination to 150-170m (at the 95% confidence level), for SPS users. SA has been implemented since 25 March, 1990, and consists of two different components:

(a) The so-called "epsilon" (ε) component consists of the *truncation* of the orbital information transmitted within the Navigation Message so that the WGS84 coordinates of the satellites cannot be computed correctly. *There is, however, no evidence to date of this effect on the broadcast ephemerides.*

(b) The so-called "delta" (δ) component which is achieved by *dithering* the satellite clock output frequency. The varying errors in the satellite clock's fundamental frequency has the same direct impact (as an artificially induced "satellite clock bias") on both the pseudo-ranges and carrier phase observations -- see §6.2. The variations in range have amplitudes of as much as 50m, with periods of several minutes (see Figure 2.4-7).

It should be emphasised that the situation as far as SA is concerned, is now under almost continuous review due to the proliferation of real-time Differential GPS services which have effectively replaced the SPS for many navigation-type applications (§2.4). LACHAPELLE (1995) and SANDLIN et al (1995) report that *it has been recommended that SA be deactivated within the next few years.* For the GPS surveying community however, SA was never a serious problem because it was very effectively nullified through the use of the relative positioning mode.

6.2.2 RECEIVER DEPENDENT BIASES

Receiver Clock Uncertainties

GPS receivers are equipped with quartz crystal oscillators, which have the advantages that they are small, consume little power and are relatively inexpensive. In addition they have good short-term frequency (or time-keeping) stability (§1.3). However, some receivers are equipped with I/O ports to permit the connection of an external frequency standard such as a cesium, rubidium and even a hydrogen maser, for specialist applications.

Although the time scale defined by individual receiver clocks have essentially *arbitrary* origins, they can be tied to a well established time scale, such as GPST (§2.1), in a number of ways. Generally, the time origin of a GPS receiver is set automatically as soon as sufficient satellites are tracked to carry out a single point pseudo-range navigation solution, using the solution strategy described for "Receiver-Biased Range Positioning" in §1.4. The subsequent time scale defined by the corrected receiver clock is then *nominally* that of GPST because:

(1) The *synchronisation at some epoch* (that is, the process of defining the time origin) is susceptible to error. Generally, it can be carried out only at the 0.1 microsecond level under Selective Availability (and perhaps to the 0.01 microsecond level with SA off).

(2) The *stability of the time scale* is directly related to the quality of oscillator used, and how often the current clock time is synchronised with GPST through the use of GPS pseudo-range observations. This "clock drift and reset" is illustrated in Figure 6.2-5.

RECEIVER CLOCK BIAS

MAGNITUDE:

The receiver clock is synchronised to GPST through the normal operation of code-correlating receivers to about 0.1 msec accuracy under SA. Therefore residual biases of the order of a dekametre (tens of metres) remain, and must be accounted for in some way.

OPTIONS:

☞ **Construct a range-like observable from which the receiver clock error has been eliminated -- DIFFERENCE between-satellites.**

☞ **Model the satellite clock error as a "random process" -- ESTIMATE an additional parameter.**

These are the same options available as in the case of the satellite clock bias.

STRATEGY:

Either option requires that two or more satellites are tracked simultaneously by the GPS receiver. Both options are very effective at COMPLETELY eliminating the effect of the receiver clock bias.

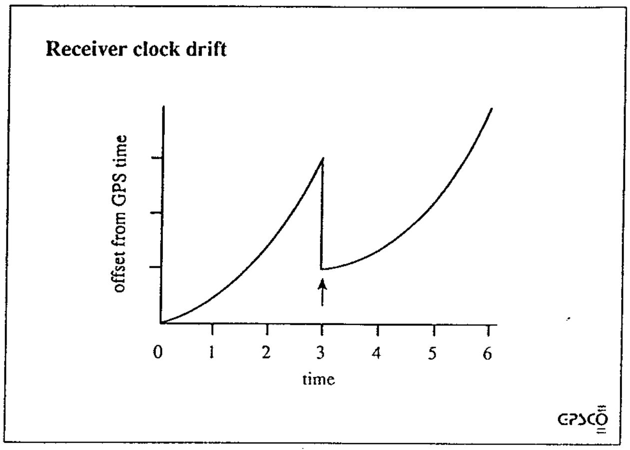


Figure 6.2-5. Clock drift and periodic reset to "true" time.

Reference Station Bias

Differential GPS positioning (essential for overcoming the satellite dependent biases described above) requires that the coordinates of one of the stations upon which a GPS receiver is located be held fixed during data processing. This station is the base or reference receiver. Hence, in effect, only the relative position (or baseline components) relating the second receiver to the first (base or reference) receiver are estimated. Any error in these fixed reference station coordinates therefore will cause a bias in the solution.

How well must the fixed base station's coordinates be known? The impact of the bias on relative positioning can be considered in a similar manner as the satellite orbit bias. The same "rule-of-thumb" may be applied (Figure 6.2-3). The more accurate the relative position is required, the more precisely the base station's coordinates must be known *in the coordinate reference of the satellite ephemerides*. Hence, the accuracy of the base station coordinates must be commensurate with the accuracy of the GPS satellite orbits.

REFERENCE STATION BIAS

EFFECT:

Analogous to an orbit bias, hence "rule-of-thumb" implies that 20m station coordinate uncertainty is acceptable for 1ppm relative accuracies (Figure 6.2-3).

OPTIONS:

☞ **ADJUST all station coordinates -- not feasible with most commercial GPS software, *leads to weak datum.***

☞ **AVERAGING of pseudo-range point position solution at reference station is usually satisfactory -- *even under SA.***

☞ **TRANSFORM local coordinates to WGS84 using published transformation parameters -- *adequate for 1ppm accuracy GPS surveys.***

6.2.3 RECEIVER-SATELLITE (OBSERVATION) DEPENDENT BIASES

These are the biases that relate mainly to the propagation link, and affect both pseudo-range and carrier phase measurements. However, the special characteristic of the integrated carrier phase observable leads to an additional special type of *constant* bias known as the phase ambiguity.

Ionospheric Delay

The ionosphere is that band of atmosphere extending from about 50 to 1000 kilometres above the earth's surface. In this layer the sun's ultraviolet radiation ionises gas molecules which then lose an electron. These free electrons in the ionosphere influence the propagation of microwave signals (speed, direction and polarisation) as they pass through the layer. The largest effect is on the speed of the signal, and hence the ionosphere primarily affects the measured range.

The **refractive index** of microwaves is a function of frequency (and hence the ionosphere has the property of "dispersion") and the density of free electrons, and may be expressed, to a first-order approximation, by (SEEBER, 1993; HOFMANN-WELLENHOF et al, 1994):

$$n = 1 \pm \frac{A \,.\, N_e}{f^2} \tag{6.2-1}$$

where:
- A is a constant,
- N_e is the total electron density (el/m^3), and
- f is the frequency.

The sign will depend on whether the **range** (+) or the **phase** (–) **refractive index** is required. The propagation speed v is related to the refractive index according to:

$$v = \frac{c}{n} \tag{6.2-2}$$

where c is the speed of electromagnetic radiation (EMR) in a vacuum.

Eqns (6.2-2) and (6.2-1) imply that the speed of the carrier wave (the "phase velocity") is actually increased, or "advanced", hence the phase refractive index is *less than unity*. However, the speed of the ranging codes is decreased (the so-called "group velocity") and therefore the pseudo-range is considered "delayed", and hence the range (or group) refractive index is *greater than unity*. (The ranging codes modulated on the carrier waves are considered a "group" of waves because they have different frequencies.)

The implication is therefore that **the distance as implied by the integrated carrier phase is *too short*, but the pseudo-range is *too long*. The correction terms are, of course, quantities with a reversed sign, that is, the carrier phase correction is *positive*, while the pseudo-range correction is *negative*** (§6.4).

An adequate expression for the ionospheric group delay d_{ion} and the ionospheric phase delay ϕ_{ion} for a microwave propagating from a satellite to the ground is:

$$d_{ion} = -\phi_{ion} \cdot \frac{c}{f} \approx 40.28 \cdot \frac{STEC}{f^2}$$
$$\approx \frac{40.28}{\operatorname{cosec}\zeta} \cdot \frac{VTEC}{f^2} \qquad (6.2\text{-}3)$$

where:

d_{ion}	is the ionospheric group delay in units of metres,
ϕ_{ion}	is the ionospheric phase delay in units of cycles,
c	is the speed of EMR in a vacuum(m/sec),
$\operatorname{cosec}\zeta$	is the cosecant of the zenith angle of the line-of-sight to the satellite,
f	is the signal frequency (Hz),
STEC	is the Slant Total Electron Content, expressed as the number of free electrons per square metre (el/m^2), and
VTEC	is the Vertical Total Electron Content (zenith direction) (el/m^2).

Note, from eqn (6.2-3), **the higher the frequency, the smaller the ionospheric delay effect.** As mentioned earlier, the pseudo-range measurements appear to be too long (d_{ion} must be subtracted), while the "phase-range" measurements (phase expressed in metric units) appear to be too short (d_{ion} must be added, or ϕ_{ion} must be subtracted). The constant A in eqn (6.2-1) is therefore equal to the line integral:

$$A = 40.28 \int STEC.ds \qquad (6.2\text{-}4)$$

There are a number of factors which influence the magnitude of the TEC (either STEC or VTEC), including (KLOBUCHAR, 1991):

- the latitude of the receiver,
- the season,
- the time of day the signal observation is being made, and
- the level of solar activity at the time of observation (Figure 6.2-6).

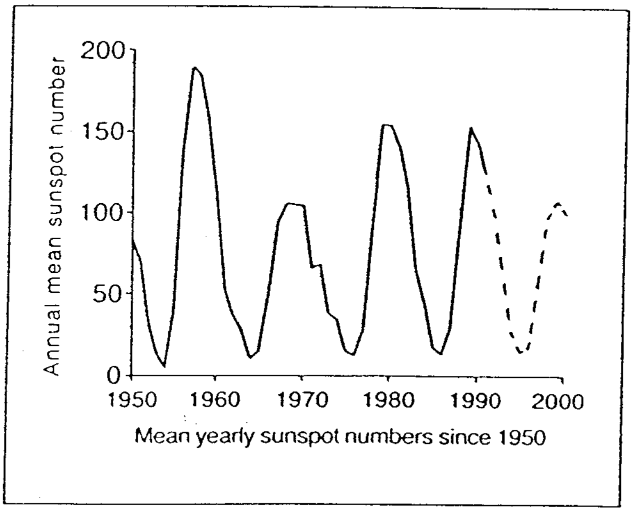


Figure 6.2-6. Sunspot activity for the last 40 years. *(KLOBUCHAR, 1991)*

The ionospheric effects therefore are subject to *spatial* and *temporal* variations. The spatial variations are usually low in frequency and generally correspond to the various ionospheric latitude zones: tropical, mid-latitude and auroral. The temporal variations can have high frequencies (the so-called "scintillations"), medium frequency (diurnal and seasonal effects), and low frequency (the 11 year solar cycle -- Figure 6.2-6). The 11 year solar activity has recently been characterised by minimums during 1986 and 1995, and a severe maximum in 1991. *The next solar maximum is not expected until the turn of the century.*

TEC is a maximum at low latitudes (the tropical zone) and at the poles (the auroral zone), and is a minimum at mid-latitudes. *At night the ionospheric delay is approximately five times less than for day time observations.* The diurnal cycle for TEC is such that the maximum occurs two hours after solar noon, and is a minimum before dawn. Ionospheric disturbances, which can occur suddenly and be very severe, also affect the value of STEC (WANNINGER, 1993). One of the phenomena responsible for these are "travelling ionospheric disturbances", another is due to irregularities in the ionosphere causing "scintillations" (especially in the tropical and auroral zones). Under such conditions the ionosphere is so perturbed that single frequency operations may be impossible because the GPS receiver loses lock on the satellite signals. Where tracking is possible, the likelihood of cycle slips and interrupted tracking is increased, both making ambiguity resolution a more difficult and unreliable task (§8.2).

The range of observed TEC is from about 10^{16} to 10^{19} el/m^2. **The maximum value of *vertical* range bias caused by the ionosphere is about 30m on L1 observations, and about 50m on L2 observations** (the magnitude of d_{ion} is found by scaling by a "mapping function" such as $1 / \mathrm{cosec}\zeta$). The effect on pseudo-range point positioning can therefore be quite severe. To aid single receiver navigation users a crude broadcast *predicted* ionospheric correction model is included within the transmitted Navigation Message (§3.3). However, this model can only reduce the RMS of the measurements (comparing observation residuals after solution, with and without including the ionospheric model correction) by approximately 50%. It is also possible to apply a correction derived from the International Reference Ionosphere.

What is the effect on relative positioning of neglecting to account for the ionospheric delay? As expected, it is a function of baseline length. For short baselines, the ionosphere that the signals to two receivers travel through would be essentially the same, hence the ionospheric delay on measurements made by the receivers to the same satellite would be very similar, and effectively cancel in between-receiver differencing (§6.3). According to BEUTLER et al (1989), when observing L1 phase data only, a scale effect will be introduced which will *shorten* baselines if the ionospheric signal delay is neglected. This effect expressed as a ratio of baseline length, in parts per million (ppm), is represented by the following "rule-of-thumb":

$$\frac{\Delta L}{L} \approx -0.7 \times 10^{-17}\,\mathrm{VTEC} \tag{6.2-5}$$

where L is the length of the baseline, and ΔL is the error.

The effect of this scale error can range from about 0.4ppm to over 3ppm for baselines determined from L1 observations alone, corresponding to low and high solar activity respectively. Hence this error is only significant for very high accuracy long baseline determination, or in instances where ambiguity resolution is critical (again, in the case of long baselines, or for very short observation sessions as in the "rapid static" techniques, or for "on-the-fly" ambiguity resolution).

The **dispersive** nature of the ionosphere can be used in actually removing most of the ionospheric delay effect by making pseudo-range and/or carrier phase measurements on both

L-band frequencies, and combining them in a special linear relation that results in an "ionosphere-free", or L3, observable (§6.4). However, there are several disadvantages in using this "ionosphere-free" combination:

- The "noise" on the resultant L3 observable is higher than on either of the original L1 and L2 measurements.
- Both L1 and L2 measurements are required, which could be a problem for some receivers during periods of high solar activity.
- The resultant L3 observable does not have an integer carrier phase ambiguity.
- The neglected second-order components are of the order of a few centimetres, hence possibly of some concern for very high accuracy applications involving long baselines.

IONOSPHERIC DELAY BIAS

MAGNITUDE:

- **Extreme at zenith ≈ 50m.**
- **Extreme at horizon ≈ 3 times zenith value.**
- **Extreme in day ≈ 5 times night value.**
- **Annual and sunspot (11 year) variation.**

Presently in a sunspot minimum period.

OPTIONS:

☞ **IGNORE the bias -- makes cycle slip editing and ambiguity resolution more difficult, and also introduces scale errors, *though only significant for long baselines.***

☞ **OBSERVE AT NIGHT if possible -- during minimum ionospheric activity.**

☞ **Use IONOSPHERE PREDICTION MODELS -- broadcast model generally <50% accuracy, *may be useful for point positioning users.***

☞ **Use DUAL-FREQUENCY receivers -- form "ionosphere-free" L1/L2 data combination (§6.4):**

$$\phi_{(L3)} = 2.546 \cdot \phi_{(L1)} - 1.984 \cdot \phi_{(L2)}$$

☞ **DIFFERENCE data between sites -- effect of error is minimised due to its high correlation over short to medium baselines, *typically 1-2ppm residual effect.***

Tropospheric Delay

Caused by the signal refraction in the electrically *neutral* (or non-ionised) atmospheric layer called the **troposphere**, extending from the earth's surface to about 8km (though thicker at the equator). Another component of the neutral atmosphere is the **stratosphere**, extending up to an altitude of about 50km, to the base of the ionospheric layer. For the purposes of discussing neutral atmospheric delay, under the term "tropospheric delay" will be included both the components due to the troposphere and the stratosphere because the troposphere, although being relatively thin, contains most of the mass of the neutral atmosphere and practically all of the water vapour.

Tropospheric delay is a function of the satellite elevation angle and the altitude of the receiver, and is dependent on the atmospheric pressure, temperature, and water vapour pressure (BRUNNER & WELSCH, 1993). However, a good starting point is to define it in terms of the refractive index, integrated along the signal ray path:

$$d_{trop} = \int (n - 1).ds \qquad (6.2\text{-}6a)$$

or in terms of the **refractivity** of the troposphere $N_{trop} = 10^6(n - 1)$:

$$d_{trop} = 10^{-6} \int N_{trop}.ds \qquad (6.2\text{-}6b)$$

The tropospheric refractivity can be partitioned into the two components, one for the <u>dry part</u> of the atmosphere and the other for the <u>wet part</u>:

$$N_{trop} = N_{wet} + N_{dry} \qquad (6.2\text{-}7a)$$

and the total tropospheric delay is:

$$d_{trop} = d_{dry} + d_{wet} \qquad (6.2\text{-}7b)$$

The d_{trop} can then be estimated by separately considering its two constituents d_{dry} and d_{wet}. **About 90% of the magnitude of the tropospheric delay arises from the dry component, and the remaining 10% from the wet component.** There are several models of the wet and dry refractivities, and a number of models for the tropospheric delay based on the numerical or analytical integration of eqn (6.2-6) -- see HOFMANN-WELLENHOF et al (1994) for details. The following treatment is mostly taken from this reference.

A model for the dry and wet refractivity *at the earth's surface* is:

$$N_{dry,0} = (77.64) . \frac{p}{T} \qquad (6.2\text{-}8a)$$

$$N_{wet,0} = -(12.96) . \frac{e}{T} + (3.718 \times 10^5) . \frac{e}{T^2} \qquad (6.2\text{-}8b)$$

where p is the atmospheric pressure in millibars (mb), e is the partial pressure of water vapour in millibars and T is the temperature in degrees Kelvin. Note that these coefficients are empirically determined and cannot fully describe the local situation even when locally measured met parameters are available.

The variation of $N_{dry,h}$ with height must also be empirically estimated, perhaps from radiosonde data. For example, **Hopfield** (IBID, 1994) derived a representation of the dry refractivity as a function of the height h (in metres) above the surface as:

$$N_{dry,h} = N_{dry,0} \cdot \left[\frac{h_d - h}{h_d}\right]^4 \tag{6.2-9}$$

where $h_d = 40136 + 148.72(T - 273.16)$ (in metres). Hence the dry part of the tropospheric delay (sometimes referred to as the hydrostatic delay) can be obtained by applying eqn (6.2-6b), solving the integral along the zenith using the variation of N_{dry} along the signal path defined by eqn (6.2-9) and neglecting the curvature of the signal path, leading to an expression for the dry zenith delay:

$$d_{dry}^z = \frac{10^{-6}}{5} \cdot N_{dry,0} \cdot h_d \tag{6.2-10}$$

This is scaled by an appropriate **mapping function** to any arbitrary elevation angle:

$$d_{dry} = \frac{1}{\sin(\sqrt{E^2 + 6.25})} \cdot d_{dry}^z \tag{6.2-11}$$

where E is the satellite elevation angle in degrees.

The wet component is much more difficult to model because of the strong variations in the distribution of atmospheric water vapour in space and time. Hence, due to a lack of an appropriate alternative, the Hopfield model assumes the same functional model for the wet component as for the dry component (eqn (6.2-9)):

$$N_{wet,h} = N_{wet,0} \cdot \left[\frac{h_w - h}{h_w}\right]^4 \tag{6.2-12}$$

where h_w is set to a value of between 11000 to 12000m. The wet zenith delay is:

$$d_{wet}^z = \frac{10^{-6}}{5} \cdot N_{wet,0} \cdot h_w \tag{6.2-13}$$

This is also scaled by a mapping function, to give the delay at any arbitrary elevation angle:

$$d_{wet} = \frac{1}{\sin(\sqrt{E^2 + 2.25})} \cdot d_{wet}^z \tag{6.2-14}$$

Alternative functions for the wet and dry refractivities at the earth's surface (eqn (6.2-8)), and for the upward continuation of these refractivities (eqns (6.2-9) and (6.2-12)) result in new zenith tropospheric delay models. The mapping functions may also be changed. For example, **modified Hopfield** models can be produced by varying several of these functions (IBID, 1994). Another popular model for GPS reductions is that due to **Saastamoinen** (SEEBER, 1993):

$$d_{trop} = \frac{0.002277}{\sin E} \cdot \left[p + \left(\frac{1255}{T} + 0.05 \right) . e - B . \cot^2 E \right] + \delta R \qquad (6.2\text{-}15)$$

where the parameters p, e , T and E have been defined earlier. The B term is dependent upon the altitude of the station (and is interpolated from a table of values), and the δR correction term is dependent upon the altitude of the station and the elevation angle of the satellite (and is also interpolated from a table of values) -- see HOFMANN-WELLENHOF et al (1994).

Note that in the models described here (and other commonly used models such as the **Lanyi** model, the **Chao** model, **Marini and Murray** model, etc.), measured values of p, e and T are assumed to be available at the GPS receiver site. Alternatively, it is possible to use profile functions to express pressure, water vapour pressure and temperature as a function of the height H above mean sea level (in kms). Examples of such profile functions are:

$$p = p_0 . (1 - 0.0226 . H)^{5.225}$$

$$T = T_0 - 6.5 . H$$

$$e = e_0 . 10^{H.(1+H/8)/8}$$

$$r = e . \frac{100}{e_{sat}}$$

$$e_{sat} = 6.107 . 10^{(7.5 . t/(238+t))} \qquad (6.2\text{-}16)$$

where p_0, e_0 and T_0 are "standard" values at sea level (say, 1013.25mb, 15mb, and 291.2°K respectively), r is relative humidity (in %), t is temperature (in degrees Celsius) and e_{sat} is the saturation water vapour pressure (in mbs)

The choice of zenith tropospheric delay model is somewhat arbitrary. All are very good at modelling the zenith dry tropospheric delay, but can only reasonably model the wet delay when it is very small (that is, when the atmosphere is very dry). For example, the zenith dry tropospheric delay at sea level is of the order of 2.3m. The zenith wet tropospheric delay, however, may vary from a few millimetres to as much as 40cm. **The variability of the dry component is relatively low and can be estimated with a precision approaching 1%. On the other hand, the wet component of the delay is notoriously difficult to estimate and errors of 10-20% are common.**

There are several mapping functions to chose from and all give line-of-sight tropospheric delay values which are reasonably close to each other (less than 5mm difference) for satellite elevation angles greater than about 30°. Some models are superior to others when elevation angles below 15° are considered. However, even the best may produce errors of several decimetres at elevation angles below about 5°, in a total line-of-sight tropospheric delay of the order of 20-30m! **The tracking of low elevation angle satellites is therefore to be avoided because the uncertainties in modelling both the wet and dry tropospheric delay are *amplified* at low elevation angles.**

There are a number of further comments that can be made regarding the tropospheric delay and its impact on GPS positioning:

- With respect to microwaves up to frequencies of 15GHz the neutral atmosphere is a non-dispersive medium, and hence the tropospheric delay is not frequency-dependent. *It cannot therefore be eliminated through linear combinations of L1 and L2 observations as in the case of the ionospheric delay.*
- The magnitude of the tropospheric delay is the same for both L1 and L2 observations, and for pseudo-range and carrier phase measurements.
- There is significantly less tropospheric delay at high altitude than at sea level. For example, if two receivers are deployed, one at sealevel and the other at 3000m, and assuming a satellite elevation angle of 20°, the tropospheric refraction error for the receiver at sea level will be approximately 7.8m, while for the receiver positioned at high altitude the tropospheric refraction error is of the order of 4.8m.
- The tropospheric delay can be predicted using values of temperature, pressure, and humidity, input into one of a number of atmospheric refraction models. Such models can account for approximately 90% of the delay (corresponding mainly to the dry troposphere component), however the remaining 10% (largely due to the wet troposphere component) will still seriously bias high accuracy positioning.
- For surveys of less than a few tens of kilometres in extent, the tropospheric delay will tend to be the same at both ends of a baseline and thus assuming: (a) the same tropospheric model is applied uniformly for the campaign processing, or (b) no model is applied; the delay effect will largely cancel. *The error introduced in baseline components under a wide range of conditions has been estimated as being of the order of 1ppm.*
- Neglecting to apply tropospheric refraction results in an absolute scale error. For example, a 1metre tropospheric delay in the zenith direction causes a *scale effect* of 0.4ppm -- that is, the baselines appear longer (BEUTLER et al, 1989). (As the total zenith value is of the order of 2.3m, neglecting tropospheric refraction introduces 1ppm relative error, as stated above.)
- In practice, surface meteorological observations are rarely made as it appears that the baseline results are worse when they are used! For small scale surveys (less than a few tens of kilometres in extent) it is often recommended that *modelled* values of met parameters be used (upward continuation of "standard" sea level values using eqn (6.2-16)), instead of observed meteorological values because small systematic errors in the meteorological values (such as due to calibration errors of the measuring instruments and observer error in the field) can often introduce larger biases than the use of a "standard atmosphere". *Furthermore, most commercial GPS software will not accept input of field measured met values.*
- GPS surveyors should always be mindful of the influences of local meteorological conditions that may exist at individual stations (for example, fog, temperature inversions, precipitation, etc.), which could lead to significant differential (or residual) tropospheric refraction biases.
- Any uncertainty in modelling the differential tropospheric refraction bias results mostly in a degradation of the height component in the solution. IBID (1989) have suggested the "rule-of-thumb": 1mm *differential* tropospheric bias causes a height error of about 3mm. *There is only a minimal effect on latitude and longitude.*

TROPOSPHERIC DELAY BIAS

MAGNITUDE:

- **Zenith value at sea level ≈ 2.3m.**
- **Near horizon at sea level ≈ 20 - 30m.**

90% due to dry atmosphere which can be modelled very well. 10% due to wet component of atmosphere is quite difficult to account for.

OPTIONS:

☞ **IGNORE the bias -- but avoid tracking low elevation satellites, *generally no observations taken below 20°.***

☞ **CORRECT data using a STANDARD TROPOSPHERIC REFRACTION MODEL (Saastamoinen, Hopfield, etc.) -- with or without surface meteorological readings.**

☞ **ESTIMATE residual tropospheric effect as an additional parameter -- *only for very high precision work.***

☞ **DIFFERENCE data between sites -- effect of error is minimised due to high correlation over short to medium baselines.**

☞ **MEASURE wet path component directly using a Water Vapour Radiometer -- too expensive, *no longer considered a viable option.***

Phase Ambiguity

The phase ambiguity bias is unique to integrated carrier beat phase measurements and is illustrated, for an instantaneous phase observation, in Figure 6.2-7.

A model for the measured GPS carrier beat phase measurement is (eqn (6.1-12):

$$\phi_{bj}{}^{i}(T_j) = \phi_{fj}{}^{i}(T_j) + C_R(T_j)$$

$$= (\, f_0 / c\,).\rho_j{}^{i}(T_j) + \phi_{bias}(T_j) + n_j{}^{i} \qquad (6.2\text{-}17)$$

where :

$\phi_{bj}{}^{i}(T_j)$ is the carrier beat phase for receiver j, satellite i, at reception time T_j,

$\phi_{fj}{}^{i}(T_j)$ is the measured *fractional* phase (in the range: 0 to 2π, 0° to 360°, 0 to 1 cycle),

$C_R(T_j)$ is the measured *integer number of "zero-crossing"* of the fractional phase during continuous tracking,

$\rho_j{}^{i}(T_j)$ is the geometric range from receiver j to satellite i, at reception time T_j,

f_0 is the signal frequency (L1 or L2),

$\phi_{bias}(T_j)$ is the term containing all the remaining biases that are not of interest in this discussion, and

$n_j{}^{i}$ is the ambiguity term, the *unknown* integer number of cycles at the first observation epoch (representing the "missing" constant component in the satellite-receiver range).

Note that the phase cycle ambiguity is a critical part of the unambiguous "phase-range" measurement model (if it can be determined, then it is no longer included in the above equation, and the observation equation degenerates to one that is virtually identical to that of a standard GPS pseudo-range measurement). The following comments may be made in relation to Figure 6.2-7 and eqn (6.2-17):

- The ambiguity is an *integer* number (a multiple of the carrier wavelength).
- The ambiguity is different for the L1 and L2 phase observations.
- The ambiguity is different for each *satellite-receiver pair.*
- The ambiguity is a constant for a particular satellite-receiver pair for all epochs of *continuous* tracking (that is, as long as no **cycle slips** occur -- see below).
- The carrier phase measurement from one observation epoch to another is a measure of the *change* in satellite-receiver range.
- The determination of the cycle ambiguity integer is known as **ambiguity resolution**, and is generally not an easy task because of the presence of other biases and errors in the carrier phase measurement.

PHASE AMBIGUITY BIAS

EFFECT:

- The existence of phase ambiguities means that instantaneous positioning using phase data, at a single epoch, is not possible.
- Including ambiguities as extra estimable parameters makes for a more complex positioning solution.
- Theoretically all phase ambiguities should be integer values.

OPTIONS:

☞ **ADJUST all ambiguities as free parameters, together with station coordinates, in a so-called ambiguity-free solution.**

☞ **"RESOLVE" estimated ambiguity parameters to their nearest integer values, and iterate solution, in a so-called ambiguity-fixed solution -- *in effect use precise unambiguous phase-range data.***

☞ **DIFFERENCE data between consecutive epochs to ELIMINATE the phase ambiguity bias.**

STRATEGY:

The strongest baseline solution is provided by "carrier-range" (or "phase-range") double-differenced data, formed when ambiguous carrier phase data is "corrected" so as to generate a high precision range-like observable.

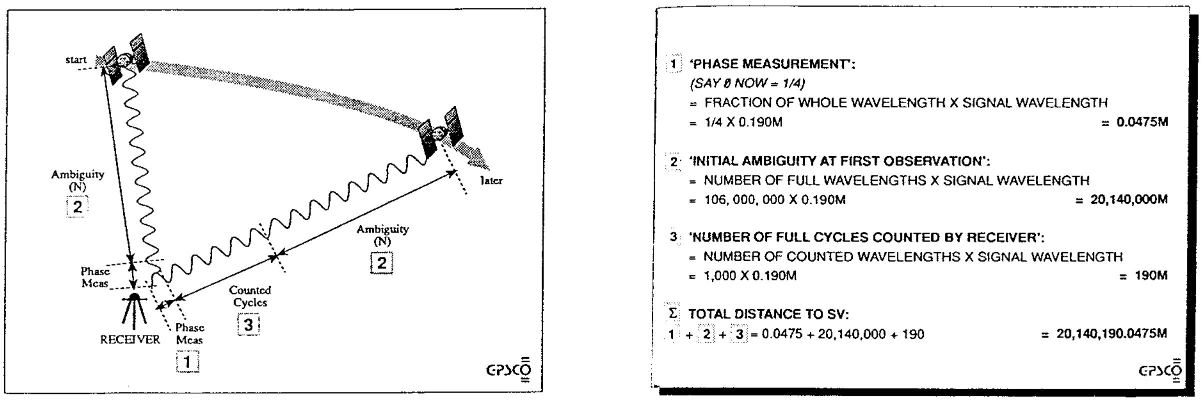


Figure 6.2-7. The phase ambiguity term and a range measurement.

6.2.4 COMMENTS ON RESIDUAL BIASES AND GPS ERRORS

The residual biases remaining after the appropriate differencing of phase data, collected simultaneously by two GPS receivers (§6.3), are primarily due to the:

(a) differential ionospheric refraction between the two sites (if not eliminated in the L3 "ionosphere-free" combination),

(b) differential tropospheric refraction between the two sites, and

(c) muted, second-order effect of orbit error on a baseline solution.

The uncertainty in the baseline solution caused by a combination of these effects is unlikely to be more than a *few parts per million for relatively short baselines* (less than about 30km in length). This uncertainty may have different characteristics -- for example, be manifest as a scale error, or a height error -- but is considered to be acceptable for standard GPS surveying as the level of uncertainty is still much lower than that generally required for conventional geodetic surveying. *Attention will instead be focussed on several sources of error that can dramatically impact on baseline accuracy.* Random measurement error is far too small to be considered of significant concern even for the highest accuracy GPS applications.

Carrier Phase Cycle Slips

Signal reception by a GPS receiver can be *interrupted* by: (a) the shadowing, or large acceleration of the antenna, both are particularly a problem in the case of kinematic positioning; or (b) during a severe signal disturbance such as due to ionospheric activity and multipath interference; or (c) low signal-to-noise ratio. After the signal tracking interruption, the ambiguity n_j^i has a different value than before.

How is the measurement affected? As indicated in Figure 6.2-7, the integrated carrier phase measurement at time T_j involving satellite i and receiver j consists of the three components:

(1) The fraction of a cycle, for example 0.25cyc (0.0475m on L1), denoted by $\phi_{fj}^i(T_j)$ in

eqn (6.2-17).

(2) The arbitrarily assigned integer ambiguity at signal lock-on, for example 106000000cyc, denoted by n_j^i in eqn (6.2-17).

(3) A count of the whole cycles by the receiver, for example 1000cyc, denoted by $C_R(T_j)$ in eqn (6.2-17).

Following a loss-of-lock, on resumption of tracking to the satellite, the accurate fractional part of the carrier phase can be measured again, however, the integer part ($n_j^i + C_R(T_j)$) will no longer provide the correction to the fractional phase measurement that yields the true satellite-receiver range. In essence, the "zero-crossing" measurement of $C_R(T_j)$ has not "kept up" with the actual change in satellite-receiver range. Therefore, to account for this "jump" in the integrated carrier phase in the interval between the epoch before loss-of-lock, and when the signal is again acquired (Figure 6.2-8), the unknown ambiguity term must be assigned a new value n_j^{i*} (= $n_j^i + S(T_e)$) in the mathematical model (eqn (6.2-17). The term $S(T_e)$ is a **cycle slip** that has occurred at epoch T_e (the epoch when signal tracking resumed), and is a *constant* additional term that affects all observations after epoch T_e (Figure 6.2-8).

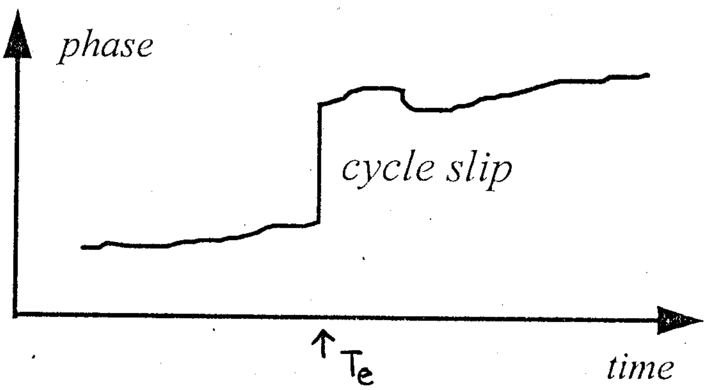


Figure 6.2-8. A "jump" the sequence of carrier phase measurements due to the occurrence of a cycle slip.

In practice, cycle slips are usually detected and repaired in a data pre-processing step (§7.3). Many techniques have been developed to perform this task dependent such things as: the positioning mode being used, the baseline length, the type of data available, the antenna dynamics, etc. The following comments can be made regarding this process:

- Cycle slip *detection* is easier than its *correction*, especially if the slip is "large" (say, greater than about 10 cycles).
- However, cycle slip *correction* involves the determination of the exact number of slipped carrier cycles at epoch T_e. This is very difficult in one-way carrier phase observations because of the presence of biases, in particular the clock biases.
- Cycle slip *correction* is easier when the clock biases have been eliminated during between-receiver and/or between-satellite differencing (§6.3).
- Cycle slip *correction* of dual-frequency data is easier than for single frequency data (§6.4 and §8.4).
- Cycle slip *correction* of *static* data is easier than the case of *kinematic* data.

- Cycle slip *correction* of data in the post-mission mode is easier than for the case of *real-time* data processing.

Multipath Disturbance

The carrier wave propagates along a *straight* line (not quite, there are small bending effects due to the presence of the atmosphere). **Multipath** is caused by extraneous reflections from nearby metallic objects, ground or water surfaces reaching the antenna. This has a number of effects: it may cause signal interference between the direct and reflected signal (Figure 6.2-9) leading to *noisier* measurement, or it may confuse the tracking electronics of the hardware resulting in a *biased* measurement that is the sum of the satellite-to-reflector distance and the reflector-to-antenna distance.

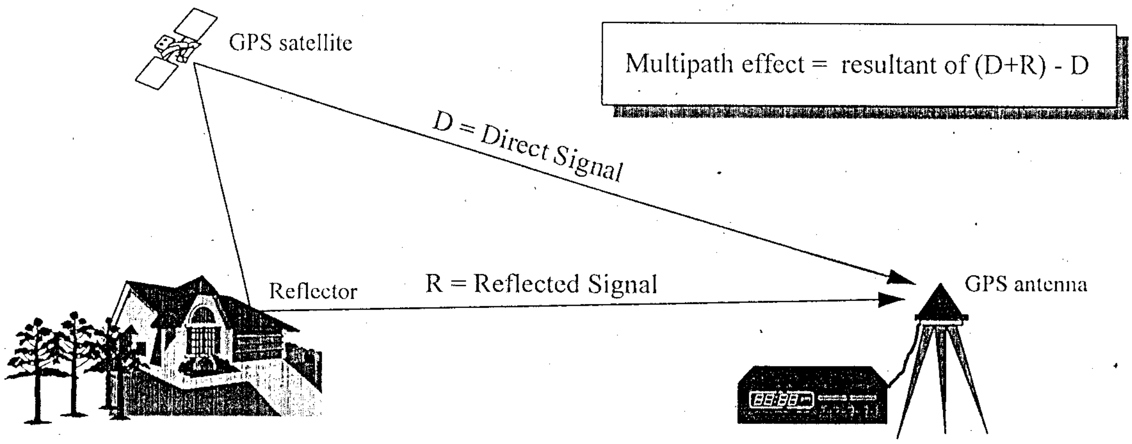


Figure 6.2-9. The multipath effect is a result of signals from a satellite reaching the antenna over more than one path.

The magnitude of the multipath effect on a phase observation can be estimated from the following mathematical relation (HOFMANN-WELLENHOF et al, 1994):

$$\tan\Delta\phi_m = \frac{\beta.\sin\Delta\phi}{1 + \beta.\cos\Delta\phi} \tag{6.2-18}$$

where: $\Delta\phi_m$ is the shift in carrier phase of the combined signal received at the antenna due to multipath,

$\Delta\phi$ is the phase shift of the reflected signal with respect to the direct signal, and

β is a damping factor which varies between 0 (no reflection) and 1 (reflected signal as strong as direct signal).

Some characteristics of multipath are:

- Multipath can cause "jumps" in the signal measurement which are a function of frequency. Therefore, the maximum multipath bias that can occur in pseudo-range data is one chip length of the code, that is, *293m for C/A code ranges* and *29.3m for P (Y) code ranges*. The carrier phase multipath *does not exceed about one-quarter of the wavelength* (when β = 1 and $\Delta\phi$ = 180° in eqn (6.2-18)) -- 5-6cm for L1 or L2, 20cm for L5, etc.
- Multipath is receiver-satellite geometry dependent, and the causes of multipath tend to be horizontal, vertical, or oblique planes/objects (such as metallic fences, buildings, chimneys, superstructure, water surfaces, etc.), hence the effect of the multipath error on

positioning will generally repeat on a daily basis for the same baseline (Figure 8.1-2).

- As the receiver-satellite geometry changes (and hence the angle of incidence and reflection of the signal with respect to the reflective surface changes), the multipath effect changes in a sinusoidal pattern, and generally "averages out" over a period from several minutes to a quarter of an hour, or more (Figure 6.2-10).
- There is no general mathematical model to *determine* or *predict* the effect of multipath on a position solution, however its effect on a range observation may be *measured* using a combination of L1 and L2 carrier phase and pseudo-range data (§6.4).
- Multipath will bias the baseline results as a function of the *percentage* of observed data that is multipath-affected, and the *amplitude* of the multipath phase disturbance.
- The impact of multipath error on kinematic positioning is greater than that on static positioning, because in the former the error propagates into an incorrect position solution (Figure 8.1-2), whereas in the case of static positioning the multipath error propagates into the residuals (implying a less precise solution).

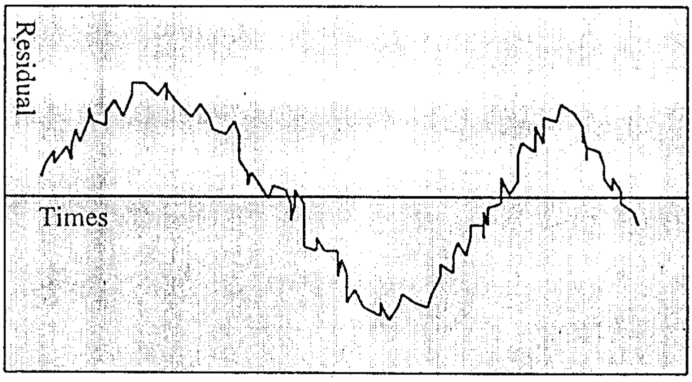


Figure 6.2-10. Sinusoidal signature of multipath effect on measurement residuals.

Some options for reducing the effect of multipath are:

- Make a careful selection of antenna site in order to avoid reflective environments.
- Use a good quality antenna that is multipath-resistant.
- Use an antenna groundplane or choke-ring assembly (§4.1).
- Use a receiver that can internally digitally filter out the effect of multipath signal disturbance.
- Do not observe low elevation satellites (signals are more susceptible to multipath).
- In the case of *pseudo-range positioning* (single point or differential), averaging the computed results over a period of time will reduce the contribution of multipath errors on the averaged pseudo-range solution.
- In the case of *carrier phase positioning*, longer observation sessions will tend to diminish the impact of multipath on the final baseline results.

Signal Jamming and Interference

This is concern about the effect of signal jamming and interference on critical GPS navigation applications. There is much anecdotal evidence, as well as some test results (SLUITER & HAAGMANS, 1995), to indicate that a problem exists in this regard. The disturbance to the incoming signal is a function of the frequency of the disturbing signal (it must be at or close to the GPS carrier frequencies), the distance from the jamming transmitter, and its power. In the surveying context, such signal disturbances are likely to manifest themselves as noisier than normal observations or, in extreme cases, the occasional loss-of-lock on the signal (and subsequent cycle slips). Hence, be suspicious of TV and microwave transmission towers, and various types of radar (especially at the critical GPS frequencies).

Antenna Phase Centre Offset

All GPS measurements relate to the distance from the electrical centre of the satellite's transmitter to the electrical centre of the receiving antenna. Ideally the *physical* centre should coincide with the *electrical* centre, however there may be a constant offset. This is an antenna manufacturing problem, but if the antenna is always oriented in the same direction, the impact on the groundmark-to-groundmark solution will be a systematic shortening or lengthening of the baseline. Because the electrical centre varies with the direction and strength of the incoming signal, a variation in the satellite-receiver geometry will cause the position of the electrical centre to also vary. Tests have shown that this effect is comparatively insignificant for microstrip antennas (well below 1cm), though it is larger in the case of quadrifilar and helical antennas (§4.1). For high precision applications care has to be taken not to mix antenna types, or to swap antennas between sites and receivers during a survey.

ERRORS

CARRIER PHASE CYCLE SLIPS:

- ☞ **SOLVE for in a pre-processing step --** ***aim is to obtain a "clean" dataset.***
- ☞ **DIFFERENCE between-epochs to obtain a more "robust" triple-differenced observable --** ***solution will be less influenced by "slipped" data.***

MULTIPATH & OTHER DISTURBANCES:

- ☞ **AVOID reflective or jamming environments.**
- ☞ **OBSERVE at a site for more than 15 minutes --** ***to average out multipath effect.***
- ☞ **USE large ground planes, RF absorbing material or choke-ring antennas.**

ANTENNA PHASE CENTRE OFFSET:

- ☞ **IGNORE --** ***generally below 1cm.***
- ☞ **BETTER antennas -- use stable micro-strip or dipole antennas.**
- ☞ **Keep antenna and receiver unit together during survey, and always orient antenna in the same direction --** ***effect on baseline is constant (hence indistinguishable from scale error).***

RESIDUAL, UNMODELLED BIASES:

- ☞ **IGNORE -- residual effect on solutions is a function of baseline length,** ***but should be of the order of only a few parts per million.***

RANDOM MEASUREMENT ERROR:

- ☞ **Measurement noise is typically at the millimetre level.**

6.3

GPS OBSERVATION MODELLING

6.3.1 OVERVIEW

All GPS pseudo-range and carrier phase observations may be modelled as (§6.1 and §6.2):

$$\mathbf{Obs_j^i = \rho_j^i + b_j + b^i + b_j^i + v_j^i}$$

where ρ_j^i is **geometric range** from stn.j to sat.i,

b^i are the **satellite dependent biases**,

b_j are the **station dependent biases**,

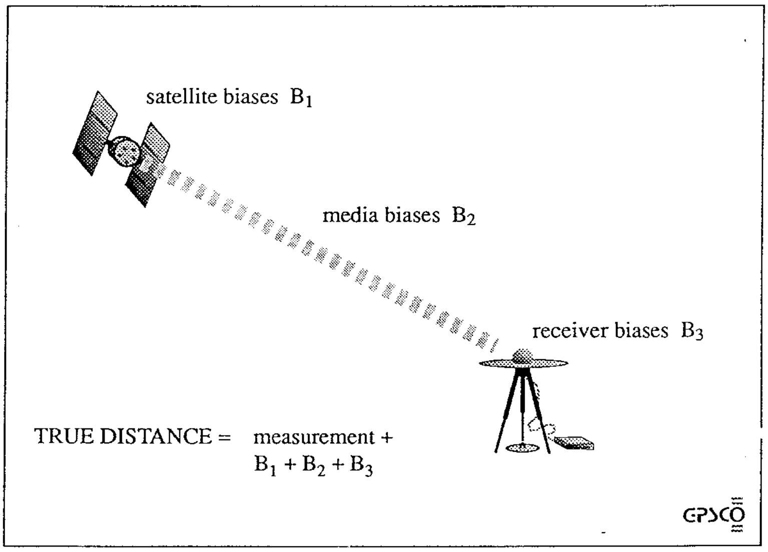
b_j^i are the **observation dependent biases**, and

v_j^i is the measurement **noise**.

Only the geometric range contains the coordinate parameters of interest.

All biases basically influence both the pseudo-range and integrated carrier phase observations by the same amount (there are some frequency-dependent effects), however only the integrated carrier phase observations contain the ambiguity bias, which is a constant for a satellite-receiver pair as long as the instrument remains locked onto the satellite.

The measurement noise is about 2-3 orders of magnitude higher in pseudo-range data than for integrated carrier phase data.



A sample data file record for one epoch (date 26/7/92, time 6:52:30 UT) to the five satellites PRN2, 11, 18, 19, 28 (one record of five observation data types to each satellite) will help illustrate the impact of some of the measurement biases and errors:

Epoch marker--> 92 7 26 6 52 30.0000000 0 5 2 11 18 19 28

	C/A pseudo-range (m)	L1 phase (L1 cycles)	L2 phase (L2 cycles)	L1 pseudo-range (m)	L2 pseudo-range (m)
PRN2 -->	22333042.96600	-12176065.17700	-9487835.16600	22333025.72200	22333027.39100
PRN11-->	22934353.22700	-7227959.70200	-5632169.61500	22934338.21900	22934340.52000
PRN18-->	20485466.29600	-23339757.04000	-18186808.56800	20485447.35300	20485447.55500
PRN19-->	20609091.79300	-20568272.40900	-16027208.73300	20609081.52200	20609081.81200
PRN28-->	23894196.98000	154816.73100	120640.79300	23894184.83300	23894185.36800

The following comments can be made concerning the various quantities:

- All five observation types are biased by the same amount (equivalent range) by the receiver and satellite clock errors, and the tropospheric delay.
- The phase observations have negligible noise. The P code pseudo-ranges have observation noise of a few decimetres, while the C/A code pseudo-ranges are the "noisiest". The multipath error (if present) is greatest for the C/A pseudo-ranges, and least for the phase measurements.
- The ionosphere accounts for most of the difference in the pseudo-range measurements on L1 and L2. This is equivalent to the difference in the L1 and L2 phase observations when they are converted to range (metric) equivalent values.
- The ionospheric delay on the C/A pseudo-range is equal to that on the L1 pseudo-range, and equal in value, though not in sign, to that of the L1 phase (when expressed in metric units).
- The ionospheric delay on pseudo-range measurements means that they measure range that is longer than "true", but that the phase observations are shorter than "true".
- The (unknown) ambiguity on the L1 phase is different to that of the L2 phase, and is different for each satellite.

Summary Remarks: Handling GPS Biases and Errors

Depending upon the level of accuracy sought, the various GPS biases and errors may be considered significant or not, and different options used in accounting for these effects. Below in Table 6.3-1 are summarised the options identified in §6.2 for those applications requiring carrier phase data processing. (In §7.2 the options appropriate for GPS **surveying** applications are discussed further.)

Table 6.3-1. Options for handling GPS measurement biases and errors.

Bias or Error	A	B	C	D	E
Satellite Clock	*1	*1			
Satellite Orbit	*2	*3			
Receiver Clock	*1	*1			
Fixed Reference Station	*4				*5
Ionospheric Delay	*6	*7	*8	*9	*10
Tropospheric Delay	*11	*12	*13	*14	*15
Phase Ambiguity 1b	*16	*17			
Cycle Slips	*18	*19			*19
Antenna Phase Centre Movement					*20
Multipath					*21
Measurement Noise					*

A parameter estimated B bias eliminated by differencing
C bias measured D bias modelled
E bias ignored, and assumed to be an error

Comments to Table 6.3-1:

1 these can be shown to be mathematically equivalent
1b only for integrated carrier phase measurements
2 only practical for high precision GPS geodesy using specialised scientific software
3 orbit bias significantly reduced in baseline solutions -- "rule-of-thumb": baseline error is 1ppm per 20m orbit error.
4 completely free solution for all station coordinates not usually feasible
5 care should be taken to ensure that errors in reference station coordinates are not larger than orbit errors; if orbit is adjusted then this bias is absorbed into estimated orbit parameters
6 feasible if dual-frequency observations are not available
7 if single frequency observations are only available, short baseline results only mildly affected by residual ionosphere (generally highly correlated at both ends of the baseline), and effect tends to be at the 1-2ppm level or below
8 dual-frequency observations are a very satisfactory option for almost entirely eliminating ionospheric bias for many applications
9 broadcast ionospheric model is rather poor, accounting for 50% of effect
10 operational procedures to minimise residual ionospheric effect on baselines, for example, observe at night
11 estimated as a constant scale factor, or more elaborate stochastic parameter
12 as in the case of ionospheric bias, short to medium baseline results only mildly affected by residual wet troposphere component (generally highly correlated at both ends of the baseline), and effect tends to be at the few ppm level
13 can be measured directly by Water Vapour Radiometers, but not feasible for standard GPS surveys
14 models of tropospheric delay accounts for about 90% of bias (at a single station)
15 ignore residual (baseline) tropospheric bias
16 estimate ambiguity as real-valued parameter in an "ambiguity-free" solution
17 elimination of ambiguity by between-epoch observation differencing
18 carried out during data "pre-processing"
19 cycle slips in triple-difference solutions can be treated as outliers and rejected from the solution
20 operational procedures can be used to eliminate effect on baseline solutions
21 avoid multipath environments

6.3.2 ELIMINATING GPS BIASES THROUGH DATA DIFFERENCING

As emphasised in Table 6.3-1, the standard approach to GPS phase data processing is to construct new observables by differencing carrier beat phase observations in such a way that some or all of the clock biases are eliminated, and the impact of several other measurement biases is significantly reduced. The various differencing options are illustrated in Figure 6.3-1.

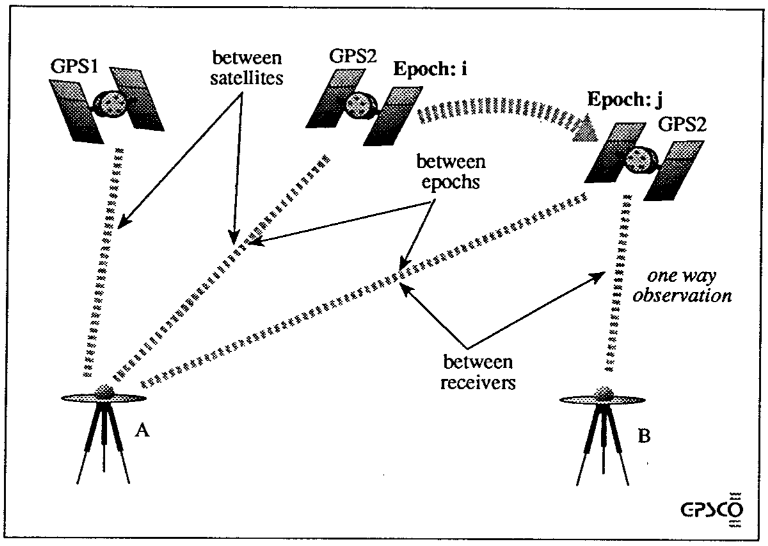


Figure 6.3-1. Differencing options for GPS observations.

Between-Satellite Differences

Receiver clock phase errors may be eliminated by forming the difference between simultaneous observations by one GPS receiver to two satellites (Figure 6.3-2). The operator ∇ indicates a **between-satellite difference**. The differenced observable can be written as:

$$\nabla\phi_j^{12}(T_j) = \phi_{bj}^{1}(T_j) - \phi_{bj}^{2}(T_j) \tag{6.3-1}$$

where it is assumed that the two observations from receiver j to satellites 1 and 2 have been made at the same time T_j. The new differenced observation is (based on eqn (6.1-13), with the inclusion of an orbit error term ε_j^{or} that represents the mapping of orbit error along the line-of-sight between receiver and satellite):

$$\begin{aligned}\nabla\phi_j^{12}(T_j) = {} & (f_0/c).[\, \rho_j^1(T_j) - \rho_j^2(T_j) + \varepsilon_j^{or1}(T_j) - \varepsilon_j^{or2}(T_j)\,] \\ & - f_0.[\, \varepsilon^{sc1}(T_j^1) - \varepsilon^{sc2}(T_j^2)\,] \\ & + n_j^1 - n_j^2 + \nabla\phi_{atmos}\end{aligned} \tag{6.3-2}$$

Note that the receiver clock phase error has been eliminated and the clock bias terms that remain are the between-satellite clock phase error and the between-satellite cycle ambiguity. As far as setting up the design matrix for a Least Squares adjustment based on between-satellite differences is concerned, there is a choice in the cycle ambiguity modelling. Either use individual ambiguity modelling for n_j^1 and n_j^2, or adopt a new definition for the ambiguity parameter consisting of the between-satellite cycle ambiguity k_j^{12} $(= n_j^1 - n_j^2)$.

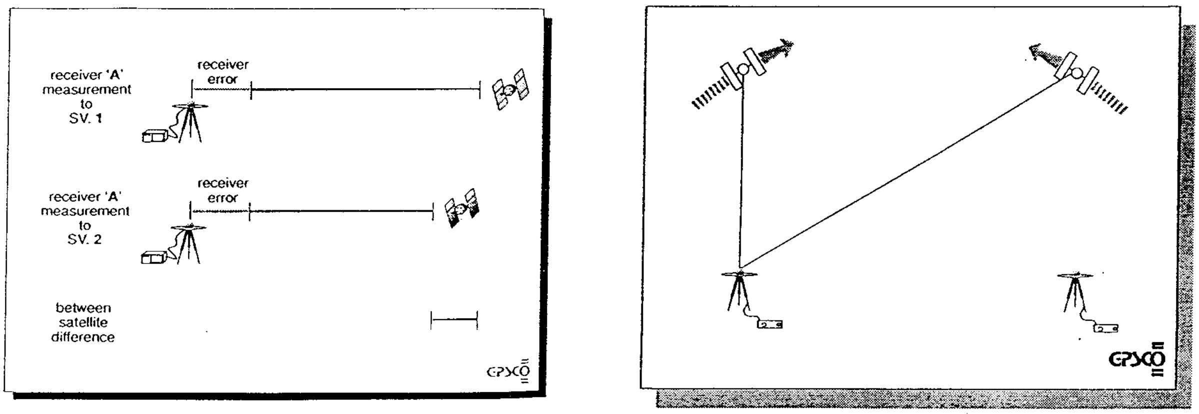


Figure 6.3-2. Between-satellite differencing eliminates receiver dependent biases.

Between-Station Differences

The satellite clock phase errors may be (almost) eliminated by forming the difference between observations by two GPS receivers to one satellite, at the same time. The operator Δ indicates a **between-station difference**, also sometimes referred to confusingly as a *single-difference*. Although the nominal observation or reception times of the two measurements may be the same, the actual time of measurement may differ between one receiver and another because of receiver *time-tag errors*. The differenced observation can be constructed from:

$$\Delta\phi_{12}^i(t) = \phi_{b1}^i(T_1) - \phi_{b2}^i(T_2) \qquad (6.3\text{-}3)$$

where t is the nominal receiver clock time, and T_j is the true time-of-reception of the signal at receiver i.

Using eqn (6.1-13), the between-station difference involving receivers 1 and 2 is:

$$\begin{aligned}\Delta\phi_{12}^i(t) = {} & (f_0/c).[\,\rho_1^i(T_1) - \rho_2^i(T_2)\,] \\ & - f_0.[\,\varepsilon^{sci}(T_1^i) - \varepsilon^{sci}(T_2^i)\,] + f_0.[\,\varepsilon_{rc1}(T_1) - \varepsilon_{rc2}(T_2)\,] \\ & + n_1^i - n_2^i + \Delta\phi_{atmos} \qquad (6.3\text{-}4)\end{aligned}$$

The satellite clock phase errors may not completely cancel because they refer to different transmission times. The transit times of the satellite signals are not equal because of the different satellite - receiver ranges, and this difference may be up to 1 millisecond for a 300km baseline. However, as the satellites use stable atomic oscillators, it is usual to assume that the satellite clock phase errors are identical and thus cancel when the between-station difference is formed (RIZOS & GRANT, 1990):

$$\begin{aligned}\Delta\phi_{12}{}^{i}(t) = (f_0/c).[\ \rho_1{}^{i}(T_1) - \rho_2{}^{i}(T_2)\ ] \\ + f_0.[\ \varepsilon_{rc1}(T_1) - \varepsilon_{rc2}(T_2)\ ] \\ + n_1{}^{i} - n_2{}^{i} + \Delta\phi_{atmos}\end{aligned} \qquad (6.3\text{-}5)$$

Note that the clock bias terms that remain are the between-station clock phase errors and the between-station cycle ambiguity. (The influence of the orbit error terms $\varepsilon_1{}^{ori}$ and $\varepsilon_2{}^{ori}$ for satellite i is such that they have approximately equal magnitude, and will therefore cancel in between-receiver differencing; though this assumption starts to break down with increasing receiver separation.) As in the case of the between-satellite difference, it is possible to either use individual ambiguity modelling for $n_1{}^{i}$ and $n_2{}^{i}$, or to adopt a new definition for the ambiguity parameter consisting of the between-station cycle ambiguity $k_{12}{}^{i}$ (= $n_1{}^{i} - n_2{}^{i}$).

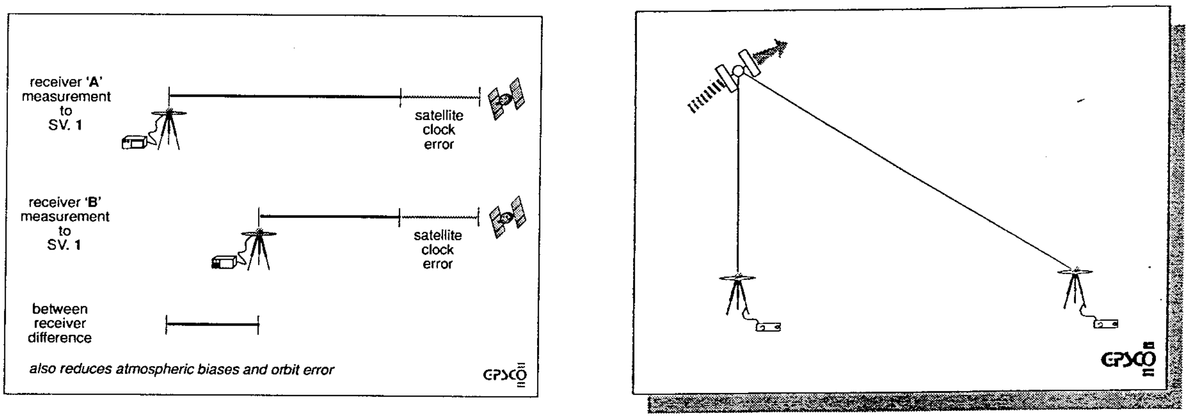


Figure 6.3-3. Between-station differencing to eliminate or reduce satellite dependent biases.

Double-Difference Observable

In these notes the term "double-difference" will refer to the observable which has been formed by differencing between satellites and between stations (Figure 6.3-4). (It is possible to generate other types of double-differences, involving different combinations of between satellite, between receiver, between epoch, and even between frequency, differencing strategies.) This *receiver-satellite* double-differenced phase may be created by forming the between-satellite difference phase and then differencing between stations, or by first forming the between-station difference and then differencing between satellites:

$$\begin{aligned}\phi_{DD}(t) &= \nabla\Delta\phi_{12}{}^{12}(t) \\ &= \nabla\phi_1{}^{12}(T_1) - \nabla\phi_2{}^{12}(T_2) \\ &= \Delta\phi_{12}{}^{1}(t) - \Delta\phi_{12}{}^{2}(t)\end{aligned} \qquad (6.3\text{-}6)$$

Assuming that the satellite clock phase errors cancel in a between-station difference, then it is possible to use eqn (6.3-5) to derive the double-difference observation equation:

$$\phi_{DD}(t) = (f_0/c).[\ \rho_1{}^1(T_1) - \rho_1{}^2(T_1) - \rho_2{}^1(T_2) + \rho_2{}^2(T_2)\ ]$$

$$+ n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2 + \Delta\nabla\phi_{atmos} \qquad (6.3\text{-}7)$$

The only clock biases remaining in this equation are the integer cycle ambiguities. The four individual cycle ambiguities $n_1{}^1$, $n_1{}^2$, $n_2{}^1$ and $n_2{}^2$ may be parameterised separately, or a new definition adopted for the ambiguity parameter consisting of the double-differenced cycle ambiguity $K_{12}{}^{12}$ $(= n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2)$.

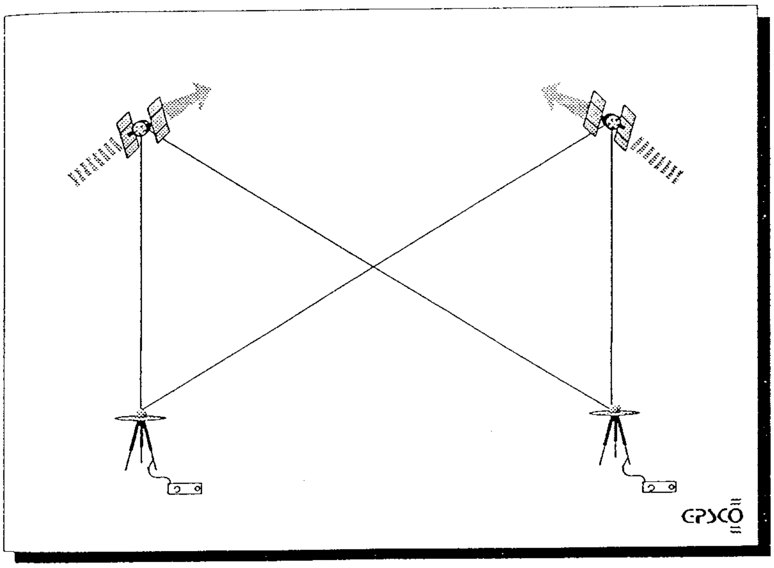


Figure 6.3-4. The double-difference observable.

Triple-Difference Observable

Bias parameters which are constant with time may be eliminated by forming the **between-epoch difference**. The operator δ is used to represent this difference, and is applied to double-difference observables to obtain the **triple-difference** observables (Figure 6.3-5). If the nominal reception times, in receiver clock time, are t_a and t_b, then applying eqn (6.3-7):

$$\phi_{TD}(t_{ab}) = \delta\Delta\nabla\phi_{12}{}^{12}(t_{ab}) = \phi_{DD}(t_a) - \phi_{DD}(t_b)$$

$$= (f_0/c).[\ \rho_1{}^1(T_{a1}) - \rho_1{}^2(T_{a1}) - \rho_2{}^1(T_{a2}) + \rho_2{}^2(T_{a2})\ ]$$

$$- (f_0/c).[\ \rho_1{}^1(T_{b1}) - \rho_1{}^2(T_{b1}) - \rho_2{}^1(T_{b2}) + \rho_2{}^2(T_{b2})\ ]$$

$$+ \delta\Delta\nabla\phi_{atmos} \qquad (6.3\text{-}8)$$

Note that all the clock bias terms, including the integer ambiguities, have been eliminated from this model. If the assumption that the integer ambiguity is constant in time is wrong then an extra term will be required to account for any cycle slip. **In fact, not only have the (unknown) clock biases been eliminated, but in addition, the effect of other biases arising from atmospheric delay, satellite ephemeris error and receiver**

coordinate error, have been substantially reduced by the process of double- and triple-differencing. *However, the noise has been increased and the geometry for positioning has been significantly weakened.*

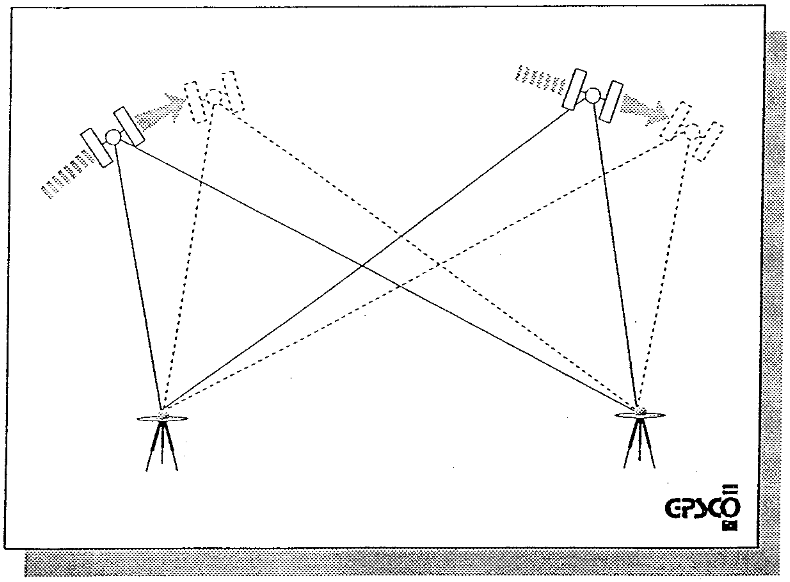


Figure 6.3-5. The triple-difference observable.

The Roles and Merits of Various Phase Observables in GPS Adjustments

The "undifferenced" approach to GPS phase adjustment is identical to using single-, double- or triple-differences only if the *mathematical correlations* introduced during the differencing operation are included in the VCV matrix and any explicitly modelled clock biases are assumed to be offsets (on an epoch-by-epoch basis in the case of the clock phase errors, and on a session-by-session basis for the cycle ambiguity parameters) -- §6.1. Often, however, these conditions for exact equivalence are not met. *Nevertheless the various differenced observables do have their role to play.*

The between-satellite and between-station differences, for example, are useful for cycle slip editing, while triple-differences are often used to obtain preliminary site coordinate solutions because of their relative insensitivity to cycle slips in the phase data (this arises because the mathematical correlations are ignored!). On the other hand, it is the double-differenced observable that is most commonly used in precise GPS phase adjustment. However, a solution based on double-differenced data is relatively intolerant of cycle slips. Furthermore, the mathematical correlations are generally neglected, and hence the results are not strictly equivalent to using the "undifferenced" approach, but they are nevertheless suitable for most survey applications requiring relative accuracies of a few parts per million (BEUTLER et al, 1987).

The impact of cycle slips on the various differenced observables is illustrated in Table 6.3-2. A cycle slip is assumed to have occurred at receiver j while observing satellite i between epochs e-1 and e. All double-differences starting with epoch e are offset by the same amount s, while only one triple-difference is affected by the cycle slip. Another thing to note is that it is not possible to identify which of the four one-way observations contains the original cycle slip, only that a cycle sip has occurred and is noticeable in the double-differenced data from epoch e onwards.

Table 6.3-2. Effect of cycle slip on carrier phase observables.

One-way Phase				Double-Difference	Triple-Difference
Receiver j		Receiver l			
$\phi_j^i(e\text{-}2)$	$\phi_j^k(e\text{-}2)$	$\phi_l^i(e\text{-}2)$	$\phi_l^k(e\text{-}2)$	$\Delta\nabla\phi_{jl}^{ik}(e\text{-}2)$	
$\phi_j^i(e\text{-}1)$	$\phi_j^k(e\text{-}1)$	$\phi_l^i(e\text{-}1)$	$\phi_l^k(e\text{-}1)$	$\Delta\nabla\phi_{jl}^{ik}(e\text{-}1)$	$\delta\Delta\nabla\phi_{jl}^{ik}(e\text{-}2,e\text{-}1)$
$\phi_j^i(e)+s$	$\phi_j^k(e)$	$\phi_l^i(e)$	$\phi_l^k(e)$	$\Delta\nabla\phi_{jl}^{ik}(e)+s$	$\delta\Delta\nabla\phi_{jl}^{ik}(e\text{-}1,e)\text{-}s$
$\phi_j^i(e+1)+s$	$\phi_j^k(e+1)$	$\phi_l^i(e+1)$	$\phi_l^k(e+1)$	$\Delta\nabla\phi_{jl}^{ik}(e+1)+s$	$\delta\Delta\nabla\phi_{jl}^{ik}(e,e+1)$
$\phi_j^i(e+2)+s$	$\phi_j^k(e+2)$	$\phi_l^i(e+2)$	$\phi_l^k(e+2)$	$\Delta\nabla\phi_{jl}^{ik}(e+2)+s$	$\delta\Delta\nabla\phi_{jl}^{ik}(e+1,e+2)$

6.3.3 SIMULTANEITY CONSIDERATIONS FOR GPS PHASE DATA REDUCTIONS

The construction of double- or triple-differenced observables requires that certain constraints apply as far as field operational procedures are concerned:

☞ It is possible to estimate how close the observations to the different satellites must be in order that the receiver clock error cancel in a between-satellite difference. For example, assume a receiver clock stability of 1 part in 10^{10}, or a drift of 0.1 nanosecond in one second, equivalent to 3cm in range. To model the between-satellite difference to an accuracy of 1mm, then **it is possible to tolerate 30 milliseconds difference in time-tags between an observation to one satellite and an observation to another satellite**.

☞ Two or more receivers must track the same satellite at the same time in order that the satellite clock error cancels in the between-receiver difference. Using the same line of reasoning as that above, assuming that the satellite clock is accurate to 1 part in 10^{10} (the satellite oscillator is generally a cesium or rubidium frequency standard and is therefore more stable than this), then **the required accuracy of synchronisation between receiver clocks is of the order of 30 milliseconds**. (There is a difference in transmission time between signals transmitted from the same satellite to two widely separated receivers. For a baseline 300km in length this will be about 1 millisecond, well within the required tolerance.) *In the case of Selective Availability, this tolerance may have to be reduced by a factor of about ten.*

☞ The **observation time-tags <u>errors</u> must also be kept below a certain level**. The time-tag is used to determine the time-of-transmission of the signal. From this information the coordinates of the satellite can be obtained (see Table 3.3-3 in the case of the Broadcast Ephemerides), and hence the calculated receiver-satellite range. The time-tag error has the effect of causing an error in the theoretical (or modelled) range used in the Least Squares adjustment of the double-differenced observation equations.

The first two requirements are relatively modest, however it is necessary to consider more closely the issue of time-tag error:

- **How large can the time-tag error be?** The maximum rate of change of range for GPS satellites (travelling 4km/sec alongtrack) is approximately 700m/sec. To model the geometric range to 1mm accuracy, it is necessary to ensure that the time-tags are accurate to a level of about *1 to 2 microseconds.*
- **How is time-tag error defined?** The Satellite Ephemeris Time scale provides the "true" time scale against which the time "scale" implied by the observation time-tags can be compared. In the case of the Broadcast Ephemerides this is the GPS Time.
- **Must all observation time-tags be related to GPST at the microsecond level?** The time-tag error can be partitioned into two components (RIZOS & GRANT, 1990):
  - a common time-tag error, and
  - a relative time-tag error.

Figures 6.3-6 and 6.3-7 illustrate the effects of *relative* and *common* time-tag errors. In these diagrams it is assumed that the observations by the two receivers have the same time-tags and, for simplicity, that the signal transit times are the same from the satellite to the two receivers. Although, this last assumption is not strictly correct, the motion of the satellite during the time of transit can be modelled accurately (and it allows for a simplification of the diagrams). It is also assumed that the satellite ephemeris is free of error.

The points T1 and T2 in these diagrams refer to the true satellite positions at the time-of-transmission to receivers 1 and 2. These are both at the same point due to the assumption made above. C1 and C2 are the respective calculated positions.

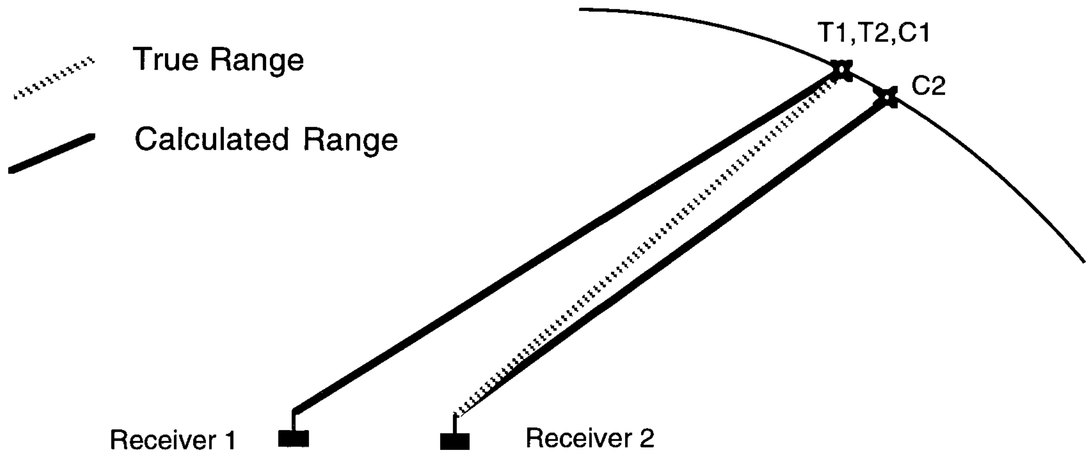


Figure 6.3-6. Relative time-tag error.

In Figure 6.3-6 it is assumed that there is no time-tag error at receiver 1 but a significant error at receiver 2. The calculated and true ranges from receiver 2 are different from each other, whereas those from receiver 1 are the same.

It is the relative time-tag that must be kept below 1-2 microseconds. That is, in order that two ranges involving different receivers and the same satellite are modelled at the 1mm level, the time-tags must be synchronised at the microsecond level.

In the case of the same time-tag error at both receivers there is a difference between calculated and true ranges for both receivers, and this difference is nearly the same for each receiver. Such a common time-tag error would arise, for example, if the time system of the GPS receivers is not the same as for the Satellite Ephemeris Time system. The "rule-of-thumb" commonly quoted for estimating the effect of ephemeris errors on adjusted baselines (§6.2) can be used to estimate the acceptable limits of the common time-tag error. (Dividing the baseline length by the satellite altitude, the ratio that propagates ephemeris error into the adjusted baseline is obtained.)

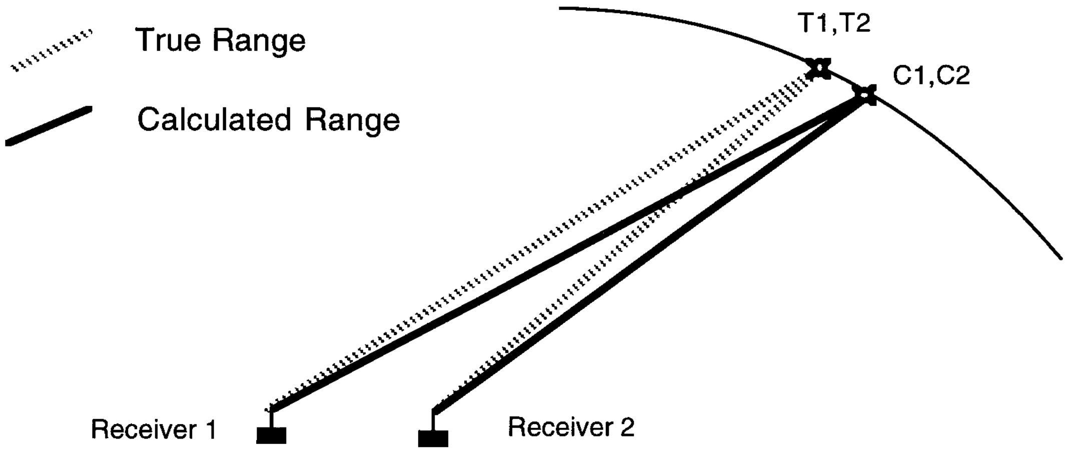


Figure 6.3-7. Common time-tag error.

For example, if the ephemeris error is of the order of 20 metres, and given that the GPS satellite altitude is about 20×10^6 metres, a 1ppm error will be introduced into the baseline (1cm in 10km, 10cm in 100km, etc.). As the GPS satellites travel 20m in approximately 5 milliseconds, and if it is assumed that the ephemeris data has errors of the order of 20m, then a common error in the receiver time-tags of 1 or 2 milliseconds will not add substantially to the systematic errors which have already been introduced by an inaccurate ephemeris.

In summary, the GPS receiver synchronisation issues are:

(1) **All receivers** ***should take observations to common-view satellites at epochs which are within 30 milliseconds of each other with SA off, or a few milliseconds with SA on,*** **to ensure that satellite clock errors cancel in between-receiver differences.**

(2) **Receivers** ***should be synchronised with each other at the microsecond level to ensure that all the observation time-tags are consistent with each other.***

(3) **All receivers** ***should be "externally "synchronised to the Satellite Ephemeris Time scale (in general, GPST) at the millisecond level.***

These various time scale / time error considerations are illustrated in Figure 6.3-8.

The time-tag conditions (points (2) and (3) above) can be met if the GPS navigation solution is used to individually synchronise the receiver clocks to GPST. The receiver clock bias (which defines the offset of the internal clock from GPST) can be determined at a better than 1 microsecond accuracy using the pseudo-range point position solution. If the clock is reset to always read GPST, code-correlating receivers can be considered as being always (automatically) synchronised to each other via GPST. On the other hand, if the receiver clock is not continuously reset to GPST using the navigation solution, then this needs to be done during post-mission analysis of the recorded data (chapter 7). Finally, in the case of earlier non-code-correlating receivers, this synchronisation had to be effected by physically bringing together the receivers at the start of the survey (see KING et al, 1987).

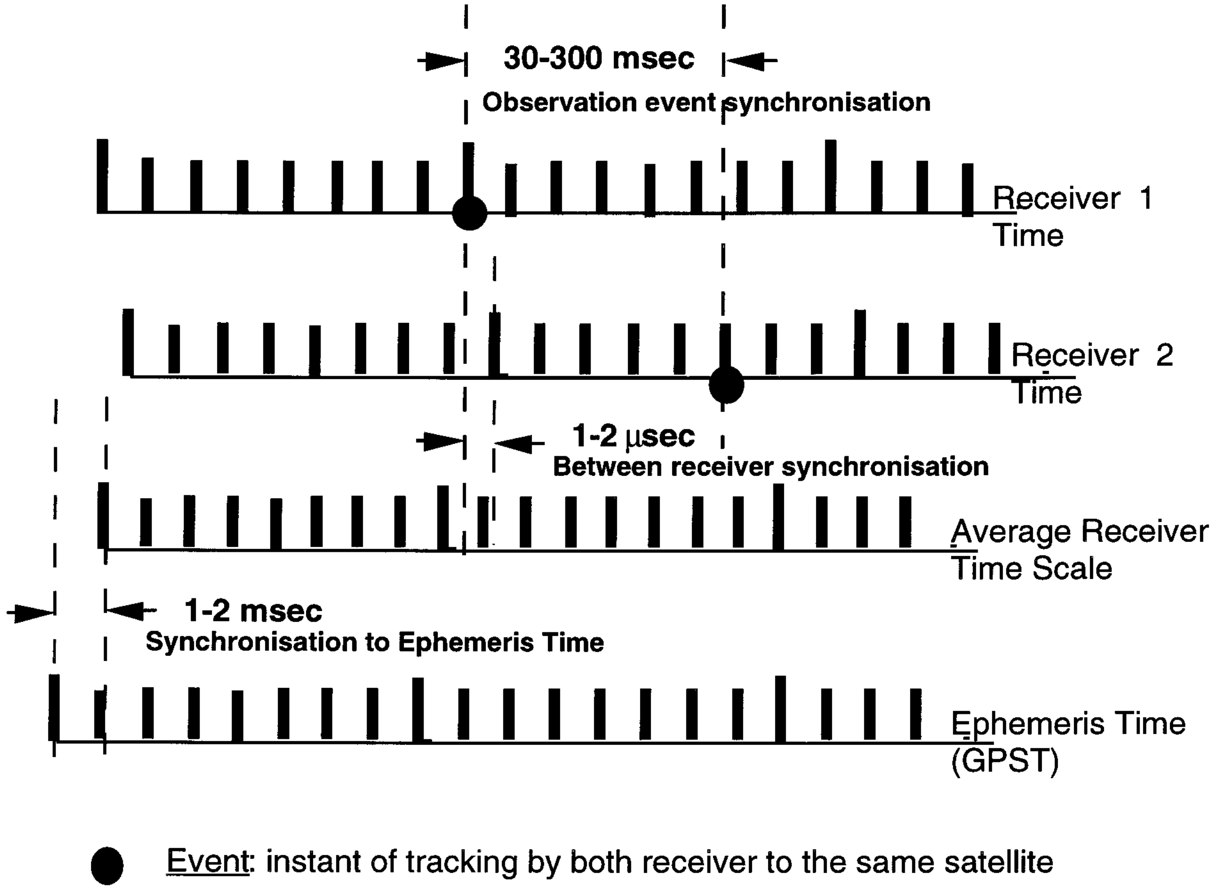


Figure 6.3-8. Elements of GPS receiver synchronisation.

Pseudo-range synchronisation of code-correlating receivers for simultaneous data recording and time-tag initialisation must not be taken for granted. The start time and data recording rate must be the same for all receivers, hence, for example, the receivers may be programmed to collect data at one minute intervals on the GPST minute.

Mixing receivers of different type could lead to problems due to different receiver clock operating scenarios. Some GPS receivers constantly reset their clocks to GPST using the navigation solution, in a process known as "clock steering" -- as in Figure 6.3-9. Other GPS receivers may not do this, simply allowing their receiver clocks to drift controllably, as illustrated in Figure 6.3-10, which shows the total time-tag error (thin curved line) and the "millisecond jumps" in the time-tag value actually attached to the observations at an epoch (thick stepped line). Over a day, GPS receiver clocks have been known to drift many tens of milliseconds! This may cause problems if the data analyst is unaware of the finer points of receiver operation, and the commercial GPS software being used cannot accommodate such time-tag inconsistencies. However, in extreme cases, even if the time-tag error is determined to the required level of accuracy required (a few microseconds), the synchronisation condition to ensure satellite clock error elimination in between-receiver observations may not be satisfied -- that is, *within 30 msec under no SA or a few milliseconds when SA is on.*

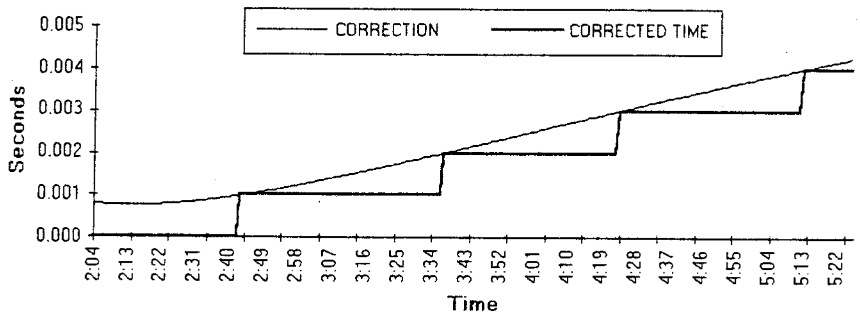


Figure 6.3-9. Drifting receiver clock error and the stepping of the time-tags.

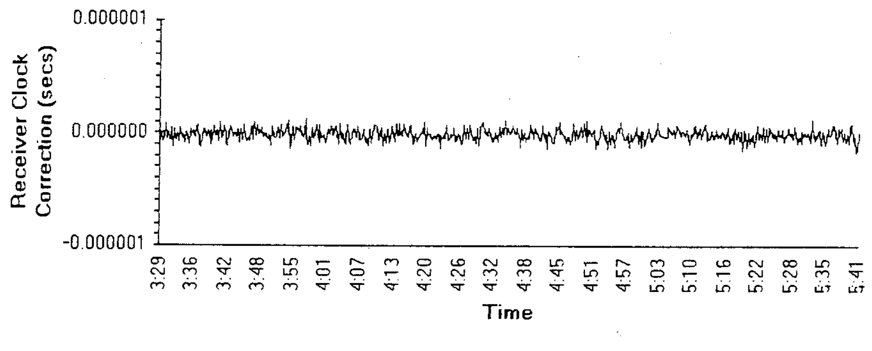


Figure 6.3-10. "Steered" receiver clock and residual clock error.

6.4
DUAL-FREQUENCY RELATIONS

As discussed in §6.2, the L-band signals transmitted by the GPS satellites are delayed as they travel through the ionosphere and the troposphere. The *residual* effects remaining after between-station differencing are generally small for short baselines, and hence are often neglected. For longer interstation distances the ionospheric and tropospheric effects cannot be ignored because:

- these unmodelled biases begin to *degrade the solution by an appreciable amount* (the magnitude of its effect tends to grow in proportion to the baseline length), and
- they make the ambiguity resolution process more difficult because they tend to "destroy" the *integer nature of the ambiguity parameters.*

The ionospheric delay is highly unpredictable, being a function of the latitude of the receiver, the elevation angle to the satellite, the time-of-day of the observations, and the level of solar activity at the time of observation (itself dependent on a number of factors such as the season, the 11 year solar cycle, etc.).

How to handle the ionospheric biases? One way is to use the corrections based on the ionospheric model transmitted by the satellites themselves in the Navigation Message. Although many navigation-style (single frequency) GPS receivers apply these corrections automatically to the pseudo-ranges (§3.3), they are likely to only model 25-50% of the daily variability. Hence, because the ionospheric range bias equivalent is of the order of tens of metres, a significant residual bias remains in the phase observations. Between-station differencing may bring this down to several decimetres in extreme cases, such as for baselines greater than 100km in length, observed in daylight, during a solar cycle maximum.

Another, more appropriate, way of accounting for the ionospheric delay is to observe on both L-band frequencies (§6.2). However, dual-frequency observations offer many possibilities of generating new observables, some with interesting properties and uses. These are discussed below.

6.4.1 LINEAR COMBINATIONS OF DUAL-FREQUENCY OBSERVATIONS

The effect of the ionosphere on GPS observations can be considered in terms of the time delay (τ_{ion}), phase change (ϕ_{ion}) or range (or group) delay (d_{ion}). A first-order approximation for the ionospheric bias is (§6.2):

$$\tau_{ion} = \frac{-\phi_{ion}}{f} = \frac{d_{ion}}{c} \approx (1.35 \times 10^{-7}) \cdot \frac{STEC}{f^2} \qquad (6.4\text{-}1)$$

where:
τ_{ion} is the ionospheric delay in seconds,
ϕ_{ion} is the ionospheric phase delay in cycles,
d_{ion} is the ionospheric group delay in metres,

c is the speed of electromagnetic radiation in a vacuum (m/s),
f is the signal frequency (Hz), and
STEC is the Slant Total Electron Content of a column of ionosphere condensed onto a disc, expressed as the number of free electrons per square metre (el/m^2).

The ionosphere causes the integrated carrier phase count to *decrease* (that is, the apparent phase velocity is greater than the velocity of light!), but causes the pseudo-range to appear longer than the geometric range. ***In the remainder of these notes only the group delay term d_{ion} will be used in the pseudo-range and phase observation equations.*** **Note that the time delay is proportional to the inverse of the frequency squared.** *That is, higher frequencies are less affected by the ionosphere, and hence the ionospheric time delay for L1 observations (1575.42MHz) is less than for L2 observations (1227.60MHz).* From eqn (6.4-1) the relationship between the time delays on the two frequencies f_1 and f_2 due to the ionosphere is:

$$f_1^2 \,.\, d_{ion(L1)} = f_2^2 \,.\, d_{ion(L2)} \tag{6.4-2}$$

Hence the L2 ionospheric effect is approximately 1.647 times that on L1 ($1.647 \approx f_1^2 / f_2^2$).

Ionosphere-Free Combination

This is also sometimes known as the "L3" combination. The ionospheric refraction bias can be eliminated by constructing a combined ionosphere-free phase or pseudo-range observable from the L1 and L2 data.

Consider the phase observations on two frequencies. The mathematical model of the undifferenced (or one-way) carrier phase measurement has been given in eqn (6.1-13), and can be simplified (neglecting: (a) the clock error terms, which are eliminated during double-differencing, and (b) the tropospheric and orbit biases, which are significantly reduced for short baselines) for the L1 frequency as:

$$\phi_j^i(T_j)_{(L1)} = (\frac{f_1}{c}).\rho_j^i(T_j) - (\frac{f_1}{c}).d_{ion(L1)} + n_j^i{}_{(L1)} \tag{6.4-3a}$$

Similarly for the L2 frequency carrier phase:

$$\phi_j^i(T_j)_{(L2)} = (\frac{f_2}{c}).\rho_j^i(T_j) - (\frac{f_2}{c}).d_{ion(L2)} + n_j^i{}_{(L2)} \tag{6.4-3b}$$

Note that $n_j^i{}_{(L1)}$ is not equal to $n_j^i{}_{(L2)}$.

In the following equations the subscripts (receiver identifier), superscripts (satellite identifier) and time arguments are dropped. The L1 ambiguity term is simply expressed as n_1, and similarly the L2 ambiguity is denoted by n_2. $\phi_{(L1)}$ denotes the phase of the L1 signal in cycles, $\phi_{(L2)}$ refers to the phase of the L2 signal in cycles.

Multiplying each of the above phase observations (in units of cycles) by the signal frequency, and then differencing them gives:

$$f_1.\phi_{(L1)} - f_2.\phi_{(L2)} = \frac{f_1^2 - f_2^2}{c} . \rho$$
$$- \frac{1}{c} . (f_1^2.d_{ion(L1)} - f_2^2.d_{ion(L2)})$$
$$+ f_1.n_1 - f_2.n_2 \tag{6.4-4}$$

where the second term of right hand side of eqn (6.4-4) is equal to zero due to the relation at eqn (6.4-2). Hence the equation degenerates to:

$$\frac{f_1.\phi_{(L1)} - f_2.\phi_{(L2)}}{f_1^2 - f_2^2} = \frac{1}{c}.\rho + \frac{f_1.n_1 - f_2.n_2}{f_1^2 - f_2^2} \tag{6.4-5}$$

In order to combine the L1 and L2 phase observations, which are in units of cycles (of different wavelengths for L1 and L2), they have to be converted to the same units, for example, scaling by the L1 frequency:

$$\frac{f_1.(f_1.\phi_{(L1)} - f_2.\phi_{(L2)})}{f_1^2 - f_2^2} = \frac{f_1}{c}.\rho + \frac{f_1.(f_1.n_1 - f_2.n_2)}{f_1^2 - f_2^2} \tag{6.4-6}$$

yields the following *corrected* L1 phase measurement:

$$\phi_{(L1)ion\text{-}free} = \alpha_1.\phi_{(L1)} + \alpha_2.\phi_{(L2)}$$
$$= \phi_{(L1)} - (\frac{f_2}{f_1}).\phi_{(L2)}$$
$$= (\frac{f_1}{c}).\rho + \alpha_1.n_1 + \alpha_2.n_2 \tag{6.4-7a}$$

where

$$\alpha_1 = \frac{f_1^2}{f_1^2 - f_2^2} \approx 2.546 \quad \text{and} \quad \alpha_2 = \frac{- f_1 f_2}{f_1^2 - f_2^2} \approx -1.984$$

Alternatively, if eqn (6.4-5) is scaled by the L2 frequency, the *corrected* L2 phase measurement is obtained:

$$\phi_{(L2)ion\text{-}free} = \beta_1.\phi_{(L1)} + \beta_2.\phi_{(L2)}$$
$$= \phi_{(L2)} - (\frac{f_1}{f_2}).\phi_{(L1)}$$
$$= (\frac{f_2}{c}).\rho + \beta_1.n_1 + \beta_2.n_2 \tag{6.4-7b}$$

where

$$\beta_1 = \frac{f_1 f_2}{f_1^2 - f_2^2} \approx 1.984 \quad \text{and} \quad \beta_2 = \frac{- f_2^2}{f_1^2 - f_2^2} \approx -1.54$$

Eqn (6.4-7) has the exact form of the original raw carrier beat phase observation eqn (6.1-13), except that the integer ambiguity term is replaced by the linear combination of the L1 and L2 ambiguities. This is also called the <u>L3 linear combination</u> of the phase data, or simply the *L3 ionosphere-free observable*. The relationship between the L1, L2 and L3 cycles, measured in units of L1 wavelengths, is:

$$L3[(L1)cycles] \approx 2.546L1[(L1)cycles] - 1.984L2[(L2)cycles] \qquad (6.4\text{-}8a)$$

and measured in L2 wavelengths is:

$$L3[(L2)cycles] \approx 1.984L1[(L1)cycles] - 1.54L2[(L2)cycles] \qquad (6.4\text{-}8b)$$

The n_3 ambiguity is related to the L1 and L2 ambiguities as follows:

$$n_3[(L1)cycles] \approx 2.546n_1 - 1.984n_2 \qquad (6.4\text{-}9a)$$

$$n_3[(L2)cycles] \approx 1.984n_1 - 1.54n_2 \qquad (6.4\text{-}9b)$$

Hence n_3 is not an integer combination of L1 and L2 ambiguities! An alternate form with integer coefficients is introduced in the discussion following Table 6.4-1.

It should be emphasised that $\phi_{(L1)}$ is the measured L1 phase in units of λ_1 wavelengths (≈0.19m), while $\phi_{(L2)}$ is the measured L2 phase in units of λ_2 wavelengths (≈0.24m). In order to obtain the L3 observable in units of metres, both sides of eqn (6.4-7a) have to be multiplied by λ_1, or both sides of eqn (6.4-7b) have to be multiplied by λ_2.

The ionosphere-free combination of pseudo-ranges is derived in the following manner. A convenient model of the L1 and L2 pseudo-range observations is (note that the sign of the ionospheric delay is *reversed* compared to the carrier phase equation):

$$P_{(L1)} = \rho + d_{ion(L1)} \qquad (6.4\text{-}10a)$$

$$P_{(L2)} = \rho + d_{ion(L2)} \qquad (6.4\text{-}10b)$$

Multiplying eqn (6.4-10) by the signal frequency squared, and then differencing gives:

$$f_1^2.P_{(L1)} - f_2^2.P_{(L2)} = (f_1^2 - f_2^2) \,.\, \rho + (\, f_1^2.d_{ion(L1)} - f_2^2.d_{ion(L2)}\,) \qquad (6.4\text{-}11)$$

The <u>equivalent pseudo-range L3 combination</u> therefore is (last term in eqn (6.4-11) is zero):

$$P_{(L3)} = \rho = \frac{f_1^2.P_{(L1)} - f_2^2.P_{(L2)}}{f_1^2 - f_2^2} \qquad (6.4\text{-}12)$$

Unfortunately the L3 combinations (phase or pseudo-range) have approximately three times the noise of the L1 observations (Table 6.4-1).

Note: all the expressions derived here are just as valid for the double-differenced observables as they are for the one-way observations. For example, eqn (6.4-7) can be also written as:

$$\Delta\nabla\phi_{(L1)ion\text{-}free} = \left(\frac{f_1}{c}\right).\Delta\nabla\rho + \alpha_1.\Delta\nabla n_1 + \alpha_2.\Delta\nabla n_2 \qquad (6.4\text{-}13)$$

Geometry-Free or Ionosphere Combination

This is also sometimes referred to as the "L4" combination. It is possible to isolate the ionospheric component using eqns (6.4-3). Both measurements in these equations are converted to metric units (scaling eqn (6.4-3a) by $\lambda_1 = c/f_1$, and eqn (6.4-3b) by $\lambda_2 = c/f_2$), and then differenced to yield:

$$\begin{aligned}\Phi_{ion} = \Phi_{(L4)} &= \Phi_{(L1)} - \Phi_{(L2)} \\ &= \lambda_1.n_1 - \lambda_2.n_2 - d_{ion(L1)} + d_{ion(L2)}\end{aligned} \qquad (6.4\text{-}14a)$$

or, in terms of only the L1 ionospheric delay (using eqn (6.4-2)):

$$\begin{aligned}\Phi_{ion} = \Phi_{(L4)} &= \Phi_{(L1)} - \Phi_{(L2)} \\ &= \lambda_1.n_1 - \lambda_2.n_2 + 0.647.d_{ion(L1)}\end{aligned} \qquad (6.4\text{-}14b)$$

Note that the first two terms are constants, *hence any variation in the L4 combination represents entirely the variation in L1 ionosphere effect* (but it is only 0.647 the effect on the L1 observation) unless there is a cycle slip on L1 (that is, n_1 is not a constant) or L2 (n_2 is not a constant). Figure 6.4-1 illustrates a typical example of a smoothly varying L4 signature derived from L1 and L2 phase data. The ionospheric delay changes slowly by about 0.9m over a four hour period. A "jump" in this signature could be interpreted as cycle slips on L1 and/or L2. Eqn (6.4-14) contains no receiver-satellite range term, hence the reason for it being referred to as a "geometry-free" quantity. The noise on the L4 combination is $\sqrt{2}$ times that of the L1 noise, in metric units -- if the L1 and L2 measurement noises when expressed in metric units are equal (Table 6.4-1).

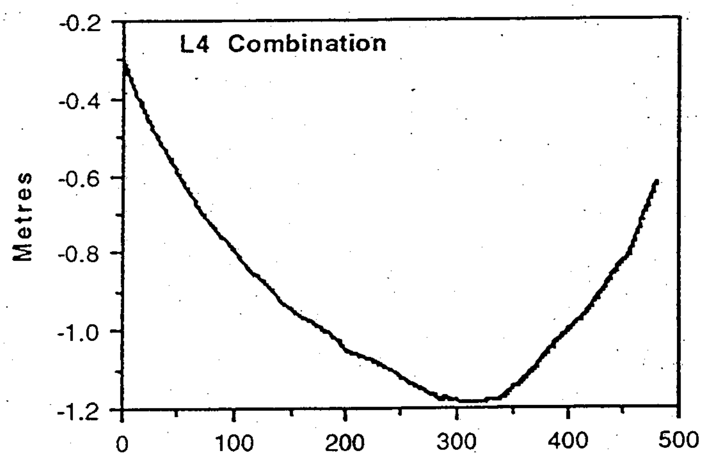


Figure 6.4-1. Typical L4 undifferenced observable signature for a four hour period. (GPS Time 7:50-11:50, day 208, 1992, PRN2, Tidbinbilla site, 30 sec data).

In the case of pseudo-range observations on L1 and L2, it is possible to write the L4 combination directly, and using eqns (6.4-2) and (6.4-10):

$$\begin{aligned} P_{(L4)} = P_{(L1)} - P_{(L2)} &= d_{ion(L1)} - d_{ion(L2)} \\ &= d_{ion(L1)} - \left(\frac{f_1^2}{f_2^2}\right).d_{ion(L1)} \\ &= -0.647.d_{ion(L1)} \end{aligned} \qquad (6.4\text{-}15)$$

Eqn (6.4-15) implies that the *ionospheric delay can be measured directly with two P code pseudo-range observations*. This is not entirely correct because what is missing from the observation model is the data "noise" and multipath, both of which can be serious disturbances at the metre level or greater. So large in fact that they may "swamp" the ionospheric signal (which is already muted by the 0.647 factor). If a large enough data series is available, then the noise and multipath can be averaged out by "smoothing" the $P_{(L4)}$ series, or by curve-fitting a polynomial to the data series.

"Wide-Lane" Combination

This is also sometimes referred to as the "L5" combination. Instead of subtracting the L2 observation from the L1 observation in metric units, the phase equations can be differenced ((6.4-3a) – (6.4-3b)):

$$\begin{aligned} \phi_{(L1)} - \phi_{(L2)} = \phi_{(L5)} &= \frac{f_1 - f_2}{c} . \rho \\ &\quad - \frac{1}{c} . (f_1.d_{ion(L1)} - f_2.d_{ion(L2)}) + n_1 - n_2 \end{aligned} \qquad (6.4\text{-}16)$$

This is known as the "wide-lane" combination because the effective wavelength of the resultant observable is $\lambda_5 = \frac{c}{f_1 - f_2} \approx 0.86$ metres.

The n_5 ambiguity is:

$$n_5[\text{cycles}] = n_1 - n_2$$

Hence n_5 is an integer, however each cycle has a 0.86m wavelength. Eqn (6.4-16) can be converted to metric units through multiplication with the wavelength λ_5:

$$\Phi_{(L5)} = \rho - \frac{1}{f_1 - f_2} . (f_1.d_{ion(L1)} - f_2.d_{ion(L2)}) + \lambda_5.n_5 \qquad (6.4\text{-}17a)$$

or in terms of the L1 ionospheric delay (using eqn (6.4-2)):

$$\Phi_{(L5)} = \rho + \frac{f_1}{f_2} . d_{ion(L1)} + \lambda_5.n_5 \qquad (6.4\text{-}17b)$$

Hence the L5 ionospheric effect is approximately 1.28 times that affecting L1 observations ($1.28 \approx f_1 / f_2$). However, the noise on L5 observables is approximately six times (Table 6.4-1) that on L1 observables!

It can be shown that eqn (6.4-17) is the same expression for the L5 observable if constructed from the P1 and P2 pseudo-range observations except that there is no ambiguity term, and the sign of the ionospheric delay is *reversed* :

$$P_{(L5)} = \rho - \frac{f_1}{f_2} \cdot d_{ion(L1)} = \frac{f_1.P_{(L1)} - f_2.P_{(L2)}}{f_1 - f_2} \tag{6.4-17c}$$

(Hint: convert eqns (6.4-10) to cycles, difference the expressions and then scale them back to metric units using λ_5). The wide-lane pseudo-range is therefore *advanced* by the ionosphere (range too short), while the wide-lane phase is *delayed* by the ionosphere (phase-range too long).

"Narrow-Lane" Combination

The L6 or "narrow-lane" observable is obtained by adding the two phase equations (L1+L2) expressed in cycles:

$$\begin{aligned} \phi_{(L1)} + \phi_{(L2)} = \phi_{(L6)} &= \frac{f_1 + f_2}{c} \cdot \rho \\ &- \frac{1}{c} \cdot (f_1.d_{ion(L1)} + f_2.d_{ion(L2)}) + n_1 + n_2 \end{aligned} \tag{6.4-18}$$

The effective wavelength of this observable is very small: $\lambda_6 = \frac{c}{f_1+f_2} \approx 0.11$ metres. The n_6 ambiguity is $n_1 + n_2$, and is therefore still an integer.

In metric units, upon multiplying eqn (6.4-18) by the wavelength λ_6:

$$\Phi_{(L6)} = \rho - \frac{1}{f_1 + f_2} \cdot (f_1.d_{ion(L1)} + f_2.d_{ion(L2)}) + \lambda_6.n_6 \tag{6.4-19a}$$

or in terms of the L1 ionospheric delay (using eqn (6.4-2)):

$$\Phi_{(L6)} = \rho - \frac{f_1}{f_2} \cdot d_{ion(L1)} + \lambda_6.n_6 \tag{6.4-19b}$$

The L6 ionospheric effect is approximately 1.28 times that on L1 ($1.28 \approx f_1 / f_2$). However the noise on the L6 observables is very small (Table 6.4-1).

Another expression for the L3 ionosphere-free observable is to sum eqns (6.4-17) and (6.4-19) and divide by two:

$$\Phi_{(L3)} = \frac{\Phi_{(L5)} + \Phi_{(L6)}}{2}$$

$$= \rho + \frac{\lambda_5}{2}.n_5 + \frac{\lambda_6}{2}.n_6 \qquad (6.4\text{-}20a)$$

Similarly, an expression for the P3 ionosphere-free combination is:

$$P_{(L3)} = \rho = \frac{P_{(L5)} + P_{(L6)}}{2} \qquad (6.4\text{-}20b)$$

In the case of pseudo-range observations the equivalent narrow-lane expression is:

$$P_{(L6)} = \rho + \frac{f_1}{f_2} \cdot d_{ion(L1)} = \frac{f_1.P_{(L1)} + f_2.P_{(L2)}}{f_1 + f_2} \qquad (6.4\text{-}21)$$

(Hint: convert eqns (6.4-10) to cycles, sum the expressions and then scale them back to metric units using λ_6). The narrow-lane pseudo-range is therefore *delayed* by the ionosphere (range too long), while the narrow-lane phase is *advanced* by the ionosphere (phase-range too short). This is the opposite to what happens with the wide-lane observables.

Other Dual-Frequency Phase Combinations

There are other linear combinations of L1 and L2 observables that could be constructed, leading to a variety of observables with different wavelengths, ionospheric amplification factors and data noise characteristics (see, for example, ABIDIN, 1993). Desirable features of the *artificial* observables that can be constructed from the L1 and L2 observations, for data processing purposes, are:

- the effective wavelength is not too short, not too long,
- the ionospheric delay is small,
- the measurement noise is small, and to aid ambiguity resolution,
- the ambiguity is an integer

The linear combinations of the L1 and L2 phases which preserve the integer nature of the cycle ambiguity, can be expressed as (in units of cycles):

$$\phi_{i,j} = i.\phi_{(L1)} + j.\phi_{(L2)} \qquad (6.4\text{-}22a)$$

where i and j are integer constants. The following are some properties of the linear combination observables (IBID, 1993).

The cycle ambiguity:

$$n_{i,j} = i.n_1 + j.n_2 \qquad (6.4\text{-}22b)$$

The frequency:

$$f_{i,j} = i.f_1 + j.f_2 \qquad (6.4\text{-}22c)$$

The effective wavelength:

$$\lambda_{i,j} = c / f_{i,j} \tag{6.4-22d}$$

The ionospheric effect:

$$d_{ion(i,j)} = \frac{i.f_1.d_{ion(L1)} + j.f_2.d_{ion(L2)}}{i.f_1 + j.f_2} \tag{6.4-22e}$$

The ionospheric scale factor isf:

$$d_{ion(i,j)} = isf.d_{ion(L1)}$$

where

$$isf = \frac{f_1}{f_2} \cdot \left(\frac{i.f_2 + j.f_1}{i.f_1 + j.f_2}\right) \tag{6.4-22f}$$

The noise scale factor nsf:

$$\sigma_{(i,j)} = nsf.\sigma_{(\phi 1)}$$

where

$$nsf = \frac{\lambda_2 \,.\, \sqrt{i^2 + j^2}}{i.\lambda_2 + j.\lambda_1} \tag{6.4-22g}$$

and $\sigma_{(\phi 1)}$ is the standard deviation of the L1 phase observation (in metres, and assuming equal noise on L1 and L2 when expressed in metric units).

The parameters in the columns in Table 6.4-1 have been computed using the relations in eqn (6.4-22).

Table 6.4-1. Some common linear combinations of L1 and L2 phase observations.

Phase Combination	$\lambda_{i,j}$ (m)	Noise nsf x L1	Ion. delay isf x $d_{ion(L1)}$	Ambiguity
L1	0.190	1.0	1.0	n_1
L2	0.244	1.28	1.65	n_2
L3	0.190	3.2	0.0	$\alpha_1 n_1 + \alpha_2 n_2$
	0.244		(or)	$\beta_1 n_1 + \beta_2 n_2$
L4		1.63	-0.65	$\lambda_1 n_1 - \lambda_2 n_2$
L5	0.862	6.4	-1.28	$n_1 - n_2$
L6	0.107	0.8	1.28	$n_1 + n_2$

nsf is the noise scale factor to relate the data "noise" of the linear combination to that of the L1 observation.
isf is the magnitude of the ionospheric scale factor needed to relate the ionospheric delay on the linear combination to the ionospheric delay on the L1 observation.

Some well known linear combinations are:

• wide-lane:	$i = 1,$	$j = -1$	(L5)
• narrow-lane:	$i = 1,$	$j = 1$	(L6)
• first GPS signal:	$i = 1,$	$j = 0$	(L1)
• second GPS signal:	$i = 0,$	$j = 1$	(L2)
• double wide-lane:	$i = -3,$	$j = 4$	
• half wide-lane:	$i = 2,$	$j = -2$	
• semi wide-lane:	$i = -1,$	$j = 2$	
• ionosphere-free:	$i = 77,$	$j = -60$	
• "monster" wide-lane:	$i = -7,$	$j = 9$	

Two interesting combinations are worth closer inspection. The "monster wide-lane" combination has the largest of all the wavelengths, being of the order of 14.65m! However, it suffers in two respects: (a) the noise scale factor is about 878 (if the noise on L1 is 1mm, the noise on $\phi_{-7,9}$ is 0.878m!), and (b) the ionospheric bias is approximately 350 times the delay on L1! An analysis eqn (6.4-22e) results in another ionosphere-free combination that can be constructed with *integer coefficients* $i = 77, j = -60$. The effective wavelength of the $\phi_{77,-60}$ combination is very small, $\lambda_{77,-60} \approx 0.006$m.

A number of linear combinations of L1 and L2 observations derived from certain codeless GPS receivers, for which the effective wavelength of the L2 measurement is half that of the L2 signal (§4.1), are useful. For example, the "half wide-lane" is the linear combination that is analogous to the standard L5 combination, constructed using the full wavelength L1 and the half wavelength L2 measurement. The effective wavelength of this combination is ≈43cm, however the ionospheric scale factor and the noise scale factor are the same as for the L5 combination.

The more "exotic" dual-frequency combinations will not be discussed in these notes. Some useful combinations of dual-frequency phase and dual-frequency pseudo-range data will be introduced below.

6.4.2 CARRIER PHASE AND PSEUDO-RANGE COMBINATIONS

The combination of single or dual-frequency phase and pseudo-range data can be useful for several purposes:

- Cycle ambiguity resolution (§8.2, §8.3 and §8.4).
- Cycle slip detection (§8.4).
- Ionospheric studies.
- Pseudo-range multipath studies.
- Smoothing of pseudo-range data.

To aid the following discussion the pseudo-range and carrier phase observation equations are expressed as follows (see eqn (6.4-3) and (6.4-10), without subscripts and superscripts, and without the explicit inclusion of the tropospheric, orbit and clock error bias terms, but with the addition of the measurement noise terms):

$$\Phi_{(L1)} = \rho - d_{ion(L1)} + \lambda_1.n_{(L1)} + \varepsilon_{(\phi1)} \quad (6.4\text{-}23a)$$

$$P_{(L1)} = \rho + d_{ion(L1)} + \varepsilon_{(P1)} \quad (6.4\text{-}23b)$$

$$\Phi_{(L2)} = \rho - d_{ion(L2)} + \lambda_2.n_{(L2)} + \varepsilon_{(\phi2)} \quad (6.4\text{-}23c)$$

$$P_{(L2)} = \rho + d_{ion(L2)} + \varepsilon_{(P2)} \quad (6.4\text{-}23d)$$

The following comments can be made:

- The *differences* between the pseudo-range data and the carrier phase data are: (a) the ambiguity terms, (b) the reversal in the sign for the ionospheric delay, and (c) the significantly larger noise term associated with pseudo-range data.
- The multipath terms are not included, however, the pseudo-range data is *more sensitive to multipath* (hence a larger magnitude term) than carrier phase data.
- The additive combination of pseudo-range and carrier phase data (made with respect to the same frequency carrier wave) leads to the *elimination of the ionospheric delay.*
- The subtractive combination of pseudo-range and carrier phase data (made with respect to the same frequency carrier wave) leads to the *elimination of the geometric range term* and hence leads to the *highlighting* of the ionospheric delay for further study.
- The additive combination of L1 pseudo-range and L1 carrier phase data results in the *elimination of the ionospheric delay on L1*, while the additive combination of L2 pseudo-range and L2 carrier phase data results in the *elimination of the ionospheric delay on L2.* The subtractive combination of these two L1 and L2 observables results in the *elimination of the geometric range term* and the *elimination of the ionospheric delay.*
- The subtractive combination of L1 pseudo-range and L2 pseudo-range data results in the *elimination of the geometric term and isolates the difference in ionospheric delay between L1 and L2*, while the subtractive combination of L1 carrier phase and L2 carrier phase data also results in the *elimination of the geometric term and isolates the difference in ionospheric delay between L1 and L2.*

Pseudo-Range & Phase Data Combinations for Ambiguity Determination

The combination of eqns (6.4-14) and (6.4-15) leads to the following "geometry-free" relation (neglecting noise terms):

$$\Phi_{(L4)} = -P_{(L4)} + \lambda_1.n_1 - \lambda_2.n_2 \quad (6.4\text{-}24)$$

Eqn (6.4-17) can be modified to include the $P_{(L6)}$ narrow-lane component (eqn (6.4-21)):

$$\Phi_{(L5)} = P_{(L6)} + \lambda_5.n_5 = P_{(L6)} + \lambda_5.(n_1 - n_2) \quad (6.4\text{-}25)$$

where the wide-lane ambiguity can be isolated (note, small noise on $P_{(L6)}$):

$$n_5 = \frac{1}{\lambda_5}.(\Phi_{(L5)} - P_{(L6)}) \quad (6.4\text{-}26a)$$

Using a similar approach leads to an ionosphere-free expression for the narrow-lane ambiguity (note, unlikely to be useful because of high noise on $P_{(L5)}$):

$$n_6 = \frac{1}{\lambda_6}.(\Phi_{(L6)} - P_{(L5)}) \tag{6.4-26b}$$

A combination of eqns (6.4-24) and (6.4-25) leads to the following "four observable" linear combination from which the L1 and L2 ambiguities have been isolated:

$$\begin{aligned} n_1 &= \frac{\lambda_5.(\Phi_{(L4)} + P_{(L4)}) - \lambda_2.(\Phi_{(L5)} - P_{(L6)})}{\lambda_5.(\lambda_1 - \lambda_2)} \\ n_2 &= \frac{\lambda_5.(\Phi_{(L4)} + P_{(L4)}) - \lambda_1.(\Phi_{(L5)} - P_{(L6)})}{\lambda_5.(\lambda_1 - \lambda_2)} \end{aligned} \tag{6.4-27}$$

This relation has a role to play both in aiding ambiguity resolution and for cycle slip detection and repair.

Under Anti-Spoofing it may not be possible to have P code pseudo-ranges from some dual-frequency instruments (some instruments are capable of outputting $P_{(L4)}$ directly but not its components $P_{(L1)}$ and $P_{(L2)}$). Often there will only be *three observables* : $P_{(L1)}$ (usually just the C/A code pseudo-range measurement), $\Phi_{(L1)}$ and $\Phi_{(L2)}$. A linear combination of these (after eliminating the geometric range and ionospheric term from the relevant observation equations) is:

$$a_3.P_{(L1)} + b_3.\Phi_{(L1)} + c_3.\Phi_{(L2)} = n_1 - d_3.n_2 \tag{6.4-28}$$

where the coefficients are:

$$\begin{aligned} a_3 &= -(\,b_3 + c_3\,) \approx -1.2844 \\ b_3 &= \frac{1}{\lambda_1} \approx 5.255 \\ c_3 &= \frac{-2b_3}{(1 + f_1^2 / f_2^2)} \approx -3.9706 \\ d_3 &= -c_3 \,.\, \lambda_2 \approx -0.9697 \end{aligned}$$

Pseudo-Range & Phase Data Combinations for Multipath Studies

The interesting characteristic about eqn (6.4-28) is that it is both "geometry-free" (as in the case of eqn (6.4-14)) and "ionosphere-free" (as in the case of eqn (6.4-7)). Furthermore, what is missing in the above expression are the components of pseudo-range and carrier phase noise, and the corresponding multipath components.

Under the assumption that multipath and noise in the carrier phase measurements is negligible in comparison to the $P_{(L1)}$ multipath and noise, this linear combination of code and phase essentially gives the C/A or P1 multipath $mp_{(P1)}$ and noise $\varepsilon_{(P1)}$ -- offset by a constant component due to the carrier phase ambiguities:

$$a_3.P_{(L1)} + b_3.\Phi_{(L1)} + c_3.\Phi_{(L2)} = a_3.(mp_{(P1)} + \varepsilon_{(P1)}) + n_1 - d_3.n_2 \qquad (6.4\text{-}29a)$$

or, setting $K_1 = (n_1 - d_3.n_2)/a_3$:

$$P_{(L1)} - 4.0915.\Phi_{(L1)} + 3.0915.\Phi_{(L2)} = mp_{(P1)} + \varepsilon_{(P1)} + K_1 \qquad (6.4\text{-}29b)$$

A similar expression for multipath on the L2 pseudo-range is:

$$P_{(L2)} - 5.0915.\Phi_{(L1)} + 4.0915.\Phi_{(L2)} = mp_{(P2)} + \varepsilon_{(P2)} + K_2 \qquad (6.4\text{-}30)$$

Figure 6.4-2 illustrates the values of ($mp_{(P1)} + \varepsilon_{(P1)}$) for two brands of GPS receivers over an approximately two hour tracking period at the same site. Note, in one case it is multipath on the C/A pseudo-range that is plotted.

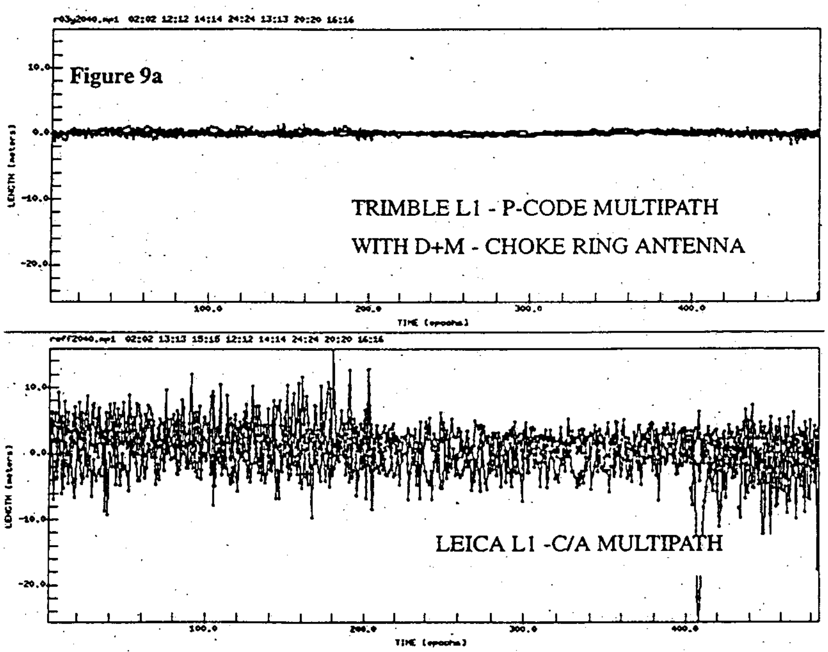


Figure 6.4-2. L1 pseudo-range multipath + noise comparison of Trimble and Leica GPS receivers.

General Form: Pseudo-Range & Phase Data "Ionosphere-Free" Combinations

A general form of the linear combination of dual-frequency pseudo-range and carrier phase data for the construction of "ionosphere-free" / "geometry-free" observables that can be used for ambiguity estimation is given in HAN & RIZOS (1995a). The integer ambiguities of any such four observable combination is given by:

$$n_{i,j} = i.\phi_{(L1)} + j.\phi_{(L2)} + a.P_{(L1)} + b.P_{(L2)} \qquad (6.4\text{-}31a)$$

where

$$n_{i,j} = i.n_1 + j.n_2 \tag{6.4-31b}$$

$$a = -\left(i.\frac{f_1^2 + f_2^2}{f_1^2 - f_2^2} + j.\frac{2f_1f_2}{f_1^2 - f_2^2} \right) \frac{1}{\lambda_1} = -\frac{9240(i+j)+289.i}{2329.\lambda_1} \tag{6.4-31c}$$

$$b = \left(i.\frac{2f_1f_2}{f_1^2 - f_2^2} + j.\frac{f_1^2 + f_2^2}{f_1^2 - f_2^2} \right) \frac{1}{\lambda_2} = \frac{9240(i+j)+289.j}{2329.\lambda_2} \tag{6.4-31d}$$

with an effective wavelength of:

$$\lambda_{i,j} = c \,/\, (i.f_1 + j.f_2) \tag{6.4-31e}$$

The critical issue with regards to such ambiguity estimation is the *standard deviation of such estimated ambiguities*, taking into account the measurement noise of the observable types and how they propagate into the combinations. Obviously, if the standard deviation is greater than the magnitude of the effective wavelength, the estimated ambiguities are weakly determined and resolving them to their nearest likely integer values is not possible.

Applying the Law of Propagation of Variances to eqn (6.4-31a), the standard deviation of the estimated ambiguities is (in units of cycles of the effective wavelength):

$$\mu_{(i,j)} = \sqrt{i^2.\sigma^2_{(\phi 1)} + j^2.\sigma^2_{(\phi 2)} + a^2.\sigma^2_{(P1)} + b^2.\sigma^2_{(P2)}} \tag{6.4-32}$$

where $\sigma_{(\phi 1)}$, $\sigma_{(\phi 2)}$, $\sigma_{(P1)}$ and $\sigma_{(P2)}$ are the standard deviations of the L1, L2 phase in cycles and L1, L2 pseudo-range measurements in metres respectively.

An expression for the *noise scale factor*, relating the data "noise" of the linear combination to that of the L1 phase observation is:

$$nsf = (\mu_{(i,j)} \,.\, \lambda_{i,j}) \,/\, (\sigma_{(\phi 1)} \,.\, \lambda_1) \tag{6.4-33}$$

Table 6.4-2 summarises the characteristics of some useful dual-frequency carrier phase / pseudo-range ionosphere-free combinations. Note that in the case of some of the combinations, the standard deviation is dominated by the uncertainty in the pseudo-ranges. In particular, note the range of values of $\mu_{(i,j)}$ in the second last column of the table. These values have been computed assuming that $\sigma_{(\phi 1)} = \sigma_{(\phi 2)} = 0.01$cyc, and $\sigma_{(P1)} = \sigma_{(P2)} = 0.3$m.

The "best" ionosphere-free / geometry-free combination is that with the smallest standard deviation $\sigma_{(i,j)}$, which is the combination $i = 1$, $j = -1$. This expression is the same as that which appears in eqn (6.4-26). Hence it appears that the wide-lane ambiguity can be well determined using this data combination technique, however it is difficult to compute any other ambiguities, even the L1 and L2 (n_1, n_2) integer ambiguities. The reason is that the estimated ionospheric delay using the pseudo-range data is not accurate enough to correct these other carrier phase combinations (compare the last column of Table 6.4-2 with the corresponding column in Table 6.4-1).

Table 6.4-2. Characteristics of dual-frequency pseudo-range and phase ionosphere-free combinations.

Data combination $\phi_{i,j}$	$\lambda_{i,j}$ (m)	a	b	$\mu_{(i,j)}$ (λ)	nsf
$\phi_{1,0}$ ($\phi_{(L1)}$)	0.190	21.501	16.246	8.085	808
$\phi_{0,1}$ ($\phi_{(L2)}$)	0.244	20.849	16.754	8.024	1028
$\phi_{77,-60}$	0.006	404.638	245.690	142.020	448
$\phi_{1,-1}$	0.862	0.652	-0.508	0.248	112
$\phi_{-1,2}$	0.341	20.197	17.262	7.971	1428
$\phi_{2,-2}$	0.431	1.304	-1.016	0.497	113
$\phi_{-3,4}$	1.628	18.892	18.278	7.886	6746
$\phi_{-7,9}$	14.653	37.133	37.065	15.740	121197

nsf is the noise scale factor to relate the data noise of the linear combination to that of the L1 phase observation.

HAN & RIZOS (1995a) present alternative formulae for eqns (6.4-31) and (6.4-32) which are based on single frequency phase and pseudo-range data. The problem with such linear combinations that give integer estimates of the ambiguity (or ambiguities) is that the ionospheric effect cannot be eliminated, hence the uncertainties associated with the ambiguity estimates are greater.

Using Carrier Phase Data to Smooth Pseudo-Range Data

One way of overcoming the two main problems associated with pseudo-range data, that is: (a) the high measurement noise, and (b) the greater multipath disturbance, in comparison to carrier phase data is to create a pseudo-range / carrier phase combination that, in effect, "smooths" the pseudo-range data. The basis of all data smoothing techniques is to derive the *rate of change of range from the carrier phase data* (the Doppler data can also be used for this purpose), and to combine this with the *absolute measurement of range provided by the pseudo-range data* (though it is a biased range).

A first example of a GPS data smoothing technique was described by HATCH (1982), and makes use of dual-frequency phase and pseudo-range data. The basis of the technique is eqn (6.4-26), which is both "geometry-free" and "ionosphere-free". The difference between the wide-lane phase combination and the narrow-lane pseudo-range combination is a constant: the wide-lane ambiguity (eqn (6.4-26a)). Differencing between epochs eliminates this ambiguity term, and hence **the *rate of change* of the (narrow-lane) pseudo-range is identical to the *rate of change* of the (wide-lane) integrated carrier phase**.

The steps involved are (HOFMANN-WELLENHOF et al, 1994):

(1) At the first observation epoch compute the quantities:

$$P_{(L6)}(t_1) = \frac{f_1.P_{(L1)}(t_1)+ f_2.P_{(L2)}(t_1)}{f_1 + f_2} \quad (6.4\text{-}34a)$$

$$\Phi_{(L5)} = \lambda_5.(\phi_{(L1)} - \phi_{(L2)}) \quad (6.4\text{-}34b)$$

Set $P_{(L6)}(t_1) = P_{(ex)}(t_1) = P_{(sm)}(t_1)$.

(2) For all observation epochs t_i after t_1, *extrapolated* values of the pseudo-range are calculated:

$$P_{(ex)}(t_i) = P_{(ex)}(t_1)+ (\Phi_{(L5)}(t_i) - \Phi_{(L5)}(t_1)) \quad (6.4\text{-}34c)$$

(3) The smoothed value is obtained as the arithmetic mean:

$$P_{(sm)}(t_i) = \frac{1}{2} \cdot (P_{(ex)}(t_i)+ P_{(L6)}(t_i)) \quad (6.4\text{-}34d)$$

(4) The above formulae can be generalised for an arbitrary epoch t_i following t_{i-1}, by modifying eqn (6.4-34c) slightly:

$$P_{(ex)}(t_i) = P_{(sm)}(t_{i-1})+ (\Phi_{(L5)}(t_i) - \Phi_{(L5)}(t_{i-1})) \quad (6.4\text{-}34e)$$

The above algorithm cannot handle cycle slips in the carrier phase data, therefore a variation to eqns (6.4-34d) and (6.4-34e) is (IBID, 1994):

$$P_{(sm)}(t_i) = \omega.P_{(L6)}(t_i) + (1 - \omega) . \left(P_{(sm)}(t_{i-1})+ (\Phi_{(L5)}(t_i) - \Phi_{(L5)}(t_{i-1}))\right) \quad (6.4\text{-}34f)$$

where ω is a time dependent weight factor. For the first epoch $\omega = 1$, thus giving full weight to the current measured pseudo-range. With the accumulation of further epochs the weight is continuously reduced, and hence increasing the influence of the carrier phase data. A reduction rate of 0.01 in the weight from epoch to epoch has been recommended for 1 second kinematic data. In such a case, after 100 epochs (100 seconds), only the smoothed value from the previous epoch (augmented by the carrier phase rate of change) is taken into account. Again, the occurrence of cycle slips would cause the algorithm to fail. (Cycle slips are detected by comparing the rate-of-change of the carrier phase data across two epochs with the measured Doppler value times the time interval between epochs.) If a cycle slip is detected, the weight is reset to $\omega = 1$ and the smoother is restarted (but the cycle slip does not have to be *repaired*).

Alternative smoothing algorithms have been developed which use Doppler data in place of carrier phase data. Furthermore, all smoothing algorithms are also applicable to single frequency data, in which case the $P_{(L6)}(t_i)$ and $\Phi_{(L5)}(t_i)$ are replaced by the observations $P_{(L1)}(t_i)$ and $\Phi_{(L1)}(t_i)$. Of course such techniques are inferior to the dual-frequency techniques because they cannot account for the ionospheric bias.

Chapter 7: Introduction to GPS Processing

7.1 GPS DATA PROCESSING

The results of GPS surveying are only obtained after extensive processing of the recorded data. The general characteristics of carrier phase data processing in support of GPS surveying applications are:

- ❑ The extensive use of Least Squares estimation procedures to process the recorded GPS data.
- ❑ The determination of three-dimensional coordinates with respect to a geocentric Cartesian coordinate system.
- ❑ The datum for GPS results is influenced by both the relative mode of GPS baseline determination, and the reference system in which any fixed (or known) geodetic quantities such as station coordinates, or the satellite orbits, are expressed.
- ❑ A GPS survey is built up in stages, by combining individual survey sessions into a complete campaign. The results may then be integrated into a previously surveyed network.
- ❑ The baseline computations begin after the data from several GPS receivers are physically brought together. There are several steps involved, and the degree of sophistication of the computer processing software is primarily a function of the final accuracy sought.

Each of these is briefly described in the following sections. A more detailed treatment of specific details of GPS data processing and aspects of final result presentation is given in the following chapters.

Chapters 7 and 8 are concerned with "Static GPS Baseline Processing" because a baseline is the smallest element of a GPS adjustment: **the determination of the relative position of one receiver with respect to another receiver, both having collected data in a coordinated manner over a finite observation session**.

7.1.1 BACKGROUND TO LEAST SQUARES ADJUSTMENTS

There are many textbooks available presenting the principles of Least Squares estimation as they apply to surveying and geodesy. Two that are particularly useful at an introductory level are CROSS (1983) and HARVEY (1994). The following is a brief summary of the basic methodology.

A Least Squares adjustment involves two models:

- The **functional model** relating the measurements and the parameters. The most common approach is to use observation equations of the general form $\mathbf{l} = f(\mathbf{x})$. To satisfy this relation, actual observations need to be corrected or "adjusted". The linearisation of the relation is performed about an approximate set of values for the parameters to be estimated $\mathring{\mathbf{x}}$:

$$\begin{aligned} l - \mathbf{v} &= f(\mathbf{x}) \\ l - \mathbf{v} &= f(\mathring{\mathbf{x}} + \delta\mathbf{x}) \\ l - \mathbf{v} &= f(\mathring{\mathbf{x}}) + \mathbf{A}\delta\mathbf{x} \\ (l - f(\mathring{\mathbf{x}})) - \mathbf{v} &= \mathbf{A}\delta\mathbf{x} \end{aligned} \tag{7.1-1}$$

The expression in brackets is the "observed minus computed" term, or approximate residual, and is denoted by $\mathring{\mathbf{v}}$. $\mathbf{l}$ and l are the *true* and *actual* observations respectively, $\mathbf{x}$ and $\mathring{\mathbf{x}}$ are the *true* and *approximate* (or apriori) parameters respectively. $\delta\mathbf{x}$ are the corrections to the approximate parameters and $\mathbf{A}$ is the design matrix containing the partial derivatives of the observations with respect to the parameters.

- The **stochastic model** describing the statistics of the measurement. This is in the form of the weight matrix $\mathbf{P}$, or its inverse the covariance matrix $\mathbf{Q}_{ll}$ of the observations. For example, all measurements could be independent (that is, a diagonal weight matrix) and have the same standard deviation.

Computational Procedures: The Classical Least Squares Approaches

Condition Method

Condition equations express properties that the observations should satisfy. The general form of a condition equation is $f(\mathbf{l}) = 0$, where $\mathbf{l}$ is the vector of *true* observations. *Actual* observations l are generally biased by a number of errors and therefore do not satisfy this condition. A vector of miscloses can be computed as $f(l) = \mathbf{w}$. The adjustment aims at computing the vector of corrections $\mathbf{v}$ to the observations such that the corrected observations satisfy both the relation $f(l - \mathbf{v}) = 0$ and the Least Squares condition $\mathbf{v}^T\mathbf{P}\mathbf{v} \rightarrow$ minimum. The linearisation of the condition equation is based on a Taylor's series expansion of the first order:

$$f(l - \mathbf{v}) = f(l) - \mathbf{B}\mathbf{v} \tag{7.1-2}$$

where $\mathbf{B}$ is the design matrix, containing partial derivatives of $f(\mathbf{l})$ with respect to $\mathbf{l}$ about the actual observations l. The variance-covariance (VCV) matrix $\mathbf{Q}_{ll}$ of the observations is assumed known. The computational procedure can be summarised in a few steps:

Linearised form: $\mathbf{B}\mathbf{v} = \mathbf{w}$ with weight matrix: $\mathbf{P} = \mathbf{Q}_{ll}^{-1}$ (7.1-3)

Solution for the residuals: $\hat{\mathbf{v}} = \mathbf{Q}_{ll}\mathbf{B}^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{w}$ (7.1-4)

VCV matrix of the residuals: $\mathbf{Q}_{\hat{v}\hat{v}} = \mathbf{Q}_{ll}\mathbf{B}^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_{ll}$ (7.1-5)

VCV of the adjusted observations: $\mathbf{Q}_{\hat{l}\hat{l}} = \mathbf{Q}_{ll} - \mathbf{Q}_{\hat{v}\hat{v}}$ (7.1-6)

The advantage of the condition method is that the unknown terms are simply corrections to the observations. The size of the matrix $\mathbf{B}\,\mathbf{Q}_{ll}\,\mathbf{B}^T$ to invert is $r \times r$, where r is the number of conditions, which is in fact equal to the number of redundant observations. However, there are a number of drawbacks. The derivation of adjusted values for functions of the observables (for example, the coordinates) is tedious, as is the derivation of their respective precisions and correlations. Furthermore the construction of eqn (7.1-2) requires a sound geometrical understanding of the situation, as only independent conditions must be used. Consequently, the setting-up of the equations is not easily automated for a computer and this method is not used for GPS adjustments. However this method was especially popular in the past for geodetic network adjustments when computers were not available.

Parametric Method

This method of adjustment makes use of observation equations, where observables are expressed as a function of some or all of the parameters, in the general form $\mathbf{l} = f(\mathbf{x})$. To satisfy this relation, actual observations need to be corrected or "adjusted". The linearisation of the relation is performed about an approximate set of parameters $\mathring{\mathbf{x}}$:

$$\begin{aligned} \boldsymbol{l} - \mathbf{v} &= f(\mathbf{x}) \\ \boldsymbol{l} - \mathbf{v} &= f(\mathring{\mathbf{x}} + \delta\mathbf{x}) \\ \boldsymbol{l} - \mathbf{v} &= f(\mathring{\mathbf{x}}) + \mathbf{A}\delta\mathbf{x} \\ (\boldsymbol{l} - f(\mathring{\mathbf{x}})) - \mathbf{v} &= \mathbf{A}\delta\mathbf{x} \end{aligned} \qquad (7.1\text{-}7)$$

The expression in brackets is the approximate residual, and is denoted by $\mathring{\mathbf{v}}$. The variance-covariance matrix $\mathbf{Q}_{ll}$ of the observations is assumed known. As $\mathring{\mathbf{v}}$ differs from $\boldsymbol{l}$ only by a constant, it has the same stochastic behaviour. The computational procedure therefore is:

Linearised form: $\mathring{\mathbf{v}} - \mathbf{v} = \mathbf{A}\delta\mathbf{x}$ with weight matrix: $\mathbf{P} = \mathbf{Q}_{ll}^{-1}$ (7.1-8)

Solution for the parameters: $\delta\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}\,\mathbf{A}^T\mathbf{P}\,\mathring{\mathbf{v}}$ (7.1-9)

with VCV matrix: $\mathbf{Q}_{\hat{x}\hat{x}} = (\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}$ (7.1-10)

The adjusted observation residuals can be computed in two different ways:

- The direct approach, by including the adjusted state in the functional model of the observations:

$$\hat{\mathbf{v}} = \boldsymbol{l} - f(\hat{\mathbf{x}}) \qquad (7.1\text{-}11)$$

- The indirect approach, by including the adjusted state in the linearised functional model of the observations eqn (7.1-8):

$$\hat{\mathbf{v}} = \mathring{\mathbf{v}} - \mathbf{A}\delta\hat{\mathbf{x}} = \mathring{\mathbf{v}} - \mathbf{A}(\hat{\mathbf{x}} - \mathring{\mathbf{x}}) \qquad (7.1\text{-}12)$$

The second method clearly illustrates the relation between the approximate and adjusted parameters. Indeed, this is the main justification for the choice of the unusual symbol $\mathring{\mathbf{v}}$ to denote the vector of approximate residuals, hence ensuring complete consistency between quantities related either to $\mathbf{x}$ or $\mathbf{v}$. The VCV matrix of the residuals is easily derived from the indirect computation of the residuals, assuming the stochastic independence of the measurements and the approximate parameter vector:

$$\mathbf{Q}_{\hat{v}\hat{v}} = \mathbf{Q}_{ll} - \mathbf{A}\,\mathbf{Q}_{\hat{x}\hat{x}}\,\mathbf{A}^{\mathrm{T}} \qquad (7.1\text{-}13)$$

The aposteriori variance factor is: $\dfrac{\mathbf{v}^{\mathrm{T}}\mathbf{P}\mathbf{v}}{(n - u)}$

where n *is the number of observations and* u *is the number of parameters.*

Although it is not strictly correct, in these notes there will be no distinction made between *co-factor* matrices and *variance-covariance* matrices. The *variance factor* scales the cofactor matrix to give the VCV matrix. If the VF is close to unity (as it should be), then the co-factor matrix and the VCV matrix are almost identical.

If the desired results (for example, the coordinates) are selected as the parameters, the solution of the system leads directly to the answer. There is exactly one equation per observation, and its form can easily be defined according to the type of observation. The size of the matrix to invert is $u \times u$, where u is the number of (unknown) parameters, and the setting-up of the equations can therefore be automated in a computer program. The linearisation of the problem requires some apriori approximate knowledge of the parameters, which is usually available. For GPS adjustments, it is generally no problem to obtain a converged solution of the non-linear problem (eqn (7.1-19)) through an *iterative* process.

Relations Between Both Approaches

If any problem can reduce to either the condition or parametric case, there is actually a choice between the two methods of setting-up and solving the Least Squares problem. Given n linear observation equations with u (unknown) parameters, the elimination of all the parameters leads to a system of $r = n - u$ condition equations. The reverse of this process is much harder because there is an almost infinite number of possible parameterisations. In some cases, for example if there are fewer redundant observations than unknown parameters, the condition method offers computational advantages. However, the dramatic improvement in the power of computers has made these advantages largely irrelevant and the adjustment by parameters is now the standard solution approach to almost all over-determined systems.

The Combined Case

It is also possible to formulate relations involving *both* observables and parameters. Functions of the observables are related to functions of the parameters, in the general form: $f(\mathbf{l}, \mathbf{x}) = 0$. For example, this method is useful when solving for transformation parameters (HARVEY, 1994). A linear relation is obtained following the usual procedure:

$$f(l, \mathring{\mathbf{x}}) = \mathbf{w}$$

$$f(l - \mathbf{v}, \mathring{\mathbf{x}} + \delta\mathbf{x}) = f(l, \mathring{\mathbf{x}}) + \mathbf{A}\delta\mathbf{x} - \mathbf{B}\mathbf{v} = 0 \qquad (7.1\text{-}14)$$

Linearised form: $\quad -\mathbf{A}\delta\mathbf{x} + \mathbf{B}\mathbf{v} = \mathbf{w}$ with weight matrix: $\mathbf{P} = \mathbf{Q}_{ll}^{-1}$ (7.1-15)

Solution for the parameters: $\quad \delta\hat{\mathbf{x}} = (\mathbf{A}^T(\mathbf{B}\,\mathbf{Q}_{ll}\,\mathbf{B}^T)^{-1}\mathbf{A})^{-1}\,\mathbf{A}^T(\mathbf{B}\,\mathbf{Q}_{ll}\,\mathbf{B}^T)^{-1}\mathbf{w}$ (7.1-16)

Variance-covariance matrix: $\quad \mathbf{Q}_{\hat{x}\hat{x}} = (\mathbf{A}^T(\mathbf{B}\,\mathbf{Q}_{ll}\,\mathbf{B}^T)^{-1}\mathbf{A})^{-1}$ (7.1-17)

Formulae for the residuals are given, for example, in CROSS (1983). To demonstrate the equivalence between these two approaches, it suffices to rearrange eqn (7.1-15) and consider particular design matrices:

Condition: $\quad \mathbf{B}\mathbf{v} = \mathbf{w} + \mathbf{A}\delta\mathbf{x}$ with $\mathbf{A} = \mathbf{0}$

Parametric: $\quad \mathbf{w} - \mathbf{B}\mathbf{v} = \mathbf{A}\delta\mathbf{x}$ with $\mathbf{B} = \mathbf{I}$

In the parametric case, the misclose vector $\mathbf{w}$ is equal to $\mathring{\mathbf{v}}$, the observation residual computed using the approximate parameters $\mathring{\mathbf{x}}$.

Bayesian Least Squares

In many cases, a fairly good apriori knowledge of the parameters is available. Thus, it is reasonable to require that the adjusted value of a parameter should not be too different from its apriori value. This condition can be imposed in a number of ways:

- The apriori values of the parameters are considered as observations. Suitable weights, relative to that of the measurements, are required. Consequently, the size of the system of parametric equations is augmented, but not that of the normal equations, as the number of parameters remains unchanged.
- The increments of the apriori values of parameters are included in the quadratic form to minimise, with appropriate weights.

The resultant normal equations are (MERMINOD & RIZOS, 1988):

$$(\mathbf{A}^T\mathbf{P}\mathbf{A} + \mathbf{P}_{\mathring{x}})\,\delta\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{P}\,\mathring{\mathbf{v}} \qquad (7.1\text{-}18)$$

The extension of the quadratic form to $\mathbf{v}^T\mathbf{P}\mathbf{v} + \delta\mathbf{x}^T\mathbf{P}_{\mathring{x}}\delta\mathbf{x}$ represents a generalisation of the classical Least Squares method, where $\mathbf{P}_{\mathring{x}}$ is the apriori weight matrix of the parameters. With $\mathbf{P}_{\mathring{x}} = \mathbf{0}$, that is, no apriori information on the parameters is available, then Bayesian Least Squares reduces to the classical definition.

Sequential Least Squares

The *batch* and *sequential* (or step-by-step) processing modes can be distinguished. The batch processing mode is the one most commonly encountered in standard GPS static data processing as well as in most geodetic adjustment problems. The Least Squares adjustment is carried out once all the data has been acquired. However, a sequential treatment of Least Squares problems may be preferable for several reasons:

- it divides a large computing burden into smaller parts, to reduce the requirements on both processing capability and storage, and
- it is the key to real-time applications.

It is therefore tempting to treat successive batches of measurements sequentially, though it should be kept in mind that the results of a sequential Least Squares adjustment will be identical to that of a batch solution once all the data has been processed. The formulae used in sequential Least Squares estimation are given in, for example, CROSS (1983).

7.1.2 LEAST SQUARES PROCEDURES APPROPRIATE FOR GPS SURVEYING

The functional models, or mathematical observation equations, appropriate for GPS survey adjustments are presented in §7.2. The degree of model sophistication can vary considerably, though the basic models for GPS survey processing are now well defined.

In the following discussions, the basic **range equation**, in which the geometric range (ρ) is a function of the satellite and receiver coordinates (Figure 7.1-1), will be considered:

$$\rho_j^i = \sqrt{(X^i-X_j)^2+(Y^i-Y_j)^2+(Z^i-Z_j)^2} \tag{7.1-19}$$

where: X^i, Y^i, Z^i are the coordinates of satellite i, and
X_j, Y_j, Z_j are the coordinates of site j.

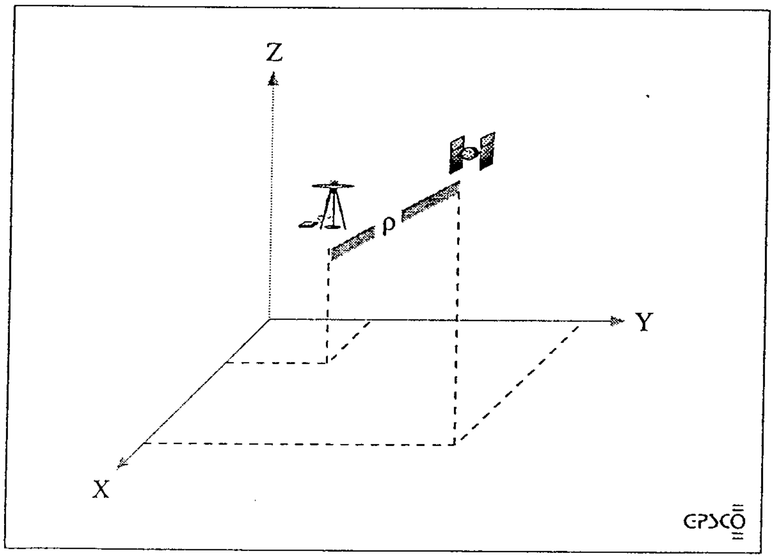


Figure 7.1-1. The geometric range: receiver to satellite.

In the case of GPS surveying, the coordinates of the satellites are assumed to be known, and hence enter the observation equations as *apriori information*. (Typically they are evaluated from the broadcast elements within the Navigation Message -- see Table 3.3-3 for an outline of the computational procedure.) It is necessary to distinguish between two classes of parameters:

(1) Geodetic parameters, the station coordinates in eqn (7.1-19), parameterised as Cartesian coordinate components (in the same reference system as the satellite coordinates), and represented geometrically in Figure 7.1-1.

(2) Non-geodetic parameters, which for GPS surveying consist of the receiver clock bias parameters, satellite clock bias parameters, and phase ambiguity parameters. These are generally considered to be "nuisance" parameters as they do not appear explicitly in the basic range equation (eqn (7.1-19)), but must be included in GPS model equations in order to account for the measurement biases present in data (§7.2). (When data is double-differenced, the clock terms are eliminated from the double-differenced observation equations.)

The data types processed in GPS survey adjustments are **pseudo-ranges** and **integrated carrier phase** (in either the differenced or undifferenced mode). Both types are measured by GPS surveying receivers, however only pseudo-range data are measured by GPS navigation receivers (LANGLEY, 1991a; LANGLEY, 1993). Because of the much higher measurement precision of carrier phase observations, the pseudo-range data plays only a minor role in GPS baseline processing (see §7.3).

The basic steps in a GPS adjustment therefore are:

- Set up the solution: select the appropriate functional model for the data type being processed and for the solution application. Compute the design matrix $\mathbf{A}$.
- Select the observations: may require some pre-processing (to correct for some known bias -- §6.2) or data manipulation (for example, data differencing -- §6.3), and sufficient in *quantity* and *quality* to estimate the parameters reliably.
- Obtain approximate values of the parameters: in particular the geodetic parameters (for example, from a previous solution or an approximate computation) $\mathring{\mathbf{x}}$.
- Define the quality of the observations: this requires the specification of the VCV matrix $\mathbf{Q}_{ll}$.
- Define the datum in such a way as to overcome any rank defects in the system of equations $\mathbf{A}^T\mathbf{PA}$.
- Iterate the solution until convergence is reached.

Calculation of the Partial Derivatives

As indicated above, the mathematical adjustment model must be linear, hence the basic GPS observation model has to be *linearised*. This requires that the parameter set consist of corrections to approximate values, rather than the parameters in the functional model themselves -- for example the coordinates of stations may be known to an accuracy of tens of metres, or even better. The elements of the design matrix in eqn (7.1-1) are the partial derivatives of the observables with respect to the geodetic and non-geodetic parameters in the functional model. (Note that the partial derivatives of the observables with respect to the clock terms need not be considered as they cancel in double-differenced and triple-differenced observations.)

The partial derivative of carrier beat phase with respect to the site coordinates is obtained from eqn (7.1-19):

$$\frac{\partial \phi_{bj}^{i}}{\partial U} = -\frac{f_o}{c} \bullet \frac{\partial \rho_j^{i}}{\partial U} \qquad (7.1\text{-}20)$$

where:

ϕ_{bj}^{i}	is the phase observation from satellite i to receiver j (in units of cycles),
U	are the site parameters, which will be defined for convenience in the Cartesian coordinate system (X, Y, Z),
f_o	is the nominally constant carrier frequency,
c	is the speed of EMR in a vacuum, and
ρ_j^{i}	is the geometric range between satellite i and receiver j (Figure 7.1-1).

From eqns (7.1-19) and (7.1-20) the partial derivative of phase with respect to site coordinates can be written as:

$$\frac{\partial \phi_{bj}^{i}}{\partial X_j} = \frac{f_o}{c} \bullet \frac{\partial \rho_j^{i}}{\partial X_j} = -\frac{f_o}{c} \bullet \frac{X^i - X_j}{\rho_j^{i}}$$

$$\frac{\partial \phi_{bj}^{i}}{\partial Y_j} = \frac{f_o}{c} \bullet \frac{\partial \rho_j^{i}}{\partial Y_j} = -\frac{f_o}{c} \bullet \frac{Y^i - Y_j}{\rho_j^{i}}$$

and

$$\frac{\partial \phi_{bj}^{i}}{\partial Z_j} = \frac{f_o}{c} \bullet \frac{\partial \rho_j^{i}}{\partial Z_j} = -\frac{f_o}{c} \bullet \frac{Z^i - Z_j}{\rho_j^{i}} \qquad (7.1\text{-}21)$$

These are the partial derivatives for the one-way phase (or range) observation equation. Partials for the double-differenced and triple-differenced observation equations are obtained by the appropriate differencing (§6.3 and §7.2).

The partial derivative of the phase observable with respect to an ambiguity parameter is <u>unity</u>.

Least Squares Solutions in GPS Surveying

The following stages in GPS phase data processing require some form of Least Squares adjustment:

☞ <u>Point-position solutions using pseudo-range data</u> to obtain *preliminary WGS84 coordinates and approximate receiver clock error estimates* (for the correction of observation time-tags at the microsecond level -- see §6.3).

☞ <u>Triple-difference carrier phase solution</u> using phase data obtained by differencing the double-differences between successive epochs. As ambiguity parameters are eliminated, such a solution can give good apriori coordinates, *but is not recommended for precise GPS phase data processing.*

☞ Perhaps some form of (polynomial) <u>curve-fitting to observation residuals</u> to permit cycle slip detection and repair during phase data pre-processing.

☞ <u>Double-differenced phase data solutions</u> estimating both the *station coordinates and* *the ambiguity parameters* (as real-valued quantities).

☞ Solutions that *combine* all GPS baseline results into a single campaign adjustment (without processing all the raw double-differenced data, as above, in a single step solution).

☞ Solutions to *integrate* the GPS results into a conventional geodetic network, involving the *distortion* of the minimally constrained GPS network to fit the surrounding control network.

☞ Determination of the transformation parameters between the GPS datum and the local geodetic datum.

7.1.3 THE NATURE OF GPS SOLUTIONS

The most distinctive feature of GPS surveying is the determination of 3-D coordinates. The following additional remarks can be made:

- ❑ The most convenient coordinate system is that provided by the rectangular Cartesian system. This is the system, in the first instance, in which the satellite coordinates are defined in the Navigation Message (Table 3.3-3). Furthermore, the coefficients of the design matrix (the partial derivatives of phase with respect to the site coordinates) are most easily computed in the Cartesian system (eqn (7.1-21)). If the results are required in any other coordinate reference system, the Cartesian components can be easily transformed at the result presentation stage (chapter 11).
- ❑ GPS phase processing result in *relative* 3-D coordinates. In the case of a two receiver scenario, it is the 3-D baseline components that are computed. In the case of more than two receivers (say R), the coordinates of the R-1 stations are determined relative to the *datum* station in the adjustment. (There is always a minimum of one station held fixed in a GPS adjustment, whether a hundred stations are involved, or just a single baseline.)
- ❑ GPS pseudo-range solutions are single point, or absolute, solutions in which the 3-D coordinates are independently determined for each receiver. These coordinates are not as accurately determined as those from carrier phase (relative) solutions (§2.4), but are often used to provide apriori values for the subsequent phase solutions.
- ❑ All results relate to the GPS antenna phase centres and must be corrected to the groundmark by applying the antenna height and eccentric station offsets.
- ❑ All results are, nominally, referred to the WGS84 reference system.

The phase data may be processed in the:

(a) **baseline mode**, taking the data of any two receivers at a time, which is the minimum configuration possible for data reduction, or

(b) **session mode**, where all data collected together during an observation session is processed together, or

(c) **campaign mode**, in which all data, collected over several sessions, is processed simultaneously.

The mode of processing is generally a function of:

- The software that is available -- *most commercial software is only able to process a single baseline at a time.*
- The degree of rigour required for the phase data reduction, and hence the accuracy sought -- *for example, scientific software can accommodate all the phase data collected during a campaign in a rigorous manner.*

7.1.4 PROPAGATION OF THE GPS SURVEY

In general, a GPS survey involves the use of a small number of receivers (two or more) to coordinate a large number of stations. **The survey must therefore proceed in stages.** The minimum GPS phase data reduction is therefore for a single baseline. However the data collected in a single session by all receivers has special characteristics. A <u>GPS session solution</u> can be obtained from a combination of separate baseline solutions, or from one simultaneous session solution. A <u>GPS network solution</u> can be obtained from combining individual session solutions, or in one simultaneous network (multi-session) solution.

GPS processing procedures may therefore be partitioned according to:

(1) **Primary adjustment** of the (raw) phase data collected during a session by the deployed GPS receivers. The software generally has the following features (§4.3):
  - Provided by the receiver manufacturer, in the case of *commercial* software.
  - Output is the coordinates, and associated VCV matrix, of the stations processed together.
  - The stations may be processed as a single baseline or, more rarely, in the multi-baseline mode.
  - The coordinate results are expressed in a global GPS satellite datum system.

(2) **Secondary adjustment** that accepts the output of the primary GPS adjustment as input (that is, as "observations"). Such software has the following features:
  - Not usually developed by the same institution as responsible for the primary GPS software.
  - Can modify and manipulate separate baseline or network solutions in order to combine together a minimally constrained adjustment from separate primary adjustments (from baselines to multi-baseline, from session to multi-session).
  - Can constrain the network adjustment to fit surrounding geodetic control.
  - Accounts for the datum relationships between global and local geodetic datums.
  - Provides the flexibility to constrain and scale individual GPS adjustments.
  - The results can only be as good as the input primary adjustment data!

The outcome of a GPS adjustment is usually a <u>minimally constrained network solution</u> in which the coordinates of only one station (the "datum" station) have been held fixed.

In subsequent sections the baseline solution will be discussed further. A discussion of session solutions in §9.2 will highlight the essential differences between single-baseline and multi-baseline session solutions. Finally in §9.3, the manner in which a total network solution is obtained from separate session solutions is described.

7.1.5 DATA PROCESSING CONSIDERATIONS

Among the most important considerations for the reliable processing of GPS data therefore are:

- Correct field procedures to ensure that good quality data is collected in the field (chapter 10): for example, avoiding multipath inducing environments, correct measurement of antenna height, adequate battery power, no signal obstructions (trees, etc.) or interference (microwave transmissions, etc.), appropriate length observation sessions, etc.
- Clean data that has been appropriately pre-processed in order to detect and repair any cycle slips.
- Appropriate observation modelling and processing software: matched to the accuracy required so that the correct strategy is used to account for the GPS measurement biases, and in particular the cycle ambiguities.
- Appropriate processing procedures: recognises that GPS phase data reduction involves a number of steps executed in sequence.

The latter two issues are discussed briefly below.

Degree of Sophistication of Data Modelling and Processing

As discussed in §5.1, the GPS relative positioning technique is employed for a wide range of applications. As a starting point, in order to sort out the variety of GPS modelling and software options that are available, the surveying applications may be categorised according to three general ranges of accuracy:

- Class A (Scientific): better than 1 ppm
- Class B (Geodetic): 1 to 10 ppm
- Class C (General Surveying): lower than 10 ppm.

Class A surveys are primarily undertaken for precise engineering, deformation analysis, and geodynamic applications. *Class B* surveys include control surveys undertaken for the purposes of geodetic network densification, mapping, and for resource development applications. *Class C* surveys primarily encompass those lower accuracy surveys undertaken for urban, cadastral, GIS and general-purpose survey applications. Users in the latter two categories comprise the majority of the GPS survey user community, and it is the applications in these categories that are collectively referred to in these notes as "GPS surveying". In §7.2 the GPS observation modelling options appropriate for GPS surveying are summarised.

Steps in GPS Data Processing

For the purpose of discussion, the data pre-processing, initial analysis and final adjustment steps as they relate to a single session are introduced here. These are described in the remaining sections of chapter 7, in chapter 8, and in parts of chapter 9. The steps relating to the combination of session adjustments and the integration of the results with terrestrial control are discussed separately in chapters 9, 11 and 12.

Data Pre-Processing

Pre-processing encompasses a number of specific tasks:

- Initial data transfer and decoding.
- Data screening and editing.
- Data reporting and database creation and entry.
- Point positioning using pseudo-range data.

These tasks may be carried out on a *single-station basis*, and can therefore be carried out in the field office. The result should be a set of appropriately formatted and pre-processed (though not entirely "clean") observations, together with ephemeris information and approximate station coordinates.

Initial Data Analysis

These tasks are carried out as a prelude to the final phase data adjustment, as soon as data from two or more GPS receivers is brought together in the field office:

- Cycle slip detection and repair.
- Preliminary baseline solution based on triple-differenced phase data or double-differenced pseudo-range data (if of suitable quality).

The former is discussed in §7.3, together with the pre-processing procedures. The latter, as it is based on the principles of baseline determination, is discussed in §8.1, together with double-differenced phase solutions. The result is a "clean" set of observations, and very good apriori station coordinates.

Final Adjustment

The adjustment of the pre-processed GPS observations similarly encompasses a number of tasks:

- The formation of the phase data differences.
- Definition of the apriori weight matrices.
- Estimation of relative station coordinates.
- Estimation of carrier beat phase ambiguities and fixing them (if possible) to integers.
- Output of estimated parameters and aposteriori covariance matrix.

The actual adjustment of GPS observations can be performed in two ways:

- by combining single baseline solutions into a network adjustment of the baseline components, or
- by a direct, simultaneous solution involving all the GPS observations for a session, or complete network.

Various types of differenced (including non-differenced) observables can be employed in GPS data processing (§7.2). *Each of these observable types have certain advantages and disadvantages with respect to bias reduction, parameter sensitivity, error modelling, and processing efficiency.* **The capability to exploit the integer nature of the carrier phase ambiguity terms is also an important consideration in the final adjustment stage**. Methods of integer ambiguity resolution appropriate to the different observable types (apart from triple-differences), and for single and dual-frequency operations are needed (chapter 8).

7.2

BASIC MODELLING CONCEPTS

There are two classes of parameters in GPS observation models:

(1) **Parameters related to geometric range:**

$$\rho_j^i(T_j) = \sqrt{[x^i(T)-x_j]^2 + [y^i(T)-y_j]^2 + [z^i(T)-z_j]^2}$$

where:

- (x^i, y^i, z^i) are the coordinates of satellite i at time T, and
- (x_j, y_j, z_j) are the coordinates of site j in the same reference system.

THIS IS THE QUANTITY OF INTEREST!

(2) **All other terms are biases, which must be accounted for during observation modelling. The most common strategy is to difference the data, and construct an observable that is, to a large extent, free of biases. The additional considerations for eliminating biases are:**

- *How are differences constructed from one-way observations from R receivers to S satellites?*
- *What is the form of the resultant observation weight matrix?*
- *What form do the remaining biases (not eliminated by differencing) take?*

This chapter is concerned with the following topics:

(1) The observation models for pseudo-range data and the various differenced phase observables used in GPS surveying, and in particular, the strategies for handling biases (discussed in §6.1) in commercial GPS surveying software packages.

(2) The datum definition issues for GPS processing.

(3) The mechanics of a GPS phase adjustment, and in particular the various data differencing options for double- and triple-differenced phase observables.

(4) The nature of correlations in the differenced data.

(5) The ambiguity parameter modelling options, and the associated datum defects.

(6) The Normal Equation structure for a GPS phase adjustment.

7.2.1 GPS OBSERVATION MODELS

Dealing with Biases:

To obtain a GPS solution, strategies have to been developed to overcome the problem of systematic biases.

There are a number of options:

- They can be estimated as explicit parameters.
- Those biases linearly correlated across different datasets can be eliminated (or significantly reduced) by data differencing.
- They can be directly measured.
- They can be assumed known from models.
- They can be ignored.

The most commonly used option is the second.

Observation DIFFERENCING is closely identified with the GPS surveying technique, requiring two or more GPS receivers to track the same set of satellites -- ***hence appropriate field and data analysis procedures must be used.***

GPS measurements, both pseudo-range and carrier phase, are affected by **biases** and **errors** (§6.1). Different levels of GPS accuracy are associated with a different partitioning of "biases" and "errors". At one extreme, in the case of GPS pseudo-range point positioning, all effects with the exception of the receiver and satellite clock biases are treated as errors. At the other extreme, GPS baseline determination to accuracies of 1 part in 10^8 requires that almost all measurement biases are explicitly accounted for in any solution. In the case of GPS surveying there is a *compromise* between fully accounting for all biases and keeping the computational burden down by not over-parameterising the observation model.

First the commonly used observation models in GPS surveying are described, and the reasoning behind the particular parameterisation that is adopted will be discussed. The bias categories are defined as follows (§6.2):

☞ **Satellite dependent:**
- Ephemeris uncertainties
- Satellite clock uncertainties
- Selective Availability effects

☞ **Receiver dependent:**
- Receiver clock uncertainties
- Reference station coordinate uncertainties

☞ **Receiver-Satellite (or observation) dependent:**
- Ionospheric delay
- Tropospheric delay
- Carrier phase ambiguity

It is assumed that the data has been "cleaned" (and hence, in the case of carrier phase data, there are no cycle slips present), and that all other effects are indistinguishable from data "noise", which the Least Squares adjustment must accommodate.

Accounting for Biases in GPS Surveying

The first point that needs to be made is that by adopting the observation strategy of *simultaneous tracking of the same set of satellites from a number of receivers, the most powerful technique of bias minimisation can be used.* Observation differencing, or more generally the principle of relative positioning, takes advantage of the correlated nature of GPS biases.

In the case of GPS phase data reductions:

- ❑ Between-satellite and between-receiver differencing effectively eliminates all clock biases due to oscillator errors, while significantly reducing most of the other biases (with the exception of the phase ambiguity, which requires special treatment).
- ❑ The ambiguity term in the phase observation model can be eliminated by forming between-epoch (or triple) differences (§6.3).
- ❑ The ambiguity term in the phase observation model can also be accounted for by estimating it in a preliminary "ambiguity-free" solution as a real-valued quantity, and then resolve it to its most likely integer in a subsequent "ambiguity-fixed" solution.
- ❑ Although all non-clock biases may be "significantly reduced" as a result of data differencing, some comments are worth making:
    - With the implementation of Selective Availability (§2.4), the orbit errors may become larger and baseline accuracies may suffer. Although there is no evidence at present (November 1996) that the orbits are degraded, if this were the case surveyors are likely to request the "precise ephemerides" (for example, from the International GPS Service for Geodynamics).
    - Errors in the coordinates assigned for the fixed station in a baseline, or network, reduction are not likely to have a significant effect on GPS surveying accuracies attained. The influence is similar to orbit errors, hence if seeking better than 5ppm relative positioning accuracy, coordinates derived from the averaging of pseudo-range solutions are suitable.
    - The ionospheric refraction is ignored in single frequency observations, but can be largely eliminated in dual-frequency observations for longer baselines (see §6.4). For baselines <30km in length, even when dual-frequency data is available, analysts usually only use the L1 observations.
    - The tropospheric refraction is more problematic. In general it is ignored, though some software packages apply refraction "corrections" to the data using some atmospheric model, either with or without the aid of observed met parameters (BRUNNER & WELSCH, 1993). (The TRIMVEC™ program had the option to solve for a baseline tropospheric scale parameter but it is generally conceded that this is ineffectual.)

Because pseudo-range data is of the order of a hundred to a thousand times "noisier" than phase data, there is not the same imperative to eliminate (or minimise) biases, as in the case of carrier phase data. In the case of pseudo-range data reductions:

- Explicit data differencing is not generally carried out.
- The satellite clock error is assumed known from the broadcast Navigation Message (§3.3).
- The satellite orbit is assumed known, and any error in the ephemeris (together with any unmodelled satellite clock error) has a significant effect on point positioning -- exactly the impact intended of Selective Availability.
- The receiver clock error must be estimated directly from the data.
- By differencing the point positions obtained at two (or more) GPS receivers, the effects of orbit and satellite clock errors are reduced (§2.4). Hence the relative position of the receivers derived from pseudo-range solutions is an order of magnitude more accurate than absolute position.
- Generally the ionospheric refraction effect is accounted for by using the broadcast ionospheric model within the Navigation Message.

Explicit expressions for the mathematical models for the following observations types which have a role in GPS surveying can now be developed:

- pseudo-range data
- one-way carrier phase
- double-differenced phase
- triple-differenced phase

Pseudo-Range Model

The basis for the model is the parametric eqn (6.2-13):

$$P_j^i(T_j) = \rho_j^i(T_j) + c.\varepsilon_{rcj}(T_j) \qquad (7.2\text{-}1)$$

where P_j^i is the pseudo-range measurement in metric units, ρ_j^i is the geometric range, c is the speed of electromagnetic radiation (=299799458m/s), T_j is the time-of-reception and ε_{rcj} is the receiver clock error (in seconds). *Note that the satellite clock error is not included in the above observation equation.* Hence there are four unknown parameters per receiver per epoch. A minimum of four observations would permit all the parameters to be estimated on an epoch-by-epoch basis. This of course is the basis of the standard GPS "navigation solution".

Pseudo-range observations, in the context of GPS data post-processing, are used in a slightly different manner than they would in the navigation mode (§7.3):

(1) Advantage is taken of the fact that the receiver does not move during an entire observation session. Hence there are many pseudo-range observations available to estimate the three coordinate component parameters. *The coordinate solution provides apriori values for the subsequent phase reduction step.*

(2) Because the receiver clock error is the means by which receiver clock time (and hence the observation time-tags) can be related to GPS Time (§6.3), a polynomial model of

the clock error may suffice for this purpose:

$$\varepsilon_{rcj} = a_0 + a_1.(T - T_0) + a_2.(T - T_0)^2 \tag{7.2-2}$$

One-Way Phase Model

In length units, eqn (6.2-13) becomes, for either L1 or L2 phase data:

$$\Phi_j^i(T_j) = \rho_j^i(T_j) + c.[\varepsilon_{rcj}(T_j) - \varepsilon^{sci}(T_j^i)] + \lambda.n_j^i \tag{7.2-3}$$

where ε^{sci} is the satellite clock error, T_j^i is the time of signal transmission, n_j^i is the ambiguity in the phase measurement, and λ is the wavelength of the carrier wave used to scale cycles into length units (≈19cm for L1 and ≈24cm for L2). All other terms have been defined previously. Note the absence of atmospheric delay biases -- it is assumed that after data differencing the residual biases are so small that they can be considered as data "noise".

Note that n_j^i is an integer, though $\lambda.n_j^i$ is not. Eqn (7.2-3) contains three clock biases that must be either explicitly estimated or eliminated by data differencing: n_j^i, ε_{rcj} and ε^{sci}.

Consequences of data differencing:

☞ **Differencing *increases* the measurement noise level.**

☞ **Differencing introduces *correlations* between the resulting "observables".**

☞ **Differencing *decreases* the number of parameters in the observation model.**

☞ **Differencing *decreases* the number of "observables" that need to be processed.**

☞ **Differencing possibilities:**

- between-receivers --> *Δ operator*
- between-satellites --> *∇ operator*
- between-epochs --> *δ operator*
- between-observation types --> P_1, P_2, C, Φ_1, Φ_2

Double-Differenced Phase Model

In length units, eqn (6.3-7) can be written as:

$$\nabla\Delta\Phi(t) = \rho_1{}^1(T_1) - \rho_1{}^2(T_1) - \rho_2{}^1(T_2) + \rho_2{}^2(T_2)$$
$$+ \lambda.(n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2) \qquad (7.2\text{-}4)$$

where t is the time-tag of the observable constructed from four one-way phase observations which have been made within 30 millisecond of each other (§6.3). *Note that there are now only two classes of parameters: coordinates and ambiguities.* There are a number of alternative ambiguity modelling options (see discussion later in this section). In the event that the ambiguities are resolved, eqn (7.2-4) can be rewritten as a double-differenced range equation in which the ambiguity terms are eliminated from the parameter set. A solution using this "reduced" phase observable is known as an "ambiguity-fixed" solution.

Triple-Differenced Phase Model

In length units, eqn (6.3-8) is rewritten as:

$$\delta\Delta\nabla\Phi(t_{ab}) = (\rho_1{}^1(T_{a1}) - \rho_1{}^2(T_{a1}) - \rho_2{}^1(T_{a2}) + \rho_2{}^2(T_{a2}))$$
$$- (\rho_1{}^1(T_{b1}) - \rho_1{}^2(T_{b1}) - \rho_2{}^1(T_{b2}) + \rho_2{}^2(T_{b2})) \qquad (7.2\text{-}5)$$

where a and b designate the two epochs involved. In general the epochs are consecutive, but this need not be so (for example, the between-epoch differences can be constructed always with respect to the first epoch in the session).

GPS Datum Definition Issues

The GPS satellite datum enters the GPS phase solutions through the definition of the known components of the range eqn (7.1-19) and Figure 7.1-1, that is:

(1) the known satellite coordinates X^i, Y^i, Z^i, and

(2) the fixed reference station coordinates -- this defines the origin and orientation of the local Cartesian coordinate system.

The datum in which the satellite coordinates are expressed is nominally the WGS84 system. However, the accuracy of the broadcast ephemerides is perhaps at the one or two dekametre level. The fixed station coordinates should therefore also be defined in the same (WGS84) reference system at this dekametre accuracy level. If the results are required in the local geodetic datum, the coordinates must be *transformed* using an appropriate procedure (see chapter 11).

In GPS satellite positioning, the following datums are involved:

- ❑ **World satellite datums such as WGS84 --** a geocentric reference system with axes aligned parallel to the earth's primary axes. The GPS orbit information transmitted by the satellites is referenced to this system, consequently GPS point positioning gives coordinate results in this system. It is also the system implicit in the local Cartesian system for GPS relative positioning.

- ❑ **In the case of the relative positioning mode, use is made of a local Cartesian system parallel to the global GPS satellite datum.** This system has its origin at the fixed station of a baseline or network. It is a block-shifted version of WGS84, or whatever satellite datum is being used. (For example, if post-processed GPS orbits are being used, the datum is no longer WGS84, but a satellite datum such as the International Terrestrial Reference System.)

- ❑ **In general, station coordinate information is available in the conventional geodetic system.** To integrate GPS results with terrestrial geodetic results requires the relationships between all the (nearly parallel) local and global Cartesian systems to be known, or determined.

7.2.2 ADJUSTMENT MODEL CONSIDERATIONS

By focussing on the double-differenced phase solution, the following three issues will be discussed:

- ☞ The mechanics of double-differencing.
- ☞ The modification of the observation weight matrix under differencing.
- ☞ The ambiguity modelling options.

(The former two are also applicable to triple-difference solutions.)

In §8.1, these and other issues relating to double-differenced solutions will be discussed further.

The Double-Differencing Process

<u>Observation example</u>:

☞ **R = 2 receivers**

☞ **S = 4 satellites**

☞ **E = 60 epochs**

Total of 480 one-way phase observations (R.S.E).

For each epoch:

- there are R.S observations *(here, 8).*
- there are R(R-1)S(S-1)/4 possible double-differences that can be formed *(here, 6).*
- there are only (R-1)(S-1) <u>independent</u> double-differences *(here, 3).*

There are therefore a total of 180 <u>independent</u> double-differences.

There are several ways to form, at an observation epoch, (R-1)(S-1) independent double-differences from the R.S one-way phase observations. Building differences between simultaneous one-way phase observables can be regarded as a pre-multiplication of the vector of one-way phase data by an appropriate matrix operator: the **difference operator**. Dropping super and subscripts as :

$$\Delta\nabla\phi = \mathbf{D}\cdot\phi \tag{7.2-6}$$

where **D** is the <u>Double-Difference operator</u> matrix. The dimension of the Double-Difference operator matrix is equal to [(R-1)(S-1),R.S].

Pre-multiplying both the functional model as well as the set of observations, the new system equations is:

$$\mathbf{D}\cdot\mathring{\mathbf{v}} + \mathbf{D}\cdot\mathbf{v} = \mathbf{D}\cdot\mathbf{A}\cdot\delta\mathbf{x} \tag{7.2-7}$$

or

$$\mathring{\mathbf{v}}' + \mathbf{v}' = \mathbf{A}'\delta\mathbf{x}$$

where: $\mathring{\mathbf{v}}' = \mathbf{D}\cdot\mathring{\mathbf{v}}$ is the misclose vector of double-differences,
$\mathbf{v}' = \mathbf{D}\cdot\mathbf{v}$ is the residual vector of double-differences, and
$\mathbf{A}' = \mathbf{D}\cdot\mathbf{A}$ is the design matrix of double-differences.

There are alternative forms for the Double-Difference operator. The two most common are the **base satellite** and **sequential satellite** options (Figure 7.2-1).

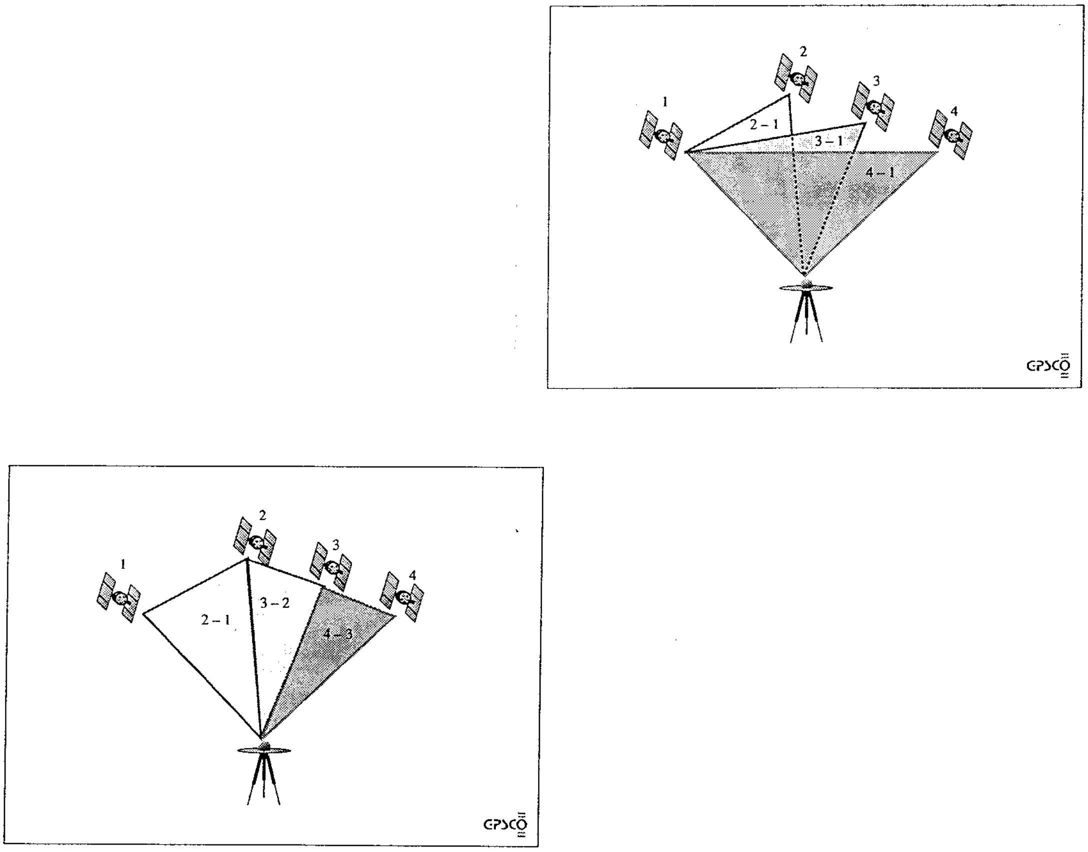


Figure 7.2-1. Between-satellite differencing strategies.

Although both approaches are mathematically equivalent (although the correlation matrices are different), in reality processing the same dataset using a different differencing scheme can lead to non-equivalent results because:

- of the presence of different characteristics of biases, errors and noise, and
- the different solution characteristics due to data dropouts, etc.

For example, if the satellite selected as the *base satellite* gives poor quality data, then this will contaminate all the double-differences. These alternative differencing schemes can be illustrated by way of an example. Assume that two receivers observe four satellites: PRN 3, 6, 9, 11.

<u>Base satellite</u>: In this option, the observable from one satellite, for example, the one with the lowest PRN number in this case, is held fixed as a reference satellite used to form the double-differences. The double-difference operator for a baseline observing the above four satellites has the form:

$$\begin{array}{cc} \mathbf{PRN} & \begin{array}{cccccccc} \mathbf{3} & \mathbf{6} & \mathbf{9} & \mathbf{11} & \mathbf{3} & \mathbf{6} & \mathbf{9} & \mathbf{11} \end{array} \\ \mathbf{D} = & \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \end{bmatrix} \end{array} \tag{7.2-8}$$

The resultant differenced observables are:

$$\begin{bmatrix} \phi_{1\text{-}2}^{3\text{-}6} \\ \phi_{1\text{-}2}^{3\text{-}9} \\ \phi_{1\text{-}2}^{3\text{-}11} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} \phi_1^3 \\ \phi_1^6 \\ \phi_1^9 \\ \phi_1^{11} \\ \phi_2^3 \\ \phi_2^6 \\ \phi_2^9 \\ \phi_2^{11} \end{bmatrix} \tag{7.2-9}$$

Note that the double-differenced observables $\phi_{1\text{-}2}^{6\text{-}9}$, $\phi_{1\text{-}2}^{6\text{-}11}$ and $\phi_{1\text{-}2}^{9\text{-}11}$ have <u>not</u> been constructed. These are dependent, and are not included in the observation set to be adjusted as the resultant Normal Equation matrix becomes singular (if the correlated nature of the double-differenced observables is reflected in the input observation weight matrix **P**). To illustrate this consider how they are formed as linear combinations of the independent double-differences:

$$\begin{aligned} \phi_{1\text{-}2}^{6\text{-}9} &= -(\phi_{1\text{-}2}^{3\text{-}6} - \phi_{1\text{-}2}^{3\text{-}9}) \\ \phi_{1\text{-}2}^{6\text{-}11} &= -(\phi_{1\text{-}2}^{3\text{-}6} - \phi_{1\text{-}2}^{3\text{-}11}) \\ \phi_{1\text{-}2}^{9\text{-}11} &= -(\phi_{1\text{-}2}^{3\text{-}9} - \phi_{1\text{-}2}^{3\text{-}11}) \end{aligned} \tag{7.2-10}$$

<u>Sequential satellite</u>: In this approach the sequential double-difference operator is of the form:

$$\begin{array}{cc} \textbf{PRN} & \begin{array}{cccccccc} \mathbf{3} & \mathbf{6} & \mathbf{9} & \mathbf{11} & \mathbf{3} & \mathbf{6} & \mathbf{9} & \mathbf{11} \end{array} \\ \mathbf{D} = & \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} \end{array} \tag{7.2-11}$$

The resultant differenced observables are:

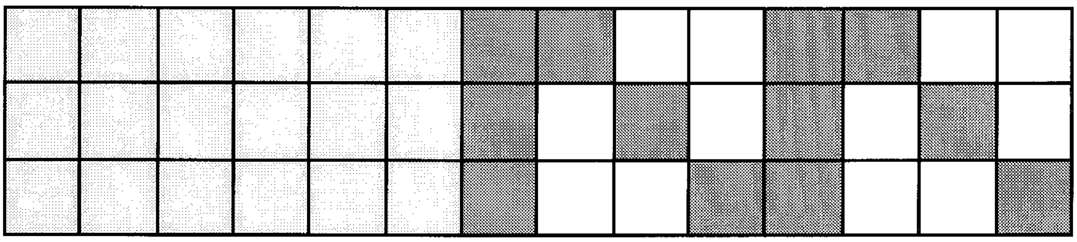
$$\begin{bmatrix} \phi_{1\text{-}2}^{3\text{-}6} \\ \phi_{1\text{-}2}^{6\text{-}9} \\ \phi_{1\text{-}2}^{9\text{-}11} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} \bullet \begin{bmatrix} \phi_1^{3} \\ \phi_1^{6} \\ \phi_1^{9} \\ \phi_1^{11} \\ \phi_2^{3} \\ \phi_2^{6} \\ \phi_2^{9} \\ \phi_2^{11} \end{bmatrix} \qquad (7.2\text{-}12)$$

Note that the double-differenced observables $\phi_{1\text{-}2}^{3\text{-}9}$, $\phi_{1\text{-}2}^{3\text{-}11}$ and $\phi_{1\text{-}2}^{6\text{-}11}$ have not been constructed. They are linearly related to the three independent double-differences through similar relations to eqn (7.2-10).

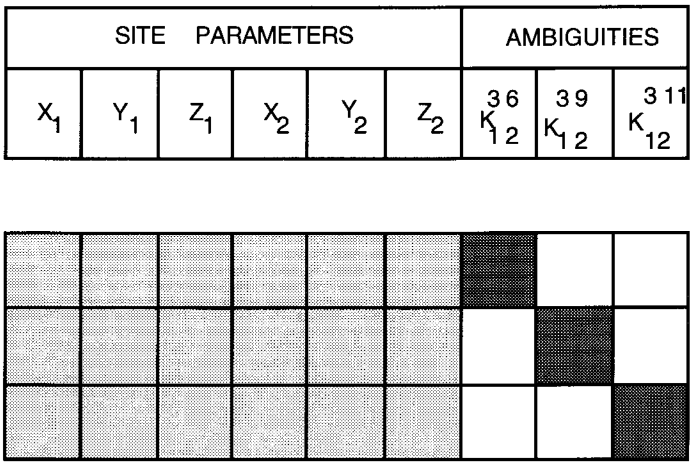
The difference operator is the same for every epoch only if the *same* constellation of satellites is tracked for the entire observation session. The situation in which satellites rise and set during the session complicates things, and requires considerable "book-keeping".

The design matrices of the base satellite double-difference and sequential double-difference for a single epoch are illustrated in Figures 7.2-2 and 7.2-3 respectively. Figures 7.2-2a and 7.2-3a show the design matrices of double-differences in the case where the ambiguity parameters are represented by the undifferenced ambiguity model. Figures 7.2-2b and 7.2-3b illustrate the design matrices of double-differences when the ambiguity parameters are in terms of the double-differenced ambiguity model. (Ambiguity models are discussed later in this section.)

SITE PARAMETERS						AMBIGUITY PARAMETERS							
X_1	Y_1	Z_1	X_2	Y_2	Z_2	n_2^1	n_2^2	n_2^3	n_2^4	n_2^1	n_2^2	n_2^3	n_2^4

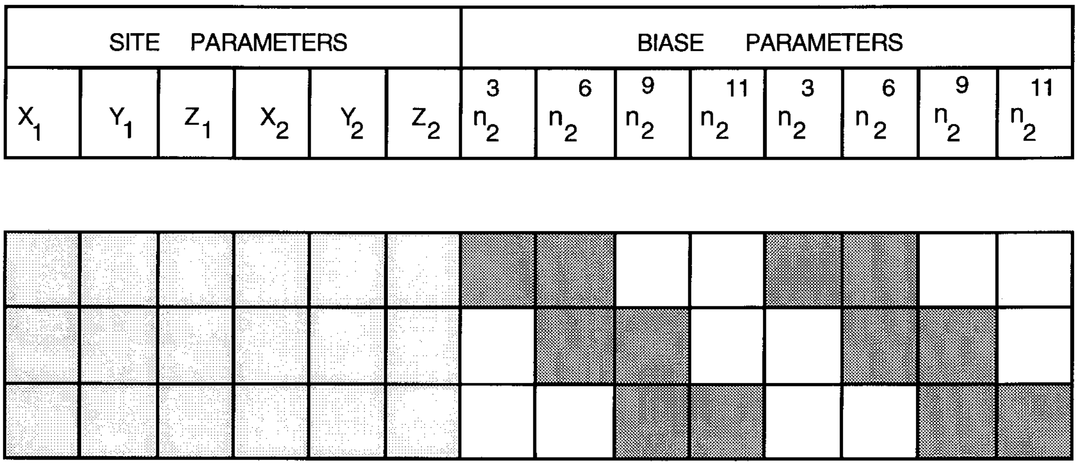


(a) Undifferenced ambiguity model case.

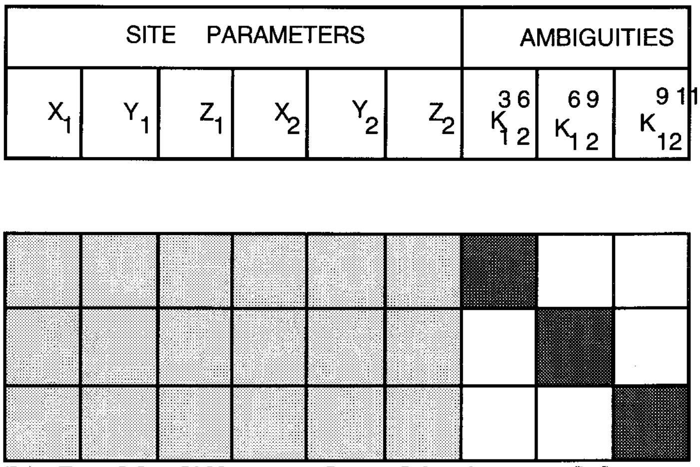


(b) Double-differenced ambiguity model.

Figure 7.2-2. Design matrices of base satellite double-differences for an epoch (2 sites, 4 satellites).



(a) Undifferenced ambiguity model case.



(b) Double-differenced ambiguity model case.

Figure 7.2-3. Design matrices of sequential satellite double-differences for an epoch (2 sites, 4 satellites).

GPS Correlation

Correlations express the inter-dependence between variables. For two variables x and y in a linear relationship, the correlation between them is defined as:

$$\chi_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{7.2-13}$$

where: σ_x is the standard deviation of x,
σ_y is the standard deviation of y, and
σ_{xy} is the covariance between x and y.

The value of correlation lies in the range –1 and +1. When the correlation of two variables approaches the maximum of ±1, the two variables are said to be highly correlated. It is worth noting that high correlation does not mean that the variations of one are caused by the variations of the others, although it may be the case. In many cases, external influences may be affecting both variables in a similar fashion. There are two types of correlation encountered in GPS measurement and data processing: **physical correlation** and **mathematical correlation.**

Physical correlation refers to the correlations between the actual field observations. It arises from the nature of the observations as well as their method of collection. If different observations or sets of observation are affected by common external influences, they are said to be physically correlated. Hence all observations made at the same time at a site may be considered physically correlated because similar atmospheric conditions and clock errors influence the measurements.

Mathematical correlation is related to the parameters in the mathematical model. It can therefore be partitioned into two further classes which correspond to the two components of the mathematical adjustment model:

(a) **Functional Correlation:** The physical correlations can be taken into account by introducing appropriate terms into the *functional model* of the observations. That is, functionally correlated quantities share the same parameter in the observation model. An example is the clock error parameter in the one-way GPS observation model, used to account for the physical correlation introduced into the measurements by the receiver clock and/or satellite clock errors.

(b) **Stochastic Correlation:** Stochastic correlation (or statistical correlation) occurs between observations when non-zero off-diagonal elements are present in the variance-covariance (VCV) matrix of the observations. This correlation also appears when functions of the observations are considered (for example, differencing), due to the Law of Propagation of Variances. *However, even if the VCV matrix of the observations is diagonal (no stochastic correlation), the VCV matrix of the resultant Least Squares estimates of the parameters will generally be full matrices, and therefore exhibit stochastic correlation.*

The accuracy of an absolute position is a function of the accuracies of the observed quantities in the functional model. Due to uncertainties in the satellite's position, satellite and receiver clock errors and propagation delays, the absolute point position using pseudo-range data will be accurate to no better than several dekametres (§2.4). However, the error sources in the GPS system (satellite orbits, satellite and receiver clocks, atmospheric propagation, etc.) will exhibit some physical correlation among the signals received at several stations that are simultaneously tracking the same set of satellites. This is why for survey applications it is necessary to use

GPS in a "relative mode", in which two or more receivers must observe the same satellites simultaneously. These physical correlations can then be physically modelled as common bias terms in the function model. Using mathematical operations such as data differencing, these effects can be eliminated or greatly reduced, leading to centimetre level accuracy. As the inter-station distances become larger, a *decorrelation* of the physical effects of orbit error and atmospheric delay will tend to occur.

The fundamental problem which is encountered when performing differencing operations is to accurately and completely determine to what extent various elements are correlated and how well they can be modelled. Physical correlations may be both spatial and temporal in nature, or a combination of both. An example of this is the refraction delay which a satellite's microwave signal experiences as it propagates through the ionosphere and troposphere. This delay is a significant source of both spatial and temporal physical correlation which is difficult to handle.

On the basis of experimental data, EL-RABBANY (1994) has derived an exponential covariance function that describes the temporal physical correlation:

$$\chi(\tau) = \exp\left(\frac{-/\tau/}{T}\right) \qquad (7.2\text{-}14)$$

where $\chi(\tau)$ is the correlation coefficient for a time lag of τ (in secs), and T is the correlation length (in secs). A typical value of T is about 250-350 seconds, implying that data collected more than 350 seconds apart, from the same receiver to the same satellite, can be considered completely uncorrelated (correlation coefficient is zero). The values of $\chi(\tau)$ for different data rates (between epoch measurement rate) quoted by IBID (1994) are: 0.996, 0.981, 0.944, 0.891, 0.794, for τ equal to 1, 5, 15, 30, 60 seconds respectively. Hence, assuming a constant data rate, the correlation coefficient between any two epochs can be expressed as:

$$\chi(i) = \chi(1)^i \qquad (7.2\text{-}15)$$

For example, for a 15 second data rate $\chi(1)$=0.944, and the correlation coefficient for data between consecutive epochs is 0.944, but it is 0.944^2 for two sets of data 30 seconds apart, and 0.944^3 for two sets of data 45 seconds apart, and so on. Hence the temporal correlation decreases significantly as the time interval between data increases. This empirically derived correlation could be introduced into the VCV matrix of one-way observations, however *no commercial software incorporates this and hence this is one of the reasons why the accuracy of baseline results will be optimistic.* This problem is discussed further in §9.2, §9.3 and §9.4. In the following development, no between-epoch correlation will be assumed to be present.

When *differencing* occurs, the original carrier phase observations and their one-way phase mathematical model are modified, hence leading to mathematical correlation of the resulting observables. This means that both the **functional** and **stochastic** models change. The new functional model has already been given in eqn (7.2-4). Generally **it is assumed that at a particular epoch all (one-way) carrier phase observations are independent and have the same variance**. (However, this is not entirely true because the magnitude of residual atmospheric biases are likely to be a function of satellite elevation angle, and hence the observation weights should vary as a function of the elevation angle.) Of interest is the *error propagation* rather than the absolute values, hence the variance can be set to unity. Therefore:

$$\mathbf{Q}_l = \mathbf{I} \qquad (7.2\text{-}16)$$

where $\mathbf{I}$ is the Identity matrix. Applying the Law of Propagation of Variances, the VCV matrix of the double-differenced observations, and hence the new stochastic model, becomes:

$$\mathbf{Q}_l' = \mathbf{D}\mathbf{Q}_l\mathbf{D}^T \tag{7.2-17}$$

The covariance matrix of the double-differenced observations is dependent on the double-difference operator used to construct the differenced observables. Using the double-difference operators (defined by eqns (7.2-8) and (7.2-11)) in eqn (7.2-17) leads to the VCV matrix of the base satellite double-differences and the sequential double-differences respectively:

$$\mathbf{Q}_l' = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} \tag{7.2-18a}$$

$$\mathbf{Q}_l' = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix} \tag{7.2-18b}$$

Both VCV matrices illustrate stochastic correlation (that is, non-zero off-diagonal terms). However, in the case of between-receiver differencing (§6.3), the resultant single-differences are *uncorrelated* (see HOFMANN-WELLENHOF et al, 1994).

The double-differencing process therefore introduces mathematical correlations into the resultant observables. As a consequence, to ensure the mathematical equivalence between processing using the differenced and undifferenced phase data approach (§6.1), it is necessary to construct a variance-covariance matrix of the double-differenced observables which expresses this correlation between observables. In practice it is not necessary to compute the VCV matrix $\mathbf{Q}_l'$ at every epoch, because the covariance matrix only changes with a change in the satellite-receiver combinations used in the construction of the double-differences. Triple-differenced data is also mathematically correlated (see, for example, IBID, 1994).

In most commercial GPS software packages the mathematical correlation arising from data differencing is usually neglected, and a diagonal VCV matrix for the double-differenced and triple-differenced data is assumed. The impact of this simplification on data adjustment is generally small for short baselines, perhaps of the order of 1 or 2ppm. Not taking into account the mathematical correlations between double-differences formed when more than two receivers are operating simultaneously (two or more independent baselines) is an important distinction between **single-baseline** and **multi-baseline processing** (see §9.2).

Double-Differenced Ambiguity Model Options

There are several options for parameterising the ambiguity component of the double-difference observation eqn (7.2-4), the most common are:

(1) Use the one-way ambiguity model option whereby each double-difference observable is modelled using four ambiguity parameters, for example n_1^3, n_1^4, n_2^3, n_2^4.

(2) Use a "lumped" single-differenced ambiguity model whereby each double-differenced observable is modelled using two ambiguity parameters, for example k_{12}^3 ($= n_1^3 - n_2^3$) and k_{12}^4 ($= n_1^4 - n_2^4$).

(3) Use a "lumped" double-differenced ambiguity model whereby each double-differenced observable is modelled by a single ambiguity parameter, for example K_{12}^{34} $(= n_1^3 - n_1^4 - n_2^3 + n_2^4)$.

GRANT et al (1990) have made a detailed study of the rank deficiencies in the GPS phase observation model. There are rank defects in both the geodetic (station coordinate) and ambiguity parameters, because the phase observable (differenced or undifferenced) does not contain datum information. (The fixed satellites do provide some datum, but it is very weak -- they are very distant, of the order of 20200km, compared to the baseline lengths -- and hence it is customary to hold a single station fixed and solve for the baseline components.) The fixed (geodetic and ambiguity) parameters are known as the **reference** or **datum coordinate / ambiguity** parameters. *The rank defect in the ambiguity parameters varies with the ambiguity model adopted.* In terms of an observation session involving R receivers tracking S satellites, the relevant conclusions are:

- For the one-way ambiguity option, out of a total of $(R.S)$ ambiguity parameters, $(R+S-1)$ of these need to be held fixed to overcome the rank defect problem. This is true for the undifferenced data (for which the one-way ambiguity model is the most appropriate) as well as for the double-differenced observables. A common means by which this is implemented is to hold fixed (at some arbitrary value such as zero) all the ambiguities (n_j^i) associated with a "base receiver" and a "base satellite".
- For the single-differenced ambiguity option (appropriate for double-differenced data), out of a total of $(R.S-S)$ ambiguity parameters, $(R-1)$ of these need to be held fixed to overcome the rank defect problem. This is usually implemented by holding fixed (at some arbitrary value such as zero) all the ambiguities (k_{rj}^i) associated with a "reference satellite".
- For the double-differenced ambiguity option (also appropriate for double-differenced data), out of a total of $(R.S-R-S)$ ambiguity parameters, none needs to be held fixed as no rank defects exist with this option.
- As an alternative to holding reference ambiguities fixed (in such an arbitrary manner), the singular Normal Equation Matrix resulting from the inclusion of all (one-way) ambiguity parameters is inverted using the Pseudo-Inverse. The resultant ambiguity estimates will not be integers, but certain linear combinations formed by double-differencing will be. (This is not an option found in commercial GPS phase data reduction software.)

In general, the ambiguity model to be used within GPS reduction software is *not selectable*. Furthermore, apart from certain algorithm design considerations, the various models are equivalent. For example, in the PoPS™ and SKI™ software the single-difference ambiguity model is used, whereas in the TRIMVEC™ and GPSurvey™ software the double-difference model is adopted.

7.2.3 THE STRUCTURE OF THE NORMAL EQUATIONS

Observation example:

☞ **R = 2 receivers, S = 4 satellites, E = 60 epochs**

There are a total of 180 independent double-differences = (R-1).(S-1).E.

Parameters:

- 6 coordinate components.
- 3 double-differenced ambiguity parameters, OR 4 single-differenced ambiguity parameters, OR 8 undifferenced ambiguity parameters.

Hence 6 parameters are to be estimated, 3 baseline components and 3 ambiguity parameters, because:

- one station has to be held fixed, and
- 1 (single-differenced model), or 5 reference (undifferenced model) ambiguities have to be held fixed.

The Undifferenced Model

In the case of double-differenced observations, using the undifferenced ambiguity model, the structure of the Normal Equation system is illustrated in Figure 7.2-4, where:

- is the station coordinate parameter partition
- is the station coordinate-ambiguity partition
- is the phase ambiguity partition

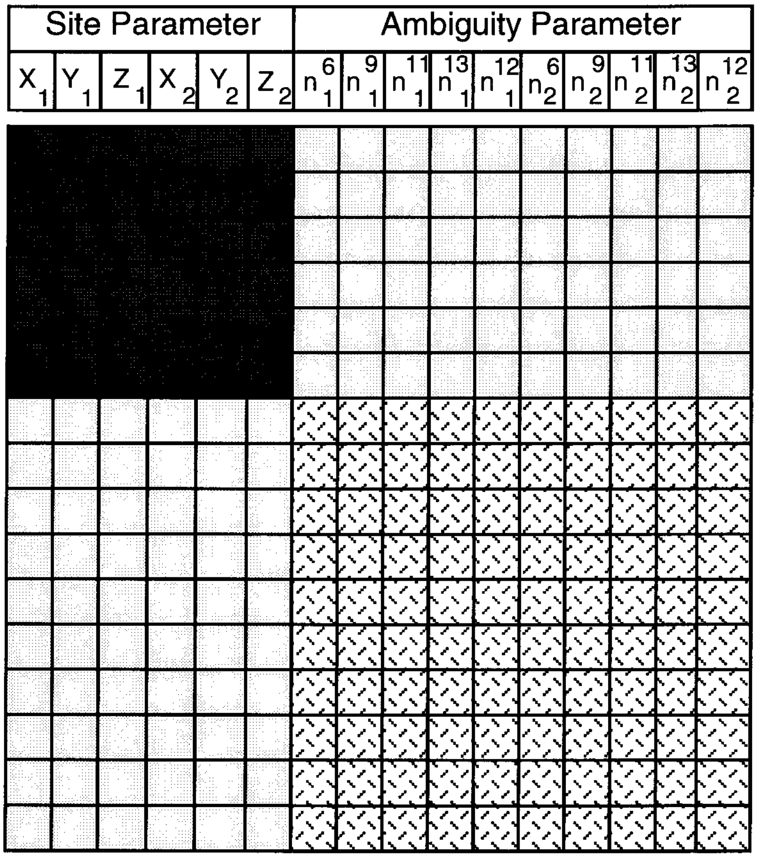


Figure 7.2-4. The structure of the Normal Equation Matrix for the undifferenced ambiguity model (2 receivers, 5 satellites).

Note that three (3) coordinate parameters and six (6) ambiguity parameters are not estimable!

The Double-Differenced Model

In the case of double-differenced observations, using the double-differenced ambiguity model, the structure of the Normal Equation system is illustrated in Figure 7.2-5.

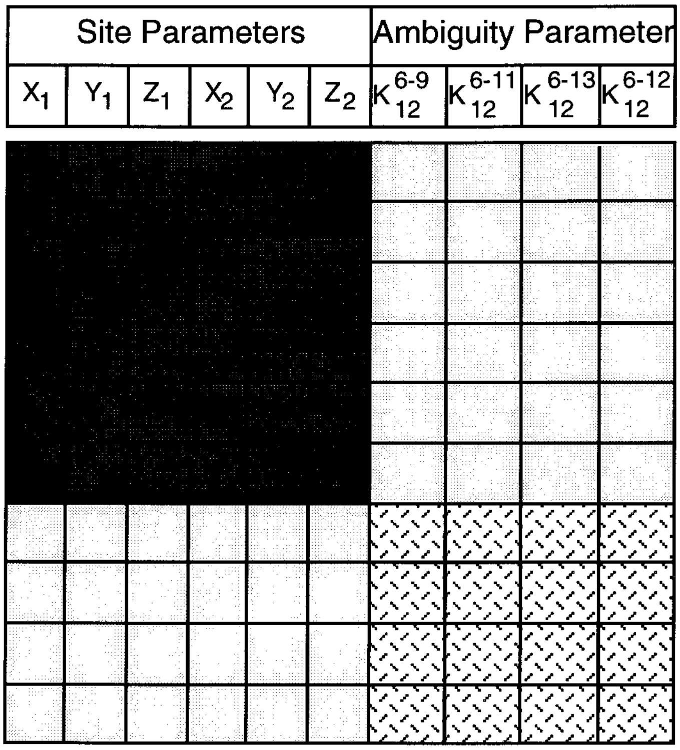


Figure 7.2-5. The structure of the Normal Equation Matrix for the double-differenced ambiguity model (2 receivers and 5 satellites).

Note that three (3) coordinate parameters are not estimable, although all ambiguity parameters are estimable!

7.3

GETTING STARTED: PRE- & INITIAL PROCESSING STEPS

The general initial / pre-processing steps for a GPS phase adjustment are:

- ❑ Transfer, compress and reformat the data (including the recorded Navigation Message) from the GPS receivers onto computer storage media. This is may be done after each observation session, for each receiver separately, using the manufacturer's software.
- ❑ Prepare data, for example form a common ephemeris file (from the multiple copies of the recorded Navigation Messages), or obtain "precise ephemerides" from an external source. If processing is to be done using third-party software (as in the case of scientific GPS processing), data files may need to be reformatted.
- ❑ Screen data: edit data according to quality flags, or if to satellites below elevation cutoff angle, unhealthy satellites, unequal observation sessions, etc. This step usually takes place when merging data from several receivers.
- ❑ Obtain preliminary point position solution, usually through the processing of pseudo-range data. This has the advantage of, in addition, giving apriori station coordinates as well as observation time-tag corrections.
- ❑ Obtain approximate baseline solution(s) using triple-differenced phase data (no ambiguities). Such a solution helps subsequent cycle slip detection and repair.
- ❑ Detect and repair cycle slips by any of a number of methods.

The most important steps are time-tag synchronisation and cycle slip editing. The triple-difference solution, although usually carried out before the cycle slip screening step, will be discussed in §8.1.

There are a number of comments that can be made in relation to data handling:

- Data files should be downloaded, backed up on suitable media, and archived as soon as possible -- *for example at the field office.*
- Data processing can commence *on a session basis* when data files from several receivers are brought together at the processing centre.
- Different receivers output one or more files per observation session *in a proprietary format.*
- Data from a particular brand of receiver is usually processed with that brand's software -- *exceptions possible using RINEXed data files.*

The RINEX Data Exchange Format

RINEX (Receiver INdependent EXchange) is an exchange format devised in 1989 for geodetic applications requiring the international exchange of GPS datasets, gathered during global campaigns, by different brands of receivers (GURTNER, 1994). It has now been adopted as the standard exchange and storage format for GPS surveying and precise navigation applications as well. RINEX Version 2, which is able to handle kinematic GPS data, is now in common use.

The RINEX format has the following characteristics:

- ASCII format, with a maximum of 80 characters per record.
- Phase data recorded in cycles of L1 and L2 carrier frequencies, pseudo-range data in metres.
- All receiver-dependent calibrations already applied to data.
- Time-tag is the time of the observation in the receiver clock time frame (not necessarily in GPS Time).
- Separate data (Table 7.3-1), Navigation Message (Table 7.3-2), and meteorological data files with header sections.

Table 7.3-1. Sample RINEX data file.

```
----|---1|0---|---2|0---|---3|0---|---4|0---|---5|0---|---6|0---|---7|0---|---8|0

 PASADENA CA 91109 USA |      LOGOS::JPLGPS                 COMMENT
TIDB                                                        MARKER NAME
DSN                 JET PROPULSION LABORATORY               OBSERVER / AGENCY
     2              DSN ROGUE           DSN ROGUE SOFTWARE  REC # / TYPE / VERS
     2              D. AND M.  C146-6-1                     ANT # / TYPE
 -4460853.8350  2682467.2310 -3674736.3630                  APPROX POSITION XYZ
        0.0000        0.0000        0.0000                  ANTENNA: DELTA H/E/N
     1     1                                                WAVELENGTH FACT L1/2
     5    C1    L1    L2    P1    P2                        # / TYPES OF OBSERV
    30                                                      INTERVAL
    92     7    26     6    52   30.000000                  TIME OF FIRST OBS
    92     7    26    23    59   30.000000                  TIME OF LAST OBS
 92  7 26  6 52 30.0000000  0  5  2 11 18 19 28
  22333042.96600 -12176065.17700  -9487835.16600  22333025.72200  22333027.39100
  22934353.22700  -7227959.70200  -5632169.61500  22934338.21900  22934340.52000
  20485466.29600 -23339757.04000 -18186808.56800  20485447.35300  20485447.55500
  20609091.79300 -20568272.40900 -16027208.73300  20609081.52200  20609081.81200
  23894196.98000     154816.73100     120640.79300  23894184.83300  23894185.36800
 92  7 26  6 53  0.0000000  0  6  2 11 16 18 19 28
  22315737.73400 -12267004.73700  -9558697.09700  22315720.51000  22315722.15200
  22922104.35200  -7292336.30500  -5682333.14300  22922087.79100  22922090.05900
  23641617.41410  -3724083.52810  -2901879.61310  23641605.66310  23641608.09510
  20491946.68800 -23305701.79300 -18160272.01900  20491927.79600  20491927.80600
  20609475.02700 -20566263.57700 -16025643.39900  20609463.71700  20609463.99400
  23903831.76500     205446.05300     160092.21000  23903819.21500  23903820.15300

----|---1|0---|---2|0---|---3|0---|---4|0---|---5|0---|---6|0---|---7|0---|---8|0
```

Table 7.3-2. Sample RINEX Navigation Message file.

```
----|---1|0---|---2|0---|---3|0---|---4|0---|---5|0---|---6|0---|---7|0---|---8|0

     1              NAVIGATION DATA                         RINEX VERSION / TYPE
DMD_TO_RINEX v1.5   JET PROPULSION LAB. 92-07-27 16:50:22   PGM / RUN BY / DATE
FILES OBTAINED FROM THIRD PARTIES MAY NOT BE IDENTICAL TO   COMMENT
 THE LATEST AVAILABLE FROM JPL. FOR INFORMATION CONTACT:    COMMENT
 THOMAS G. LOCKHART   | PHONE (818)354-6102                 COMMENT
 NASA/JPL M/S 238-625 | FAX   (818)393-4965                 COMMENT
 4800 OAK GROVE DRIVE | EMAIL jplgps@logos.dnet.nasa.gov    COMMENT
 PASADENA CA 91109 USA |      LOGOS::JPLGPS                 COMMENT

26 92  7 26  8 51 44.0 1.635868102312D-06-3.410605131648D-13 0.000000000000D+00
    2.170000000000D+02 1.128125000000D+01 4.738768932810D-09 5.863412966114D-02
    4.898756742477D-07 8.141674217768D-03 9.039416909218D-06 5.153610269547D+03
    3.190400000000D+04-7.636845111847D-08-1.371976967685D+00-4.470348358154D-08
    9.603279283700D-01 2.067812500000D+02-1.491259472097D+00-8.059264366977D-09
   -4.760912775126D-10 1.000000000000D+00 6.550000000000D+02 0.000000000000D+00
    3.200000000000D+01 0.000000000000D+00 0.000000000000D+00 2.170000000000D+02
15 92  7 26  8 51 44.0-2.961605787277D-07-4.547473508865D-13 0.000000000000D+00
    3.000000000000D+01-5.693750000000D+01 4.155887189938D-09 1.478679355308D+00
   -2.870336174965D-06 7.184803485870D-03 1.033581793308D-05 5.153687067032D+03
    3.190400000000D+04 1.918524503708D-07 2.862619155647D+00 5.215406417847D-08
    9.620121289009D-01 1.818437500000D+02 1.902453767716D+00-7.588887740440D-09
    4.989493818108D-10 1.000000000000D+00 6.550000000000D+02 0.000000000000D+00
    3.200000000000D+01 0.000000000000D+00 0.000000000000D+00 3.000000000000D+01

----|---1|0---|---2|0---|---3|0---|---4|0---|---5|0---|---6|0---|---7|0---|---8|0
```

Most commercial GPS surveying packages will *output* RINEX formatted data files. Many commercial GPS surveying packages will also permit the *input* of RINEX formatted data files for processing. These have the following implications:

☞ Data from one brand of receiver may be processed within another commercial software package.

☞ Data from campaigns using a mixed set of GPS receivers can be processed within one software package.

In reality, the processing of data collected by a certain brand of GPS receiver using software not explicitly designed to process this data can be problematical. The main problem arises because of incompatible time-tags (see Figures 6.3-9 and 6.3-10).

Time-Tag Synchronisation

There are in fact two effects (§6.3):

(a) The synchronisation of two (or more) GPS receivers, so that the assumptions made in modelling of the double-differenced observations are satisfied.

(b) The synchronisation of the observation time-tags to the time system of the satellite ephemerides.

The former condition must be satisfied at the microsecond level, while the latter only at the millisecond level. With today's code-correlating receivers this is easily accomplished. As data is collected by the receiver, an internal "navigation solution" is performed that solves for the receiver clock offset to GPS Time (ε_{rcj}) using pseudo-range data. Even under conditions of Selective Availability, such a calculation ensures that the receiver clock can be synchronised to GPST to the level of a few microseconds. The problem is, however, that the time-tag interval must be correctly initialised in all receivers, for example collect data on each minute, or even minute only, etc. Mixing receivers of different types may therefore be a problem unless it is absolutely clear that the time-tags (although individually accurate) correspond to the same epoch. **This synchronisation is therefore a task of the field procedures**. <u>It does not mean that the observations at the receivers must be made within a microsecond of each other!</u>

If time-tag synchronisation is not carried out within the receiver, the pseudo-range data can be processed at a *pre-processing step.* Unlike the standard navigation solution in which the receiver is assumed to be in motion, and a separate clock offset parameter is estimated each epoch, the pseudo-range solution has the following characteristics (§7.2):

(1) Advantage is taken of the fact that the receiver does not move during an entire observation session. Hence there are very many pseudo-range observations to estimate the three coordinate component parameters. In addition, the coordinate solution provides *apriori* values for the subsequent phase reduction step.

(2) The receiver clock error, as a time-tag error (in seconds), can be modelled as a polynomial (eqn (7.2-2):

$$\varepsilon_{rcj}(t_j) = a_{0j} + a_{1j}.(t_j - T_{0j}) + a_{2j}.(t_j - T_{0j})^2 \qquad (7.3\text{-}1)$$

The time-tags are then corrected for this offset, that is:

$$T_{corr}(t_j) = T_{obs}(t_j) + \varepsilon_{rcj}(t_j) \qquad (7.3\text{-}2)$$

However, the data at this epoch to all the satellites are not corrected for clock error as the estimates of ε_{rcj} are not accurate enough to perform the following transformation to the sub-nanosecond accuracy level:

$$P_{corrj}{}^{i}(T_j) = P_{obsj}{}^{i}(T_j) + c.\varepsilon_{rcj}(T_j) \quad \text{for pseudo-ranges to all satellites i} \qquad (7.3\text{-}3a)$$

and

$$\Phi_{corrj}{}^{i}(T_j) = \Phi_{obsj}{}^{i}(T_j) + c.\varepsilon_{rcj}(T_j) \quad \text{for carrier phase to all satellites i} \qquad (7.3\text{-}3b)$$

Apriori Site Coordinates: The Triple-Difference Solution

Often a preliminary triple-difference solution is performed to obtain the baseline solutions that form the basis for the subsequent cycle slip editing step, as well as to provide good quality apriori coordinates for the final double-difference solution.

Such solutions are <u>tolerant to cycle slips</u> in the data (slips appear as isolated "outliers", see Figure 7.3-1 and Table 6.3-2). Triple-difference solutions are discussed in §8.1.

Detection and Repair of Cycle Slips

Although triple-difference solutions are tolerant of cycle slips, double-difference solutions cannot tolerate cycle slips (see Table 6.3-2), hence all data must be "cleaned" before reaching this stage of the adjustment.

Cycle slips occur when the continuous tracking of a satellite is interrupted by an obstruction, or the antenna being moved too fast, or faulty signal processing within the receiver, or even when the ionospheric activity is too high (WANNINGER, 1993). They cause the integrated carrier phase count to become "corrupted" once signal lock-on is reacquired on the satellite. Cycle slips generally occur at a receiver tracking a satellite, and it is possible to have slips on more than one satellite at the same time. Slips can occur independently on L1 and L2. It is a characteristic of cycle slips that all observations taken after the cycle slip has occurred are shifted by the same **integer** number of cycles (§6.2).

The detection and repair of cycle slips is therefore an important pre-processing step. It can also be a labour intensive and time consuming process if the data is noisy, has gaps and has many slips on more than one satellite at a time. *Automatic procedures are generally used for standard GPS surveying applications addressed by commercial processing software.*

Several techniques have been developed to perform this task. Cycle slip *detection* is easier than its *correction*, and the difficulty is a function of: the positioning mode, the baseline length, the type of data available, the antenna dynamics, etc., hence the following comments can be made (§6.2):

- Cycle slip *detection/correction* of dual-frequency data is *easier* than for single frequency data (§6.4 and §8.4).
- Cycle slip *detection/correction* of differenced data is *easier* than for one-way (undifferenced) phase data, because of the elimination of the clock biases.
- Cycle slip *detection/correction* of static data is *easier* than the case of kinematic data.
- Cycle slip *detection/correction* of short baseline data (<30km) is *easier* than for long baseline data.
- Cycle slip *detection/correction* of data in the post-mission mode is *easier* than for the case of real-time data processing.

All cycle slip procedures are dependent on the (implicit or explicit) analysis of the residual pattern of the data from epoch to epoch. Hence they are dependent on how well the rate-of-change of phase can be predicted from epoch to epoch. Rate-of-change of phase is a function of the rate-of-change of the receiver-satellite geometric range, as well as the other biases. The ease with which the rate-of-change of phase can be predicted is a function of:

- The data collection rate -- *the higher the rate (the shorter the time between observations) the easier it is to predict the phase/range change from one epoch to the other.*
- How well the errors and biases in the observation model have been accounted for. For example, the atmospheric refraction (particularly the ionosphere in the case of single frequency observations) and multipath, and receiver and/or satellite clock biases -- *this is best accomplished through data differencing.*
- Whether the GPS receiver is stationary or moving -- *the phase "jump" between epochs may be due to antenna motion rather than a cycle slip.*
- How good the apriori coordinates of the baseline stations are -- *these must be known to the few decimetre level to ensure that cycle slips are not masked by baseline errors.*

- The baseline length -- *the longer the baseline the more difficult it is to detect cycle slips as they are masked by residual biases due to the decorrelated atmospheric biases, and the impact of orbit errors.*

In general, the cycle slip detection methods can be classified as follows, and the various test observables include the single- and double-differenced data, the raw undifferenced phase data, linear combinations of L1 and L2 phase data, and combinations of phase and pseudo-range data:

(a) Use of a low-order polynomial to fit the time series of the tested observable -- *the residual after fitting is screened for the cycle slip.* This is the most commonly used method.

(b) Use of a Kalman filter (§7.4) to forward predict the phase observable -- *the difference between the predicted and measured test observable is used for cycle slip detection.* Often used with kinematic phase data.

(c) Use of a first, second, third and even fourth between-epoch differences to highlight any anomalous single epoch slips -- *at some level of differencing the resulting values are almost a **constant**, and any slips are amplified.* This is illustrated in Table 7.3-3.

(d) Use of a precise pseudo-range data allows the ambiguity to be independently determined at each epoch, at some level of accuracy (§6.4 and §8.4) -- *any variation in the value of the ambiguity from epoch to epoch can be construed as a cycle slip.* A common strategy for kinematic data and precise static GPS geodesy, as both applications use top-of-the-range receivers able to measure all observables (see §4.3).

(e) Use of dual-frequency phase and/or pseudo-range data to independently determine the ambiguity at each epoch -- *any variation in the value of the ambiguity from epoch to epoch can be construed as a cycle slip.* This is particularly appropriate for kinematic data where "on-the-fly" ambiguity resolution techniques are used (§8.3).

(f) Use of external data within an integrated system such as when GPS is combined with an Inertial Navigation System (INS) -- *the INS can measure change in position of the antenna very precisely and this can be compared to that implied by the phase measurements.* This is particularly appropriate for real-time kinematic airborne GPS applications.

Table 7.3-3. First, second and third between-epoch differences.

Single Epoch Observable	First Difference (across two epochs)	Second Difference (across two 1st diffs)	Third Difference (across two 2nd diffs)
$\phi_j^i{}_{(e-2)}$			
	$\delta\phi_j^i{}_{(e-2,e-1)}$		
$\phi_j^i{}_{(e-1)}$		$\delta^2\phi_l^i{}_{(e-2,e-1,e)}$	
	$\delta\phi_j^i{}_{(e-1,e)}$		$\delta^3\phi_l^i{}_{(e-2,e-1,e,e+1)}$
$\phi_j^i{}_{(e)}$		$\delta^2\phi_l^i{}_{(e-1,e,e+1)}$	
	$\delta\phi_j^i{}_{(e,e+1)}$		$\delta^3\phi_l^i{}_{(e-1,e,e+1,e+2)}$
$\phi_j^i{}_{(e+1)}$		$\delta^2\phi_l^i{}_{(e,e+1,e+2)}$	
	$\delta\phi_j^i{}_{(e+1,e+2)}$		
$\phi_j^i{}_{(e+2)}$			

Note that only large cycle slips are likely to show up in the one-way residuals (because of the unpredictable size of the receiver and satellite clock's erratic behaviour), whereas all cycle slips should be visible in the double-differences (Figure 7.3-1a). Figure 7.3-1b illustrates the residuals after between-epoch differencing. Note the "outlier" effect illustrated in Table 6.3-2.

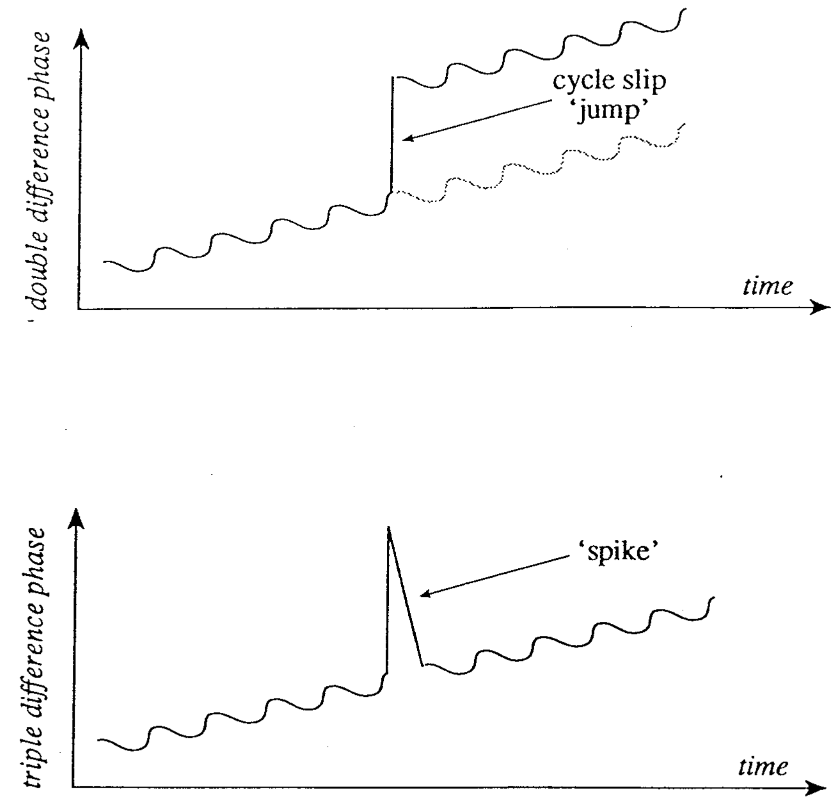


Figure 7.3-1. Cycle slip signature in double-differenced phase (a "jump") and triple-differenced phase data (an "outlier") series.

Cycle slip detection and repair procedures therefore consist essentially of the following steps:

(1) Identifying a baseline and observable (undifferenced or differenced) to be processed.

(2) Obtaining good *apriori* coordinates of the baseline stations using, for example, a cycle slip tolerant triple-difference solution. If a triple-differenced observable is constructed from a phase observation contaminated by a cycle slip it could be detected in the residuals *after adjustment*, as the large residual (a multiple of the wavelength) would be located at the epoch at which the cycle slip had occurred. It's magnitude may be subsequently determined through an appropriate screening strategy.

(3) The change in one-way residuals from epoch to epoch can indicate the location of a cycle slip. This change may be modelled by a polynomial or determined from successive between-epoch differences as illustrated in Table 7.3-3. Alternatively, the range-rate estimated from pseudo-range data can be used to repair the "big" slips, leaving the "fine" detection for a subsequent step involving either single- or double-differences.

(4) Changes in single-difference residuals may then be scanned. Using some criteria, a "jump" in the residuals can be identified and the slip repaired. This step does not

usually unambiguously determine the value of the cycle slip at the one cycle level (approximately 20cm).

(5) Changes in the double-difference residuals are much smoother than in the case of single-differenced or one-way phase data, as all the clock errors have been eliminated. Hence the precise value of the cycle slip can be determined, and the data repaired. *Note however that the repair is to the double-difference and not to the particular receiver-satellite observation that was the source of the slip in the first place.*

(6) If dual-frequency data is scanned, the process may be repeated on the L2 observable, or linear combinations of the L1 and L2 observable, as discussed below.

If scanning dual-frequency observations, this process would take place independently on L1 and L2, as well for certain linear combinations of L1 and L2 data (see §6.4, §8.4, and HOFMANN-WELLENHOF et al, 1994, for further details). It may also be necessary to screen for "half-cycle" slips on L2 if "squaring" receivers have been used.

The "clean" double-differenced data is now ready for the main processing step.

Chapter 8: GPS Baseline Processing

8.1 BASELINE PROCESSING PRINCIPLES

> **A GPS campaign generally involves the use of several GPS receivers deployed over a network of points for a number of observation sessions.**

This chapter will focus on single baseline processing as this is the fundamental unit of a GPS solution. Most commercial GPS data processing software will accept only the simultaneous phase data collected by two GPS receivers. This is because the observable modelling necessary for GPS phase data reduction involves two stations -- defining a baseline. This is obvious for the double-differenced and triple-differenced phase models, but perhaps less obviously for the case of the undifferenced model. It is however stated, without proof, that in order to estimate the clock parameters, appearing explicitly in the undifferenced observation eqns (6.1-13), data from several sites to several satellites must be collected and processed together. In §6.1 comments were made regarding the equivalence of the "undifferenced data processing approach" and the "double-differenced data processing approach" to GPS phase data reduction. (It is beyond the scope of these notes to algebraically prove that the result is exactly equivalent to processing double-differenced data.) To differentiate between the two approaches, the former may be referred to as involving **implicit differencing** (carried out in the solution algorithm), while the conventional double- and triple-differences are generated by **explicit differencing**.

As many of the remarks about double-differenced data (baseline) solutions hold equally whether the differencing is "implicit" or "explicit", it is not necessary to distinguish between undifferenced and double-differenced data models for baseline solutions. They will be simply referred to as the **double-differenced mode** of data processing to differentiate it from the **triple-differenced solution mode**.

An entire network may be built up either from a large number of independently processed baselines, or, more efficiently, in a simultaneous adjustment in which the set of differenced observations is generated in a mathematically correlated fashion. This aspect of GPS data reduction will be discussed in chapter 9 under the heading of "Network Processing". An understanding of the nature of a single baseline solution is nevertheless vital as many of the concepts, for example "ambiguity resolution", are central even to network solutions.

§7.3 described some of the data pre-processing procedures. It will now be assumed that all the data is "clean", and free of cycle slips. The next step is the phase data processing in single baseline mode. The main steps of a baseline solution using phase data are:

(1) **SETUP**
- Preparation: apriori coordinates, ephemeris file to be used, baseline to be processed

(if more than one observed in a session), data file names, antenna height and offsets, etc.

- Selection of parameters: this is dependent on the baseline to be processed, the ambiguity model used in the software, differencing scheme adopted, etc.
- Selection of options: apriori standard deviation of parameters and observations, criteria for data rejection, data reduction flags, whether correlations are to be considered, elevation cutoff, satellites to be excluded from solution, differencing strategy to be used, etc.

(2) **PROCESSING**

There are three types of phase solutions:

- Triple-Difference Solution: provides good apriori baseline solution even in the presence of cycle slips in the phase data.
- Double-Difference Solution (Ambiguity-free): where the ambiguity parameters are "free" to take whatever values the Least Squares solution gives.
- Double-Difference Solution (Ambiguity-fixed): where the ambiguity parameter values are fixed to integer values (and in the process converting ambiguous phase data to range data).

(3) **OUTPUT**

- Coordinate parameters as: Cartesian (x,y,z) or geodetic (ϕ,λ,h) values in the WGS84 datum, baseline components, for ground mark and/or antenna centres.
- Estimated standard deviation and correlation matrix (or VCV matrix) of parameters.
- Indicators of quality of the solution.

There are two types of double-difference phase data solutions:

- ❑ AMBIGUITY-FREE SOLUTION *(also known as a bias or ambiguity "float" solution)*, in which the parameters to be estimated are the **non-reference coordinates** (one end of a baseline) and the **ambiguity parameters**.
- ❑ AMBIGUITY-FIXED SOLUTION *(or "bias-fixed" solution)*, in which some or all the ambiguity parameters have been **"resolved"** to their integer values.

The AMBIGUITY-FREE solution is generally the first step. If after this solution the values of the ambiguity parameters *(which theoretically should be integers)* may be close to integer values *(but not exactly due to the presence of residual biases such as atmospheric refraction, orbit error, multipath, etc.).*

If the correct ambiguity values are identified, they can be held fixed in the subsequent AMBIGUITY-FIXED solution. Such a solution is very strong as it only contains the station coordinate parameters *(hence it is equivalent to processing unambiguous and precise ranges).*

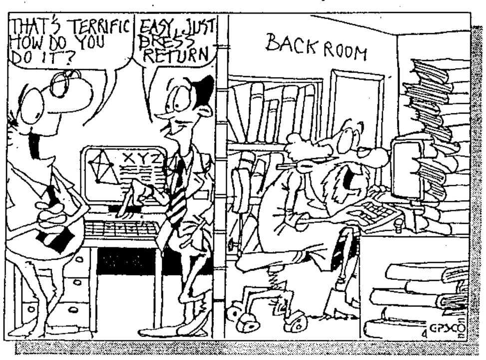
Note: it may not always be possible to attempt the ambiguity-fixed solution!

8.1.1 SETTING UP THE SOLUTION

With regards to initialising a triple- or double-differenced data solution, the following usually can be influenced by the analyst:

- Defining the apriori groundmark coordinates, including that for the "datum" station to be held fixed (for example, from a pseudo-range point position solution, or a triple-difference phase solution, or a previous solution when "chaining" baselines). This involves correctly setting up the station file (usually from the recorded field data), with information on antenna height, etc.
- Identifying the ephemeris file to be used (may be a Navigation Message file, or the Precise Ephemerides).
- Any satellites to be excluded from solution (for example, because of known health problems).
- Identifying the baseline to be processed, by selecting the data files to be used (generally from a database of GPS files).
- Inputting the standard deviation of the differenced observations.
- If option is available for taking correlations into account, this may be exercised.
- Minimum elevation cutoff angle for data culling to low satellites.
- Data selection for solution (all data or some sample rate, for example every 5th data epoch).
- Tropospheric refraction model for bias may be activated, based on input met data or "standard atmosphere" values.
- Dual-frequency processing options to be exercised.
- Whether to attempt ambiguity resolution or not (in the case of double-differenced solution).

Generally, standard data processing in an operational environment is largely automatic, offering the analyst little choice. Such "blind" or "batch" processing is usually controlled by default processing options. These can be changed from time to time by experienced operators to activate appropriate processing options when the survey conditions demand it (for example in the case of very short baselines, or dual-frequency observations, or short session lengths, sessions with bad satellite geometry, etc.).



There is, however, a limit to the options offered by commercial software packages. This can lead to some frustration in a small number of cases, when the processing strikes problems and "trouble-shooting" through the exercise of radically different options is not possible. The best option in such cases may be to reobserve the baseline! (Research or scientific software used in academic or government organisations offer many more options.)

As the best solution results are normally provided by a double-difference solution, the "hard-wired" processing options for double-differenced data solutions are the ones of greatest interest. The following are likely to be internally defined by the software:

- Differencing strategy (between-satellites) to be used.
- Ambiguity parameter model, usually single- or double-differenced model.
- Criteria for judging success of ambiguity resolution on a parameter-by-parameter basis.
- Solution convergence criteria.
- Internal modelling of the satellite orbit.
- Ordering of the satellites (influences the between-satellite differencing process).

8.1.2 PROCESSING OF DIFFERENCED DATA

There are a number of comments that can be made with regards to the various differenced phase data solutions:

❑ Triple- and double-differenced data solutions have different strengths and weaknesses, and are therefore offered as a "package" of processing options by the GPS instrument manufacturer. All three processing strategies are generally used (although it may not always be obvious from a study of the output results file).

❑ Triple-, double- (ambiguity-free) and double-differenced (ambiguity-fixed) solutions represent a hierarchy of processing strategies that are generally applied in sequence: first the triple-differenced solution, through to the double-differenced solutions, with the ultimate aim to obtain an ambiguity-fixed solution (Figure 8.1-1).

❑ The sequence of processing strategies: triple-difference through to the double-difference ambiguity-fixed solution, is from a robust, but relatively imprecise, algorithm to one that is capable of high precision and reliability only if the following (incomplete) list of conditions are met:
- the data is free of cycle slips,
- the observation session length is long enough (generally 30 to 60 minutes),
- Favourable receiver-satellite geometry (tracking as many satellites as are visible, with low PDOP),
- the baseline length is relatively short (<20km) for single frequency observations.

❑ The "optimum" baseline result is dependent on the quality and reliability of the double-difference solution. If an ambiguity-fixed solution is not possible (that is, it is not reliable), the double-difference ambiguity-free solution may be accepted as the "best". Under certain conditions, the triple-difference solution is the one preferred over the relatively weak double-difference solutions.

❑ GPS manufacturers provide guidelines for the selection of the "optimum" solution, on the basis of the length of the baseline. For example, the TRIMVEC™ manual recommends that the "optimum" solution for baselines: (a) shorter than 15km is the double-difference ambiguity-fixed solution (assuming ambiguity resolution was possible), (b) for baseline lengths greater than 15km and less than 50km it is the double-difference ambiguity-free solution, and (c) for baselines greater than 50km it is the triple-difference solution. *(Note: recommendations vary from manufacturer to manufacturer.)*

- Baseline reductions are invariably performed using the broadcast ephemerides. The main reasons are that it is instantly available (via the Navigation Message), and it is of sufficient accuracy. Investigations have shown that the accuracy of the broadcast ephemerides is of the order of ten metres, often much higher, hence contributing less than 1-2ppm to the baseline error. *(Note: precise IGS ephemerides are now available with very little delay over the Internet.)*

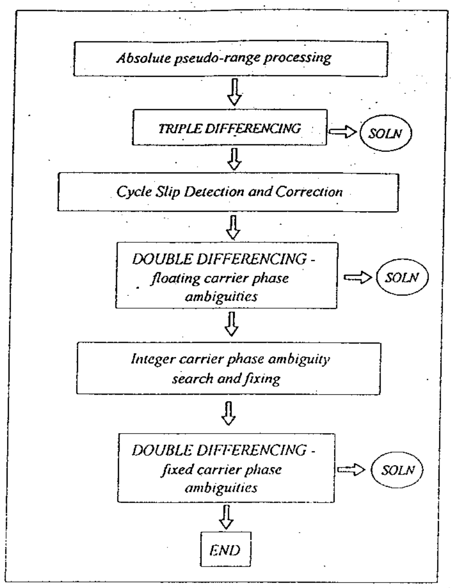


Figure 8.1-1. GPS phase data processing sequence.

Triple-Differenced Phase Solution

The following are some characteristics of triple-difference solutions:

- The functional model for the solution is eqn (7.2-5), containing only coordinate parameters (the ambiguity and clock phase errors were eliminated during differencing).
- Triple-difference solutions are "robust", being relatively immune to the effect of cycle slips in the data, which have the characteristics of "spikes" in the data (see Table 6.3-2 and Figure 7.3-1).
- This low susceptibility to data that is not free from cycle slips is due to the correlations in the differenced data not being taken into account (assume a diagonal Weight Matrix **P**).
- The algorithm used to construct the triple-differenced observables is ideally suited for detecting and repairing cycle slips in the double-differenced data. Hence these solutions are generally carried out as part of the overall data cleaning (pre-)process (see §7.3). An automatic procedure would be based on scanning the residuals of the triple-difference solution for those close to an integer value of one or more cycles.
- Relatively simple algorithm that can easily handle a changing satellite constellation.

- The triple-difference solution provides good apriori values for the baseline components.
- Under extreme circumstances the triple-difference solution may be the only one that is reliable.

The triple-difference solution algorithm:

☞ Difference epoch data between-satellites, *form double-differences.*

☞ Difference double-differences between epochs at some sample rate (for example, every 5th observation epoch), *form triple-differences.*

☞ Assume all triple-difference observations are independent when forming Weight Matrix (no correlations taken into account), *define* $\mathbf{P}$ *matrix.*

☞ Form Observation Equations, *construct the* $\mathbf{A}$ *matrix.*

☞ Accumulate Normal Equations, *scaled by the Weight Matrix* $\mathbf{A^TPA}$.

☞ At end of data set, invert Normal Matrix and obtain geodetic parameter solution, $\delta\hat{\mathbf{x}} = (\mathbf{A^TPA})^{-1}.\mathbf{A^TP}\,\mathring{\mathbf{v}}$.

☞ Update parameters.

☞ Optionally scan triple-difference residuals for cycle slips in double-difference observables.

Double-Differenced Phase Solution (Ambiguity-Free)

The following are some characteristics of double-differenced phase (ambiguity-free) solutions:

- The functional model for the solution is eqn (7.2-4), containing both coordinate parameters and ambiguity parameters (the exact form depending upon the ambiguity parameter model used -- see §7.2).
- Double-difference solutions are vulnerable to cycle slips in the (double-differenced) data.
- The solution can be quite sensitive to a number of internal software factors such as:
  - between-satellite differencing strategy (see below),
  - data rejection criteria,
  - whether correlations are taken into account during differencing (see below),
  - whether the observation time-tags are in the GPS Time system (§6.3 and §7.3).
- The solution is also sensitive to such external factors as:
  - length of observation session,
  - receiver-satellite geometry (including the number of simultaneously tracked satellites),
  - residual biases in the double-differenced data due to such things as atmospheric disturbances, multipath, etc.,
  - the length of the baseline.
- Only the independent epoch double-differences are constructed: $(S-1)$ double-differences per baseline per epoch, where S is the number of satellites tracked.
- The algorithm used to construct the independent double-differenced observables must

take into account the situations such as the rising and setting of a satellite during an observation session (and the appropriate definition of the ambiguity parameters in such a case).

The double-difference solution algorithm:

- ☞ Difference epoch data between-satellites, *form double-differences.*
- ☞ Apply data reductions, *such as a troposphere bias model.*
- ☞ Construct Weight Matrix (depending on whether correlations are to be taken into account), *define the* **P** *matrix.*
- ☞ Form Observation Equations --> *construct the* **A** *matrix.*
- ☞ Accumulate Normal Equations, *scaled by the Weight Matrix* $\mathbf{A^TPA}$.
- ☞ At end of session, invert Normal Matrix and obtain geodetic and ambiguity parameter solution, $\delta\hat{x} = (\mathbf{A^TPA})^{-1}.\mathbf{A^TP}\,\mathring{v}$.
- ☞ Update parameters.
- ☞ Decide (a) iterate solution, or (b) iterate solution only after ambiguity resolution attempted.

Double-Differenced Phase Solution (Ambiguity-Fixed)

The following are some characteristics of double-differenced phase (ambiguity-fixed) solutions:

- The functional model for the solution is eqn (7.2-4), containing coordinate parameters and any unresolved ambiguity parameters. As ambiguities are resolved the (integer) value of the ambiguity becomes part of the apriori known information, and this has the effect of converting *ambiguous* phase observations into *unambiguous* range observations.
- Such a double-difference solution is comparatively strong (there are less parameters to estimate!), but is reliable only if the correct integer values of the ambiguities have been identified.
- The solution can be quite sensitive to the strategy used to resolve the ambiguities, for example:
  - whether all ambiguities are to be resolved *as a set*, or only a subset,
  - the resolution criteria used for decision making,
  - the search strategy used for integer values.
- The ambiguity resolution process is also sensitive to such external factors as:
  - length of observation session,
  - receiver-satellite geometry,
  - residual biases in the double-differences due to such things as atmospheric disturbances, multipath, etc.,
  - whether satellites rise or set during the session,
  - the length of the baseline.

The ambiguity-fixed solution algorithm:

- ☞ Difference epoch data between-satellites, *form double-differences as before but without ambiguities as solve-for parameters.*
- ☞ Apply data reductions, *such as a troposphere bias model.*
- ☞ Construct Weight Matrix (depending on whether correlations are to be taken into account), *define the* **P** *matrix.*
- ☞ Form Observation Equations, *construct the* **A** *matrix.*
- ☞ Accumulate Normal Equations, *scaled by the Weight Matrix* $\mathbf{A^TPA}$.
- ☞ At end of session, invert Normal Matrix and obtain geodetic parameter solution, $\delta\hat{x} = (\mathbf{A^TPA})^{-1}.\mathbf{A^TP}\,\mathring{v}$.
- ☞ Update parameters.
- ☞ This process can be iterated to resolve other ambiguities until (a) all have been resolved (and "fixed" to integers), or (b) no more can be reliably resolved.

Once ambiguities have been resolved, the ambiguous phase measurements are converted to precise range observations. As in conventional GPS navigation, single epoch positioning is now possible and hence "carrier-range" observations are ideal for kinematic positioning applications.

"Carrier-Range" Positioning

To appreciate the *power* of the "carrier-range" positioning solution it is necessary to rewrite eqn (7.2-4) to change the balance between known and unknown quantities in the double-differenced observation equation:

$$\begin{aligned}\nabla\Delta\Phi(t) - \rho_1{}^1(T_1) + \rho_1{}^2(T_1) - \lambda.(\,n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2\,) \\ = -\rho_2{}^1(T_2) + \rho_2{}^2(T_2) \qquad (8.1\text{-}1)\end{aligned}$$

The quantities on the first line are known. The only unknown quantities are the coordinates of receiver 2. If the measurement noise is at the few millimetre level (no multipath and residual atmospheric biases are assumed to be present in the double-differences), then it would be possible to determine the three coordinate components of receiver 2 to centimetre level accuracy with just three independent double-differences (from the simultaneous tracking of four GPS satellites)!

Unambiguous carrier phase (or "carrier-range") positioning has all the advantages of pseudo-range positioning, such as instantaneous single epoch results, but with unprecedented precision. Figure 8.1-2 illustrates a series of repeated single epoch results (the "carrier-range" data was processed in the "kinematic" mode) for the length of a static baseline. Note that the baseline length variability is of the order of a centimetre (similarly for the other components). The following comments can be made:

- The signature in Figure 8.1-2 is largely due to the impact of multipath on the phase observations -- *the multipath effect on pseudo-range would be up to several orders of magnitude greater.*
- A change in the constellation of satellites from one epoch to the next will cause a "jump" in the solution.
- As the baseline was static, the redundant measurements are not contributing to the solution (solutions are carried out on a single epoch basis) -- *redundant observations increase precision according to the "averaging law" ($\sqrt{"n"}$).*
- The precision is influenced by the instantaneous satellite geometry as represented by navigation DOPs such as PDOP -- *this varies smoothly over time except when satellites set below the tracking horizon or new satellites rise above it.*
- For each epoch the repeated baseline estimates are derived from double-differenced precise "carrier-range" observations (eqn (8.1-1) as long as the ambiguities used in the previous epoch remain valid -- *hence GPS hardware must continue to track the satellites (new satellites require new ambiguities to be estimated) and cycle slips must be avoided.*

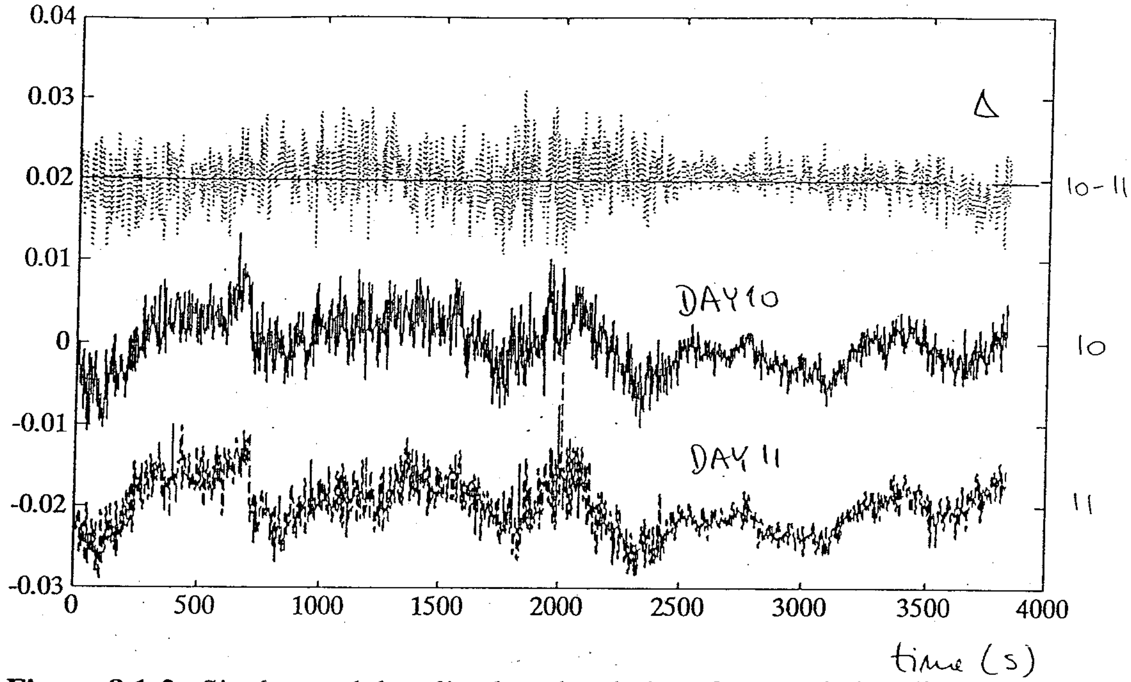


Figure 8.1-2. Single epoch baseline length solutions for a static baseline using carrier phase data with resolved ambiguities.

Some of the issues that must be addressed in order to make "carrier-range" positioning reliable are:

❑ How to prevent cycle slips?

☞ *Requires new ambiguity resolution to be carried out, but how to do this quickly, reliably and without too great a nuisance to the surveyor?*

❑ How to improve the precision of the solution and increase its "robustness"?

☞ *Extra satellites, Kalman filter algorithms, dual-frequency observations, precise pseudo-range data, multipath resistant antennas and receiver electronics, etc.*

❑ Is *real-time* "carrier-range" positioning possible, or desirable?

☞ *It is possible for pseudo-range positioning. Would help indicate to field staff if something goes wrong (cycle slip, etc.).*

Ambiguity Resolution

What is ambiguity resolution?

☞ ***The mathematical process of converting ambiguous ranges (integrated carrier phase) to unambiguous ranges of millimetre measurement precision ...***

☞ ***For conventional GPS surveying, corresponds to converting real-valued ambiguity parameter values to the likeliest integer values ...***

As has been emphasised several times, if the carrier beat phase observations were not "accumulated" or "integrated", by measuring the change-in-phase, or range, from the epoch of initial signal lock-on, the value on n would change with each epoch. This would make carrier phase observations all but useless. Hence determining the value of this unknown initial (integer) ambiguity is an important task of GPS data reduction software.

In an ambiguity-free solution, no advantage is taken of the integer nature of n as it is indistinguishable from other, non-integer, biases such as orbit uncertainties, multipath and atmospheric refraction (ionosphere and troposphere), and is in fact "contaminated" by them. Thus, "ambiguity resolution" as it is generally known, is only possible after all biases are eliminated or otherwise accounted for to better than one cycle (≈20cm wavelength on L1). Constraint of an ambiguity value to its correct integer will improve the estimation of the remaining geodetic (station coordinate) parameters, as an inspection of Figure 8.1-3 indicates.

Figure 8.1-3 illustrates what happens in a sequential transition from an 100% ambiguity-free solution to an 100% ambiguity-fixed solution. For the first ten epochs the solution is an ambiguity-free one, but after 10-15 epochs, when the ambiguities have been resolved, the precision of the remaining estimable coordinate parameters improves significantly. It should be noted that the precisions (as well as the numeric values of the parameters) have virtually converged to their final values immediately after all ambiguities have been resolved. As a corollary therefore, **phase data collected beyond the minimum necessary to ensure an ambiguity-fixed solution is obtained has almost no influence on the final results**.

Considerable R&D effort has been invested in so-called "rapid ambiguity resolution" algorithms, which are the basis of the "rapid static" GPS surveying techniques (§5.5 and §8.3).

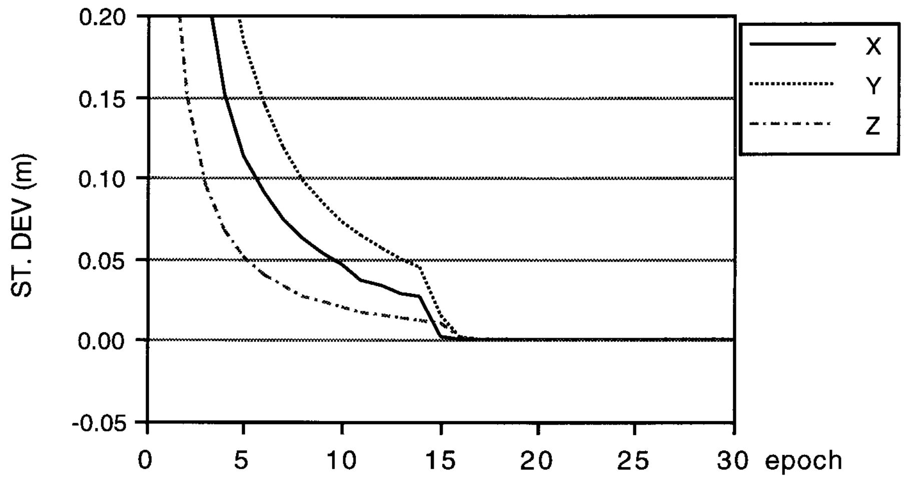


Figure 8.1-3. The change in quality of baseline components in an ambiguity-free compared to an ambiguity-fixed solution.

Clearly, *an ambiguity-fixed solution is very desirable, and all efforts should be made to obtain one.* There are several steps involved in obtaining an ambiguity-fixed solution:

(1) Define the apriori values of the ambiguity parameters.

(2) Use a search algorithm to identify likely integer values.

(3) Employ a decision-making algorithm to select the "best" set of integer values.

(4) Apply ambiguities to the new (ambiguity-fixed) solution.

Ambiguity resolution is discussed in detail in §8.2 and §8.3. Although ambiguity resolution can be considered largely an *optional* solution strategy for conventional (static) GPS surveying, it is a *vital* operation for some of the modern high productivity GPS surveying techniques.

Maximising the Chances of Successful Ambiguity Resolution

There are some well-known strategies that can be employed for maximising the chances of resolving ambiguities (though it cannot be guaranteed) for conventional static GPS surveying, including:

- Single frequency observations for short baselines (< 20km).
- Dual-frequency observations for longer baselines.
- Adequate length observation session (1/2 hour —> 2 hours).
- Minimise interference (for example, no multipath, low unmodelled ionosphere, etc.), by good selection of sites, observing at night, etc.
- Observe as many satellites as possible to ensure good receiver-satellite geometry.
- Use precise pseudo-range data if available (and the ambiguity resolution algorithm can make use of this data -- see §8.2, §8.3 and §8.4).

Factors making ambiguity resolution difficult include (§8.2):

- The degree to which the geodetic parameters are reliably separated from the ambiguity parameters.
- The magnitude of any unmodelled biases present in the double-differenced phase data.
- The length of the baseline.
- The quality of the receiver-satellite geometry, and how much it has changed during the observation session.
- The data quality.
- Sub-optimal algorithms.

RELIABLE ambiguity resolution is essential ...

Incorrect resolution of some ambiguities will lead to a poor solution (worse than ambiguity-free or triple-difference solution). *Hint: has the solution changed by more than 10cm?*

How does one know when Ambiguity Resolution is possible?

IT IS NOT POSSIBLE!! ***but with 1hour sessions to > 4 satellites & baselines < 20km & rapidly changing PDOP it should be possible to resolve ambiguities.***

Forming Differences and Correlated Observables

Most software packages do not permit the between-satellite differencing strategy to be defined by the user. In general, one of two strategies is "hard-wired" (§7.2, Figure 7.2-1):

(1) Base satellite option.

(2) Sequential satellite option.

At first glance there would appear to be no major difference in these approaches, and the one implemented in a package has generally been selected on the basis of personal preference. For example, the TRIMVEC™ program uses the base satellite approach, while others programs such as PoPS™ and SKI™ employ the sequential satellite differencing approach. Both have to contend with the "book-keeping" problem of satellites that may rise or set during an observation session. (This is more critical in the base satellite option if the base satellite itself sets before the end of the session.)

However, one remark concerning the influence of differencing strategy on the coordinate results should be made. In the case of the base satellite option, the choice of "base satellite" is critical. A poor choice (for example, a satellite that is low on the horizon, or one with an unstable onboard oscillator, or a satellite whose track relative to the receiver induces multipath reflections) will "corrupt" all double-differences formed. This same influence would not be as widespread in the case of the sequential satellite differencing approach, as it will "corrupt" only those double-differences that include that satellite.

To induce a change in the differencing strategy that is employed (or at least the order in which the satellite differences are generated), one option may be to reverse the order of the stations. The first solution would identify station "A" as receiver 1 and station "B" as receiver 2. The order of the satellites recorded at station "A" may be different to those at station "B", and the order of satellites for receiver 1 generally define the differencing pattern. A second solution, in which station "B" is nominated as receiver 1 and station "A" as receiver 2, may lead to a better (or worse) baseline solution than previously. In extreme cases, it may be possible to even successfully resolve ambiguities in one case, but not if the station order is reversed!

Another test of solution "sensitivity" is to attempt double-difference solutions first taking the correlations into account (if it is an option), and then repeating the solution with the correlations "turned off". By replacing the diagonal VCV matrix of the double-difference observations with the correct (non-diagonal) form (eqn (7.2-18)), two effects may be mentioned:

☞ a change in the baseline solution --> *of the order of a few parts per million*

☞ an apparent worsening of the solution standard deviations --> *more realistic values?*

For baseline solutions, not taking into account the observation correlations does not have much effect. In the case of network solutions for scientific applications (for example, high accuracy over long baselines), observation correlations must be considered as they can be employed to advantage in multi-station ambiguity resolution.

8.1.3 SOLUTION OUTPUT

The first remark that must be made is that there is, as yet, **no standard form for the baseline solution output**. Different software packages produce differently formatted output files. Not only are the coordinate results likely to be variously labelled, but in addition, different terminology is often used (this is partly cultural as well, as most software is developed in the U.S.). For example, it may not be clear if the "baseline components" are antenna-to-antenna or groundmark-to-groundmark, "ambiguities" may be labelled "biases", and the source of the apriori coordinates may not be given. Furthermore, the types of statistical tests applied and the exact (mathematical) definition of some of the "quality" indicators may not be given. (They may be found in the software manual, but these generally suffer the problem of all manuals, that is they may be user-unfriendly and sometimes the source of misinformation, or at least out-of-date advice.)

There is a draft SINEX (Solution INdependent EXchange) format presently being tested (§9.2).

With regards to the output of a baseline solution, the following information is usually provided in some form or another:

- Type of solution: whether triple-, ambiguity-free or ambiguity-fixed.
- Input and output coordinates (solve-for and fixed), in various systems, for example Cartesian, geodetic, baseline components.
- Echo of receiver (serial numbers, etc.) and station information (site I.D., antenna height, etc.).

- Standard deviation of estimated coordinate components.
- The correlation matrix or variance-covariance (VCV) matrix for all coordinate parameters (and perhaps for the ambiguities as well).
- Optional estimate of quality of satellite geometry, for example the RDOP value (derived from the VCV matrix).
- Tracking data acquired, logging times at individual sites, tracking channels used, satellites tracked, signal quality flags, etc.
- Number of observations used in solution, as well as those rejected, the sampling rate used, and the data edit criteria (usually some factor times "RMS tracking").
- Summary of ephemeris information, health warning flags in the Navigation Message, etc.
- Any data pre-processing performed (for example, tropospheric delay model).
- Indicator of fit of observations to final solution (that is, the residuals),usually in the form of an "RMS tracking" value (though the label may vary for different processing software). *The magnitude is an indicator of the quality and reliability of solution, and whether ambiguity resolution is possible (or resolution was successful).*
- Results of statistical tests of the residuals may also be displayed.
- If an ambiguity-fixed solution was attempted, then a summary of the number of ambiguity parameters resolved.
- Possibly a summary recommendation on the quality of the solution.

Solution output may be scanned to assess the quality of the results, but in order to be able to compare one baseline solution to another, some "rule-of-thumb" remarks may be kept in mind:

- ❑ In conventional static GPS surveying successful ambiguity resolution is basically a function of baseline length. For baselines over 15km in length it can be a problem. If ambiguity resolution was not possible for baselines < 15km, check the continuity of tracking, length of observation session, data "noise", and take steps such as reduce the resolution decision making criteria, and perhaps rerun the solution.

- ❑ Having sessions of the same length (each day and throughout the day) is not sufficient to ensure similar quality results. Nor are sessions observed at the same time each day a guarantee of consistent quality results, even though the receiver-satellite geometry are the same. Residual biases due to orbit error, atmospheric effects, etc. are <u>dynamic in nature</u> and will influence the results as much as (and often more than) the receiver-satellite geometry during a session.

- ❑ The "RMS of tracking" quantity tends to increase with increasing baseline length. (This quantity may have a different label in some software.)

- ❑ The accuracy of the baseline components, expressed in metres, will degrade with increasing baseline length. *Expressed as "parts per million" it should be a constant.*

- ❑ A "total" quality indicator, generally on the basis of a number of internal "tests" such as whether the ambiguities were resolved, the "RMS tracking", etc. may be provided, but is invariably software dependent -- *therefore READ THE MANUAL.*

- ❑ The coordinate standard deviations are lower for an ambiguity-fixed solution than for an ambiguity-free solution. (Though that does not mean that the ambiguity-fixed solution is correct!) For example, the following results were obtained for a baseline of 6.8km length:

	σ_x	σ_y	σ_z
A-free soln.	.398E-02	.108E-01	.327E-02
A-fix soln.	.352E-02	.175E-02	.182E-02

- The solution stochastic information can be presented in one of two forms, for example, in the case of the above 6.8km baseline (ambiguity-fixed solution):
  - as standard deviations and correlations:

σ_x = .352E-02	σ_y = .175E-02	σ_z = .182E-02
χ_{xy} = -.52198	χ_{xz} = .65125	χ_{yz} = -.72810

  - or as the full variance-covariance matrix:

.12461E-04		
-.32245E-05	.30625E-05	
.41840E-05	-.23190E-05	.33124E-05

- Solution statistics based on the VCV information are often optimistic. *Beware of software packages that may artificially inflate the uncertainties of the parameters in order for them to appear more "realistic".*

- There is no measure for "reliability", the output VCV information will not reflect the influence of systematic biases.

- It is good practice to first carry out preliminary baseline reductions with a minimum of options. This would permit the output to be assessed for bad data, residual cycle slips, etc., and the final adjustment may be carried out using the best dataset.

- If there is any doubt about the quality of the ambiguity-fixed solution, it is preferable to accept the ambiguity-free solution in its place. If the ambiguity-free solution indicates high a "RMS tracking" value (for example because the baseline > 50km), the triple-difference solution may be the preferred solution. *However, check that the recommended standards & practices for a certain class of GPS survey will accept such a solution (see §10.2).*

- Improvement in the modelling of biases (for example, through the use of dual-frequency observations), or increase in the sophistication of the solution ("rigorous" adjustment, inclusion of additional parameters, etc.) leads to better and more reliable results for the same length baseline and observation session compared with the basic single baseline processing strategy.

- Additional statistical testing may be carried out (for example, using "chi squared" and "variance factor" tests). In addition, external tests may be based on loop closure statistics, or the results of a network adjustment. This is dealt with further in chapter 9.

How Good is the Solution?

There are a number of "Quality Indicators" that may be monitored, including:

- RMS of observation residuals.
- Number of rejected observations.
- Statistical tests on residuals or parameters.
- Aposteriori variance factor.
- VCV matrix of solution.
- The type of "optimal" solution obtained.
- "Trustworthiness" of solution.
- Measures of reliability of selected ambiguity parameter set.

The following comments may be made with respect to the "RMS of residuals" and "rejected observations":

- A "low" RMS value and a "low" number of rejected observations often indicates that both the data and solution quality are OK.
- Manufacturers often give recommended maximum values of RMS. *Generally a function of baseline length, observation type (L1, L2, L3), etc.*
- In general, an RMS value below 0.1 cycles is considered acceptable.
- Data editing is often carried out during solution iterations. *Generally based on some factor (say 3) x RMS.*
- Possible reasons for high RMS and data rejection rates are the presence of multipath and uncorrected cycle slips. *Residual plots are good tools to verify this.*
- Some phase data processing software permits the residuals to be plotted. *Residuals should be examined.*

The following comments may be made with respect to the "statistical tests" and "VCV information":

- In general, little statistical testing is carried out on parameters or residuals.
- If the aposteriori variance factor is unity then it is likely that the VCV matrix has been *adaptively* scaled to ensure this happens.
- In general however, the output VCV matrix is too *optimistic* , suggesting higher precisions for the parameters than is warranted. *Does not take into account unmodelled systematic biases (atmospheric refraction, satellite orbit and fixed station errors, etc.).*
- The standard deviations of baseline components vary considerably as a function of the *type* of phase solution (triple-difference, double-difference ambiguity-free, double-difference ambiguity-fixed).

There are other several Quality Indicators related to "solution characteristics", including:

- What is the "optimal" solution? *Was an ambiguity-fixed solution obtained? Was it expected?*
- If an ambiguity-fixed solution was obtained, check the absolute and relative RMS of the residuals. *Are resolved ambiguities reliable?*

- If an ambiguity-fixed solution was obtained, check baseline components. *Did baseline solution change by more than 10cm compared to the ambiguity-free solution?*
- The formal accuracy estimates for the vertical component is usually twice the size of the horizontal components.
- Verify solution characteristics such as:
  - satellites used --> *any health problems?*
  - sample data rate
  - common tracking period --> *as planned?*
  - apriori station coordinates --> *were the correct WGS84 values used?*
  - antenna heights
  - elevation mask angle
  - troposphere reduction applied?
  - satellite geometry indicator --> *PDOP, RDOP, etc.*

Advantage can also be taken of "external evidence", including:

- Hierarchy of solutions --> *Comparison of triple-difference and double-difference solutions.*
  - "decimetre" triple-difference solution precisions
  - "centimetre" double-difference ambiguity-free solution precisions
  - "millimetre" double-difference ambiguity-fixed solution precisions
- Dual-frequency solutions -->*Comparison of L1, L2 and L3 solutions.*
- Single baseline vs multi-baseline solutions **-->***Different analysts? Process using different software packages?*
- Repeat baseline solutions from different sessions.
- Solutions involving tracking to 4 or more satellites, over a period of 30-60 minutes, for baselines less than about 15km, should be high quality ambiguity-fixed solutions.
- Compare with ground control **-->***Usually just distance.*
- Check BBS or Integrity Monitoring Service.

How does one really tell the quality of individual solutions?

Combine the baselines into a network adjustment !

8.2
INTRODUCTION TO AMBIGUITY RESOLUTION

What is ambiguity resolution?

☞ *A means of improving the accuracy of GPS surveying ...*

☞ *The mathematical process of converting ambiguous ranges (integrated carrier phase) to unambiguous ranges of millimetre measurement precision ...*

☞ *For conventional GPS surveying, corresponds to converting real-valued ambiguity parameter values to the likeliest integer values ...*

☞ *For modern GPS surveying, corresponds to discriminating the likeliest set of integer values from many alternative sets ...*

Determining the value of the unknown initial (integer) ambiguity for a GPS double-differenced phase solution is an important task of GPS phase reduction software. **Ambiguity resolution is probably the most uniquely identifiable characteristic of high precision GPS positioning.**

Two scenarios for ambiguity resolution can be distinguished:

(1) In the case of *conventional static GPS surveying*, as developed since the early 1980's, the lengths of the observation sessions are sufficient to ensure a reliable ambiguity-free solution, and the successful resolution of the ambiguities to their integer values is a useful "bonus".

(2) For *modern GPS surveying techniques* (see §5.5), ambiguity resolution is a critical operation, and if not successfully carried out, the integrity and reliability of the baseline solution can be degraded significantly.

Considerable R&D has therefore been invested in developing ambiguity resolution techniques for modern GPS surveying. Before considering the strategies developed specifically for "rapid static" and "kinematic" techniques (including some dual-frequency procedures -- §8.4), it is useful to study the steps involved in ambiguity resolution for conventional -- long observation session -- GPS surveying.

8.2.1 GENERAL OVERVIEW OF AMBIGUITY RESOLUTION

Several steps in the ambiguity resolution process can be identified:

(1) Define the apriori values of the ambiguity parameters.

(2) Use a search algorithm to identify likely integer values.

(3) Employ a decision-making algorithm to select the "best" set of integer values.

(4) Apply ambiguities to the new (ambiguity-fixed) solution.

Each of these are discussed in the following sections.

Apriori Values of the Ambiguity Parameters

The apriori values of the ambiguities are generally provided by a double-difference solution in the form of *real-valued quantities plus the variance-covariance information.* The likeliest values of the ambiguities therefore are the nearest, "round-off" integer values. In some cases the estimate may be very near an integer, but in other cases the real-valued estimate is not *obviously* near an integer. The reliability of these estimates is a function of the baseline length, satellite-receiver geometry and length of observation session, and they are affected by multipath, residual biases and cycle slips.

There are several approaches which can be used, some of which are particularly useful for modern GPS surveying techniques. These include:

- Estimate ambiguities with the aid of pseudo-range data:

$$\nabla\Delta\Phi_{ij}^{kl}(t) - \nabla\Delta P_{ij}^{kl}(t) = \lambda . n_{ij}^{kl} \tag{8.2-1}$$

where $n_{ij}^{kl} = n_i^k - n_i^l - n_j^k + n_j^l$

- Use other geometric information, such as the known length of the baseline:

$$\nabla\Delta\Phi_{ij}^{kl}(t) - (\, \rho_i^k(T_i) - \rho_i^l(T_i) - \rho_j^k(T_j) + \rho_j^l(T_j)\,) = \lambda . n_{ij}^{kl} \tag{8.2-2}$$

- Make use of dual-frequency relationships that permit the L1 (or L2) ambiguities to be estimated from the wide-lane or ionosphere-free "lumped" ambiguity terms (§8.4).

Searching Integer Ambiguity Sets

All techniques rely on some "search" technique that tests a range of neighbouring values around the initial ambiguity values (Figure 8.2-1). For example, in the case of six tracked satellites there are five (double-differenced) ambiguities to be resolved. If the search window is three integers wide (one on either side of the round-off value), then there are 3^5 ambiguity sets to be tested (here = 243).

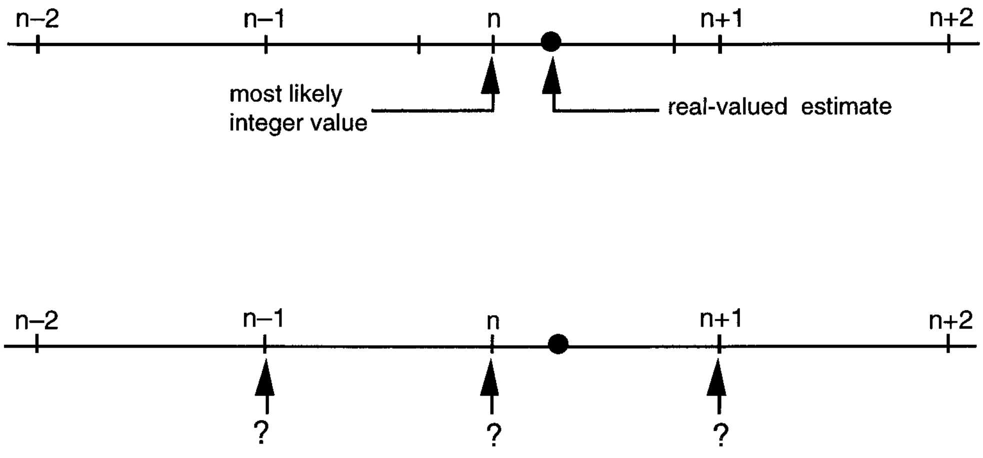


Figure 8.2-1. Real-valued ambiguity estimates and candidate integer values to be tested.

The following is a partial listing of the 243 candidate ambiguity sets in the above example, where n^{kl} is the double-differenced ambiguity for a baseline ij (subscript deleted) involving satellites k and l, $n(-)^{kl}$ is one integer value less, and $n(+)^{kl}$ is one integer greater than the round-off value (Figure 8.2-1). Each line lists five possible integer ambiguity values that could *reasonably* be expected to be the correct values that convert the ambiguous phase observations to unambiguous carrier-range measurement:

$n(-)^{12}$	$n(-)^{23}$	$n(-)^{34}$	$n(-)^{45}$	$n(-)^{56}$
$n(-)^{12}$	$n(-)^{23}$	$n(-)^{34}$	$n(-)^{45}$	n^{56}
$n(-)^{12}$	$n(-)^{23}$	$n(-)^{34}$	n^{45}	$n(-)^{56}$
$n(-)^{12}$	$n(-)^{23}$	n^{34}	$n(-)^{45}$	$n(-)^{56}$
$n(-)^{12}$	n^{23}	$n(-)^{34}$	$n(-)^{45}$	$n(-)^{56}$
n^{12}	$n(-)^{23}$	$n(-)^{34}$	$n(-)^{45}$	$n(-)^{56}$
$n(-)^{12}$	$n(-)^{23}$	$n(-)^{34}$	n^{45}	n^{56}
$n(-)^{12}$	$n(-)^{23}$	n^{34}	$n(-)^{45}$	n^{56}
$n(-)^{12}$	n^{23}	$n(-)^{34}$	$n(-)^{45}$	n^{56}
n^{12}	$n(-)^{23}$	$n(-)^{34}$	$n(-)^{45}$	n^{56}
$n(-)^{12}$	n^{23}	n^{34}	n^{45}	$n(-)^{56}$
$n(-)^{12}$	n^{23}	$n(-)^{34}$	n^{45}	$n(-)^{56}$
n^{12}	$n(-)^{23}$	$n(-)^{34}$	n^{45}	$n(-)^{56}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n^{12}	$n(+)^{23}$	$n(+)^{34}$	$n(+)^{45}$	$n(+)^{56}$
$n(+)^{12}$	n^{23}	$n(+)^{34}$	$n(+)^{45}$	$n(+)^{56}$
$n(+)^{12}$	$n(+)^{23}$	n^{34}	$n(+)^{45}$	$n(+)^{56}$
$n(+)^{12}$	$n(+)^{23}$	$n(+)^{34}$	n^{45}	$n(+)^{56}$
$n(+)^{12}$	$n(+)^{23}$	$n(+)^{34}$	$n(+)^{45}$	n^{56}
$n(+)^{12}$	$n(+)^{23}$	$n(+)^{34}$	$n(+)^{45}$	$n(+)^{56}$

Selecting the Best Set of Integer Ambiguities

The standard criteria for successful ambiguity resolution is if the identified ambiguity set (n^*) clearly fits the double-differenced phase data better than any other ambiguity set:

$$\nabla\Delta\Phi_{ij}^{kl}(t) - \nabla\Delta\rho^*{}_{ij}^{kl}(t) - \lambda.n^*{}_{ij}^{kl} = \upsilon_{ij}^{kl}(t) \tag{8.2-3}$$

The **testing criteria** is generally the lowest weighted root-sum-of-squares (RSS) of the double-differenced data residuals:

$$RSS = \sqrt{\mathbf{v}^T\mathbf{P}\mathbf{v}} \tag{8.2-4}$$

where $\mathbf{v}$ is the vector of residuals, $\mathbf{P}$ is the observation weight matrix.

This procedure requires:

- The computation of residuals for each ambiguity-fixed solution being tested -- *estimate a new baseline for each ambiguity set.*
- Assumes <u>unbiased</u> (for example, no residual atmospheric refraction), <u>clean</u> (that is, no cycle slips), <u>low noise</u> (for example, no multipath) data.
- Observation data series long enough for reliable residual testing.

Such a strategy seeks to find the "best" set of ambiguities, the one that is clearly better than the "second best" set by some **rejection criteria** (which can be varied). In the TRIMVEC™ program the "ratio" value is the *ratio of the RSS obtained using the second best ambiguity set to the RSS for the best set.* This value should be as large as possible. Ambiguity resolution is obviously least reliable when there is no obvious "best" set of ambiguities (that is, the "ratio" value is small, perhaps less than 2 or 3). This is generally implemented as an "all or nothing" process.

An alternative approach is to resolve only some ambiguities. Using the round-off strategy, only those ambiguities that are near integers, and which have low standard errors (Figure 8.2-2a), are assumed to have been reliably resolved. Another approach is a <u>sequential</u> ambiguity resolution search procedure, also based on minimising the estimated "weighted RSS". However, in this procedure the best determined (the one with the lowest standard error) ambiguity parameter is used as the test value about which the search window is defined. Once that ambiguity has been resolved, new ambiguity values are derived for the remaining estimable parameters, and the search continues using the best determined of the remaining ambiguities, and so on. This process is more flexible and can be halted when no further ambiguities can be resolved reliably.

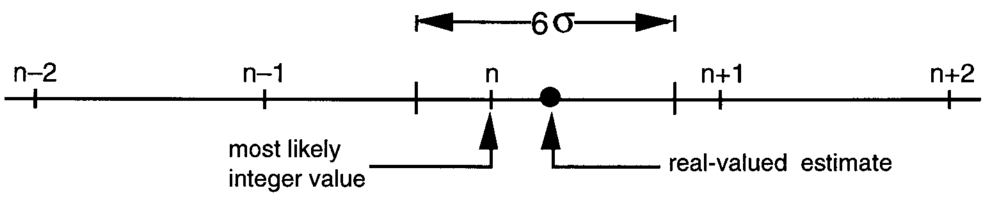


Figure 8.2-2a. Scenarios in ambiguity resolution: ambiguity well determined and close to an integer.

There are several factors that make ambiguity resolution difficult, including:

- The degree to which the geodetic parameters are reliably *separated* from the ambiguity parameters.
- The magnitude of any *unmodelled biases* still present in the double-differenced phase data.
- The length of the baseline.
- The quality of the receiver-satellite geometry, and how much it has changed during the observation session.
- The data quality.
- Sub-optimal algorithms.

When the baseline is comparatively long (>20 km), or there are some unmodelled biases present (high ionospheric activity, etc.), some ambiguities may be far from an integer value but still have small standard deviations. Or, alternatively, the ambiguity values may be close to integers but the standard deviations may be too large (arising from poor receiver-satellite geometry, data outages, etc.). These two situations are illustrated in Figures 8.2-2b and 8.2-2c.

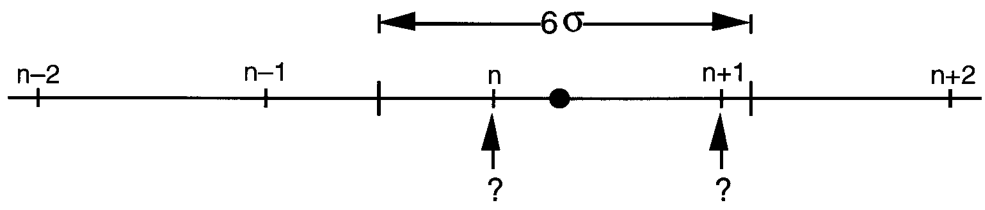


Figure 8.2-2b. Scenarios in ambiguity resolution:
Session not long enough? Poor geometry?

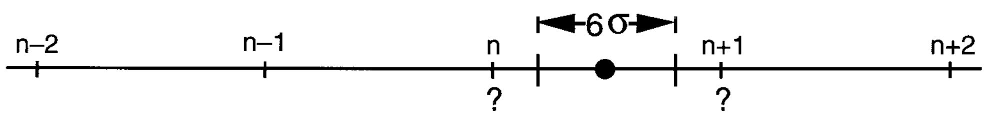


Figure 8.2-2c. Scenarios in ambiguity resolution:
Biased data? Baseline too long? Over-optimistic std. dev.?

The situations illustrated in Figures 8.2-2b and 8.2-2c mean that the rejection criteria cannot discriminate between more than one candidate ambiguity set, and hence ambiguity resolution cannot proceed reliably. Hence, because an ambiguity-fixed solution is not possible, the ambiguity-free solution may be considered to be the "optimal" one.

Another useful way of visualising this ambiguity search procedure is from a geometric viewpoint, as discussed below.

Ambiguity Resolution and Wavefront Geometry

Imagine the carrier phase wavefronts from satellites 1 and 2, as illustrated in Figure 8.2-3a in a 2-D representation. The grid has a mesh which is λ wide ($\approx$19cm wavelength on L1). (In reality these wavefronts can be considered to be the result of between-receiver differencing of data to each of the satellites in turn.)

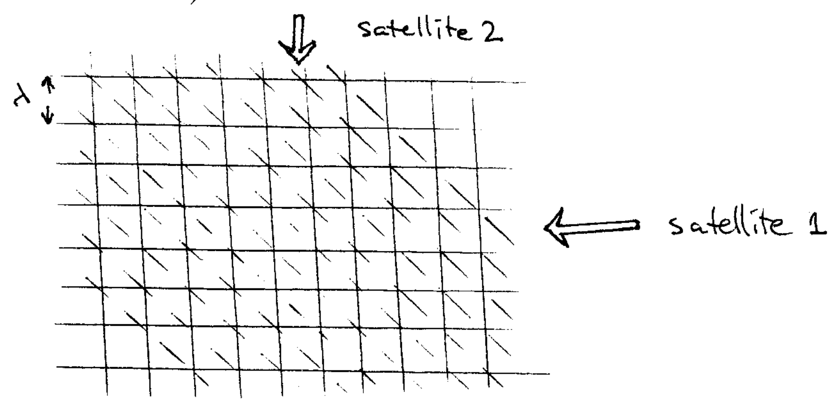


Figure 8.2-3a. Wavefront grid formed from two satellites.

The two sets of parallel lines (in 3-D they are surfaces) can be combined into lines of double-differenced ambiguities (each line is the intersection of two wavefronts, and represents a constant double-differenced integer ambiguity value), as illustrated in Figure 8.2-3b.

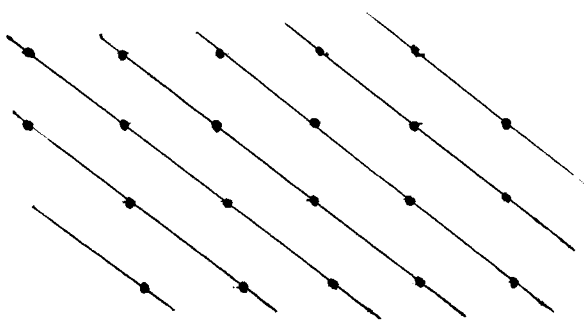


Figure 8.2-3b. Constant double-differenced "lines-of-ambiguities" involving two satellites.

In the case of $(m+1)$ satellites, the geometric lattice is formed by the intersection of m sets of "lines-of-ambiguities". In Figure 8.2-3c, pairs of candidate ambiguities n^{*12} and n^{*23} are located at the intersection of the resulting lattice formed from two sets of "lines-of-ambiguities".

However, there is no redundant information to allow for the unambiguous selection of the correct pair of ambiguities (corresponding to one intersection point of the "lines-of-ambiguities"). The observations from a fourth satellite would permit another set of parallel "lines-of-ambiguities" to be overlain on Figure 8.2-3c. **There may be one intersection that satisfies all geometric conditions, or more likely the case, several which are "close" intersections.** As data is accumulated and the satellite geometry changes (due to the motion of the satellites), each set of "lines-of-ambiguities" (involving a pair of satellites) rotates by a different amount. Hence the total lattice pattern changes in a manner similar to interference fringe lines, and the one correct ambiguity set may become steadily more obvious (it is the only intersection point about which all the grids rotate).

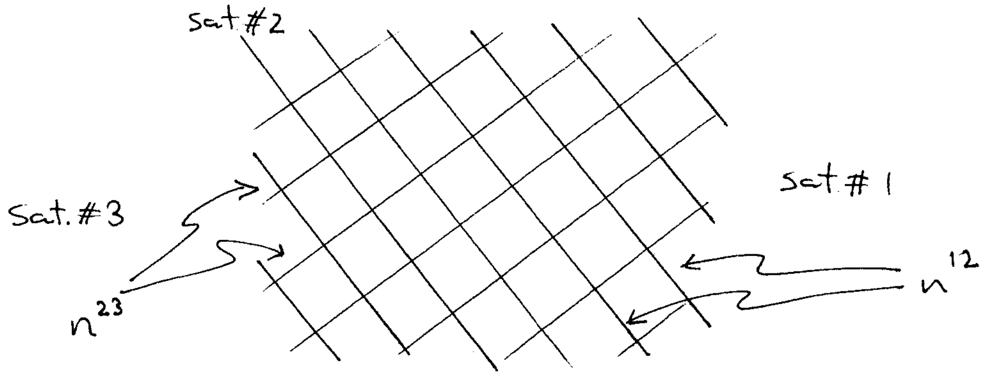


Figure 8.2-3c. Two sets of constant double-differenced "lines-of-ambiguities" involving three satellites.

Hence what is required is either:

☞ **a significant change in satellite-receiver geometry over an observation session** so that the intersection point representing the correct resolved integer ambiguity values becomes obvious. *This is generally the situation for conventional static GPS surveying with long observation sessions.*

☞ **good geometry at a single epoch, or over a very short time period** (a matter of minutes), when there is close to orthogonal intersection of the "lines-of-ambiguities" and sufficient redundancy so that there is only one candidate intersection point within the grid. *This is the requirement for modern GPS surveying techniques.*

Geometric issues in relation to ambiguity search procedures is discussed further in §8.3.

8.3

AMBIGUITY RESOLUTION & MODERN GPS SURVEYING

Introductory Remarks: Modern GPS Surveying Techniques

The basis of high precision GPS positioning is the double-differenced observable (§6.3 and §7.2). Expressing eqn (7.2-4) in a form that highlights the known quantities:

$$\nabla\Delta\Phi(t) - \rho_1{}^1(T_1) + \rho_1{}^2(T_1) = -\rho_2{}^1(T_2) + \rho_2{}^2(T_2) + \lambda.(\, n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2\,) \qquad (8.3\text{-}1)$$

The coordinates of receiver 1 are assumed known. The unknown parameters consist of the ambiguity parameters (which are constants for the entire observation session) and the coordinates of receiver 2 (embedded in the $\rho_2{}^i$ range quantities). The processing of these observation equations in a conventional Least Squares scheme permits the values of the unknown parameters (and their uncertainties) to be estimated. This, of course, is the basis of the double-difference **ambiguity-free solution**. The evolution of the uncertainties of the coordinate parameters is illustrated in Figure 8.3-1.

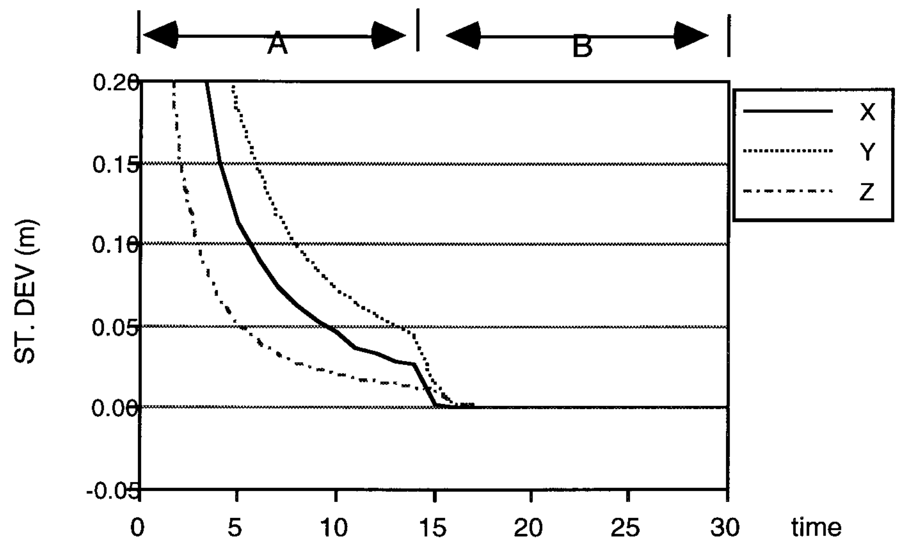


Figure 8.3-1. The evolution in quality of a baseline solution as ambiguities are resolved.

With regard to Figure 8.3-1 there are several comments that may be made:

(1) In region A the precision (and accuracy) of the coordinates steadily improves as more data is collected.

(2) As soon as sufficient data is available to resolve the ambiguities (at epoch 15) a dramatic

improvement in the coordinate parameter precisions is evident.

(3) In region B, when the unambiguous range data (based on precise phase measurements now converted to precise range observations) are processed, there is no improvement in the quality of the coordinate solution, and in effect there is no real justification for continuing to collect data past epoch 15.

(4) If enough data is collected over an observation session, the precision of the ambiguity-free baseline solution will steadily improve, converging to that obtained from an ambiguity-fixed solution.

In conventional static GPS surveying the data is post-processed and it is therefore not known apriori at what point (or even if) sufficient data has been collected to ensure an ambiguity-fixed solution is obtained. Hence conservative observation session lengths (not less than 30 minutes, and usually 60 minutes) are recommended (§5.2).

The basis of all modern GPS surveying techniques has been the ability to address the following questions, that suggest themselves from an inspection of Figure 8.3-1 and the comments made above:

- ❑ How can the length of the observation span required to ensure ambiguity resolution be made significantly shorter than for the case of conventional GPS surveying?
- ❑ Are there any "tricks" to improving ambiguity resolution efficiency, particularly when observation sessions are short?
- ❑ How can ambiguity resolution be made more reliable, particularly when observation sessions are short?
- ❑ How can positioning using phase data be best carried out after ambiguities have been resolved? *That is, how to minimise the number of times ambiguity resolution must be carried out?*
- ❑ How can the ambiguity resolution procedure be made so "transparent" that it may be carried out automatically, even with the receiver in motion, and whenever it is required?

In response to GPS survey-user pressure for: (a) increased productivity (that is, shorter observation sessions), and (b) increased operational flexibility (particularly in order to address kinematic applications); manufacturers have developed several modern GPS surveying techniques (§5.5):

- **Rapid static** positioning techniques.
- **Reoccupation** techniques.
- **"Stop & Go"** techniques.
- **Kinematic positioning** techniques.

Each technique has its advantages, and disadvantages, as far as field and office operations are concerned. Furthermore, each technique addresses the two issues of: (1) ambiguity resolution for short observation sessions, and (2) "carrier-range" positioning (§8.1), in different ways, for example:

(1) Rapid Static Procedure: employs a sophisticated ambiguity search procedure to test many sets of candidate ambiguity sets and select the (most likely) correct one, using only a small amount of data. (Hence this method strives to narrow the width of region A in Figure 8.3-1, and to predict the point at which ambiguity resolution occurs with high reliability.) *If the ambiguity search procedure fails, the technique gives poor baseline results because there is insufficient data to obtain a good quality ambiguity-free solution.* The method can only be applied on static baselines.

(2) Reoccupation Procedure: *simulates* a long observation session via two short observation sessions over the same baseline, but separated in time by an hour or more. (It is not the length of the observation session but the change in satellite-baseline geometry over a session that is important for ambiguity resolution, hence the data can be "thinned" during the middle of the session, or deleted entirely!) It shares some of the advantages of conventional (long observation session) GPS surveying in that ambiguity resolution is optional. Good quality baseline results can be obtained from an ambiguity-free solution, but an ambiguity-fixed solution is preferable. Obviously the method can only be employed on static baselines.

(3) "Stop & Go" Procedure: is a combination of static positioning (the "stop" part), and kinematic movement of the antenna (the "go" part). The ambiguities must be determined by some initialisation process, so that all positioning takes place with carrier-range observations (that is, within region B of Figure 8.3-1). There are several methods of ambiguity resolution, including standard static baseline determination, observing a known baseline, "antenna swap" method, and determination of ambiguities "on-the-fly" (that is, as the antenna moves). Such ambiguity resolution (or initialisation) takes place at the start of the survey (before moving to the first point to be surveyed), and at any time loss-of-lock occurs.

(4) Kinematic Procedures: are those when the entire process of ambiguity resolution (or initialisation) and "carrier-range" positioning takes place while the antenna is in motion (see Figure 8.1-2). Otherwise it is identical to the "stop & go" procedure.

8.3.1 AMBIGUITY RESOLUTION REVISITED

For all modern GPS surveying procedures ambiguity resolution plays a critical role. In the case of "rapid static" techniques the aim is to consistently resolve ambiguities for very short observation sessions. For the "stop & go" and "kinematic" techniques, ambiguity resolution is a *precondition* for "carrier-range" positioning. Although in the case of these two techniques *the method* of resolving ambiguities is not important, obviously if it can be done as quickly as possible then the survey operation (using "carrier-range") can be commenced sooner. Further time savings can be had if the ambiguities could be resolved *while the receiver is tracking satellites and moving* to the first survey point, in the so-called "on-the-fly" ambiguity method. It is necessary to re-examine the ambiguity resolution process from these perspectives.

The following strategies can be used to decrease the length of the observation session and increase the reliability of the ambiguity resolution process:

- ❑ Keep baselines short!
  - ☞ *Chances for ambiguity resolution are much higher for baselines < 20km in length.*
- ❑ Track as many satellites as possible and ensure good satellite geometry.

☞ *Generally a minimum of five visible satellites is recommended, with a rapidly changing PDOP.*

- ❑ Improve ambiguity search, selection and testing algorithms.
  ☞ *Better statistical testing, improved search techniques such as Least Squares searching and Ambiguity Function technique, better apriori baseline knowledge, sequential or boot-strapping procedures.*
- ❑ Use dual-frequency observations.
  ☞ *It is possible to boot-strap from resolved "wide-lane" ambiguities to L1/L2 ambiguities.*
- ❑ Use precise pseudo-range data.
  ☞ *Replacing the known geometric range quantities in eqn (8.2-2) by measured pseudo-ranges, leading to eqn (8.2-1).*
- ❑ Use *apriori* information such as known baseline length.
  ☞ *If baseline components are known then the double-differenced ambiguities can be directly estimated from eqn (8.2-2).*

The first two strategies are essential preconditions for reliable ambiguity resolution, but are largely beyond the control of the surveyor. The last strategy is only useful for *redetermination* of the ambiguities when cycle slips occur as the GPS antenna is moved from one survey point to another, in the "stop & go" technique (§5.5).

The ambiguity search, selection and testing algorithms for the "rapid static" and "kinematic" techniques are re-examined.

Apriori Values of the Ambiguity Parameters

There are several possible sources for apriori ambiguity values (§8.2):

- ❑ Determine ambiguities with the aid of pseudo-range data (for example, via eqn (8.2-1)). *The challenge is to average out the pseudo-range noise and multipath signal.*
- ❑ Determine ambiguities using linear combinations of dual-frequency phase and pseudo-range data (for example, the "four observable" combination -- §8.4). *The weakness is always the noise and multipath biases in the pseudo-range data.*
- ❑ Estimate ambiguities using linear combinations of dual-frequency phase data in Least Squares solution schemes, and then extract the L1 and L2 ambiguities using linear combinations of the other estimated ambiguities (see §8.4). *This will give good apriori ambiguities and their uncertainties.*
- ❑ An approximate baseline estimate can be used to *infer* the likely range of ambiguities (taking into account the uncertainty in the baseline components). *Baseline components may be estimated from triple-differenced phase solutions, or double-differenced pseudo-range solutions, or predicted by a Kalman filter (particularly useful in kinematic applications in which the baseline changes with time).*
- ❑ *Apriori* ambiguities may be obtained from an ambiguity-free solution. *For short observation sessions the estimated ambiguities and their uncertainties define a volume containing candidate ambiguity sets that must be searched.*

Note that the above procedures may give the likely ambiguities directly, or merely define candidate ambiguity sets that must be searched. The procedures may be applied to one epoch of data or take into account the data from the whole observation session. They may also be implemented in either the kinematic (moving antenna) mode, or the static baseline mode.

Improving the Efficiency of the Ambiguity Search Procedure

At this point the apriori information may be:

☞ approximate baseline components, with relatively high accuracy.

☞ a range of ambiguity values corresponding to the apriori baseline knowledge.

The more accurate the apriori information, the smaller the search volume (as in Figure 8.3-2) and the greater the likelihood that the correct position is within the search volume. Furthermore, the smaller the search space, the lower the computational burden.

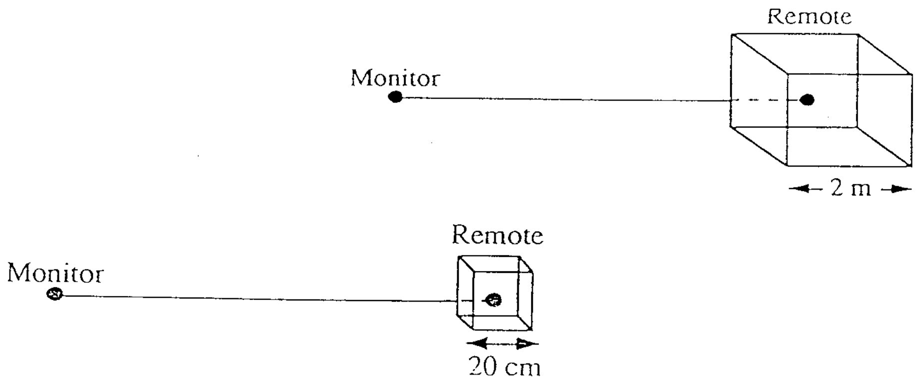


Figure 8.3-2. The search volume as a function of apriori baseline component accuracy.

There are a number of search procedures, each with their own geometric conditions:

(1) *Search volume in the "ambiguity domain"*, consisting of a hyper-ellipsoid in n-dimensional space (n being the number of double-differenced ambiguities), as used in the Fast Ambiguity Resolution Approach employed in the SKI™ software (FREI & BEUTLER, 1990). The *apriori* ambiguities and their uncertainties (determined from a preliminary ambiguity-free solution) define the location, shape and size of the hyper-ellipsoid (see Figure 8.3-3).

(2) *Three-dimensional search volume in the "ambiguity domain"*, consisting of an ellipsoid or cube within which different sets of "lines-of-ambiguities" intersect. The apriori baseline information and its uncertainty define the location, shape and size of the search volume (Figure 8.2-3c is a 2-D example).

(3) Cubic search volume in which the coordinates of the free or "remote" station are expected to lie (the other end of the baseline is the fixed or "monitor" station). The apriori baseline information and its uncertainty define the location and size of the search cube (Figure 8.3-2). *This method is sometimes referred to as a search technique in the "coordinate domain".*

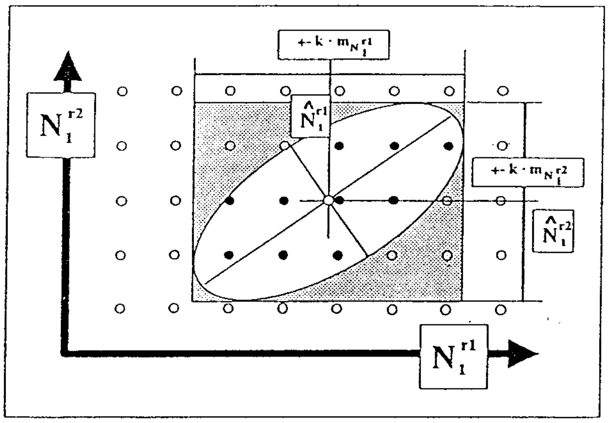


Figure 8.3-3. N-dimensional ambiguity search space used for the FARA technique.
(FREI & BEUTLER, 1990)

(4) Determination of the candidate ambiguity sets directly from the observations, using a combination of carrier phase and pseudo-range measurements. *This method is sometimes referred to as a search technique in the "measurement domain".*

The Least Squares search procedures for "rapid static" and "on-the-fly" GPS surveying techniques, although similar to those employed in the conventional ambiguity resolution techniques, are a more sophisticated implementation of one or more search and validation procedures.

Ambiguity Domain Techniques

Methods (1) and (2) are two geometric descriptions of "ambiguity domain" search procedures. However, there are quite a number of practical techniques described in the literature:

- Fast Ambiguity Resolution Approach (FARA) (FREI & BEUTLER, 1990).
- Cholesky Decomposition (LANDAU & EULER, 1992).
- Spectral Decomposition (ABIDIN, 1993).
- Least Squares Ambiguity Search Technique (HATCH, 1990).
- Fast Ambiguity Search Filter (FASF) (CHEN, 1993).
- Least-squares AMBiguity Decorrelation Adjustment (LAMBDA) (TEUNISSEN, 1994).

HAN (1995) discusses the mathematical basis for each of these techniques, and explains how they are all related to each other and to the general theory of Integer Least Squares estimation.

However, there are several factors which affect the geometry of ambiguity intersections and which may be manipulated in order to change the geometry to make the Least Squares search in the 3-D "ambiguity domain" more efficient (that is, fewer candidate ambiguity intersections):

- Order in which the sequential between-satellite differences are made (Figure 8.3-4a).
- Whether the between-satellite differencing is sequential, or with respect to a fixed base satellite (Figure 7.2-1 and 8.3-4b).

- Choice of base satellite, if the fixed base satellite strategy for between-satellite differencing is used (Figure 8.3-4c).
- Number of satellites observed (Figure 8.3-4d).
- Effective wavelength of the ambiguity cycles (Figure 8.3-4e).

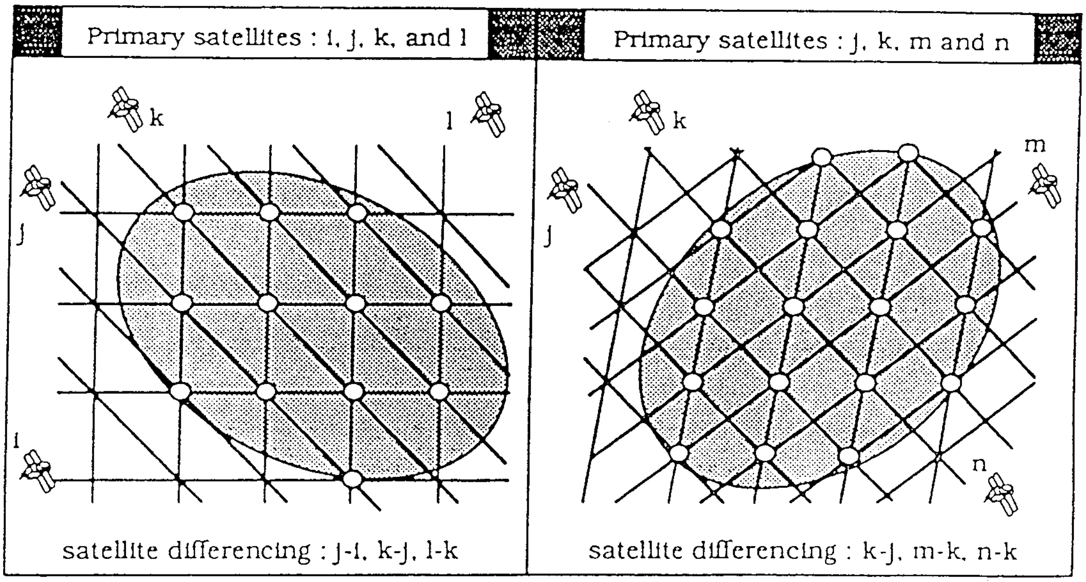


Figure 8.3-4a. Impact of between-satellite differencing order on candidate ambiguity sets. *(ABIDIN, 1993)*

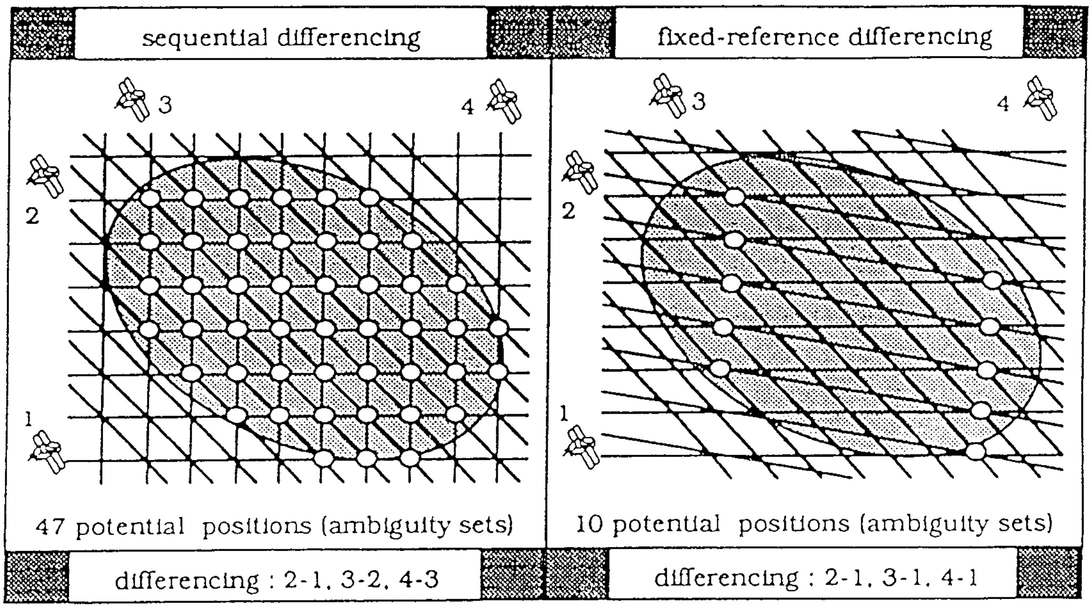


Figure 8.3-4b. Impact of between-satellite differencing strategy on candidate ambiguity sets. *(ABIDIN, 1993)*

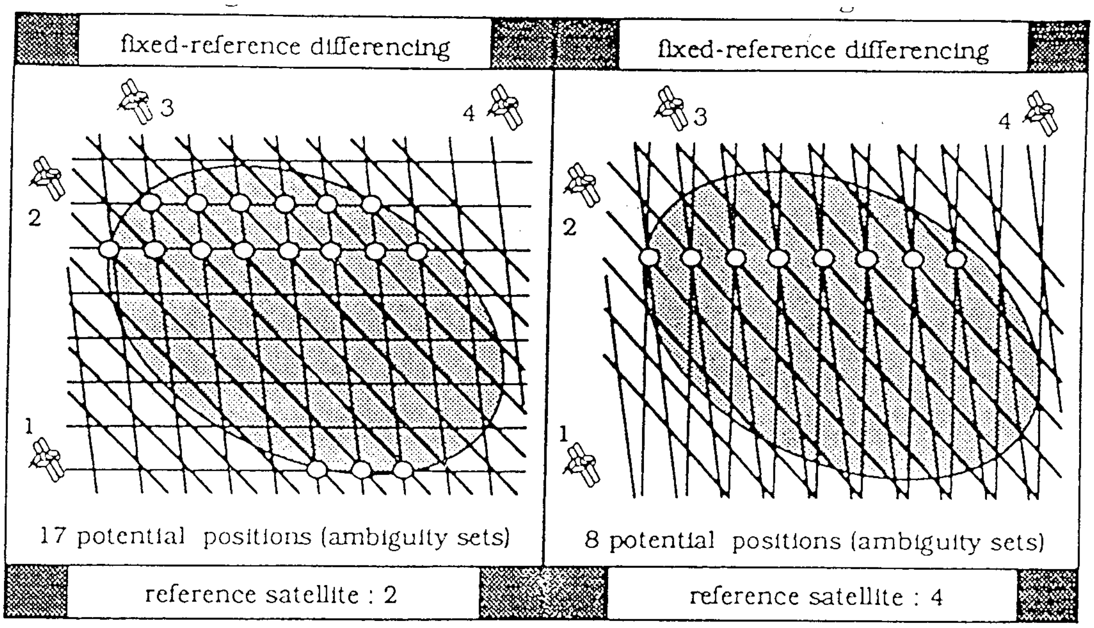


Figure 8.3-4c. Impact of selection of base satellite on candidate ambiguity sets. *(ABIDIN, 1993)*

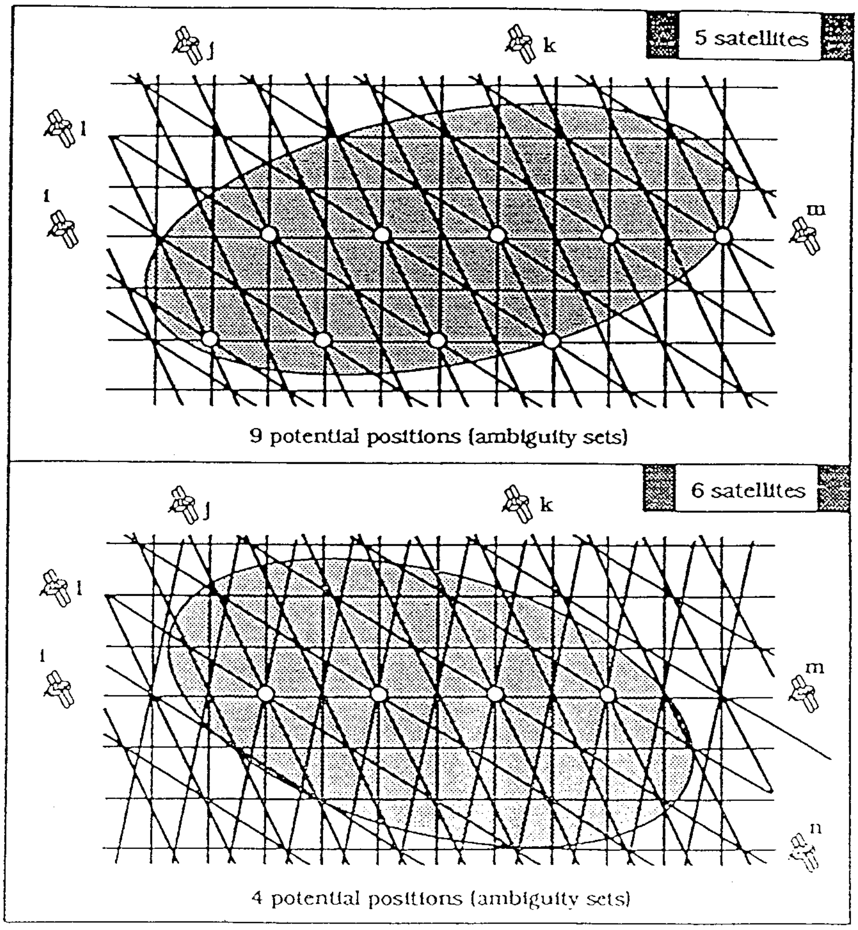


Figure 8.3-4d. Impact of number of satellites on candidate ambiguity sets. *(ABIDIN, 1993)*

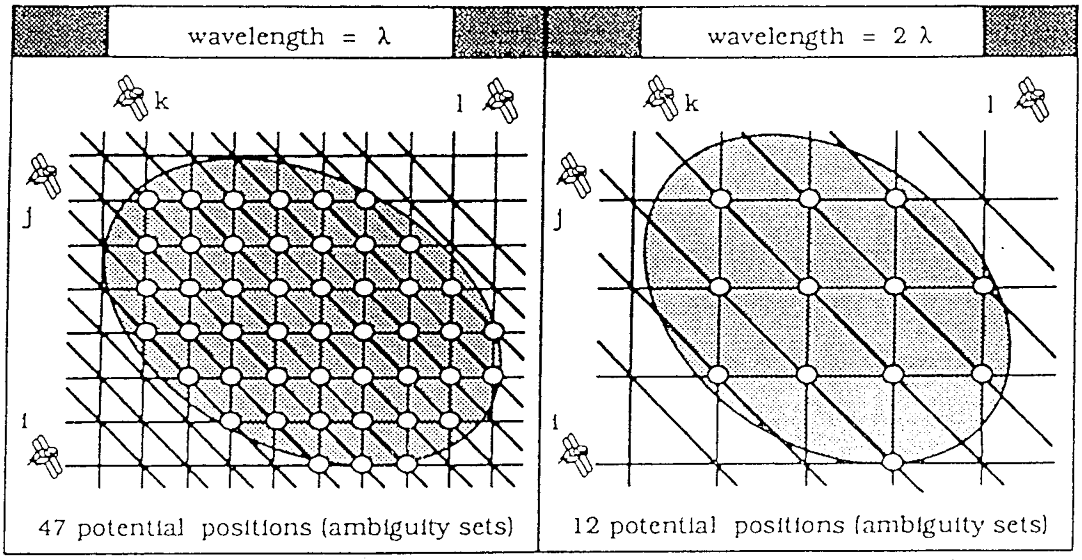


Figure 8.3-4e. Impact of observation wavelength on candidate ambiguity sets. *(ABIDIN, 1993)*

In Figure 8.3-4, the candidate ambiguity sets are those represented by the open circles, where all "lines-of-ambiguities" intersect at one point (or very close to a single point). The fewer such candidate ambiguity sets within the search volume (either the 3-D space here, or the n-dimensional ambiguity space in the case of the FARA technique), the less testing is necessary to eliminate the *bogus* ambiguity sets. However, there is still the issue of how *obvious* is the correct ambiguity set from all possible candidate sets. Very often it is hard to discriminate between the "best" set and the "second best" ambiguity sets when very short observation sessions are processed, hence care is necessary in order to design sensitive testing and rejection criteria that reliably identify only the correct ambiguity set. This is discussed later in the chapter.

Coordinate Domain Techniques

The Ambiguity Function Method (ERICKSON, 1992) is an example of this class of search technique. The AFM has some unique characteristics:

- The AF is defined by the function:

$$AF(x,y,z) = \sum_{k=1}^{m-1} \sum_{l=1}^{j} \cos(\nabla\Delta\phi^{kl} - \nabla\Delta\rho^{kl}) \qquad (8.3\text{-}2)$$

where:
- m is the number of satellites,
- j is the number of observation epochs,
- x,y,z is the coordinate of the trial point,
- $\nabla\Delta\phi$ is the observed double-differenced phase (in cycles), and
- $\nabla\Delta\rho$ is the calculated double-differenced range based on the trial point coordinates (in cycles).

- ❑ Maximum value of AF = (m-1).j
- ❑ Create a fine grid over the search cube and test all trial points within the cube, and compare largest to second largest AF value. The 3-D intersection point (trial coordinate) where the maximum value of the AF is found is the final coordinate value for the "remote" station (unknown end of the baseline).
- ❑ Ambiguity values are not explicitly obtained, though under ideal conditions the maximum AF value occurs at an intersection of integer "lines-of-ambiguities".
- ❑ The method is insensitive to cycle slips, and can be used for static and kinematic applications, with one or more epochs of data.
- ❑ The AFM can be applied to any linear combination of L1 and L2 carrier phase observations, and in fact can be applied sequentially in a boot-strapping approach using observables with successively smaller wavelengths (corresponding to the "fineness" of the search grid).
- ❑ The AFM technique is computationally intensive (particularly if the search volume is large), and is not well suited for real-time applications.

<u>The AFM is mathematically equivalent to the Least Squares search methods</u>, and hence provides no better discrimination between the "best" and "second best" candidate ambiguity sets (or the coordinates corresponding to the largest and second largest AF values) than other search techniques. One important difference between the AF method and the Least Squares search methods is that apart from the size of the search volume and the resolution or "fineness" of the grid of trial coordinates, the AFM does not take into account any of the geometric conditions that influence the number and distribution of ambiguity intersections.

Measurement Domain Techniques

The basis of this technique is the combination of (double-differenced) carrier phase and pseudo-range data (§6.4) as exemplified by eqn (8.2-1):

$$\nabla\Delta\Phi_{ij}^{kl}(t) - \nabla\Delta P_{ij}^{kl}(t) = \lambda.n_{ij}^{kl} + \text{residual biases} + \text{errors} \qquad (8.3\text{-}3)$$

This relation is valid for either L1 or L2, as well as linear combinations of the two frequencies (such as the "wide-lane" or "narrow-lane" combinations -- §6.4). There are several comments that must be made:

- This is a "geometry-free" or "orbit-free" technique as the geometric range is missing in eqn (8.3-3).
- Using L1 or L2 observations in eqn (8.3-3) is problematic as the ionospheric delay for phase or pseudo-range is opposite in sign, and hence does not cancel in the combination (certain dual-frequency combinations can overcome this).
- The pseudo-range data is quite "noisy", with measurement errors that may be as large as several wavelengths.
- The pseudo-range data is contaminated by multipath errors.

Despite these problems this technique appears to be the simplest and most "direct" of the ambiguity search techniques, and is particularly useful on its own, or in combination with another search procedure. However, it can only be used with top-of-the-line GPS receivers able to measure all four observables (Φ_{L1}, Φ_{L2}, P_{L1}, P_{L2}).

Test and Rejection Criteria for Ambiguity Resolution

The standard test criteria is based on the ratio of "best" RSS of the observation residuals to the "second best" (eqn (8.2-4)). The correct ambiguity set is assumed to have been identified if the ratio is larger than some specified threshold value. A more formal approach is to define the test statistic and to then test this statistic against an expected value, at some confidence level. For example, the test statistic may be defined as (HAN & RIZOS, 1996):

$$T_i = \frac{m^2_i}{m^2_s} \tag{8.3-4}$$

where the aim of the test is to discriminate between the smallest RSS value (m_s) and any other RSS value (m_i). The rejection criteria for the ambiguity set corresponding to m_i being if:

$$T_i > \xi_{F_{n-t,n-t};1-\alpha} \tag{8.3-5}$$

where:

	n	is the number of double-differenced observations,
	t	is the number of estimated parameters,
	$F_{n\text{-}t,n\text{-}t}$	is the F distribution with (n-t,n-t) degrees of freedom, and
	$\xi_{F_{n-t,n-t};1-\alpha}$	is the boundary of the 1-α confidence interval,

where α is usually taken as 5%.

All candidate ambiguity sets generate an RSS value m_i, and the aim of the test is to see if there is a "statistically significant" difference between any of these RSS values and the smallest RSS value. If after testing all the candidate ambiguity sets against the prime candidate set (the one generating the smallest RSS value) it is found that no other ambiguity set can be considered "statistically close" the prime candidate set (by all failing the test at eqn (8.3-4)), *then the prime ambiguity set is indeed the correct set and the ambiguity resolution process has concluded successfully.*

There are several comments to be made with regard to such ambiguity resolution tests:

- The definition of the test statistic (eqn (8.3-4)) can vary.
- Under conditions of short observation sessions, weak satellite-receiver geometry, the presence of multipath and other biases, the correct ambiguities may in fact not generate the smallest RSS value. In this case, it would be unfortunate if ambiguity resolution was unsuccessful (that is, there is an insignificant difference between the smallest and the second or third smallest RSS values), but disastrous if the ambiguity resolution process was thought to have concluded successfully and the false ambiguity set was selected!
- The *sensitivity* of the test is somewhat arbitrary as the aim is for the test to be sensitive enough to eliminate many of the bogus candidate ambiguity sets, but not *too* sensitive to cause a false ambiguity set to be selected as the correct one.

- The best procedure may be to incorporate a number of different tests (each with a different "power" for eliminating bogus ambiguity sets, but which do not reject the correct set), in a certain order as suggested by ABIDIN (1993, 1994) in his "integrated on-the-fly ambiguity resolution approach" -- Figure 8.3-5.

There is much R&D being carried out at present in this area.

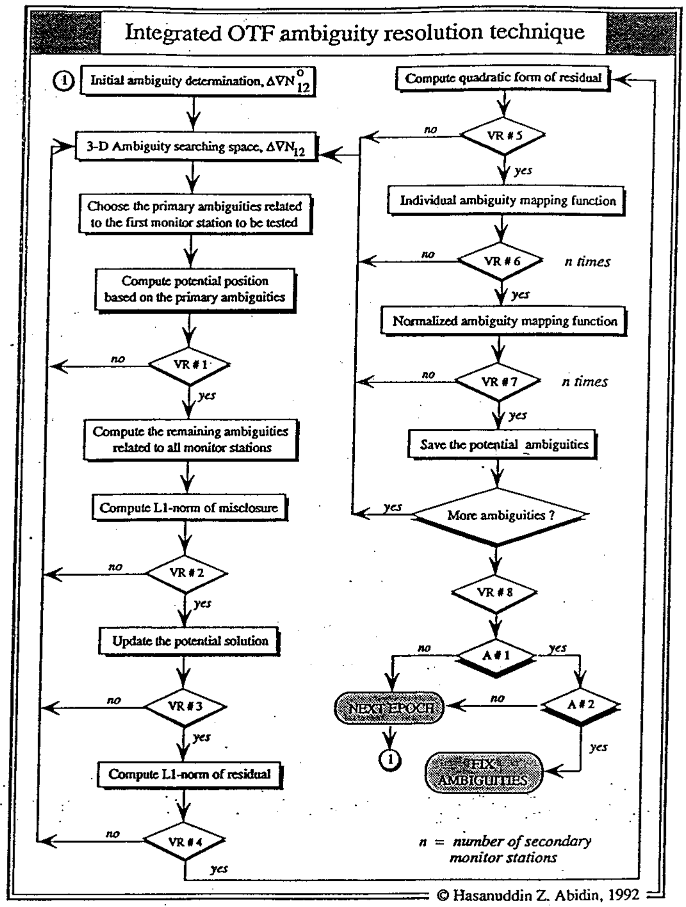


Figure 8.3-5. Ambiguity resolution multiple testing procedure as suggested by ABIDIN (1993).

Despite the tremendous progress made in the development of ambiguity resolution algorithms in the past few years, it is important to keep in mind the inherent weaknesses of short observation session techniques, as ambiguity resolution under such circumstances is not a totally foolproof procedure:

- Increased multipath susceptibility, of both phase and pseudo-range data.
- Very susceptible to the quality of satellite-receiver geometry.
- Only applicable for short baselines, *as baseline increases then observation session length must also increase in an attempt to overcome the increasing influence of residual biases.*
- There is always the risk that not enough data is collected in the field to obtain an ambiguity-fixed solution when data is post-processed, resulting in degraded baseline accuracy. *Real-time processing is therefore an attractive alternative.*
- In the event that the ambiguities are incorrectly resolved it is the baseline linking the base and rover receiver that is degraded. *If signal lock is maintained (such as when the "stop & go" technique is used), then the positions of visited rover sites relative to each other are correct. This factor may be useful for certain observational scenarios.*
- Techniques that rely on full wavelength L2 phase observations and/or dual-frequency pseudo-range are vulnerable with Anti-Spoofing now permanently on (see §4.2), *posing an even greater challenge to the designers of ambiguity resolution algorithms.*
- An increased emphasis should be placed on survey planning as more complex logistics may be required, and in order *to take advantage of the greater productivity of these techniques* (see §5.5), and *to ensure that good Quality Control practices are always followed* (§10.1 and §10.2).

8.4 SOME DUAL-FREQUENCY PROCEDURES

There are a number of strategies for processing dual-frequency data. The ones that are most often used in the context of GPS surveying are:

(1) Processing L1 and L2 data separately.

(2) Processing the L3 ionosphere-free observable.

(3) Processing the L5 "wide-lane" observable, possibly in an iterative procedure with other types of observables.

(4) Using L3, L4, L5 and L6 in certain combinations to aid ambiguity resolution, or cycle slip detection and repair.

Before discussing these it should be noted that the processing procedures (1), (2), (3) attempt to either eliminate the ionospheric delay term from phase data, or at least reduce its impact on certain dual-frequency phase combinations. There are alternative approaches to dual-frequency processing that attempt to handle the ionospheric biases by modelling them in some way (generally with the aid of the L4 combination), or by estimating the ionospheric delay as some form of stochastic process within a Kalman filter. Whatever linear combination of phase (and pseudo-range) data is used, the following characteristics of the resulting linear combinations is desired (eqn (6.4-22)):

- The wavelength is reasonable, not too short, preferably long.
- The residual ionospheric effect is small.
- The measurement noise is kept small.
- The resulting ambiguity is an integer.

Using L1 and L2 Observations

This is the simplest procedure, requiring a minimum of algorithm development. The double-differences (or triple-differences) are formed as discussed earlier, but for the L1 observations independently of the L2 phase observations. The differenced observables are then processed separately either:

☞ to give a single solution in which the L2 observables are merely "extra" observations, strengthening the solution by virtue of an increase in redundancy (but not strengthening the geometry -- this is influenced by the length of observation session not by the "density" of observations),

☞ to give two independent solutions, one an L1 only solution, the other an L2 only solution, the mean of which may be considered the "optimum" solution.

Both of these approaches are tantamount to assuming that between-station differencing eliminates the ionospheric biases. In the case of double-differences, the two types of observables are:

$$\Delta\nabla\phi(t)_{(L1)} = (\frac{f_1}{c}).(\ \rho_1{}^1(T_1) - \rho_1{}^2(T_1) - \rho_2{}^1(T_2) + \rho_2{}^2(T_2)\ )$$
$$+ n_1{}^1{}_{(L1)} - n_1{}^2{}_{(L1)} - n_2{}^1{}_{(L1)} + n_2{}^2{}_{(L1)}$$
$$- (\frac{f_1}{c}).\Delta\nabla d_{ion(L1)} \qquad (8.4\text{-}1a)$$

and

$$\Delta\nabla\phi(t)_{(L2)} = (\frac{f_2}{c}).(\ \rho_1{}^1(T_1) - \rho_1{}^2(T_1) - \rho_2{}^1(T_2) + \rho_2{}^2(T_2)\ )$$
$$+ n_1{}^1{}_{(L2)} - n_1{}^2{}_{(L2)} - n_2{}^1{}_{(L2)} + n_2{}^2{}_{(L2)}$$
$$- (\frac{f_2}{c}).\Delta\nabla d_{ion(L2)} \qquad (8.4\text{-}1b)$$

The $n_1{}^1{}_{(L2)}$ ambiguities as well as the $n_1{}^1{}_{(L1)}$ ambiguities have to be estimated. It is assumed that $\Delta\nabla d_{ion(L1)}$ and $\Delta\nabla d_{ion(L2)}$ are negligible and need no longer be considered. This approach suffers from a number of problems:

- $\Delta\nabla d_{ion(L2)}$ is larger than $\Delta\nabla d_{ion(L1)}$ because the L2 frequency is lower than the L1 frequency.
- The L2 observations tend to be "noisier" than the L1 observations, especially if the GPS receivers employ some variation of the "codeless" (or squaring) technique to track L2 (§3.2).
- Under Anti-Spoofing the L2 wavelength may not be $\frac{f_2}{c}$, but half this value, that is ≈12cm.
- $\Delta\nabla d_{ion}$ tends to grow with increasing baseline length, as the ionospheric conditions at the two antennas decorrelate.

The first three problems make ambiguity resolution more difficult in L2-only double-difference solutions. The last point is the crunch. Because the ionospheric bias is not adequately handled for interstation distances of the order of 20km or greater, ambiguity resolution is often difficult, or not possible at all. There are better strategies for using dual-frequency observations for baselines longer than about 20km.

Among the other possibilities are to treat the L1 and L2 observations as two separate observation equations (as mentioned earlier), but to introduce a common ionospheric parameter linking the L1 and L2 observations (using eqn (6.4-2)) to be estimated as an epoch parameter in much the same way that clock errors are accounted for when processing undifferenced phase data. More specifically, at each epoch, the ionospheric parameters are estimated as weighted parameters with an apriori sigma σ_i. When $\sigma_i \rightarrow 0$, the ionospheric bias contaminating L1 and L2 is assumed to have the characteristics of "white noise", and when $\sigma_i \rightarrow \infty$, the solution is identical to the L3 ionosphere-free linear combination (BOCK et al, 1986).

Using the L3 Ionosphere-Free Observable

To process ionosphere-corrected double- or triple-differenced data requires very few algorithm changes as the L1 and L2 data are linearly combined into one "pseudo-measurement". Once combined, the data reduction continues as before, as for the case of the ambiguity-free solution using single-frequency data. However, the estimated results from the L3 combination are real-valued ambiguity parameters $\Delta\nabla n_3$ that are a combination of $\Delta\nabla n_1$ and $\Delta\nabla n_2$ (eqn (6.4-9).

The wavelength of the n_3 ambiguity is of the order of 6mm, hence this makes the cycle ambiguity resolution of the L3 signal more complicated than the separate resolution of the L1 or L2 ambiguities because there are several combinations of integer $\Delta\nabla n_1$ and $\Delta\nabla n_2$ ambiguities that produce almost the same $\Delta\nabla n_3$ ambiguity. An inspection of Table 8.4-1 shows how several combinations of even small $\Delta\nabla n_1$ and $\Delta\nabla n_2$ cycles (c_1 and c_2) will produce values of $\Delta\nabla n_3$ cycles (c_3) that are near unity. Hence it is often difficult to reliably decorrelate the integer ambiguities on L1 and L2 using the L3 observable alone. The use of L3 is usually justified so as to provide a better quality ambiguity-free solution than that which would have been obtained from single-frequency observations, or by processing L1 and L2 data separately. Commercial software capable of dual-frequency data processing usually allows, in addition to separate L1 and L2 solutions, for an L3 option to be used for medium and long baselines (> 20km).

Using the L5 Wide-Lane Observable

The L5 observable approach is only appropriate for double-difference solutions. First an ambiguity-free solution is obtained using L5 observables. Because the L5 observable has a relatively long wavelength ($\lambda_5 = \frac{c}{f_1 - f_2} \approx 0.86$ metres), the L5 ambiguities can be more easily resolved than either the L1 or L2 ambiguities for short and medium length baselines, even in the presence of the ionospheric biases (they are still present in the L5 observable -- Table 6.4-1). An ambiguity-fixed solution is then obtained where perhaps one would not have been possible if only single-frequency data had been processed.

How good is an L5 solution? An L5 ambiguity-free solution is inferior to an L1 (or L2) ambiguity-free solution. If the L5 ambiguities can be resolved then the ambiguity-fixed solution is superior to an L1 or L2 ambiguity-free solution. Some commercial GPS software make use of the L5 observable. There are however a number of more sophisticated procedures that have been developed for high precision applications (in which ambiguity resolution is desired for baselines in excess of 100km in length), and which now use the L5 observable to aid ambiguity resolution from very short observation sessions ("rapid static" techniques, or "on-the-fly" ambiguity resolution -- §5.5), or to assist in cycle slip detection and repair.

One approach is to form double-differences of the geometry-free L4 observable from the L1 and L2 phase observations and to decouple the L1 and L2 terms using eqn (6.4-27):

$$\Delta\nabla n_1 = \frac{\lambda_5.\Delta\nabla L4 - \lambda_2.\Delta\nabla L5}{\lambda_5.(\lambda_1 - \lambda_2)}$$

$$\Delta\nabla n_2 = \frac{\lambda_5.\Delta\nabla L4 - \lambda_1.\Delta\nabla L5}{\lambda_5.(\lambda_1 - \lambda_2)} \qquad (8.4\text{-}2a)$$

Several comments can be made with respect to this relation:

- The L4 values could come from a number of sources. For example, they could be the average for the entire observation session, or just a few observations. It could be derived from the L1 and L2 phase data, or the P1 and P2 pseudo-range data, or a combination of all as in eqn (6.4-27).
- P4 cannot be directly substituted for L4 derived from phase data because pseudo-range data is too noisy and susceptible to multipath. However, the P4 observable could be smoothed in some way.
- The L5 values may be residual quantities (the L5 observable after subtracting the modelled geometric range), or the value $\Delta\nabla n_5$ determined from an L5 solution in which case eqn (8.4-2a) would be rewritten in the form:

$$\Delta\nabla n_1 = \frac{\Delta\nabla L4 - \lambda_2.\Delta\nabla n_5}{(\lambda_1 - \lambda_2)}$$

$$\Delta\nabla n_2 = \frac{\Delta\nabla L4 - \lambda_1.\Delta\nabla n_5}{(\lambda_1 - \lambda_2)} \tag{8.4-2b}$$

- Eqn (8.4-2) can also be used for cycle slip detection/repair, by using a procedure in which the variation of the L4 and L5 quantities are separately screened for any cycle slips (by whatever method -- see §7.3) and any jumps c_4 or c_5 can be decoupled using the relation:

$$\Delta\nabla c_1 = \frac{\lambda_5.\Delta\nabla c_4 - \lambda_2.\Delta\nabla c_5}{\lambda_5.(\lambda_1 - \lambda_2)}$$

$$\Delta\nabla c_2 = \frac{\lambda_5.\Delta\nabla c_4 - \lambda_1.\Delta\nabla c_5}{\lambda_5.(\lambda_1 - \lambda_2)} \tag{8.4-2c}$$

Table 8.4-1 lists combinations of c_4 and c_5 for small values of c_1 and c_2. For example, a 1 cycle slip on L1 and L2 will cause only a 0.283 cycles (L1 wavelength size) jump in L4, making that combination difficult to spot in the L4 data series.

Another approach for cycle slip detection/repair, or ambiguity resolution, is to use L3 in combination with L5, and to decouple the L1 and L2 ambiguities using the relation:

$$\Delta\nabla n_1 = \frac{\Delta\nabla n_3 + \alpha_2.\Delta\nabla n_5}{(\alpha_1 + \alpha_2)}$$

$$\Delta\nabla n_2 = \frac{\Delta\nabla n_3 - \alpha_1.\Delta\nabla n_5}{(\alpha_1 + \alpha_2)} \tag{8.4-3}$$

where $\alpha_1 = \frac{f_1^2}{f_1^2 - f_2^2} \approx 2.546$ and $\alpha_2 = \frac{-f_1 f_2}{f_1^2 - f_2^2} \approx -1.984$

The values of $\Delta\nabla n_5$ and $\Delta\nabla n_3$ can be assumed to have been derived from two separate Least Squares ambiguity-free solutions. Alternatively, $\Delta\nabla n_5$ can be obtained from eqn (6.4-

26a). This expression is ionosphere-free, and hence the combination of L3 and L5 processing can be used for ambiguity resolution and cycle slip detection/repair for very long baselines (as in the case of GPS crustal motion surveys) -- see discussion below following eqn (8.4-6). However, dual-frequency pseudo-range data would be required in order to determine the $P_{(L6)}$ quantity.

On the other hand, a similar expression to eqn (8.4-3) can be written for cycle slip detection and repair, by substituting c_5 for $\Delta\nabla n_5$, c_3 for $\Delta\nabla n_3$, etc. In this case the L3 and L5 data series are screened and when a "jump" is detected, the L1 and L2 cycle slips can be decoupled. Table 8.4-1 shows the expected jumps in L3 and L5 for small jumps in L1 and L2. Note that for certain combinations of L1 and L2 slips (for example, 4 cycles on L1 and 5 cycles on L2), the L3 signature is very small (being only 0.264 cycles of L1 size). **An ideal cycle slip detection algorithm should therefore screen all data series: L1, L2, L3, L4 and L5.**

Figure 8.4-1. L3 and L4 signatures (in units of L1 cycles) for small L1 and L2 cycle slips.

$L5 = c_1 - c_2$ $\quad$ $L3 = 2.546c_1 - 1.984c_2$ $\quad$ $L4 = c_1 - 1.283c_2$

$\lvert c_1\text{-}c_2 \rvert$	c_1	c_2	**L3**	**L4**
0	± 1	± 1	**0.562**	**0.283**
0	± 2	± 2	1.124	0.567
0.5	± 0	± 0.5	0.992	0.642
0.5	± 1	± 0.5	**1.554**	**0.358**
1	± 1	± 2	1.422	1.567
1	± 2	± 3	0.860	1.850
1	± 3	± 4	0.298	2.133
1	± 4	± 5	0.264	2.417
1	± 5	± 4	**4.794**	**0.132**
1	± 5	± 6	0.827	2.700
1	± 6	± 7	1.389	2.983
2	± 5	± 7	1.157	3.983
2	± 6	± 8	0.595	4.267
2	± 7	± 9	0.033	4.550
2	± 8	± 10	0.529	4.833

Another procedure for resolving L1 and L2 integer ambiguities from the resolved wide-lane ambiguities $\Delta\nabla n_5$ is to attempt to resolve the narrow-lane ambiguities $\Delta\nabla n_6$, and then to use the relations:

$$\Delta\nabla n_1 = \tfrac{1}{2}.(\Delta\nabla n_6 + \Delta\nabla n_5)$$

$$\Delta\nabla n_2 = \tfrac{1}{2}.(\Delta\nabla n_6 - \Delta\nabla n_5) \tag{8.4-4}$$

This is does not appear to be an easy task because of the short wavelength of the L6 observables ($\lambda_6 = \frac{c}{f_1+f_2} \approx 0.11$ metres). However, there are two positive aspects to using L5 and L6 combinations in this way:

- The L6 observable has very low noise (see Table 6.4-1).
- Advantage can be taken of the "even-odd" relation:

  If $\Delta\nabla n_5$ is even --> $\Delta\nabla n_6$ must be even
  If $\Delta\nabla n_5$ is odd --> $\Delta\nabla n_6$ must be odd

  In this way the effective wavelength of the L6 ambiguity is doubled, and is ≈ 0.21 metres, hence making the resolution of the $\Delta\nabla n_6$ much easier.

Use can also be made of the L4 geometry-free / ionosphere-free observable (eqn (6.4-27)) and the ionosphere-free $\nabla\Delta n_3$ ambiguity values, and then decouple them using the relation:

$$\nabla\Delta n_1 = \frac{\lambda_2.\nabla\Delta n_3 + \alpha_2.(\nabla\Delta\Phi_{(L4)} + \nabla\Delta P_{(L4)})}{(\alpha_1.\lambda_2 + \alpha_2.\lambda_1)}$$

$$\nabla\Delta n_2 = \frac{\lambda_1.\nabla\Delta n_3 - \alpha_1.(\nabla\Delta\Phi_{(L4)} + \nabla\Delta P_{(L4)})}{(\alpha_2.\lambda_1 + \alpha_1.\lambda_2)} \tag{8.4-5}$$

If dual-frequency pseudo-range data is <u>not</u> available, eqn (8.4-5) can still be used but only after making the assumption that $\nabla\Delta P_{(L4)}=0$. This technique is often used for baselines up to 100km, if observed using C/A code dual-frequency receivers (no $P_{(L2)}$ observations). (Note, that $\nabla\Delta L4$ in eqn (8.4-2) is equivalent to $\nabla\Delta\Phi_{(L4)} + \nabla\Delta P_{(L4)}$ in eqn (8.4-5).)

If the narrow-lane pseudo-range combination is available then, through the combination of eqns (6.4-17b) and (6.4-21):

$$\Delta\nabla n_5 = \frac{1}{\lambda_5}(\Delta\nabla\Phi_{(L5)}(t) - \Delta\nabla P_{(L6)}(t)) \tag{8.4-6}$$

the wide-lane ambiguities are obtained directly, without the need for a Least Squares search procedure. Furthermore, due to the peculiar influence of the ionosphere on phase and pseudo-range data (§6.2), the combination above has the unique property of being free of any ionospheric delay. It is advisable however to average the data combination in eqn (8.4-6) in order to beat down the multipath errors in the pseudo-range data. Once the wide-lane ambiguities have been determined, then one of several methods described earlier can be used to decouple the L1 and L2 ambiguities (using either the L4, L3 or L6 observables).

Chapter 9: Elements of GPS Network Processing

9.1 INTRODUCTION TO GPS NETWORK PROCESSING

In general, a GPS survey campaign involves the use of a small number of receivers to coordinate a large number of stations. The area of survey operations may span distances of merely a few kilometres (as on an engineering site), to several hundred kilometres, or even thousands of kilometres in the case of geodynamical surveys. A typical GPS survey, such as for mapping or control densification, involves distances of the order of several tens of kilometres. The survey may be carried out using conventional static GPS survey techniques, or the modern "high productivity" techniques described in §5.5. **However, the principles of network processing are the same whether a small number of baselines were observed over several days using conventional GPS, or many baselines observed in a matter of a few hours.**

In a GPS network survey a number of processing strategies are possible:

(1) If the number of receivers that have been deployed during an observation session is greater than two, then the appropriate **single session processing strategy** must take into account multiple baselines.

(2) As the survey cannot be completed during a single session of observations, a suitable **multi-session processing strategy**, involving the propagation of the results of one session solution into another (and eventually across the entire network) has to be used.

Network Processing Terminology

- A GPS "network solution" is a generic term for any GPS solution involving two or more stations. It may be: (a) as small as a subset of one session (for example, a single baseline), (b) an entire session, or (c) a number of sessions. It may be derived from a reduction of the phase data itself (as in the baseline processing described in chapters 7 and 8), or computed from a secondary adjustment of GPS baseline results.
- A GPS "campaign solution" involves all the stations within a survey network that: (a) have been coordinated in a single field operation (usually involving a number of observation sessions) and identifiable as a single "job" for a client, and (b) which are adjusted together, transformed and perhaps integrated into an existing geodetic network.
- It is possible to distinguish between directly-connected stations (those observed in the same session, and adjusted together in a single session solution), and the indirectly connected stations whose relative coordinates are derived from a multi-session solution.
- Both single and multi-session network solutions may be obtained from either: (a) a

primary adjustment of the raw GPS phase data (if the appropriate phase reduction software is available), or (b) a secondary adjustment of the GPS baseline results (themselves the output of a primary phase adjustment).

❑ It is possible to distinguish between a minimally constrained network solution in which only one station (the so-called "datum" station) has been held fixed, and a GPS network solution that has been constrained to fit an existing geodetic network.

Single Session Processing

The following are the **single session** processing strategies:

❑ The **SINGLE BASELINE** processing mode. In this case the primary GPS reduction software can only handle single baselines. The individual baselines are processed one by one, and the output coordinates and variance-covariance (VCV) matrices are input into the secondary network adjustment software. The correlations between the baselines and between the differenced data are neglected.

❑ The **MULTI-BASELINE** processing mode. This is the mathematically rigorous mode of single session phase data processing because the correlations between the baselines are taken into account. However, in forming the double-differences only the independent baselines are used. This mode of processing therefore takes away the "arbitrariness" of single session processing.

There is therefore an increase in mathematical rigour for session adjustments, from the single baseline mode to the multi-station mode. Furthermore, ambiguity resolution may be easier in the context of multi-baseline and multi-station session processing than if the baselines are processed independently. This is because the correlations in double-differenced observables between baselines aid the resolution procedure.

Multi-Session Processing

In relation to GPS surveying involving more than one session, the following points should be noted:

❑ *The GPS data collected during one session has special properties.* The dataset for a particular GPS session is independent of any other data collected during any other session.

☞ *It may therefore be processed separately.*

❑ However, if there are stations that are common to two (or more) observation sessions, the station coordinates, or baselines, are (functionally) correlated and the coordinate results of one session solution will influence the results of another solution.

☞ *If there is only one common station between sessions, then each session solution may be considered to represent a separate minimally constrained GPS solution.*

❑ To combine session solutions, a minimum of one station must be common to two sessions in order to provide the connection between sessions, and to maintain the "minimally constrained" nature of the GPS network as it is built up session by session. More than the minimum number of connections increases the redundancies in the survey (Figure 5.2-6).

☞ *This has two effects: (a) it improves the overall quality and reliability of the* *network, and (b) it means that the most rigorous form of adjustment must be a simultaneous multi-session reduction of GPS phase data.*

- ❑ The manner in which each session is linked to the previous session (and hence back to the original Datum Station) and to the next: (a) is an important consideration during the network design stage, and (b) has implications for the processing strategies for the total (multi-session) dataset.

The combination of separate session solutions can be carried out in a number of ways:

(1) Using multi-session GPS phase data reduction software. This ensures that stations that are common to more than one session are correctly weighted in the solution, and that datum transfer is carried out within the matrix operations of the Least Squares adjustment. PoPS™ was an example of a commercial package capable of this. High precision geodetic (or "scientific") GPS software invariably have multi-session processing capability.

(2) Secondary network adjustment software can use the results of individual session solutions as input. The multi-session adjustment results and VCV matrices reflect the (independent) contributions of the sessions. The matrices are altered to maintain the minimally constrained nature of the total adjustment. GEOLAB™ is a well-known network adjustment program which can accept GPS coordinates and baselines as observations, as well as traditional survey measurements such as slope distances, theodolite directions, etc.

(3) The task can also be performed with specially written software to handle only the GPS output (that is, no conventional distance and direction data to be used, unlike the case of the GEOLAB™ program). The software is relatively simple as it merely *concatenates* the individual session VCV matrices.

The most rigorous processing of multi-session data, collected in a field campaign in which redundant station occupations were made, is therefore the simultaneous reduction of all phase data in one step.

9.1.1 SECONDARY NETWORK ADJUSTMENT CONCEPTS

From the remarks made above, for an average GPS campaign solution there are two types of solutions: a primary adjustment requiring the appropriate modelling of the GPS observables and the development of the processing strategies that permit the geodetic parameters of interest to be estimated (discussed at the single baseline level in chapters 7 and 8), and **a secondary adjustment that treats the outcome of the primary adjustment as an observation**.

The basis for the following discussion is the theory of Least Squares adjustment. The mathematical models involved are much simpler than for the reduction of GPS phase data. In fact, the methodology is very similar to conventional network adjustments of geodetic observations such as distances, except for the fact that 3-D quantities are involved. The reader is referred to texts such as CROSS (1983) and HARVEY (1994) for background material on conventional geodetic adjustments.

The Input

The minimum input required for a Least Squares adjustment are the observations and their precision. In general these are output by the GPS (primary) phase reduction solution. The output is generally written to a "list" file which must be scanned to extract the relevant information (either automatically by the network adjustment program itself, for example as GEOLAB™ does, or manually by the analyst using a text-editor program). *There is now a trend for commercial GPS software to incorporate the network adjustment module within the main program, as in the case of the SKI™, GPSurvey™ and PRISM™ programs.*

The Nature of the Observations

3-D baseline components, as would be output by single or multi-baseline GPS phase solutions, either triple-differenced or double-differenced phase solutions (ambiguity-free or ambiguity-fixed).

Observation Weights and Re-scaling of VCVs

Each of the above "observations" has a corresponding VCV matrix $\mathbf{Q}_{ll}$ (the inverse of the weight matrix $\mathbf{P}$ multiplied by the variance of unit weight). In the case of a baseline "observation", the VCV matrix is a 3x3 matrix:

$$VCV_{Bij} = \begin{bmatrix} \sigma^2_{Bx} & \sigma_{BxBy} & \sigma_{BxBz} \\ \sigma_{ByBx} & \sigma^2_{By} & \sigma_{ByBz} \\ \sigma_{BzBx} & \sigma_{BzBy} & \sigma^2_{Bz} \end{bmatrix} \tag{9.1-1}$$

This relates station i to station j, and does not involve any other GPS station. The total VCV matrix $\mathbf{Q}_{ll}$ formed from combining separate baseline solutions is likely to be *block diagonal.* Bx, By, Bz refer to the three baseline components. Using the VCV matrix as the basis for assigning observation weight matrix implies an assumption that the errors in the observations (the baseline components) are *normally distributed.* This may not be valid assumption! (However, there is little alternative to accepting this.)

In the case of multi-baseline "observations" the total VCV of a group of m stations (or m-1 independent baselines) is available. The VCV is of dimension 3(m-1)x3(m-1) and is generally full, with the following structure:

$$VCV = \begin{bmatrix} \sigma^2_{Bx1} & \sigma_{Bx1By1} & \sigma_{Bx1Bz1} & \sigma_{Bx1Bx2} & \sigma_{Bx1By2} & \sigma_{Bx1Bz2} & \cdots \\ \sigma_{By1Bx1} & \sigma^2_{By1} & \sigma_{By1Bz1} & \sigma_{By1Bx2} & \sigma_{By1By2} & \sigma_{By1Bz2} & \cdots \\ \sigma_{Bz1Bx1} & \sigma_{Bz1By1} & \sigma^2_{Bz1} & \sigma_{Bz1Bx2} & \sigma_{Bz1By2} & \sigma_{Bz1Bz2} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{9.1-2}$$

The weight matrix ensures that each observation contributes to the adjustment in accordance with its measurement accuracy. Very accurate observations should therefore have higher weights than less accurate observations. Hence, it should be noted that if the weight/VCV matrix is the output of a GPS phase reduction the precisions implied for the coordinate components are likely to be over-optimistic (that is, imply higher accuracies than are warranted). *Invariably millimetre accuracies are claimed, particularly for the double-difference ambiguity-fixed solutions.* The reason that the internal precision implied by the output VCV matrix (from the primary GPS phase reduction) is not a true measure of the external accuracy is that GPS phase observations are affected more by systematic errors (§6.2 and §6.3) than by random measurement errors, as well as by physical correlations. *When such an unrealistic VCV is input into a network adjustment program the resulting solution will generally fail the* ***Variance Factor Test*** (see below). However, the outcome of the adjustment (as far as the parameters is concerned) is dependent on the relative magnitudes of the weights, not just the absolute values.

If it were possible to determine the true VCV (including the effect of the unmodelled systematic errors), then this could be used *in place of the VCV output by commercial GPS software.* Unfortunately there is no (straightforward) means of obtaining such a VCV. *In practice, analysts resort to various empirical techniques to modify the VCV matrix.* This is usually done in an iterative manner, using the variance factor test (or some other suitable statistical test) to guide the analyst.

A scale factor w could be applied (say, 5), making all variances and covariances appear worse in eqns (9.1-1) and (9.1-2):

$$VCV^c = w \, . \, VCV \qquad (9.1\text{-}3)$$

However, a popular method is to describe the total error (that is, combined "internal" and "external" errors) in GPS session results through the use of relative weighting of the session observations according to the baseline lengths. The *true* errors in the observations are assumed to have constant (a) and length dependent (b) components, and a suitable population variance can be developed according to:

$$s^2 = (a + b.L)^2 \qquad (9.1\text{-}4)$$

This may then be used in place of the original diagonal values, or added to the ones already computed. The off-diagonal elements maybe left unaltered or scaled so that the correlations are preserved. Further discussion on this procedure is given in §9.4.

The Solution

The solution procedure is relatively straightforward. The sets of parametric equations (scaled by the appropriate weight matrices) are combined into the Normal Equations, the datum defect is accounted for in some way, and the Normal Equations inverted. The corrections to the apriori coordinates are obtained and the residuals are tested.

The Parameters

In the case of the secondary network adjustment there is only one class of parameters: the geodetic coordinates. Further, the "natural" coordinate parameter system is in the form of 3-D Cartesian components (x_j, y_j, z_j), nominally expressed in the WGS84 geodetic reference system, though some processing software carries out the adjustment in the (ϕ, λ, h) system. The standard Least Squares parametric equation has the form (CROSS, 1983; HARVEY, 1994):

$$(l - f(\overset{\circ}{\mathbf{x}})) - \mathbf{v} = \mathbf{A}\delta\mathbf{x} \tag{9.1-5}$$

The expression in brackets is the so-called "observed minus computed" term. l is the vector of the actual observations, $\overset{\circ}{\mathbf{x}}$ are the approximate parameters, $\delta\mathbf{x}$ are the corrections to the approximate parameters and $\mathbf{A}$ is the design matrix containing the partial derivatives of the observations with respect to the parameters. It is possible to write the explicit form of this parametric equation for a baseline "observation" linking stations k and j as:

$$\begin{aligned} [\, Bx_{kj} - \overset{\circ}{x}_k + \overset{\circ}{x}_j \,] &= \delta x_k - \delta x_j \\ [\, By_{kj} - \overset{\circ}{y}_k + \overset{\circ}{y}_j \,] &= \delta y_k - \delta y_j \\ [\, Bz_{kj} - \overset{\circ}{z}_k + \overset{\circ}{z}_j \,] &= \delta z_k - \delta z_j \end{aligned} \tag{9.1-6}$$

where the RHS contains the parameters and Bx_{kj}, By_{kj}, Bz_{kj} are baseline component "observations". Note that there are in fact three "observations" per baseline, one for each Cartesian component (or ϕ, λ, h components if the ellipsoidal adjustment model is used). However the three "observations" are correlated (hence the weight matrix of the observations $\mathbf{P}$ is full).

The Solution Procedure

There are several characteristics of GPS secondary adjustments worth mentioning here:

- As has been mentioned several times, apart from the unique nature of the observations, a distinguishing feature of a GPS secondary network solution is that it is a <u>three-dimensional adjustment</u>.
- Furthermore <u>the adjustment is strictly "geometric"</u> in the sense that no gravity field influence is modelled, as is usually the case for conventional adjustments (through deflection of the vertical corrections, incorporation of zenith distance observations, levelling data, etc.).
- <u>The structure of the design matrix $\mathbf{A}$ of an adjustment of baselines or station coordinate "observations" is very simple</u>, comprising +1 and –1 (eqn (9.1-6)).
- The <u>greater the redundancies</u> (baseline connections, or multiple station occupations), the <u>greater the usefulness of the secondary adjustment as a measure of the internal reliability and repeatability of the GPS survey</u>. With no redundancy (that is, only one common station between sessions) the network solution degenerates to an algorithm for "concatenating" sessions.
- <u>A minimally constrained solution assumes that only one station is held "fixed"</u> (that is, the coordinates are not permitted to adjust). Having one Datum Station is the minimum necessary to obtain a solution. (If no coordinate parameters are held fixed, then the resulting normal equation matrix $\mathbf{A}^T\mathbf{PA}$ is singular and cannot be inverted.) It is possible to use alternate Least Squares algorithms that overcome this requirement, such as by specifying a suitable apriori VCV of the parameters, or by performing a pseudo-inverse, however, the majority of network adjustment software packages do not have such options.

The Output

As with any Least Squares adjustment the output consists of the estimated parameters (actually the corrections $\delta\hat{\mathbf{x}}$ to the *apriori* values of the parameters $\mathring{\mathbf{x}}$) and their precisions (inferred from the solution variance-covariance matrix $\mathbf{Q}_{\hat{x}\hat{x}}$). What is not output is the true *accuracy* of the parameters, hence the impact of unmodelled systematic errors is not obvious.

The output coordinates can be converted from Cartesian coordinates into ellipsoidal coordinates, or into map projection coordinates, and perhaps transformed to a local geodetic datum (chapter 11). The solution VCV matrix contains the following information:

- **standard deviations** of the estimated parameters,
- **correlations** between the parameters,
- **absolute or point error ellipses** (or ellipsoids), and
- **line or relative error ellipses** (or ellipsoids).

In addition to the parameters and their precisions, other useful information that can be obtained from the solution is contained in the residuals $\mathring{\mathbf{v}}$ and the VCV of the residuals. Least Squares theory does not require that the observation residuals be normally distributed. However, if the observation errors have a Gaussian Normal Distribution, normally distributed residuals may be expected. Several statistical tests can therefore be carried out on the residuals. The statistical tests may be applied to assess the quality of the observations and assist in outlier detection ("bad data") -- a useful summary of this topic can be found in HARVEY (1994). Tests also permit an assessment of the quality of the adjusted parameters and of the validity of the mathematical (functional) model to be made.

Interpreting Solution VCV Matrices

The solution VCV matrix is influenced by:

- the **geometry** of the observations,
- the **reference system** used to relate the parameters to the observations, and
- the **stochastic component** (the observation and apriori parameter VCVs).

Any interpretation of the VCV must be made with an awareness of these factors.

The output VCV matrix of the estimated parameters ($\mathbf{Q}_{\hat{x}\hat{x}}$) is quoted in the same reference system as the parameters, that is, as Cartesian components in the WGS84 system. Conventional geodetic adjustments are essentially two-dimensional, utilising ellipsoidal components (ϕ, λ) or local topocentric components (East, North), with the vertical parameters being estimated in a separate procedure. In order to aid interpretation of the results of a GPS adjustment, it may therefore be useful to modify the Cartesian formulation to an ellipsoidal or topocentric coordinate formulation. The process of changing the VCV from the Cartesian form to the equivalent ellipsoidal and topocentric form is described in IBID (1994) and in §11.1. Figure 9.1-1 is an illustration of the structure of a 2-D VCV matrix (the rows and columns correspond to the horizontal components and their correlations).

Error Ellipses

In all subsequent discussions relating to error figures, attention will be restricted to the horizontal precisions as exemplified by the error ellipses. In the case of planimetric coordinates, the VCV matrix of the coordinates of a point extracted from the whole VCV (Figure 9.1-1) is:

$$\mathrm{VCV} = \begin{bmatrix} \sigma^2_E & \sigma_{EN} \\ \sigma_{EN} & \sigma^2_N \end{bmatrix} = \begin{bmatrix} \sigma^2_E & r\sigma_E\sigma_N \\ r\sigma_E\sigma_N & \sigma^2_N \end{bmatrix} \qquad (9.1\text{-}7)$$

where r is the coefficient of correlation $= \dfrac{\sigma_{EN}}{\sigma_E\sigma_N}$, σ_E and σ_N are the standard deviations of the parameters, and σ_{EN} is the covariance between the parameters.

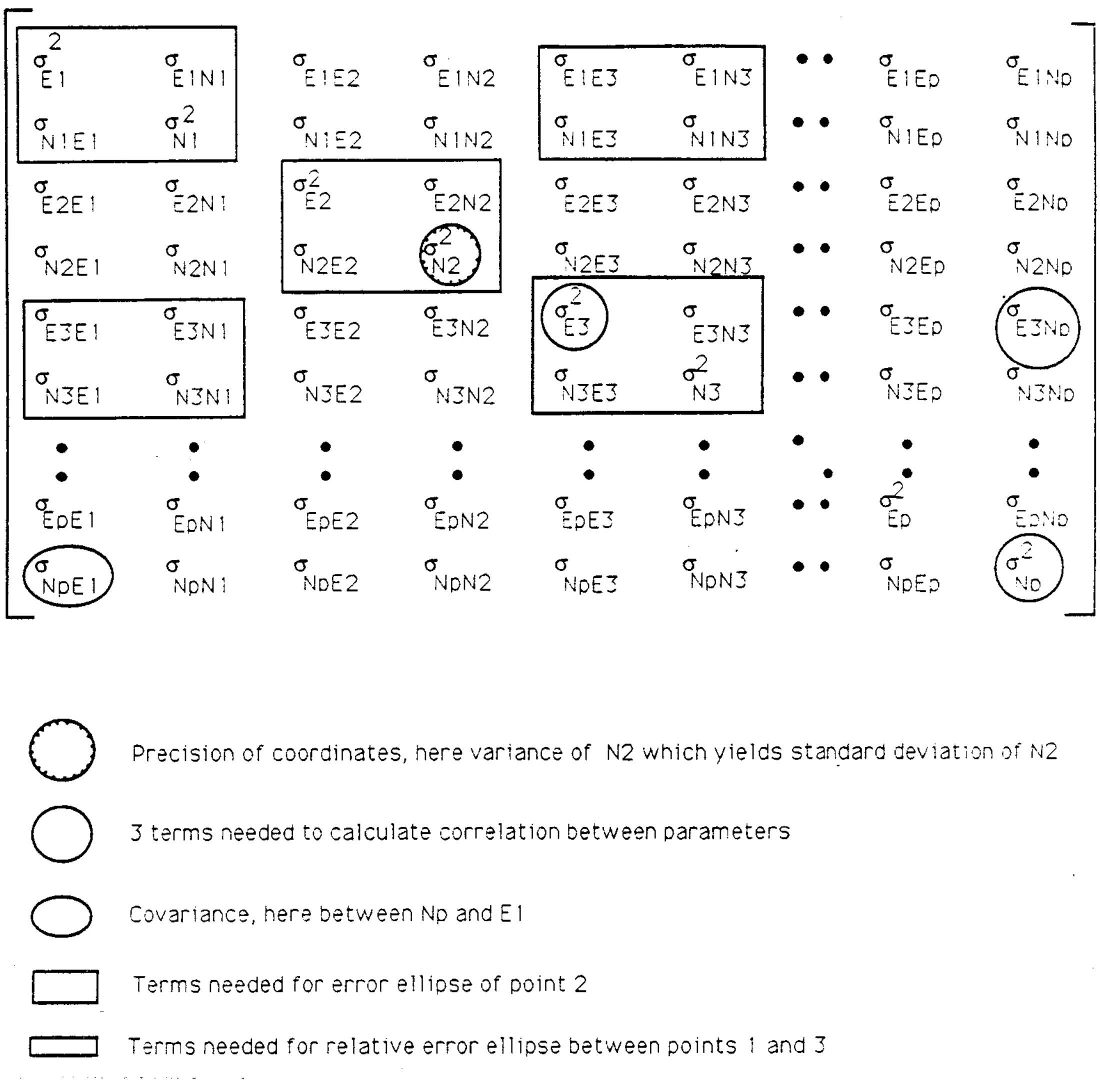


Figure 9.1-1. VCV components of a 2-D survey network.
(HARVEY, 1994)

The "absolute" or "point" error ellipse gives the confidence region of the coordinates of a single point, independent of any other points in the adjustment. To define the error ellipse it is necessary to define the size and orientation of the semi-major and semi-minor axes of the ellipse. The bearing α for which σ^2_x is maximised is:

$$\tan 2\alpha = \frac{2\sigma_{EN}}{\sigma^2_N - \sigma^2_E} \tag{9.1-8}$$

The formulae for the semi-major and semi-minor axes are:

$$\text{semi major axis: } \sqrt{\frac{1}{2}\left\{\sigma^2_E + \sigma^2_N + \sqrt{(\sigma^2_E - \sigma^2_N)^2} + 4(\sigma_{EN})\right\}^2)}$$

$$\text{semi minor axis: } \sqrt{\frac{1}{2}\left\{\sigma^2_E + \sigma^2_N - \sqrt{(\sigma^2_E - \sigma^2_N)^2} + 4(\sigma_{EN})\right\}^2)} \tag{9.1-9}$$

The absolute error ellipses will vary with the choice of origin of a network, as illustrated in Figure 9.1-2. When the origin of the minimally constrained network is located approximately 30km to the south-east of the network (about 15-20km in east-west and north-south extent), the error ellipses are larger (Figure 9.1-2a) than if the origin station is located at PM43494, in the centre of the network (Figure 9.1-2b).

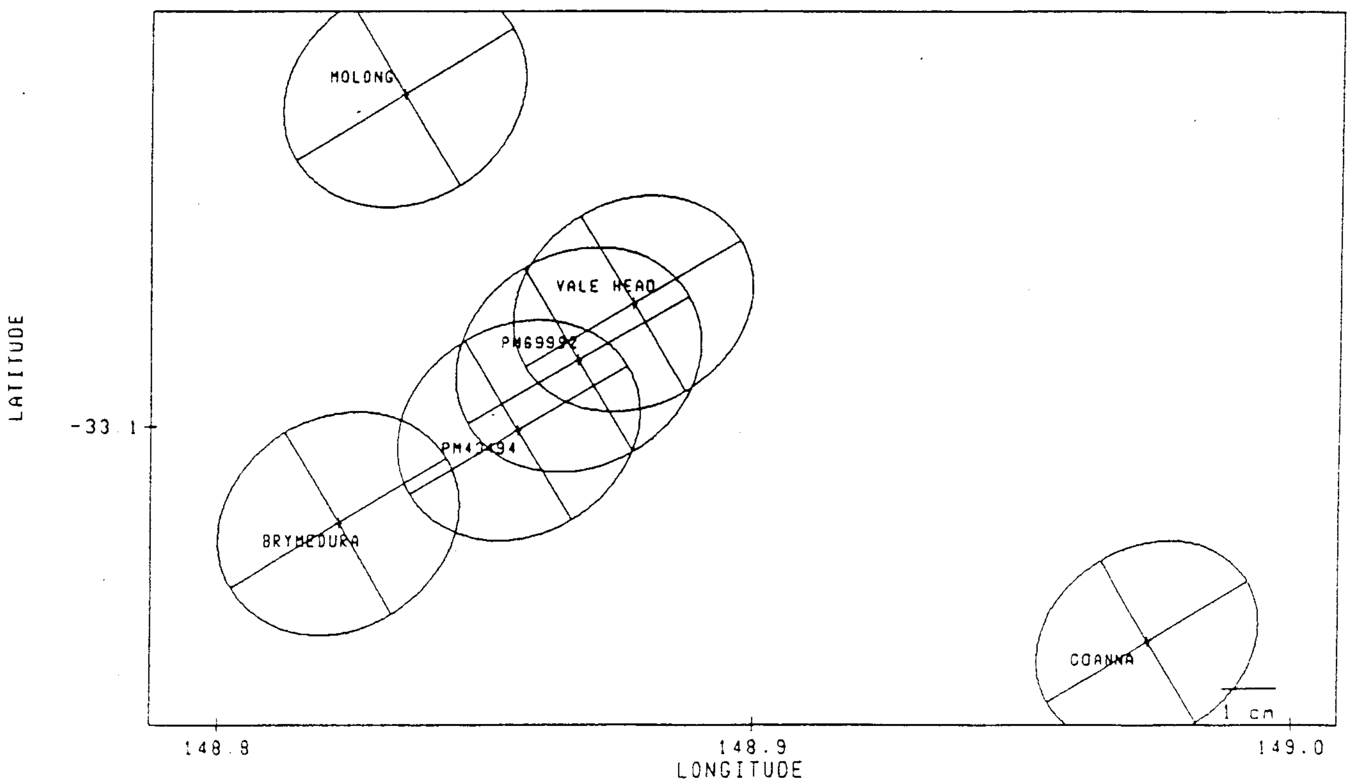


Figure 9.1-2a. Effect of datum selection on absolute 2-D error ellipses: Canobolas T.S. datum (approximately 30km S.E. of centroid of figure).

The error ellipses in Figure 9.1-2 are also referred to as "standard" error ellipses. The probability of a point being in the ellipse for the 2-D case is 39%. It is common practice to draw 95% confidence ellipses at 2.45 times their "standard" size.

Often it is more important to obtain estimates of the precision of the relative positions of two points, rather than of their absolute positions. These estimates can also be found from the VCV matrix of the coordinates. Consider two points A and B, the relevant section of the VCV matrix is:

$$VCV_{EN} = \begin{bmatrix} \sigma^2_{E_A} & \sigma_{E_A N_A} & \sigma_{E_A E_B} & \sigma_{E_A N_B} \\ \cdots & \sigma^2_{N_A} & \sigma_{N_A E_B} & \sigma_{N_A N_B} \\ \cdots & \cdots & \sigma^2_{E_B} & \sigma_{E_B N_B} \\ \text{sym} & \cdots & \cdots & \sigma^2_{N_B} \end{bmatrix} \qquad (9.1\text{-}10)$$

Then $\sigma_{DEDE} = \sigma^2_{E_A} - 2.\sigma_{E_A E_B} + \sigma^2_{E_B}$ (9.1-11a)

$$\sigma_{DNDN} = \sigma^2_{N_A} - 2.\sigma_{N_A N_B} + \sigma^2_{N_B} \qquad (9.1\text{-}11b)$$

$$\sigma_{DEDN} = \sigma_{N_B E_B} + \sigma_{N_A E_A} - \sigma_{N_B E_A} - \sigma_{N_A E_B} \qquad (9.1\text{-}11c)$$

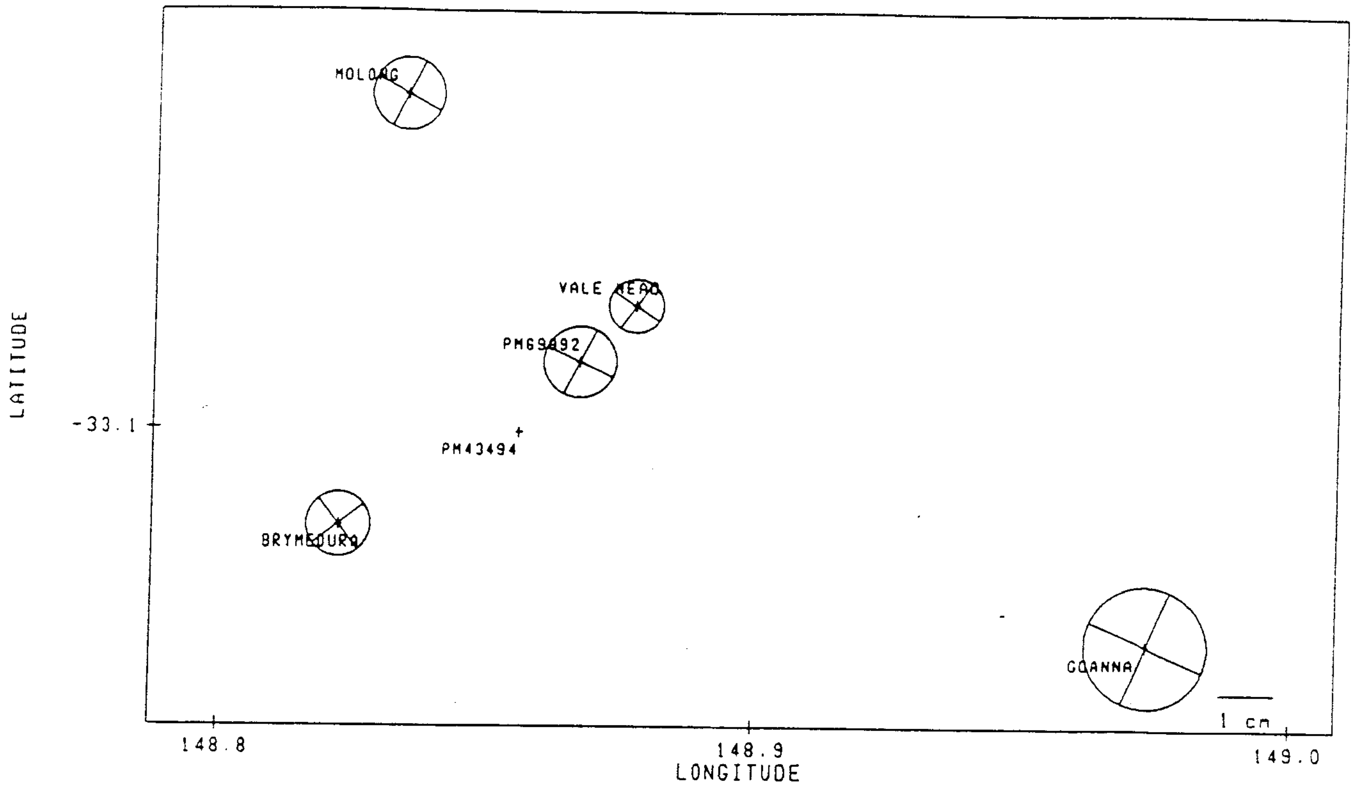


Figure 9.1-2b. Effect of datum selection on absolute 2-D error ellipses: PM43494 datum (approximate centroid of network).

The orientation and lengths of the axes of the semi-major axis can then be obtained from eqns (9.1-8) and (9.1-9), in the same way as for the standard point ellipses. Figure 9.1-3 shows the relative or line error ellipses between points in a minimally constrained network. **The relative error ellipses are unaffected by the choice of origin.**

In the case of a GPS baseline, the semi-major axis of the (relative) error ellipse would be expected to be oriented approximately east-west, and the error ellipse to be nearly circular if the ambiguities had been resolved (although λ is a little weaker than ϕ), but more *elongated* in the case of double-difference ambiguity-free solutions.

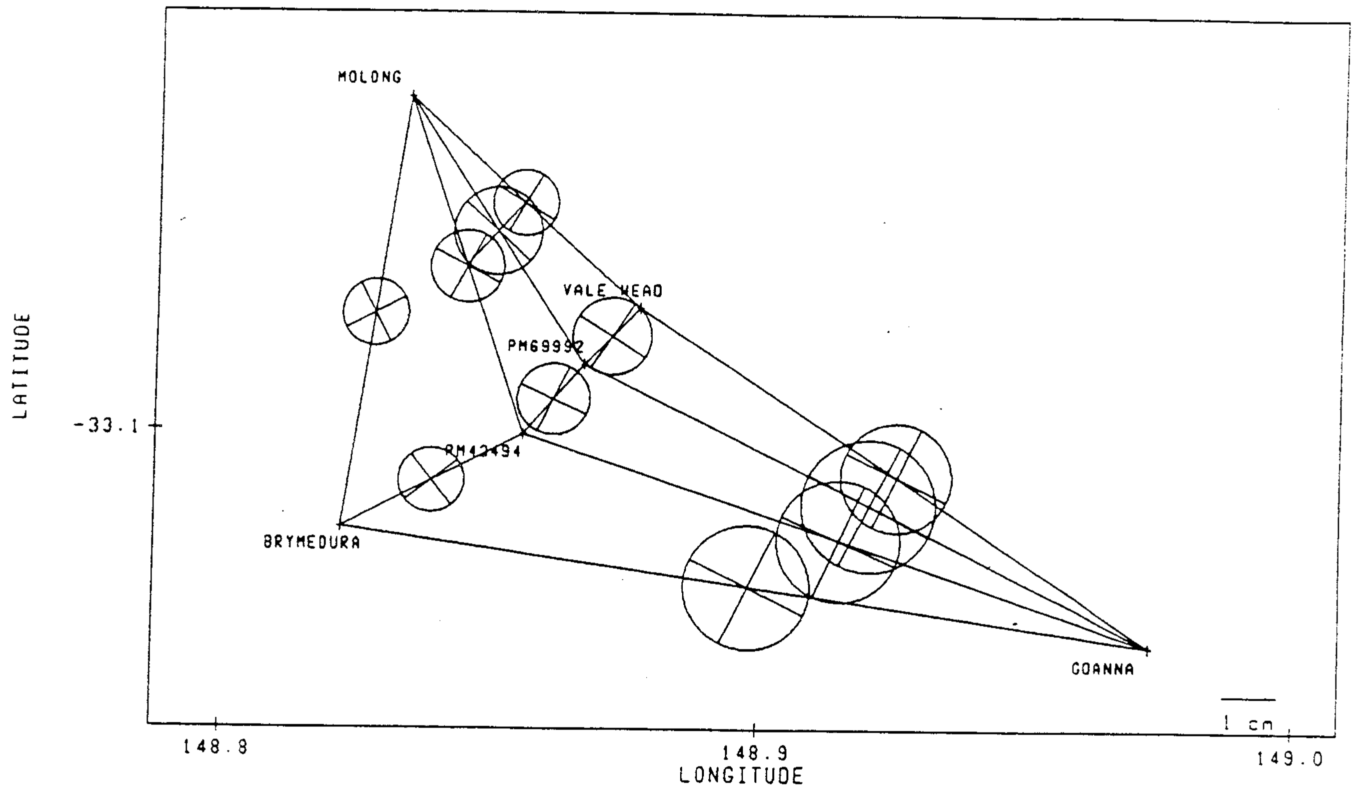


Figure 9.1-3. Example of relative 2-D error ellipses.

Variance Factor Test

The most common test involves the variance factor $\mathrm{VF} = \dfrac{\mathbf{v^T P v}}{(n - u)}$ (also known as the **estimated variance of unit weight**), where n is the number of observations, and u is the number of parameters. This test generally involves comparing the computed (or aposteriori) variance factor against a test statistic from the chi squared distribution, and is usually performed at the 95% confidence level (IBID, 1994):

$$\frac{\chi^2_{1-\alpha/2,n-u}}{n-u} < \mathrm{VF} < \frac{\chi^2_{\alpha/2,n-u}}{n-u} \qquad (9.1\text{-}12)$$

(Sometimes the "F test" is applied instead of the chi squared test.)

If the observation residuals are consistent with their accuracy estimates (VCV matrices), and the residuals are normally distributed, then the estimated variance factor would be expected to take a value between about 0.6 and 1.6, depending on the degrees of freedom $(n - u)$ in the adjustment (see Figure 9.1-4).

If the test fails for GPS session adjustments it may indicate the presence of one or more of the following problems:

- Poor estimate of standard deviations and correlations of observations. The observation residuals are significantly larger or smaller than those implied by the *apriori* standard deviations. The *apriori* covariance matrix $\mathbf{Q}_{ll}$ is inappropriate and may require re-scaling.
- Model or systematic error. *Are enough parameters held fixed?*
- Significant gross errors in one or more of the observations, for example, blunders in the measurement of antenna height propagating into an incorrect baseline solution.
- Poor quality (low precision) observations.
- The calculations are incorrect, for example, there is a "bug" in the program.
- Typing error in input (observations, approximate parameters, or option switches). *Have all the observations been included?*

Note: It is possible to make an error of the type listed above and the solution residuals may still pass the variance factor test!

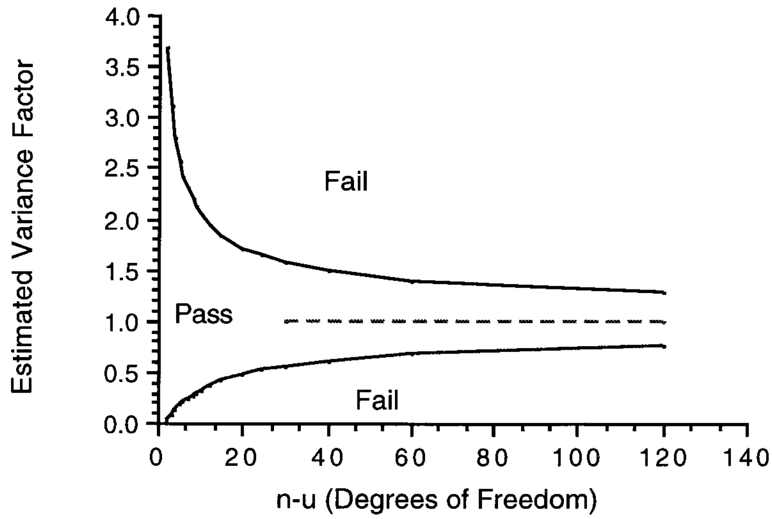


Figure 9.1-4. Variance factor limits. *(HARVEY, 1994)*

Outlier Testing and Residuals

The following is mostly taken from HARVEY (1994), although any textbook on Least Squares adjustment or statistics would give similar information and advice. The residuals are given by:

$$\mathbf{v} = (l - f(\hat{\mathbf{x}})) \qquad (9.1\text{-}13)$$

where $\hat{\mathbf{x}}$ are the adjusted parameters.

Residual testing in general assumes that the errors in observations, and the residuals, are normally distributed. Hence, before statistical tests can be applied it may be necessary to check/test that the residuals are normally distributed.

The familiar bell-shape of the Normal Distribution frequency curve indicates that relatively large residuals can be expected, although these should occur much less frequently than relatively small residuals. For example, 99.7% of all residuals should be less than ±3 times the "root-mean-square" value of the residuals (= the square root of the sum of squares of the residuals divided by the number of residuals), which can be considered an estimate of the square root of the variance (standard deviation) of the observations σ. Thus the chance of a residual exceeding 3σ is very small. (This is the basis of the oft used "rule-of-thumb" that rejects any observation with a residual exceeding 3 times the standard deviation of the observations.)

If one or more of the residuals are significantly larger than either the other residuals in the set, or the residuals obtained from similar adjustments in the past, then it must be decided whether:

- the anomalous observation represents an observation at the extremity of the Normal Distribution, in which case it should be retained, or
- it is indicative of an observation containing a gross error (or blunder), known as an "outlier", in which case it should be rejected.

There is no clear cut boundary between a "small" error (expected in any observation, a "normal" occurrence!), and a "large" error which can be considered "unnatural". *At what cutoff point is an error assumed to belong to a Normal Distribution (ND), or to an "Alternative" (unknown) Distribution (AD)?* This cutoff point is known as the critical value (CV), hence *below* the CV the errors belong to the ND and *above* the CV the errors belong to the AD.

The CV is based on the standard deviation, hence figures such as 1.96, 2.58 and 3.29 correspond to probabilities of 95%, 99% and 99.9% respectively. The figure chosen for the CV will determine what percentage of good observations will be *incorrectly rejected.* If 2.58 times the standard deviation (99% confidence level) is selected as the CV, it is expected that 1% of good data is rejected (together with any observations with "true" gross errors) -- this is a so-called Type I error. The CV figure defines the **level of significance of the test** α (=0.05, 0.01 and 0.001, corresponding to the confidence levels 95%, 99% and 99.9% respectively), and the probability of making a Type I is therefore a function of the CV (=5%, 1% and 0.1%, corresponding to α =0.05, 0.01 and 0.001 respectively).

Second type of false outcome of observation/residual testing is to *accept* a bad observation (that is, assume it belongs to a Normal Distribution), when it should be *rejected* (that is, it belongs to an Alternative Distribution) -- in the statistics literature this is referred to as a Type II error. The probability of making a Type II error is referred to as the **power of the test** β (=0.30, 0.20 and 0.10, corresponding to probabilities 30%, 20% and 10% respectively). Hence, if the power is set to 20%, this means that there is a 20% chance of incorrectly accepting an observation that should have been rejected (or, an 80% chance of correctly detecting an outlier when one occurs).

The Alternative Distribution may also be a Normal Distribution, but with a different mean and standard deviation -- see Figure 9.1-5. This would be the situation if the observations are systematically *biased* in some way. These observations may still be considered outliers.

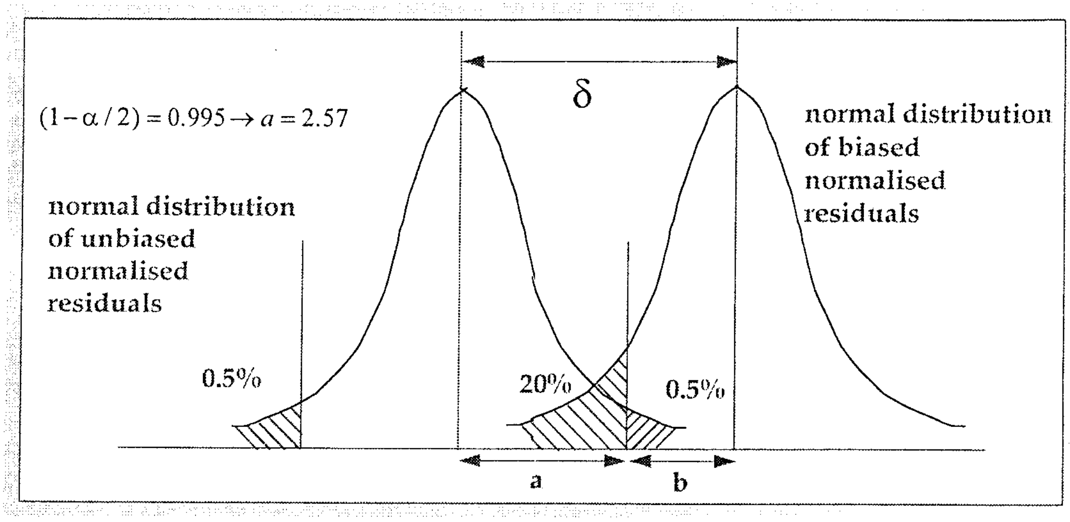


Figure 9.1-5. Residuals may belong to either a biased (RH) or unbiased (LH) Normal Distribution.

Residual testing is usually carried out not on the residual itself, but on a dimensionless quantity known as the **"normalised residual"** $u_i = v_i/\sigma_{vi}$, where σ_{vi} is the square root of the diagonal of the cofactor matrix of the residuals $\mathbf{Q}_{\hat{v}\hat{v}}$ (eqn (7.1-13), in the case of the Least Squares parametric method):

$$\mathbf{Q}_{\hat{v}\hat{v}} = \mathbf{Q}_{ll} - \mathbf{A}\,\mathbf{Q}_{\hat{x}\hat{x}}\,\mathbf{A}^T \qquad (9.1\text{-}14)$$

When observations are *unbiased* (that is, contain no gross error), the normalised residuals are centred around the lefthand ND (Figure 9.1-5). This observation is accepted within the band set by the choice of the level of significance (here 1%, or 0.5% either side of the mean). However, if the observation is *biased* then the normalised residual will also be biased, and its distribution will be centred around another mean (righthand ND). There is still a chance that the value of the anomalous normalised residual will fall within the band between -2.58 and +2.58 standard deviations of the mean of the unbiased residuals, and would be incorrectly accepted as an unbiased residual (and therefore an unbiased observation). The probability of this happening is 20%. The value δ is referred to as the **upper bound** (UB) and its magnitude is the sum of a and b, where a is a function of the parameter α, and b is a function of the parameter β. For example, if α =0.01 and β=0.20, then a=2.58 and b=0.84, resulting in δ=3.42.

How to detect an outlier? Apart from using the above mentioned "rule-of-thumb", there are several statistical tests that can be applied to the residuals, which *do not require a modification of the secondary adjustment process.* The two most common outlier detection techniques are (CROSS, 1983; HARVEY, 1994):

(a) Baarda's Data Snooping method:

- Assumes that the VCV of the observations $\mathbf{Q}_{ll}$ is known.
- Compute the normalised residuals: u_i .
- Considers both the significance level (α) and power (β) of the statistical test.
- Observation i is an outlier if $u_i > 3.42$ (α=0.01, β=0.20).
- Assumes that only one outlier occurs at a time, and that all the observations are uncorrelated.

(b) Pope's Tau Test:

- Assumes that the VCV of the observations $\mathbf{Q}_{ll}$ is not known reliability.
- Compute the standardised residuals: $u_i = v_i/s_i$, where s_i is the square root of the diagonal of the cofactor matrix of the residuals $\mathbf{Q}_{\hat{v}\hat{v}}$ multiplied by $\sqrt{\mathbf{v}^T\mathbf{P}\mathbf{v}/n}$.
- Observation i is an outlier if $u_i > \tau_{\alpha_0/2,n-u}$, a Tau distribution where $(n-u)$ is the degrees of freedom (n is the number of observations, u is the number of parameters), and $\alpha_0 = 1-(1-\alpha)^{1/n}$ (α is the significance level).
- The value of τ depends on the degrees of freedom, the number of observations and the significance level. However, the main effect is the magnitude of $(n-u)$, for if it is small then τ is about 2 or 3, otherwise it is about 4.

The following comments can be made regarding the detection, and subsequent elimination, of observation outliers in GPS Least Squares secondary adjustments:

(1) The *impact* of a certain size of error on the baseline component parameter estimates may vary significantly. For example, a very "sensitive" adjustment will not tolerate even a small observation error because its effect is amplified on the parameter estimates. On the other hand, even quite large errors may be tolerated if they have only a marginal effect on the adjustment.

(2) Marginally Detectable Errors (MDE) is the term used for the magnitude of the error can be only just marginally detected as an outlier, and hence any error smaller will probably remain undetected (and still affect the results). For example, if the outlier test flags residuals greater than 3σ, then this is the MDE and anything smaller will not be detected. If the power of the outlier detection test is 20% then MDE is the smallest size of an outlier that it is possible to detect at the 80% probability level, which according to Figure 9.1-5 is 3.42σ (where $\sigma = \sigma_{vi}$, the square root of the appropriate diagonal element of the matrix $\mathbf{Q}_{\hat{v}\hat{v}}$ in eqn (9.1-14)).

(3) The greater the uncertainty in the estimate of the residual (the larger the value of σ_{vi}), the more difficult it is to detect small outliers, and hence the larger the value of the MDE for that observation. For example, in the case of one observation when outlier detection is carried out with a level of significance of 1%, there may be a 20% chance of an outlier of 10cm (or less) remaining undetected. Yet for another observation the value of the MDE, based on the same level of significance and power of the test, may

be much greater, say 50cm.

(4) The sensitivity of an adjustment in detecting outliers is referred to as the **"internal reliability"**. The internal reliability of an adjustment is therefore quantified by the MDE, and a highly reliable adjustment is one where the MDEs are very small. An overall measure may be taken as the *largest* of the MDEs.

(5) A measure of the **"external reliability"** of an adjustment is obtained by calculating the effect of an undetected outlier on the estimated parameters, assuming that the magnitude of the outlier is equal to the MDE of the corresponding observation. In effect, a new adjustment is carried out n times (the number of observations), each time propagating the MDE of the observation. In this way, an external reliability vector (each of dimension u, the number of parameters) is generated for each observation. An overall measure may be the worst case, that is, the *largest* effect caused by an MDE (which may not necessarily, though usually is, that due to the observation with the largest MDE).

(6) An alternative approach is to *define* the external reliability tolerance, the maximum bias in the estimated parameters that is acceptable, and eliminate only those observations that cause the external reliability to be greater than the specified tolerance.

(7) In the case of a single outlier in a GPS adjustment, representing one baseline component, it would be necessary to delete not just that component "observation" but the entire baseline (that is, all three components). *A further problem is that the three baseline "observations" are correlated, making statistical testing unreliable.*

(8) In the event that there are *multiple* outliers in a population of residuals, the best strategy is to iterate the outlier detection procedure, commencing with the largest residual and if it is flagged an outlier then it is removed from the vector of observations, the adjustment is re-run, and the residual testing is carried out again.

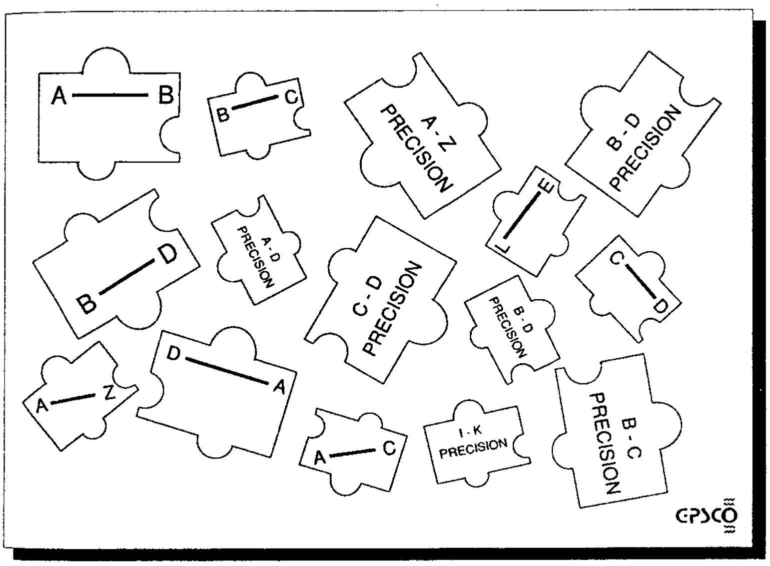
Not mentioned here are procedures based on modifying the Least Squares adjustment process itself, to either make it easier to detect outliers or to make the adjustment procedure less sensitive to the presence of outliers, as in the case of "robust" Least Squares.

9.2

SINGLE-SESSION NETWORK PROCESSING

There are, generally speaking, two classes of processing strategies for obtaining a GPS single session solution:

- **Using a SINGLE BASELINE mode GPS data reduction, for some or all of the possible baseline combinations.**
- **Using a one-step GPS data reduction in the MULTI-BASELINE mode.**



Single Baseline vs Multi-Baseline Solutions

A multi-baseline phase data reduction procedure accounts for the inherent correlation between stations observing simultaneously through the functional model for the double-differences (§6.3) containing parameters linking independent baselines. For example, if baselines 1-2 and 2-3 are to be used to generate double-differences, then the coordinate parameters for station 2 appear in the parametric equations for both of the baselines 1-2 and 2-3, hence the baselines are functionally correlated. Further, if the between-station correlations are include in the weight matrix of the observations, the stochastic correlations between the differenced observations are also taken into account.

It must be stressed from the outset however that, **in practice, there is often little discernible effect at the few parts per million level, on GPS solutions, between**

processing single baselines and processing all session observations in a simultaneous multi-baseline adjustment. However, such a statement must carry several riders:

- ❑ For example, multi-baseline solutions are theoretically superior with regard to ambiguity resolution for baselines of the order of tens of kilometres, than are single baseline solutions.
- ❑ High precision GPS geodesy, particularly for medium and long baselines, demands that all correlations are taken into account.

(With regard to the former, some discussion is given in this chapter. As far as GPS geodesy is concerned, this topic is beyond the scope of these notes and will not be dealt with further.)

The two approaches to GPS single session adjustment: (a) the single baseline processing of individual baselines, and (b) the one-step multi-baseline processing; are discussed in this chapter. Although they may be viewed as being in many respects (competitive) alternatives for processing session data, the two approaches have strengths and weaknesses that can be used to advantage in certain applications. Some of these are summarised below.

Single Baseline Solutions:

- Permit close examination of the *quality of a solution* (for example, such aspects as cycle slip detection/repair, and outlier detection).
- Requires *less sophisticated* GPS data reduction software.
- The results are more easily visualised than in a multi-baseline adjustment.
- All baselines can be processed, *not just the independent ones.*
- Each session's results can be processed at the field office as the survey progresses, and the baseline results accumulated within the network program.
- This non-rigorous approach may furnish the final results (in the case of lower accuracy surveys), or alternatively, the entire dataset may be reprocessed in the office at the end of the campaign.

Multi-Baseline Solutions:

- The solution is *mathematically rigorous.*
- Requires suitable GPS data reduction software, however, *most commercial GPS reduction software is not able to carry out a multi-baseline reduction.*
- The element of "arbitrariness" is reduced, as there is *no need to select independent baselines to process.*
- Ambiguity resolution may be made easier (or at least more reliable) using the multi-baseline approach.
- Does not require secondary network adjustment software to combine the session solutions.
- May be more difficult to "trouble-shoot" problems such as bad data, antenna height errors, etc.

9.2.1 MULTI-BASELINE PROCESSING: PRIMARY GPS PHASE REDUCTION

The procedures for GPS phase reduction described in chapters 7 and 8 are appropriate at the single baseline level only. If more than two GPS receivers (say R) have collected simultaneous phase data during an observation session, then an alternative to processing the R.(R-1)/2 single baselines is to process together all the double-differences that can be formed between the R stations and the S tracked satellites. The essential difference between a one-step phase reduction such as this, and processing the baselines separately is that the double-differences are correlated, and this correlation has the following implications for single baseline processing:

- ignoring these correlations affects both the secondary solution network and its precision estimates,
- the "implied" baselines used for the double-differencing process are considered independent, and
- ambiguity resolution is sub-optimal unless all correlations have been considered.

Independent and Trivial Baselines

Consider a session involving three stations (Figure 9.2-1).

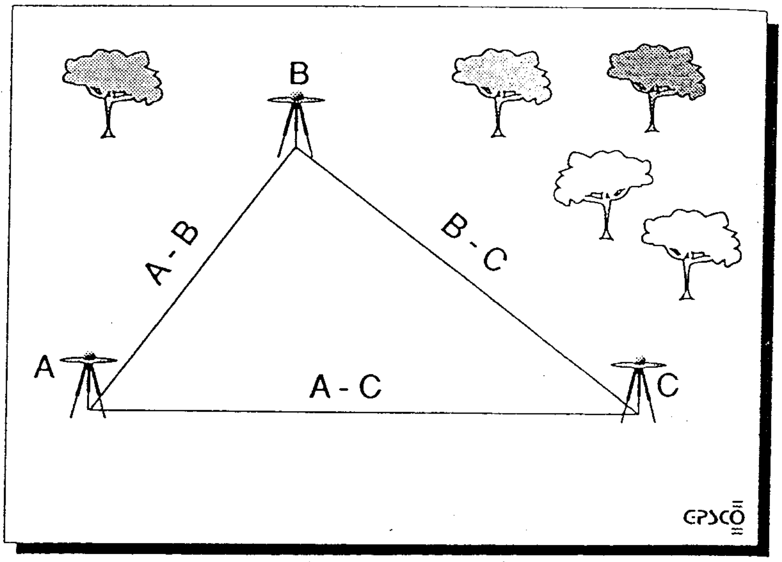


Figure 9.2-1. Multi-receiver session configuration.

One baseline is "trivial", as it can be derived from any other two independent baselines. The choice of which two to process is theoretically arbitrary: A-C & C-B, or A-B & A-C, or A-B & B-C; *but certain combinations may be better for ambiguity resolution.* Alternatively, all baselines may be individually processed and input into the secondary network adjustment (see subsequent discussion on this topic). **It can be argued that there are NO trivial baselines if the baselines have been reduced separately, as the ambiguity resolution process imparts an "independent" quality to all baselines for which it is carried out.**

What role do trivial and independent baselines have in multi-baseline data reduction?

"Baselines" in a multi-baseline reduction of phase data have meaning only in the context of forming the double-differenced observable. In §7.2 independent double-differences for a single baseline were discussed. That is, if S satellites are tracked, then there are $(S-1)$ possible double-differences that can be constructed and used in the baseline reduction software <u>if the correlations introduced by differencing were taken into account</u>. Otherwise a singular normal matrix would result. In a similar fashion, only $(R-1)(S-1)$ double-differences can be constructed if the correlations between observables is taken into account (as in a multi-baseline solution, through the use of the appropriate observation VCV matrix).

Independent double-differences can be obtained by pre-multiplying both the functional model, as well as the set of one-way phase observations, by a **differencing operator** $\mathbf{D}$, as in eqn (7.2-7):

$$\mathbf{D}\bullet\mathring{\mathbf{v}} + \mathbf{D}\bullet\mathbf{v} = \mathbf{D}\bullet\mathbf{A}\bullet\delta\mathbf{x} \qquad (9.2\text{-}1)$$

or

$$\mathring{\mathbf{v}}' + \mathbf{v}' = \mathbf{A}'\delta\mathbf{x}$$

where $\mathring{\mathbf{v}}' = \mathbf{D}\bullet\mathring{\mathbf{v}}$ is the misclose vector of double-differences,
$\mathbf{v}' = \mathbf{D}\bullet\mathbf{v}$ is the residual vector of double-differences, and
$\mathbf{A}' = \mathbf{D}\bullet\mathbf{A}$ is the design matrix of double-differences.

For example, consider 3 receivers tracking 5 satellites, then the differencing operator $\mathbf{D}$ for the case of <u>fixed base satellite differencing</u> and two independent baselines containing station 1 has the form:

$$D = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The $\mathbf{D}$ matrix has dimension 15x8. The structure of the new design matrix $\mathbf{A'}$ is illustrated in Figure 9.2-2 for a single epoch. (The double-difference ambiguities, for example $K_{12}^{6\text{-}9}$, were defined in §6.3 and §7.2.)

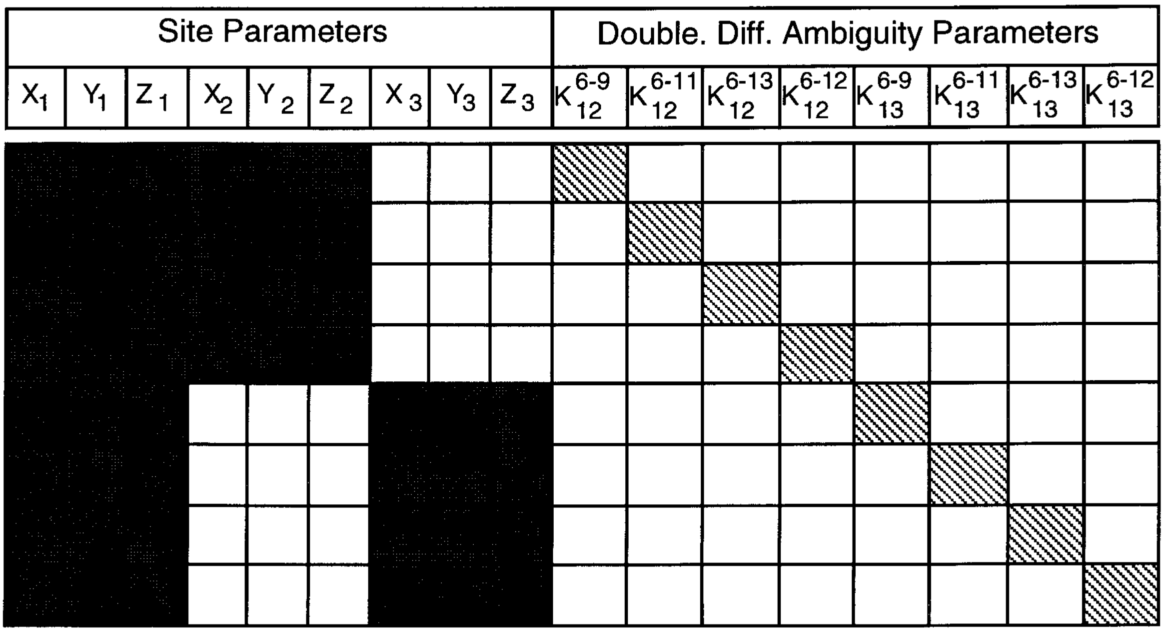


Figure 9.2-2. Structure of the Design Matrix $\mathbf{A'}$ for double-differenced observations (R=3, S=5; base station - base satellite differencing).

The VCV matrix of the resulting 8 double-differenced observations can be formed by applying the Law of Propagation of Variances (eqn (7.2-17)):

$$\mathbf{Q}_{ll}' = \mathbf{D}\mathbf{Q}_{ll}\mathbf{D}^{\mathrm{T}} \tag{9.2-2}$$

where $\mathbf{Q}_{ll}$ is a diagonal matrix. $\mathbf{Q}_{ll}'$ is however a full matrix. In the construction of the normal equations, if $\mathbf{Q}_{ll}'$ is used in place of $\mathbf{Q}_{ll}$ (that is, correlated double-differences between baselines, rather than uncorrelated observables) then the Normal Equation matrix will be full. **When using correlated double-difference observations, the same normal equation matrix $\mathbf{A}^{\mathrm{T}}\mathbf{PA}$ will be obtained no matter which between-satellite and between-receiver differencing strategy is used.**

Selecting the Baselines

There are a number of strategies for determining which $(R-1)(S-1)$ combinations of two satellites and two receivers will be selected to form a double-difference observable. The most common are:

- BETWEEN-SATELLITE differencing using the base satellite or sequential satellite differencing mode. This results in $(S-1)$ independent double-differences per baseline (§7.2).
- BETWEEN-RECEIVER differencing using the base receiver or sequential receiver differencing mode. This results in $(R-1)$ independent baselines (Figure 9.2-3).

For example, any of the between-satellite and between-receiver differencing strategies for R=3 and S=5 will lead to 8 double-difference observations per epoch.

Although the between-satellite differencing strategy can be easily automated, between-receiver differencing can be more complex, sometimes requiring analyst input. For example, the options available for selecting independent baselines from which to form double-differences in the PoPS™ and TRIMMBP™ software were:

(1) User selected reference (or base) receiver.

(2) User selected baselines.

(3) Automatic selection of base receiver, such that the sum of all baseline lengths radiating from this station is a minimum.

(4) Automatic selection of baselines, such that only the shortest (R-1) baselines are used.

The selection of the shortest baselines is a deliberate one, made with a view to maximising the chances of resolving (all or some of) the ambiguities subsequent to obtaining the ambiguity-free solution.

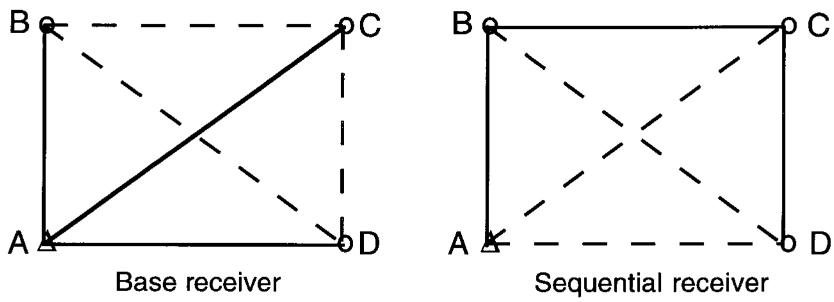
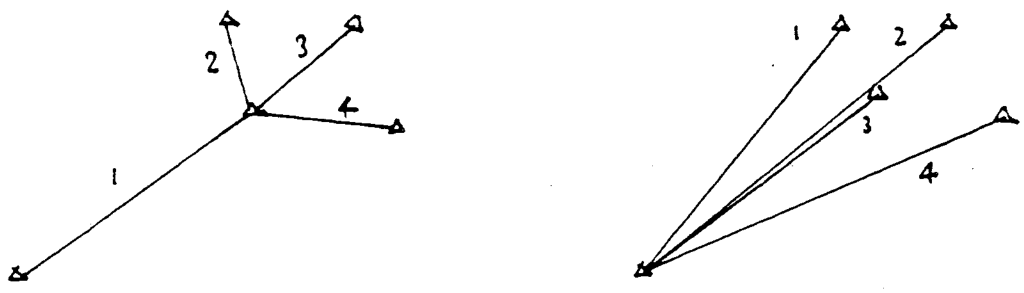


Figure 9.2-3. Options for selecting independent baselines.

Ambiguity Resolution

Ambiguity resolution in a session adjustment requires particular attention. As the double-differenced observations are stochastically correlated and the baseline parameters are functionally correlated, the estimated ambiguity parameters are therefore also correlated. Only in a MULTI-BASELINE adjustment is it possible to resolve ambiguities across baselines by making use of these correlations, through a process of "boot-strapping" from one baseline to another. For example, if ambiguities on a short baseline are resolved, then the subsequent (partially ambiguity-free, partially ambiguity-fixed) solution for the ambiguity parameters on the longer baselines (those not yet resolved) may allow these remaining ambiguities also to be resolved. Therefore, although the selection of baselines (for forming correlated double-differences) is arbitrary, as the same Normal Equation system will result, different $\mathbf{v}'$ vectors can be obtained, and hence a different RHS to the Normal Equation matrix.

Hence, for multi-baseline ambiguity resolution it is good practice to select the independent baselines so that they are the shortest possible. This could be through the use of the base receiver or sequential receiver differencing strategy (or some combination of these). Examples of "good" and "bad" baseline selection are illustrated in Figure 9.2-4.



"Good" baseline selection "Bad" baseline selection

Figure 9.2-4. Baseline selection for multi-station ambiguity resolution.

9.2.2 BASELINE PROCESSING & SECONDARY NETWORK ADJUSTMENT

In chapters 7 and 8 the principles of GPS phase reduction at the single baseline level were presented. **The most recent generation of commercial GPS phase reduction software is not capable of multi-baseline processing.** Secondary network adjustments must be used to obtain a "fit" of the baseline solutions into a single session. Note, the "observations" are the individual GPS baseline components, and associated VCV information, as provided by the baseline reduction software, and not the raw GPS carrier phase data.

Although a true multi-baseline solution is in fact "minimally constrained", as only one station is held fixed, the situation with baseline solutions is more complex. Each baseline solution required one end of the baseline to be left fixed to its *apriori* coordinate value (for example, obtained from a previous adjustment). This means that each baseline solution has a 3x3 VCV matrix associated with it, and the parameters which are treated as "observations" in the secondary network solution are in fact the baseline components, and <u>not the coordinates of individual stations</u>. Some issues that arise in regards to the secondary adjustment of GPS networks are:

- Which baselines should be included in the secondary adjustment. *All or merely the independent ones?*
- Is a "minimally constrained" adjustment required? *Or will several stations be held fixed to their apriori coordinates?*
- Will the single session adjustment be extended to the multi-session case? *To build a campaign network solution.*

These have implications for the processing strategies adopted. Two types of secondary network "adjustments" can therefore be identified:

(1) **The Arithmetic Approach**: This is the trivial form of "adjustment", based on simply adding the baseline vectors together to propagate from the single known (datum) station out to the other stations in the network. The VCV matrices are concatenated into a single network VCV. This can be used in a subsequent network adjustment.

(2) **The Least Squares Approach**: This type of adjustment uses the functional model defined by eqn (9.1-5), and because of its flexibility can process any number of baselines, use any stochastic information and accommodate more than one fixed station. The software can therefore be easily made to handle multi-session adjustments (through input of individual baselines).

Combining Independent Baselines: Using the Arithmetic Approach

If the two conditions: (a) that there is only one fixed station whose coordinates provide the datum, and (b) that no closed loops can be formed using the baseline data (that is, there are no redundant baselines in the network), are met then the single session solution is easily obtained by joining the baseline vectors radiating from the fixed datum station. Generating the total network VCV matrix from the separate 3x3 baseline VCVs is done in a similar manner, and illustrated by using the two examples of baseline selection:

- ❑ If all the baselines radiate from the same (fixed base) station, as in a "cartwheel" pattern, it is possible to simply concatenate the VCV matrices by inserting the appropriate 3x3 submatrix into the network VCV matrix.
- ❑ In the event of more complex baseline schemes (for example "end-on-end" traversing, as used in the sequential receiver deployment mode), the VCV matrix of the total session coordinate parameters involves adding the appropriate 3x3 VCV submatrices together for the station common to both baselines, so that the combined network VCV matrix relates to the new (single) datum station.

Combining Redundant Baselines: Using the Least Squares Network Adjustment Approach

As indicated previously, the use of the method of Least Squares allows for the incorporation of all forms of data into an adjustment in an optimal manner. In particular, its utility for network adjustments, for both the single session case (discussed here) and the multi-session case (§9.3), is that it can handle redundant data. In the single session case these arise from holding more than one station fixed as a datum, and/or by including redundant baselines connecting stations that have already been linked to other stations by another route within the adjustment. *They can also accommodate trivial baselines.*

Consider the example of a four station network observed in a single session with four GPS receivers. The following comments may be made concerning the desirability, or otherwise, of processing all six baselines (three independent, three trivial):

- **In theory**, checking the close of various closed figures (triangles and quadrilaterals) formed by joining baselines is unnecessary for ambiguity-free baseline solutions, as they will always close no matter what systematic errors are present. *It may therefore be a false quality control measure.*
- **In practice**, it may be useful to check the close of figures formed by joining

ambiguity-free baseline solutions, in order to verify that the correct station coordinates, height of antenna, eccentric offsets, dataset, etc., have been used in the independently reduced baselines. *It may therefore be considered a form of quality control.*

- Separate reduction of each baseline can be attempted in order to check if there is a problem with the data, and to check whether ambiguity resolution is possible (it is usually easier on short baselines). It may be preferable to omit from the network solution those trivial baselines *where ambiguity resolution was not possible*, and include only those independent (and trivial baselines) for which ambiguity-fixed baseline solutions were obtained.
- Checking the close of a figure formed by joining the baselines obtained from ambiguity-fixed solutions may detect a falsely resolved ambiguity set on one of the baselines (see §10.3).
- Secondary network adjustment software provides a convenient tool for checking the "fit" of various closed figures, as well as giving the "optimum" network solution for further analysis. An example of such a solution is given below. All six baselines are comparatively short and were derived from ambiguity-fixed solutions. The results indicate that there appear to be no serious errors in the baseline solutions, in as far as can be ascertained from a single session solution (for quality control discussion see §10.3).
- However, if all reduced baselines (independent and trivial) are input into the secondary network adjustment software, the resultant solution statistics are over-optimistic. This is discussed below.
- The resulting VCV matrix of the parameters has the following structure:
  - A block-diagonal structure in the case of the inclusion of only independent baselines (equivalent to the arithmetic session adjustment outcome), indicating that some coordinate parameters are not correlated with others.
  - A full matrix in the case of all baselines, independent and trivial, being included in the adjustment, indicating that the coordinate parameters are correlated with each other.

Should Trivial Baselines be Included in a Single Session Adjustment?

The main argument against incorporation of all baselines (independent and trivial) into any adjustment, single session or multi-session, is because the resulting solution statistics are "over-optimistic", that is, the length of the axes of the error ellipses (or ellipsoids) are shorter than they should be. However, there is strong support for including so-called "quasi-independent" baselines in the single session network adjustment (§9.4):

(1) It can be argued that all baselines reduced separately (and particularly if an ambiguity-fixed solution is obtained) *are in fact independent.*

(2) Including all baselines from a session, if reduced separately, would (partly) *account for the between-baseline correlations* that have been ignored at the single baseline reduction step.

(3) Increasing the number of observations, and hence the redundancies (or degrees of freedom), permits *more reliable statistical testing.*

(4) The use of all baselines, independent and quasi-independent, removes some of the arbitrariness involved in the selection of only independent baselines (different sets of independent baselines may lead to differing results). *The outcome is therefore a*

unique solution.

Example of a Six Baseline Single Session Redundant Network Adjustment

Each baseline has been reduced separately. The VCV matrices have been modified in the manner described in §9.4.

```
                  BASELINE OBSERVATION DATA
                  =========================
BASELINE:    1
"FROM" SITE: PM9920       -4579195.624    2765270.042    -3462837.303
"TO" SITE  : VALE HEAD    -4580210.441    2765452.961    -3461449.908
VCV MATRIX : .5712100000E-05
            -.3688782947E-05   .4368100000E-05
             .2837932189E-05  -.1688775945E-05   .3027600000E-05
BASELINE:    2
"FROM" SITE: PM9920       -4579195.674    2765270.071    -3462837.341
"TO" SITE  : PM43517      -4578660.700    2766320.048    -3462753.140
VCV MATRIX : .4000000000E-05
            -.2535651540E-05   .3027600000E-05
             .2007931506E-05  -.1184518459E-05   .2160900000E-05
BASELINE:    3
"FROM" SITE: PM43517      -4578670.756    2766305.138    -3462754.685
"TO" SITE  : PM43494      -4577941.012    2766364.667    -3463643.425
VCV MATRIX : .3534400000E-05
            -.2572191966E-05   .3204100000E-05
             .1639849702E-05  -.9694237787E-05   .1690000000E-05
BASELINE:    4
"FROM" SITE: PM43517      -4578670.707    2766305.108    -3462754.647
"TO" SITE  : VALE HEAD    -4580220.498    2765438.053    -3461451.453
VCV MATRIX : .4972900000E-05
            -.3268030165E-05   .3880900000E-05
             .2473092755E-05  -.1476933645E-05   .2624400000E-05
BASELINE:    5
"FROM" SITE: PM43494      -4577920.320    2766368.883    -3463628.044
"TO" SITE  : PM9920       -4579185.034    2765259.374    -3462823.502
VCV MATRIX : .4622500000E-05
            -.3282697398E-05   .4040100000E-05
             .2170544325E-05  -.1276909785E-05   .2250000000E-05
BASELINE:    6
"FROM" SITE: VALE HEAD    -4580219.138    2765436.774    -3461450.389
"TO" SITE  : PM43494      -4577939.605    2766363.360    -3463642.324
VCV MATRIX : .8643600000E-05
            -.6053084591E-05   .7290000000E-05
             .4175140182E-05  -.2562779736E-05   .4368100000E-05

    ALTER INPUT VCV MATRICES:
    -------------------------
    CONSTANT HORIZONTAL OFFSET (m)=  .010
    CONSTANT VERTICAL OFFSET (m)  =  .010
    SCALE VCV STD.DEVS...         =1.0000
    LINE-LENGTH ERROR (ppm)       = 1.000
```

APRIORI STATION COORDINATES

1: PM9920	FIXED	X: -4579195.674	Y: 2765270.071	Z: -3462837.341
2: VALE HEAD	FREE	X: -4580210.441	Y: 2765452.961	Z: -3461449.908
3: PM43517	FREE	X: -4578668.689	Y: 2766303.359	Z: -3462751.234
4: PM43494	FREE	X: -4577923.488	Y: 2766359.771	Z: -3463627.595

PROCESSING RESULTS

NUMBER FREE PARAMETERS = 9
NUMBER PARAMETERS (TOTAL)= 12

PARAMETER	NAME	DIFF. CORR.	FINAL VALUE	ST.DV.
*SITE 1 X	PM9920	.000	-4579195.674	.000
*SITE 1 Y	PM9920	.000	2765270.071	.000
*SITE 1 Z	PM9920	.000	-3462837.341	.000
SITE 2 X	VALE HEAD	-.051	-4580210.492	.007
SITE 2 Y	VALE HEAD	.031	2765452.992	.006
SITE 2 Z	VALE HEAD	-.039	-3461449.947	.007
SITE 3 X	PM43517	7.988	-4578660.701	.007
SITE 3 Y	PM43517	16.689	2766320.048	.006
SITE 3 Z	PM43517	-1.907	-3462753.141	.007
SITE 4 X	PM43494	-7.470	-4577930.958	.007
SITE 4 Y	PM43494	19.807	2766379.578	.006
SITE 4 Z	PM43494	-14.287	-3463641.882	.007

OBSERVATION RESIDUALS

BASELINE:	1	SITE 1: PM9920	SITE 2: VALE HEAD	
VECTOR:		-1014.817	182.919	1387.395
RESID :		.001	-.002	.001
BASELINE:	2	SITE 1: PM9920	SITE 2: PM43517	
VECTOR:		534.974	1049.977	84.201
RESID :		.001	.000	.001
BASELINE:	3	SITE 1: PM43517	SITE 2: PM43494	
VECTOR:		729.744	59.529	-888.740
RESID :		.001	-.001	.001
BASELINE:	4	SITE 1: PM43517	SITE 2: VALE HEAD	
VECTOR:		-1549.791	-867.055	1303.194
RESID :		-.001	.001	.000
BASELINE:	5	SITE 1: PM43494	SITE 2: PM9920	
VECTOR:		-1264.714	-1109.509	804.542
RESID :		.002	-.002	.001
BASELINE:	6	SITE 1: VALE HEAD	SITE 2: PM43494	
VECTOR:		2279.533	926.586	-2191.935
RESID :		.000	.000	.000

9.3 MULTI-SESSION NETWORK PROCESSING

The GPS data collected during one session has special properties (essentially, the data is correlated across baselines -- as discussed in §9.2). However, GPS data from different sessions are stochastically uncorrelated -- **no double-differences are formed between sessions!** This characteristic of session data means that the use of secondary network adjustment procedures to combine different session solutions is largely warranted, without sacrificing accuracy, at the part per million level.

However, before proceeding some comments concerning multi-session GPS networks may be made:

(a) As in the case of single session network solutions, there are two general classes of multi-session processing techniques:
- **Primary GPS reduction** in which all raw carrier phase data, collected over several sessions, is reduced in a one step procedure.
- **Secondary network adjustment** using previous single session baseline results as "input".

(b) A "minimally constrained" (multi-session) network means that only a single datum station is held fixed.

(c) To combine session solutions, a minimum of one station must be common to two sessions in order to provide the connection between sessions as the network is built up session-by-session (Figure 9.3-1). More than the minimum number of connections:
- increases the redundancies in the survey, and hence improves the accuracy and reliability of the solution, and allows quality control measures to be applied (Figure 9.3-2), and
- results in the secondary network adjustment being sub-optimal *vis a vis* a primary GPS reduction strategy.

(d) The manner in which each session is linked to the previous session (and hence back to the original datum station), and to the next session, is an important network design consideration.

(e) In order to obtain realistic estimates of the coordinate precisions (in fact the entire VCV matrix) it is often necessary to empirically re-scale the stochastic information provided by the primary GPS phase reduction.

(f) The combination of single session or multi-session solutions (either from primary phase reduction procedures, or secondary network adjustment), into larger campaign solutions will be facilitated through the adoption of a standard solution output format such as SINEX.

(g) The datum for the resulting multi-session network solution can be further modified, through the determination of the appropriate transformation parameters, so that it may be compatible with that of a local geodetic datum. The network may then be constrained to fit local control.

Point (g) is discussed in §12.1. The other issues raised above will be discussed in this chapter. However it must be emphasised that it is harder to obtain the "best" solution in a one-off adjustment as the number of stations in the network grows. Often different strategies, involving combinations of single baseline and session (with and without redundant baselines) solutions, must be used in an iterative process before arriving at the optimum solution. **As with any geodetic network adjustment, coaxing the best results from the available data is as much a "black art" as the rigid application of rules.**

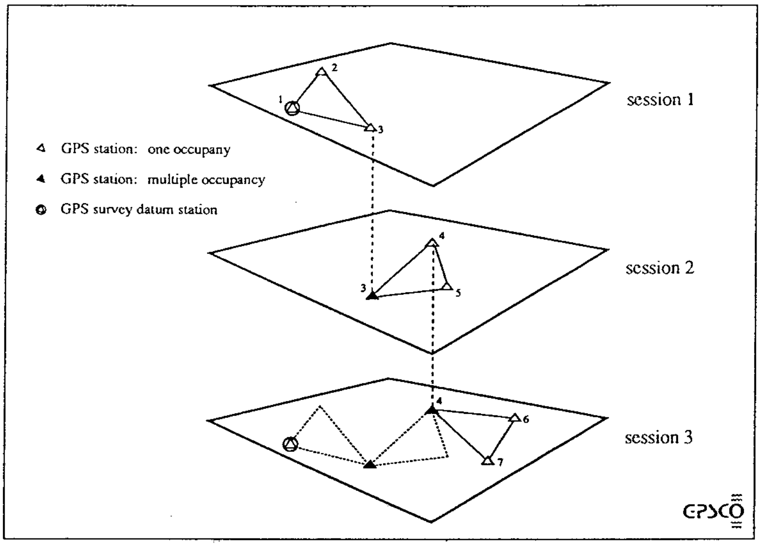


Figure 9.3-1. A multi-session assembly with no redundancy.

9.3.1 OBTAINING A MULTI-SESSION SOLUTION

As in the case of single session solution (§9.2), there are a several strategies for obtaining a multi-session solution:

- ☞ Using multi-session GPS data reduction software. This ensures that stations that are common to more than one session are correctly weighted in the solution, and that datum transfer is carried out within the matrix operations of the Least Squares adjustment. The PoPS™ software was the only commercial package capable of this. *High precision scientific GPS software invariably has a multi-session capability.*

- ☞ Secondary network adjustment software uses the results of individual session solutions as input. *This is the same software that is used at the single session processing level.* General geodetic adjustment packages include such programs as GEOLAB™, NEWGAN, TURBO-NET, COMPUNET, etc., which can, in addition to processing GPS baselines as observations, permit input of conventional survey observations such as distances, theodolite directions, etc. There are many other software packages that will handle only the GPS baselines as input.

In the case of single session adjustments, it is clear that a secondary adjustment of single baselines is sub-optimal compared with a multi-baseline GPS reduction. *This is not so obvious in the case for multi-session adjustments because of the uncorrelated nature of double-*

differences between sessions. **However, there are circumstances where the functional correlation of parameters between sessions would suggest that a simultaneous reduction of the GPS phase data collected over a number of sessions is the superior approach.** Two can be mentioned:

- In the case of the adjustment of common parameters, such as satellite orbital elements, that span several sessions. Commercial GPS software packages do not have this capability. Packages designed for GPS geodesy applications invariably have a multi-session processing capability because of the orbit adjustment requirement.
- Where many redundancies exist, by virtue of more than one common station linking sessions and/or the reoccupation of several stations in different sessions (Figure 9.3-2), a secondary network adjustment is sub-optimal.

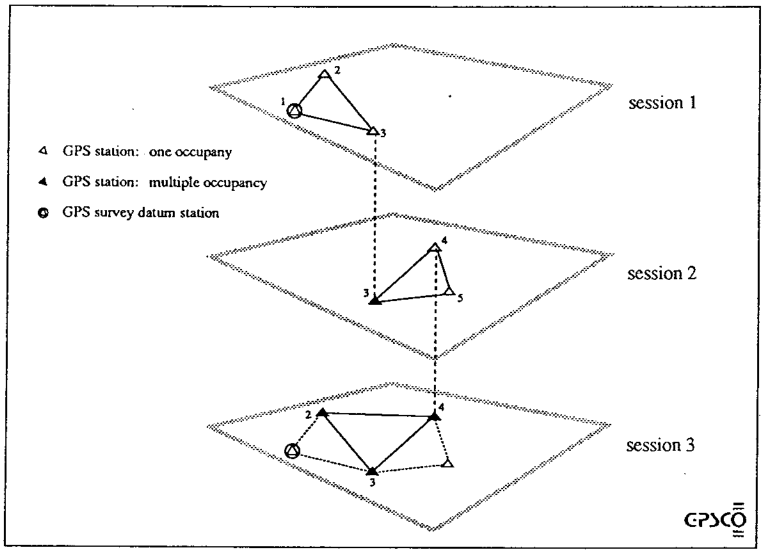


Figure 9.3-2. A multi-session station configuration with redundancies.

However, it must be emphasised, perhaps even more than in the case of single session processing, that **in practice there is often little discernible effect, at the few parts per million accuracy level, on GPS solutions, between: (a) processing single sessions, then combining them in a secondary adjustment, and (b) processing all observations from several sessions in one multi-session GPS phase adjustment**. In fact, there are advantages in combining both approaches where possible. For example, single session solutions could be obtained as an initial office procedure to verify the data quality, followed by a rigorous multi-session solution only when the analyst is satisfied that the network "fits" adequately.

9.3.2 SECONDARY NETWORK ADJUSTMENT: THE MODELS

Two levels of complexity can be identified (as has been done in §9.2):

(1) **The Arithmetic Approach**: The trivial form of "adjustment" where there are no redundancies, based on simply adding the single session solutions together to propagate from the single known (datum) station out to the other stations in the network. The VCV matrices are concatenated into a single network VCV. The final

VCV structure consists of full block diagonal submatrices for each session within a matrix containing null submatrices for the covariances between stations occupied in separate sessions.

(2) **The Least Squares Approach**: The more common type of adjustment can accommodate redundancies, to produce an optimal solution with any degree of constraint, from the minimal (single datum), to the over-constraint (many fixed stations, as discussed in §12.1).

The latter approach, because of its flexibility, is the one universally favoured and will be referred to here by the generic label of "secondary network adjustment approach". This is the one used by software packages such as GEOLAB™.

Choosing a network adjustment program.

Some features to consider:

- **Can it automatically read output files of GPS baseline reduction software?**
- **Can it handle vector correlations?**
- **Can it handle multi-baseline input?**
- **Can it handle non-GPS observation types?**
- **Can it carry out conventional 2-D adjustments as well?**
- **Coordinate Geometry (COGO) capability?**
- **Does it use a geoid model? Can it be changed?**
- **Can it solve for transformation parameters?**
- **Can it hold several stations fixed in an adjustment?**
- **Can it carry out sequential processing (of subnets)?**
- **Does it have graphical output?**
- **What features for altering input VCVs are there?**
- **Result presentation? Variety of coordinate systems?**
- **Maximum number of parameters?**
- **Maximum number of observations?**
- **Computer resources required?**
- **Training requirements? Support and upgrade?**

There are many additional features that can aid Quality Control, error checking and general "trouble-shooting". Some of these are:

- **What statistical testing can it do?**
- **Compute Tau values?**
- **Error ellipse/ellipsoid computations?**
- **What observation editing features are there? Error checking features? Blunder detection?**
- **Can it detect singularities?**
- **Can it detect multiple networks?**
- **Can it flag "no check" measurements?**
- **Can it output vectors and residuals in projection system?**
- **Can it aid users in detecting antenna height errors?**
- **Can it sort stations by name?**
- **Can it sort baselines by length?**
- **Can it sort coordinates by duplicate station names?**

The output of the primary GPS data processing of an observation session is similar to reduced field data in conventional terrestrial surveys. The fact that the GPS carrier phase observations have already gone through a Least Squares process to determine the baseline parameters makes little difference to the development of a network adjustment model, except that the functional and stochastic model of these "reduced" session observations must be correct. There are two options:

- ❑ The output of single baseline solutions consists of the baseline vector and associated 3x3 VCV matrix.
- ❑ The output of multi-baseline processing is a set of n station coordinates and the associated 3(n-1)x3(n-1) VCV matrix.

9.3.3 DESIGNING A MULTI-SESSION NETWORK

As stated in §5.2:

> *"... network design must finally be a compromise between technical requirements and economics, worked out within the framework of explicit recommended practices for GPS surveys (or at the very least, prudent practices that ensure the job gets done satisfactorily) ..."*

The "technical requirements" referred to relate to the propagation of the network from session to session. Surveying more stations than the minimum necessary for the adjustment is an added expense that must be justified on grounds such as "quality control", generally in the form of a contractual obligation, possibly arising from some specification relating to the degree of redundancy that is required (see §10.2). An example will illustrate the principles of designing redundancy into a GPS survey, and how this may impact on multi-session processing. A GPS survey was carried out to coordinate 12 survey control stations, to Australian class "A" standard. Each tracking session involved observations by four receivers, and there were a total of seven sessions. The session deployment strategy is summarised in Table 9.3-1 (the numbers in the columns under the heading "SESSION" refer to the field party).

Table 9.3-1. Station occupation strategy for the Molong (N.S.W.) class "A" GPS survey.

Station	**SESSION**						
	A	**B**	**C**	**D**	**E**	**F**	**G**
PM69992	**1**			**1**			
PM9920		**1**	**1**				
PM69993	**2**		**2**			**2**	
PM43517	**3**	**3**				**3**	
PM43494	**4**	**2**					**2**
Vale Hd. Trig		**4**	**4**				**3**
Molong Trig					**1**	**1**	**1**
Brymedura Trig			**3**			**4**	
Goanna Trig				**2**	**2**		**4**
Canobolas Trig				**3**	**3**		
Yuranigh Trig					**4**		
Nandillyan Trig				**4**			

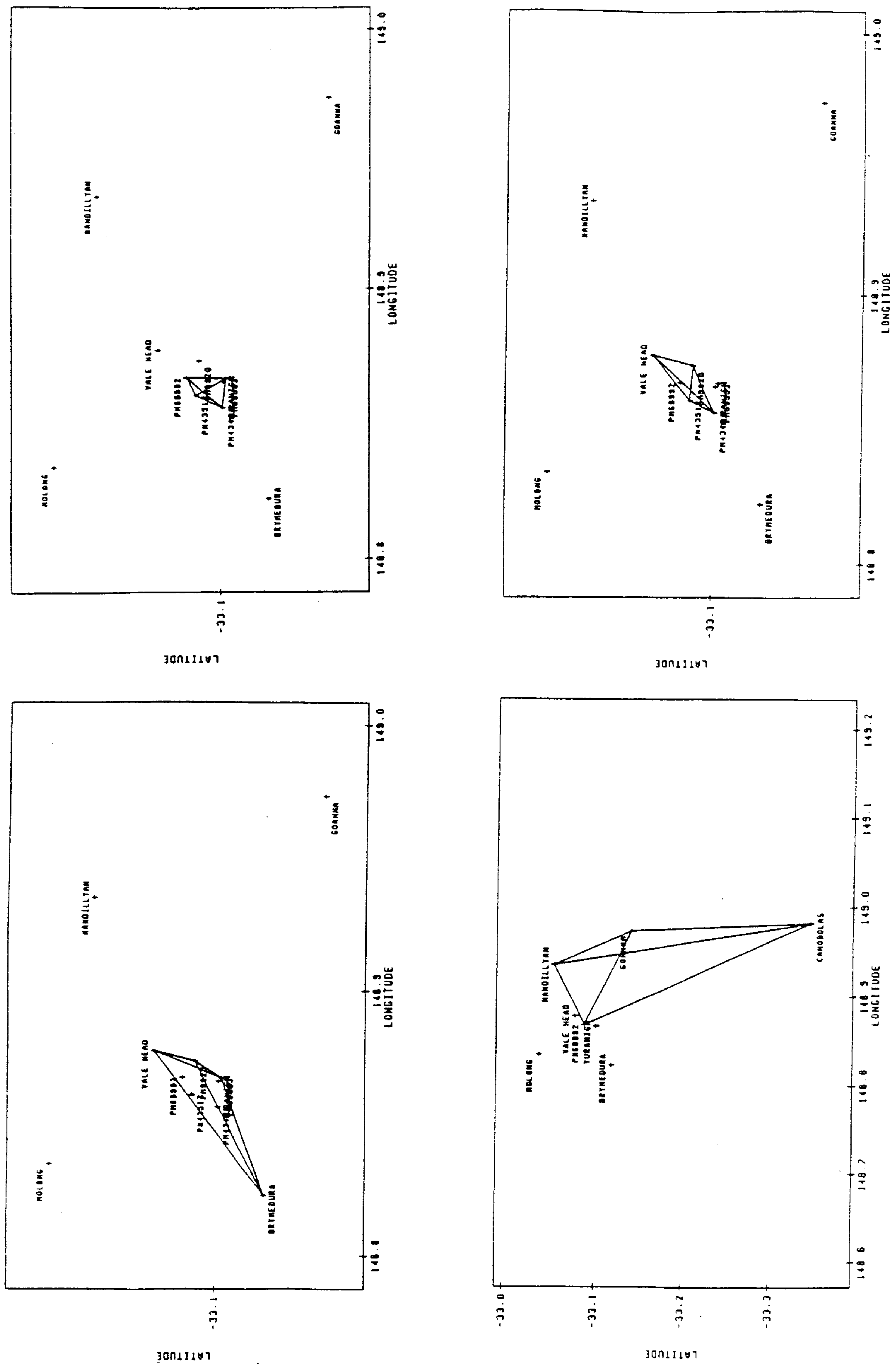
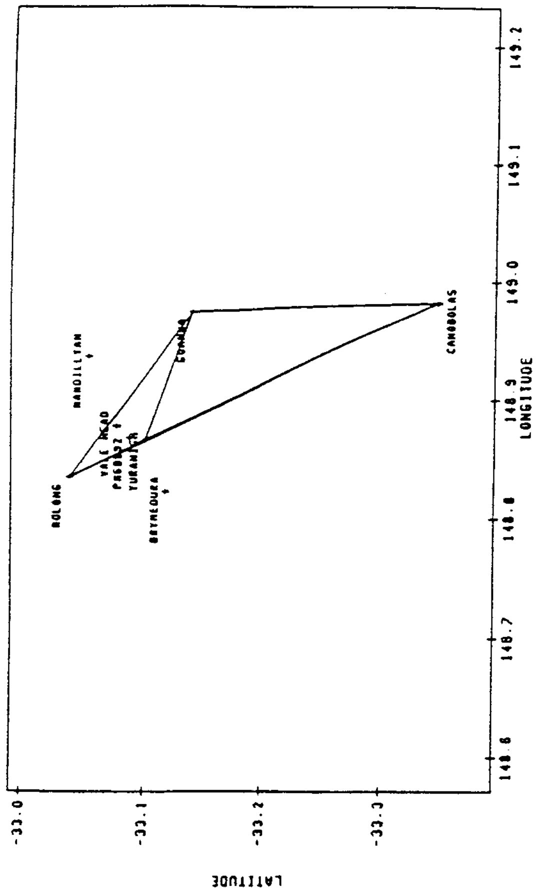
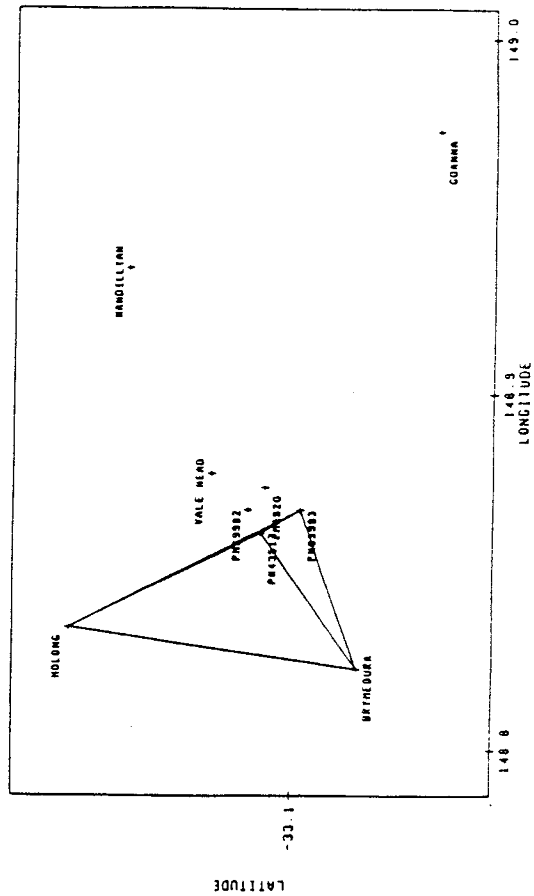


Figure 9.3-3a. The Molong (N.S.W.) GPS survey -- single session layout for session A (upper left), B (upper right), C (lower left), D (lower right).





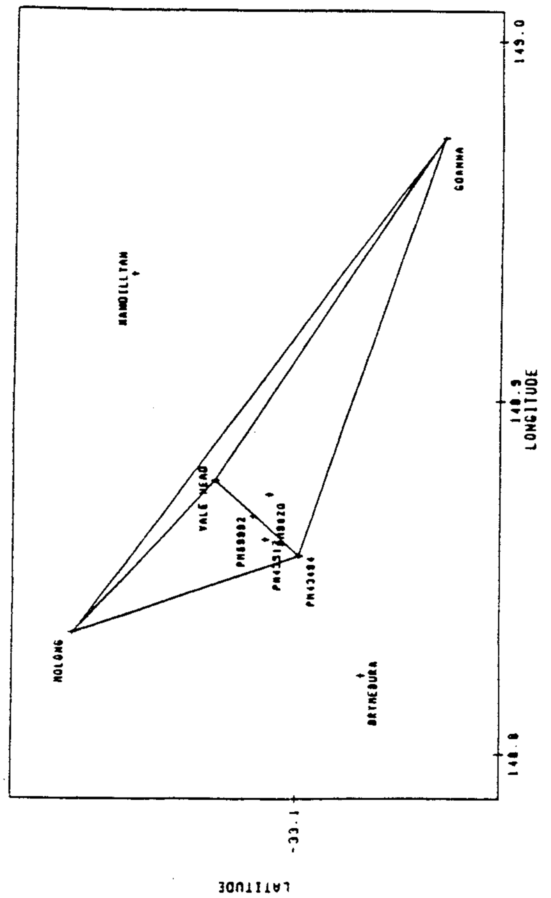


Figure 9.3-3b. The Molong (N.S.W.) GPS survey -- single session layout for session E (upper left), F (upper right), G (lower left).

Note that although there are only 11 independent baselines in this network, seven observation sessions were required in order to ensure a significant amount of redundancy. Four stations (33% of the total) were occupied twice, six stations (50%) were occupied three times, while only two stations were occupied once. The session by session progress of this survey is illustrated in Figure 9.3-3 (all 72 baselines are shown).

9.3.4 SECONDARY ADJUSTMENT OF A MULTI-SESSION NETWORK

The basic steps in a multi-session adjustment are:

(1) Assemble, adjust and analyse each session solution independently using any of the strategies described in §9.2. The aim is to be assured, as far as is possible, that the GPS reductions (single baseline or multi-baseline) are correct by carrying out some quality control tests. *The outcome is a set of station coordinates, and their VCV matrix, in which one station has been held fixed.*

(2) Assemble each session solution into a network (or multi-session) solution. This is generally done in a sequential manner: appending sessions one at a time, checking that each new session does not cause the expanding network to come apart. The tool for this is the secondary network adjustment program. (Some further comments on this are given below.)

(3) Following the successful assembly of all the sessions into a single network solution, an iterative refinement process is commenced. Step (2) aimed to identify problem data (baselines that are not consistent with the overall network). The refinement would involve close analysis of the relative weighting of individual sessions within the overall adjustment, to take into account any factors that differentiate the quality of one session solution from another. *The primary test that is applied to the overall multi-session adjustment is the variance factor test, or some variant of it.*

(4) The outcome is a set of coordinates, and their associated VCV matrix. The solution is minimally constrained (if only one datum station is involved), with the coordinate results referring to a geocentric datum that is "near" WGS84 (see §11.1).

Combining Different Session Solutions: Some Considerations

A multi-session solution can be obtained by:

☞ Combining all previously adjusted single session solutions in one-step, or

☞ Combining sessions sequentially and building up the network step-by-step.

The latter strategy is generally preferred because if a problem with the data is encountered, its source can be identified more easily than if the "trouble-shooting" had to be carried out on a one-step multi-session adjustment. The presence of more than one error, possibly in different session solutions, may "smear" the impact in such a way as to make unambiguous detection of the source of error(s) difficult.

No matter which strategy is adopted, the following must be considered:

- Network Data Types. As discussed earlier, if baseline data is input into the secondary adjustment the output is a set of coordinates, with the resulting VCV reflecting the correlations that exist between the stations. To then input the single session solution into a multi-session adjustment requires that the adjustment software be able to handle station coordinate input. Hence the software may have to handle baselines at one adjustment level (single session), and then coordinate data at the next adjustment level (multi-session).
- Datum Transfer and the Minimally Constrained Condition. This requires that the fixed station in each session solution is "unfixed", and hence variances and covariances terms inserted into the station coordinate VCV where formerly null matrices were present. This cannot be rigorous, although some estimate may be made, using the VCV information from a previous session (also containing the fixed station, but being a "free" station within that earlier session). In the case of baseline data, as with the data type problem mentioned above, the original baseline data can be input into the final multi-session adjustment, the datum may be defined in terms of the *apriori* coordinate variances of the fixed station and no approximation is made as session is appended to session.
- Choice of GPS Datum. The precisions in the VCV matrix of a (multi- or single) session solution are dependent on the geodetic datum definition. In a minimally constrained secondary session adjustment the choice of fixed station will influence the size and orientation of each station's 2-D error ellipse (or 3-D ellipsoid), as illustrated in Figure 9.1-2. The relative error ellipses/ellipsoids are insensitive to the datum choice, and it is these which ultimately define the "class" of the GPS survey (§10.2).
- Mixing Different Receivers or Solutions. Initially, there is no "re-scaling" of session VCV matrices before combination into a multi-session solution. (If one session solution is an ambiguity-free one, and the other an ambiguity-fixed one, it is assumed that this will be reflected in the VCV matrices.) Re-scaling is generally attempted during the "refinement" stage. For example, if dual-frequency GPS receivers were used during one session, and single frequency receivers in another session, re-scaling may be required. The situation is even more difficult when combining GPS surveys from different eras. A great deal of information concerning receivers used, field and processing procedures, etc., must be available to make an accurate assessment of the manner in which individual VCV matrices are to be scaled, or modified in some way -- SINEX is a development to overcome this. (Some further comments on this problem are given below, illustrated by an example.)
- Variance Factor Tests. Any geodetic adjustment requires some action to be taken when the variance factor test fails. A GPS secondary adjustment is no exception.

What To Do About the Variance Factor?

GPS reduction software invariably produce VCV matrices that imply very high precisions of the estimated parameters. The input of the resulting parameters, and their associated VCV matrices, into a secondary network adjustment will produce a network VCV and variance factor that may also not be realistic. *The statistical testing of the multi-session variance factor will therefore, in general, fail.* The variance factor test is usually performed at the 95% confidence level, requiring the estimated variance factor to be generally between 0.6 and 1.6 (depending on the degrees of freedom of the adjustment -- Figure 9.1-4).

If it is assumed that poor quality baseline observations, and blunders, have been eliminated by careful screening (§9.1) and quality control (§10.3), that the mathematical model is correct, and there are no typing errors in the network adjustment input file, there are two possible reasons why the variance factor test may fail (§9.1):

- ❑ Poor estimate of standard deviations and correlations of the observations.
- ❑ The presence of model or systematic errors.

In the case of primary GPS phase data solutions it is generally the latter (that is, the neglect of unmodelled systematic biases, principally errors in atmospheric refraction or the orbit -- §6.2) that are the problem. By neglecting to account for them, the resulting VCV of parameters will reflect only the phase observation "noise" -- which is only of the order of several millimetres, and the baseline-satellite geometry.

In the case of secondary adjustment of the output of GPS phase reduction software, the "observations" are the baselines, and their associated VCVs. Hence if the variance factor test fails it is generally because the VCV of the observations is unrealistic (for the reasons given above). It is possible to have systematic errors (for example, there may be a systematic shortening of all single frequency baselines), but in a minimally constrained adjustment such systematic errors are only likely to be important if different receiver types (or more particularly, their corresponding primary solutions) and different field techniques (conventional static, compared with "rapid static" or "stop & go" procedures) are mixed together.

What are the options in respect of GPS network adjustments?

- In conventional network adjustments it is common practice to use the *aposteriori* variance factor (VF) to scale the estimated VCV matrix of the parameters (by the VF). Is this desirable for GPS multi-session solutions? **Only if the network is "homogeneous"**, that is, if all GPS results are produced by the same software, processing data collected by the same receivers under similar field conditions (session duration, comparable interstation distances, etc.), and the network has a significant number of redundancies (number of observations >> number of estimated parameters). *The correlations are preserved, only the precisions are changed.*
- Do nothing. An option to be exercised, for example, **if the VF test passes**, or **if there is insufficient redundancy in the network**. This would be the case, for example, if the VF test is carried out at a single session network level (§9.2).
- Derive the VCV matrix from covariance analysis of the phase reduction process. Phase simulation software can be used to propagate realistic systematic biases (such as orbit error, refraction uncertainties, etc.) into the VCV of the estimated parameters. However, this is **not a viable option for GPS surveying because of the specialised software that is required**.
- Modify the observation VCV information. The solution can be iterated, changing the VCVs, until the VF test passes. The manner in which this can be carried out, and the rationale behind such a process, is discussed further in §9.4.

9.3.5 SOLUTION INDEPENDENT EXCHANGE FORMAT: SINEX

Recently, an effort has been made to design a standard output format for solutions. These solutions may be the product of a single baseline reduction program (typically commercial phase reduction software), a multi-baseline / multi-session reduction program (such as possible with GPS scientific software), or secondary network adjustment software. There are several benefits of such a standardised format:

- Secondary adjustment software would no longer need different "front-end" file reading software, one for each brand of phase data reduction software.
- There is a standard archiving format for baseline, session or network solutions.
- GPS networks from different eras can be combined with greater confidence.
- GPS baseline, session or network solutions can be exchanged with a minimum of additional documentation.

The following documentation has been produced to describe the SINEX (Solution INdependent EXchange) format.

```
%=SNX 0.05 NRC 95:123:55260 NRC 95:113:00000 95:120:00000 P 00117 1 X E
*-------------------------------------------------------------------------------
*
* This is an annotated Sinex example, based on the first submission from
* NRC (Natural Resources Canada -- Geodetic Surveys). It has been ammended and extended
* by NCL (Dept. of Surveying, University of Newcastle-upon-Tyne) to illustrate the full
* Sinex 0.05 format. Long blocks have been truncated.
*
* The first and last lines begin '%'. Only '%', '*', '+', '-' and ' ' are
* allowed in the first column, meaning 'begin/end sinex', 'comment',
* 'start block', 'end block' and 'data line' respectively.
*
* Header line explanation:
*
* '='       Solution operator code. '=' means 'resultant' and is the
*           only legal code in a header line. See INPUT/HISTORY notes.
* 'SNX'     This is a Sinex document. Other formats may use similar headers.
* '0.05'    Sinex version number. Version 1.00 will be established after the initial
*           submission period.
* 'NRC 95:122:67080'  The sinex reference for this file. Sinex files are
*           referred to by the three-character agency code, and a creation time-stamp in
*           yy:ddd:sssss format. Agency codes should have entries in
*            INPUT/ACKNOWLEDGEMENTS.
* 'NRC 95:113:00000
*           95:120:00000' The agency responsible for the data, and the overall
*           data time span. COM means multiple agencies.
* 'P'       Technique code. 'P' (GPS) 'L' (SLR) 'R' (VLBI)
*                'C' (multiple) and 'M' (LLR) are allowed.
* 00117     This solution estimates 117 parameters.
* 1         Constraint code. '0' (unconstrained), '1' (loose constraints),
*           '2' (tight constraints) are allowed.
* X E       This solution includes coordinates and erps. 'X', 'E' and 'V' (velocities)
*           are allowed.
*
```

```
*-------------------------------------------------------------------------------
*        1         2         3         4         5         6         7         8
*234567890123456789012345678901234567890123456789012345678901234567890123456789 0
*-------------------------------------------------------------------------------
+FILE/REFERENCE
* This block always contains the following six records
*
 DESCRIPTION   Natural Resources Canada / Geodetic Surveys, altered by NCL
 OUTPUT        NRCan 1995 weekly solution.
 CONTACT       ferland@gdim.geod.emr.ca
 SOFTWARE      combine v0.01
 HARDWARE      HP 750
 INPUT         NRCan daily solution
-FILE/REFERENCE
*-------------------------------------------------------------------------------
+FILE/COMMENT
* This is a free-format block for notes and comments. Substantial remarks
* should go in here, not in * lines.
*
-FILE/COMMENT
*-------------------------------------------------------------------------------
+INPUT/HISTORY
* Each input solution used to create this solution is listed here. A series
* of + lines give inputs to a combination - the = code is used for the
* resultant. A '*' code denotes a similarity transformation of the previous
* solution. THe format is identical to the header line. The last line should
* always refer to this solution, i.e. match the header line.
*
*O_FM VER_ AGY TIME_STAMP__ DAT DATA_START__ DATA_END____ T PARAM C TYPE_
 +SNX 0.04 NRC 95:123:52328 NRC 95:113:00000 95:114:00000 P 00081 1 X E
 +SNX 0.04 NRC 95:123:52590 NRC 95:114:00000 95:115:00000 P 00082 1 X E
 +SNX 0.04 NRC 95:123:52881 NRC 95:115:00000 95:116:00000 P 00082 1 X E
 +SNX 0.04 NRC 95:123:53091 NRC 95:116:00000 95:117:00000 P 00076 1 X E
 +SNX 0.04 NRC 95:123:53365 NRC 95:117:00000 95:118:00000 P 00073 1 X E
 +SNX 0.04 NRC 95:123:53646 NRC 95:118:00000 95:119:00000 P 00079 1 X E
 +SNX 0.04 NRC 95:123:53962 NRC 95:119:00000 95:120:00000 P 00082 1 X E
 +SNX 0.04 NRC 95:121:59613 NRC 95:116:00000 95:117:00000 P 00039 0 X V
 =SNX 0.05 NRC 95:123:55260 NRC 95:113:00000 95:120:00000 P 00117 1 X E
-INPUT/HISTORY
*-------------------------------------------------------------------------------
+INPUT/FILES
* Every sinex file referenced in INPUT/HISTORY should have a filename entered
* here. The last line of this block is always the name of the current file.
* Path names should be given meaningful aliases to keep them short!
*
*AGY TIME_STAMP__ FILE_NAME____________________ DESCRIPTION_____________________
 NRC 95:123:52328 1995/w_798/EMR07980.snx       NRC Daily solution
 NRC 95:123:52590 1995/w_798/EMR07981.snx       NRC Daily solution
 NRC 95:123:52881 1995/w_798/EMR07982.snx       NRC Daily solution
 NRC 95:123:53091 1995/w_798/EMR07983.snx       NRC Daily solution
 NRC 95:123:53365 1995/w_798/EMR07984.snx       NRC Daily solution
 NRC 95:123:53646 1995/w_798/EMR07985.snx       NRC Daily solution
 NRC 95:123:53962 1995/w_798/EMR07986.snx       NRC Daily solution
 NRC 95:121:59613 stacomb_sinex/950426_apr.snx  ITRF93 for 13 stations
 NRC 95:123:55260 stacomb_sinex/EMR07987.snx    Week 798 combination
-INPUT/FILES
*-------------------------------------------------------------------------------
+INPUT/ACKNOWLEDGMENTS
* Each agency three-character code used in any other block is explained here.
*
*AGY DESCRIPTION________________________________________________________________
 NRC Natural Resources Canada, Geodetic surveys
 NCL Newcastle AAC, University of Newcastle upon Tyne, England.
-INPUT/ACKNOWLEDGMENTS
*-------------------------------------------------------------------------------
```

```
+SITE/ID
* Each physical monument is known in Sinex by a four-character site code
* (standardised) and an alphabetic point code (arbitrary). Each CODE+PT
* is equivalent to an IERS DOMES code. Each monument estimated in the solution
* has an entry in this block. Unknown DOMES codes are represented as M or S
* following the IERS convention.
*
*CODE PT DOMES____ T STATION DESCRIPTION___ APPROX_LON_ APPROX_LAT_ _APP_H_
 ALBH  A 40129M003 P Albert Head, Canada    236 30 45.2  48 23 23.3    31.0
 ALGO  A 40104M002 P Algonquin Park, Canada 281 55 43.1  45 57 20.9   200.0
 AREQ  A 42202M005 P Arequipa, Peru         288 30 26.0 -16 27 55.9  2488.0
 DAV1  A 66010M001 P Davis, Antartica        77 58 21.5 -68 34 38.4    96.0
 DRAO  A 40105M002 P Dom. RAdio Obs.,Canada 240 22 30.1  49 19 21.5   541.0
 FAIR  A 40408M001 P Fairbanks, U.S.A.      212 30  2.8  64 58 40.9   319.0
 FORT  A 41602M001 P Fortaleza, Brazil      321 34 27.8  -3 52 38.9    19.0
 GOLD  B 40405S031 P Goldstone, U.S.A.      243  6 38.8  35 25 30.6   986.0
 GUAM  A 50501M002 P Dedego, Guam           144 52  6.2  13 35 21.4   206.0
 KIT3  A 12334M001 P Kitab, Uzbekistan       66 53  7.6  39  8  5.2   622.0
 KOKB  A 40424M004 P Kokee Park,Haw.,U.S.A. 200 20  6.3  22  7 34.6  1167.0
 KOSG  A 13504M003 P Kootwijk, Netherlands    5 48 34.8  52 10 42.4    96.0
 MADR  A 13407S012 P Madrid, Spain          355 45  1.3  40 25 45.0   829.0
 MCM4  A 66001M003 P McMurdo, Antartica     166 40 31.2 -77 50 55.2    -1.0
 NRC1  A         M P NRC, Ottawa, Canada    284 22 30.0  45 27 15.0    82.0
 KERG  A 91201M002 P Kerguelen Is.,Antartic  70 15 19.9 -49 21  5.3    73.0
 RCM5  A 40499S018 P Richmond, Flor. U.S.A. 279 36 57.9  25 36 49.7   -15.0
 SANT  A 41705M003 P Santiago, Chile        289 19 53.2 -33  9  1.1   723.0
 SCHE  A         M P Schefferville, Canada  293  0   .0  55  0   .0   200.0
 STJO  A 40101M001 P St-John's, Canada      307 19 20.2  47 35 42.9   152.0
 TIDB  A 50103M108 P Tidbila, Australia     148 58 48.0 -35 23 57.2   665.0
 TROM  A 10302M003 P Tromso, Norway          18 56 18.0  69 39 45.9   132.0
 TSKB  A 21730S005 P Tuskuba, Japan         140  5 15.0  36  6 20.4    67.0
 WETT  A 14201M009 P Wettzell, Germany       12 52 44.1  49  8 39.3   666.0
 YAR1  A 50107M004 P Yaragadee, Australia   115 20 49.2 -29  2 47.7   241.0
 YELL  A 40127M003 P Yellowknife, Canada    245 31  9.5  62 28 51.3   180.0
 TAIW  A 23601M001 P Taipei, Taiwan         121 32 11.6  25  1 16.8    44.0
 HART  A 30302M002 P Hartebeesthoek, S. A.   27 42 28.0 -25 53 13.6  1555.0
 CHUR  A         M P Churchill, Canada      266  0   .0  59  0   .0      .0
 WILL  A         M P Williams Lake, Canada  237 49 55.9  52 14 12.9  1097.0
-SITE/ID
*-------------------------------------------------------------------------------
+SITE/DATA
* This block contains information on the source of each station. The format
* here has changed from Sinex 0.04. Since point and solution codes are
* arbitrary, the station name (SITE+PT+SOLN codes) may be different in the
* input solution - both are given here. Stations which are estimated in
* multiple input files have several lines here.
* The information here is fictional, to illustrate the format.
* Each station is defined in SOLUTION/EPOCHS, and each file (AGY+TIME_STAMP__)
* appears in INPUT/FILES.
*
*SOLUTION____ INPUT________
*SITE PT SOLN SITE PT SOLN T DATA_START__ DATA_END____ AGY TIME_STAMP__
 ALBH  A    1 ALBH  B    1 P 95:113:00000 95:120:00000 NRC 95:123:52328
 ALBH  A    1 ALBH  A    1 P 95:113:00000 95:120:00000 NRC 95:123:52590
* etc
 ALGO  A    1 ALGO  A    1 P 95:113:00000 95:120:00000 NRC 95:123:52328
* etc
-SITE/DATA
*-------------------------------------------------------------------------------
+SITE/RECEIVER
* Here each station (SITE+PT+SOLN codes) has receiver details attached. If
* receivers change during the data span for that station, multiple lines are
* used here. These data spans must fit within the overall station span
* (given in SOLUTION/EPOCHS) and should cover the entire span for each station.
*
* Note unknown fields filled with - characters. No field is left blank.
```

```
*
*SITE PT SOLN T DATA_START__ DATA_END____ DESCRIPTION_____ S/N__ FIRMWARE_______
 ALBH  A    1 P 95:012:67680 00:000:00000 ROGUE SNR-8000     292 3.0.32.2
 ALGO  A    1 P 94:355:00000 00:000:00000 ROGUE SNR-8000    T226 3.0.32.2
* etc
-SITE/RECEIVER
*-------------------------------------------------------------------------------
+SITE/ANTENNA
*
* Here each station (SITE+PT+SOLN codes) has antenna details attached. If
* antennae change during the data span for that station, multiple lines are
* used here. These data spans must fit within the overall station span
* (given in SOLUTION/EPOCHS) and should cover the entire span for each station.
*
* Note unknown fields filled with - characters. No field is left blank.
*
* In this NRC file, there are relational problems between this block and
* the next (GPS_PHASE_CENTRE). The Description and SN fields here are
* non-unique (e.g. records 17 and 28 below refer to the 'same' antenna,
* the pco of which is given in record 13 of the following block!)
* If this is a general problem (i.e. non-unique antenna SN codes), we may
* need to introduce index numbers in this block.
*
*SITE PT SOLN T DATA_START__ DATA_END____ DESCRIPTION_____ S/N__
 ALBH  A    1 P 95:011:80100 00:000:00000 DORNE-MARGOLIN T   368
 ALGO  A    1 P 94:047:69300 00:000:00000 DORNE-MARGOLIN T   173
* etc
-SITE/ANTENNA
*-------------------------------------------------------------------------------
+SITE/GPS_PHASE_CENTER
*
* Here each antenna (DESCRIPTION + S/N fields) listed in SITE/ANTENNA has phase
* centre details attached.
*
* Note unknown fields filled with - characters. No field is left blank.
*
*                           UP____ NORTH_ EAST__ UP____ NORTH_ EAST__
*DESCRIPTION_____ S/N__ L1->ARP(m)__________ __L2->ARP(m)________ AZ_EL
 DORNE-MARGOLIN B -----  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN B   113  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN B   119  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN R -----  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN R     2  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN R     3  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN R    10  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN R    95  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN R    96  .0780  .0000  .0000  .0960  .0000  .0000 None
 DORNE-MARGOLIN T -----  .1100  .0000  .0000  .1280  .0000  .0000 None
 DORNE-MARGOLIN T   105  .1100  .0000  .0000  .1280  .0000  .0000 None
 DORNE-MARGOLIN T   119  .1100  .0000  .0000  .1280  .0000  .0000 None
 DORNE-MARGOLIN T   148  .1100  .0000  .0000  .1280  .0000  .0000 None
 DORNE-MARGOLIN T   154  .1100  .0000  .0000  .1280  .0000  .0000 None
 DORNE-MARGOLIN T   171  .1100  .0000  .0000  .1280  .0000  .0000 None
 DORNE-MARGOLIN T   172  .1100  .0000  .0000  .1280  .0000  .0000 None
-SITE/GPS_PHASE_CENTER
*-------------------------------------------------------------------------------
+SITE/ECCENTRICITY
*
* Here each station (SITE+PT+SOLN codes) has eccentricity vectors attached. If
* these change during the data span for that station, multiple lines are
* used here. These data spans must fit within the overall station span
* (given in SOLUTION/EPOCHS) and should cover the entire span for each station.
*
*                                             UP______ NORTH___ EAST____
*SITE PT SOLN T DATA_START__ DATA_END____ AXE ARP->BENCHMARK(m)_________
 ALBH  A    1 P 95:011:80100 00:000:00000 UNE   .1000    .0000    .0000
```

```
 ALGO  A    1 P 94:139:00000 00:000:00000 UNE    .1000    .0000    .0000
* etc
-SITE/ECCENTRICITY
*-------------------------------------------------------------------------------
+SOLUTION/EPOCHS
*
* This block is the logical starting-point for interpreting the file, since it
* defines the stations in the solution. A station is particular solution for
* a monument, referenced by SITE, PT and SOLN codes. Multiple integer solution
* codes may be used (arbitrarily) to give multiple solutions for a point in the
* same estimate - at different epochs, for instance.
*
* Each station invoked here should have one or more entries in each of
* SITE/RECEIVER, SITE/ANTENNA, SITE/DATA and SITE/ECCENTRICITY.
* The monument (SITE+PT) should be defined in SITE/ID.
*
*SITE PT SOLN T _DATA_START_ __DATA_END__ _MEAN_EPOCH_
 ALBH  A    1 P 95:113:00000 95:120:00000 95:116:43200
 ALGO  A    1 P 95:113:00000 95:120:00000 95:116:43200
* etc
-SOLUTION/EPOCHS
*-------------------------------------------------------------------------------
+SOLUTION/ESTIMATE
*
* The parameter estimates are written here. Parameter types STAX, STAY, STAZ,
* VELX, VELY, VELZ (coordinate and velocity x, y, z) are followed by a
* station reference. Erp types LOD, UT, XPO, YPO have no station. The
* constraint code (0, 1 or 2) is given here for each parameter - the empty
* fields are filled with a data-not-given character (-).  Units:
*         XPO  mas   (milli-arc seconds)    X POLE MOTION
*         YPO  mas   (milli-arc seconds)    Y POLE MOTION
*         XPOR ma/s  (milli-arc seconds/s)  X POLE RATE
*         YPOR ma/s  (milli-arc seconds/s)  Y POLE RATE
*         UT   ms    (milli-seconds)        UT1-UTC
*         LOD  ms    (milli-seconds)        UT1-UTC RATE
*
*INDEX TYPE CODE PT SOLN _REF_EPOCH__ UNIT S __ESTIMATED VALUE____ _STD_DEV______
     1 STAX ALBH  A    1 95:116:43200 m    2 -.234133292758691E+7 .1845776905E-2
     2 STAY ALBH  A    1 95:116:43200 m    2 -.353904953122971E+7 .1890911219E-2
     3 STAZ ALBH  A    1 95:116:43200 m    2 .4745791466277621E+7 .2075918818E-2
     4 STAX ALGO  A    1 95:116:43200 m    1 .9181294929904674E+6 .1768625434E-2
     5 STAY ALGO  A    1 95:116:43200 m    1 -.434607120901217E+7 .1797731140E-2
     6 STAZ ALGO  A    1 95:116:43200 m    1 .4561977840428489E+7 .1878956710E-2
     7 STAX AREQ  A    1 95:116:28800 m    2 .1942826687525561E+7 .6477347581E-2
     8 STAY AREQ  A    1 95:116:28800 m    2 -.580407019776578E+7 .8829387992E-2
     9 STAZ AREQ  A    1 95:116:28800 m    2 -.179689395509440E+7 .3872643191E-2
    10 STAX CHUR  A    1 95:119:00000 m    2 -.236438707221352E+6 .2190659023E-2
    11 STAY CHUR  A    1 95:119:00000 m    2 -.330761674613259E+7 .2499980216E-2
    12 STAZ CHUR  A    1 95:119:00000 m    2 .5430049170384845E+7 .3338507720E-2
    13 STAX DAV1  A    1 95:113:43200 m    2 .4868545524273632E+6 .5143560198E-2
    14 STAY DAV1  A    1 95:113:43200 m    2 .2285099364466271E+7 .5465295936E-2
    15 STAZ DAV1  A    1 95:113:43200 m    2 -.591495576584752E+7 .8718856714E-2
    16 STAX DRAO  A    1 95:116:43200 m    2 -.205916467723249E+7 .1818058836E-2
    17 STAY DRAO  A    1 95:116:43200 m    2 -.362110834605865E+7 .1859042797E-2
    18 STAZ DRAO  A    1 95:116:43200 m    2 .4814432386809346E+7 .2053716886E-2
    19 STAX FAIR  A    1 95:116:43200 m    0 -.228162142409438E+7 .2008781156E-2
    20 STAY FAIR  A    1 95:116:43200 m    0 -.145359574941003E+7 .2100198773E-2
    21 STAZ FAIR  A    1 95:116:43200 m    0 .5756961936406008E+7 .2509140868E-2
    22 STAX FORT  A    1 95:115:21600 m    2 .4985386578502384E+7 .1084655739E-1
    23 STAY FORT  A    1 95:115:21600 m    2 -.395499854274894E+7 .9229132687E-2
    24 STAZ FORT  A    1 95:115:21600 m    2 -.428426474252779E+6 .2879426693E-2
    25 STAX GOLD  B    1 95:116:43200 m    1 -.235361417310070E+7 .2060113719E-2
    26 STAY GOLD  B    1 95:116:43200 m    1 -.464138536535744E+7 .2135362677E-2
    27 STAZ GOLD  B    1 95:116:43200 m    1 .3676976474604919E+7 .2151652969E-2
    28 STAX GUAM  A    1 95:116:43200 m    2 -.507131279252173E+7 .3359775222E-2
    29 STAY GUAM  A    1 95:116:43200 m    2 .3568363515536474E+7 .3434317006E-2
```

```
    30 STAZ GUAM  A     1 95:116:43200 m     2 .1488904271291384E+7 .2465369172E-2
    31 STAX HART  A     1 95:117:28800 m     0 .5084625439996016E+7 .3117434937E-2
    32 STAY HART  A     1 95:117:28800 m     0 .2670366550990838E+7 .2988406484E-2
    33 STAZ HART  A     1 95:117:28800 m     1 -.276849396332954E+7 .2436794798E-2
    34 STAX KERG  A     1 95:116:43200 m     2 .1406337354635808E+7 .3228912790E-2
    35 STAY KERG  A     1 95:116:43200 m     2 .3918161143630010E+7 .3090251696E-2
    36 STAZ KERG  A     1 95:116:43200 m     2 -.481616739541420E+7 .2887894904E-2
    37 STAX KIT3  A     1 95:116:00000 m     2 .1944945408967126E+7 .3880638482E-2
    38 STAY KIT3  A     1 95:116:00000 m     2 .4556652228809900E+7 .4395018442E-2
    39 STAZ KIT3  A     1 95:116:00000 m     2 .4004325952269760E+7 .4075488675E-2
* etc
    91 LOD  ---- --     1 95:113:43200 ms    2 .2871055744817214E+1 .1212729066E-1
    92 LOD  ---- --     2 95:114:43200 ms    2 .2959652540110830E+1 .1131045636E-1
    93 LOD  ---- --     3 95:115:43200 ms    2 .2973492029661421E+1 .1201761141E-1
* etc
    98 UT   ---- --     1 95:114:43200 ms    2 .8722024405764063E+2 .1318171995E-1
    99 UT   ---- --     2 95:115:43200 ms    2 .8430515991559695E+2 .1575504990E-1
   100 UT   ---- --     3 95:116:43200 ms    2 .8136510032786199E+2 .1745336183E-1
* etc
   104 XPO  ---- --     1 95:113:43200 mas   2 .1029608387361842E+3 .7876117908E-1
   105 XPO  ---- --     2 95:114:43200 mas   2 .1069725602672064E+3 .7569313943E-1
   106 XPO  ---- --     3 95:115:43200 mas   2 .1113899879374726E+3 .7622913583E-1
* etc
   111 YPO  ---- --     1 95:113:43200 mas   2 .5530926512007116E+3 .9289514438E-1
   112 YPO  ---- --     2 95:114:43200 mas   2 .5521110887312243E+3 .8795571011E-1
   113 YPO  ---- --     3 95:115:43200 mas   2 .5512599272862197E+3 .8666021180E-1
* etc
-SOLUTION/ESTIMATE
*-------------------------------------------------------------------------------
+SOLUTION/APRIORI
*
* The same format as the previous block, but parameters given, and their
* order, can be different.
*
* ITRF93(1995.318) coord. constraints for the 13 stations applied (ITRF SSC+
* SSV sigmas used, responsible for correlation in APRIORI matrix)
*
*INDEX TYPE CODE PT SOLN REF_EPOCH___ UNIT S __ESTIMATED VALUE___ _STD_DEV______
     1 STAX ALGO  A    1 95:116:43200 m    2  .91812950316301E+06 .300264598E-02
     2 STAY ALGO  A    1 95:116:43200 m    2 -.43460712286616E+07 .300413333E-02
     3 STAZ ALGO  A    1 95:116:43200 m    2  .45619778480795E+07 .300413333E-02
     4 STAX FAIR  A    1 95:116:43200 m    2 -.22816214309794E+07 .300148865E-02
     5 STAY FAIR  A    1 95:116:43200 m    2 -.14535957605986E+07 .300264598E-02
     6 STAZ FAIR  A    1 95:116:43200 m    2  .57569619418178E+07 .300264598E-02
* etc
    40 VELX ALGO  A    1 95:116:43200 m/y  2 -.21700000000000E-01 .400000000E-03
    41 VELY ALGO  A    1 95:116:43200 m/y  2 -.21000000000000E-02 .500000000E-03
    42 VELZ ALGO  A    1 95:116:43200 m/y  2  .66000000000000E-02 .500000000E-03
* etc
-SOLUTION/APRIORI
*-------------------------------------------------------------------------------
+SOLUTION/MATRIX_ESTIMATE L CORR
* Lower triangular correlation matrix elements, referenced by two parameter
* index numbers from SOLUTION/ESTIMATE, are given here. Elements not given are
* taken from the identity matrix.
*
*PARA1 PARA2 ____PARA2+0_________ ____PARA2+1_________ ____PARA2+2_________
     1     1  .10000000000000E+01
     2     1  .29425156137028E-01  .10000000000000E+01
     3     1 -.26004023015799E+00 -.28793741628090E+00  .10000000000000E+01
     4     1  .54687523972711E+00 -.65452884277258E-01 -.43204669126067E-01
     4     4  .10000000000000E+01
     5     1 -.39819911260492E-01  .40433299131131E+00  .11136939059432E+00
     5     4 -.50106659384620E-02  .10000000000000E+01
     6     1 -.35321953180784E-01  .14019789950983E+00  .42813765846990E+00
     6     4  .49918344878102E-01 -.26545148121353E+00  .10000000000000E+01
```

```
      7     1  .69393516209264E-01 -.50927698251188E-01 -.18770770478848E-02
      7     4  .11249919556344E+00  .36417010197355E-01 -.96512409551535E-02
*..................2290 lines deleted .....................................
    117     1  .39554853137016E-02 -.14701327842692E+00 -.21849261652602E+00
    117     4  .61854528268407E-01 -.15582766664907E+00 -.35595192513222E+00
    117     7 -.14839162033233E-01  .10088643393429E-01 -.34620521010973E+00
    117    10  .24473678103268E-01 -.17719246220580E+00 -.11471841650685E+00
    117    13  .13124059834379E-01  .19447280189975E+00  .55372167099218E-01
    117    16  .11834483725591E-01 -.15394348548100E+00 -.23577099430936E+00
    117    19 -.69920935880248E-02 -.15152016672011E+00 -.15707791358241E-01
    117    22  .56262209482092E-02  .44664432377454E-01 -.29206116099859E+00
    117    25  .13595733163360E-01 -.36791832813222E-01 -.34302028599229E+00
    117    28  .49976730750360E-01 -.52603056350601E-02  .35995661727203E+00
    117    31  .58434167118809E-01  .79162912401483E-01  .19036198391000E+00
    117    34  .18742319369700E-01  .29877748842137E+00  .27423619442271E+00
    117    37  .37293677828651E-01 -.26370207939877E-01  .28314439940865E+00
    117    40 -.15793512947600E-01  .16902050747741E-01 -.68991757800683E-01
    117    43  .13839204459350E-01 -.90319568728349E-01  .87241280008327E-01
    117    46  .17373506530565E-02  .75995410738772E-02 -.22097307167853E-01
    117    49 -.31558372465614E-02  .33086324692108E+00 -.21155116202680E-01
    117    52  .52662048663150E-01 -.96698061270683E-01 -.38697065241955E+00
    117    55  .17854027402451E-01  .45413441850211E-01 -.39199818754031E+00
    117    58 -.20557084589204E-01  .17772427521895E+00 -.38888132782579E+00
    117    61  .51717169153988E-01 -.12774683850231E+00 -.18640329670510E+00
    117    64  .60566764029226E-01 -.96756801702711E-01 -.31097880769930E+00
    117    67  .15946573701026E-01 -.52959490859687E-01  .43877479729634E+00
    117    70 -.56965234511833E-01  .12258278357446E+00  .20637613018780E+00
    117    73  .10379256768073E-01 -.13337986646258E+00  .34656771257246E-01
    117    76  .45650296192972E-01 -.14097331082148E+00  .31904369696173E+00
    117    79 -.15366288987877E-01 -.76872083582857E-01  .14681131961463E+00
    117    82 -.71521817107470E-02 -.14261694699670E+00 -.12950060100137E+00
    117    85 -.46279787222485E-01  .17675785161574E+00  .43321519299327E+00
    117    88  .17270566844414E-02 -.17956461690783E+00 -.95965022007633E-01
    117    91  .67279754048700E-02  .17105290276520E-01  .96879240353699E-02
    117    94  .46879651728024E-02  .15162522246501E-02  .16405537485657E-01
    117    97 -.33487318865767E-01 -.15479070213606E-01 -.17088921255237E-01
    117   100 -.10651453627699E-01 -.97673708447794E-02 -.44763578708339E-02
    117   103 -.10868132563959E-01  .55392984612278E-01  .38519367197329E-01
    117   106  .56288295168491E-01  .61044503649448E-01  .62180451664303E-01
    117   109  .48458825773687E-01  .64903577511731E-01  .46603628193206E+00
    117   112  .50888540957265E+00  .49512938569870E+00  .49390806153475E+00
    117   115  .49291348118876E+00  .51840633003163E+00  .10000000000000E+01
-SOLUTION/MATRIX_ESTIMATE L CORR
*-------------------------------------------------------------------------------
+SOLUTION/MATRIX_APRIORI L CORR
*
* Same format as SOLUTION/MATRIX_ESTIMATE, but a priori values. Again, elements
* not given are taken from the identity matrix.
*
* Correlations due to velocity sigma propagation
*
*PARA1 PARA2 ____PARA2+0_________ ____PARA2+1_________ ____PARA2+2_________
      1     1  .10000000000000E+01
      2     2  .10000000000000E+01
      3     3  .10000000000000E+01
* etc
     40     1  .41972113240929E-01
     40    40  .10000000000000E+01
     41     2  .52439166139985E-01
* etc
-SOLUTION/MATRIX_APRIORI L CORR
*-------------------------------------------------------------------------------
%ENDSNX
```

9.4

MODIFYING THE STOCHASTIC MODEL

To carry out a reliable multi-session GPS network adjustment, the input VCV matrix of the baseline component "observations" must be correct. There are a number of reasons why the input VCV matrix may need to be modified, including:

- To empirically alter the VCV information in such a way as to better reflect the true GPS accuracy.
- To alter the VCV matrix in such a way as to simulate a rigorous multi-baseline adjustment.
- To "standardise" the VCV matrix so as to account for between-epoch correlation.

This chapter discusses each of these topics.

9.4.1 REALISTIC VCV MATRICES FOR SECONDARY GPS ADJUSTMENTS

As has been mentioned several times, the output VCV matrices of phase data processing software are unlikely to be the appropriate ones to input into a secondary network adjustment. *What practical methods can be employed to produce a VCV matrix that is representative of the external factors influencing the GPS results, as well as the internal "noise only" contribution?*

One method, known as Variance Component Estimation, is described in CASPARY (1987, pp.97-110). It has been used to combine GPS and terrestrial networks, and can be applied to both homogeneous and heterogeneous sets of observables in order to determine the relative uncertainties of the various components that make up the total stochastic model. In GPS session observations, the components could be a constant error plus an error that is distance dependent. In the case of heterogeneous observations, each observational group may be assigned its own error characteristics.

The VCE method uses an iterative numerical procedure starting with reasonable *apriori* estimates of the variance components. It relies on the assumption that all significant errors are randomly represented in the data set, and that a single scale factor is sufficient to account for the errors in one observation type, particularly where observations may seem consistent and yet still be affected by undetected systematic errors. This is usually true when dealing with the propagation of errors into new observation types such as is the case in GPS. Therefore it is recommended that methods such as VCE be accompanied by attempts to evaluate the effects of systematic errors.

An alternative, empirical method of modifying VCVs, uses the knowledge that session errors are likely to have both a constant and length dependent component (§2.4 and §10.2), and to apply suitable factors to the VCV matrix of the baseline vectors. Each GPS network (single session, or even individual baseline) is then combined, after altering their VCV, so as to ensure that the *aposteriori* variance factor of the combined multi-session adjustment passes the Variance Factor Test (§9.1).

In modifying the VCV of a GPS phase data reduction solution, it is necessary to take into account both the "internal" errors as well as the "external" (largely undetectable) influences. Because GPS is a new surveying technology, there is little previous experience to use as a reference for the expected precision or accuracy of each type of receiver or software reduction package (that provided by the manufacturer, see §4.2, is likely to be very sensitive to many factors such as length of observation session, satellites tracked, etc.). The internal precision of a GPS relative positioning solution could in fact be assessed by its aposteriori variance factor, as it is derived directly from the residuals and will reflect the effects of receiver errors such as measurement noise, and multipath. The VCV matrix could then be scaled through *multiplication by the aposteriori variance factor*, the solution re-run and the new VF tested. *However, scaling the VCV matrix by the aposteriori variance factor will produce estimates of the VCV that only reflect these "internal" error sources.*

The above two approaches: (a) modification of individual elements of the VCV matrix, and (b) scaling by the aposteriori variance factor; are common options implemented in commercial network adjustment software. These are discussed further, in this chapter.

A popular approach to describing the total error (that is, combined "internal" and "external" errors) in GPS session adjustment results is by constant (a) and length dependent (b) terms, and to develop a suitable population variance such as (eqn (9.1-4)):

$$s^2 = (a + b.L)^2 \tag{9.4-1}$$

where L is the interstation distance in kilometres, a is in millimetres and b is in parts per million (ppm).

The constant term can be justified, for example, **as a centring (horizontal effect) or height-of-antenna measurement (vertical effect) uncertainty.** On the other hand, **residual ionospheric and tropospheric errors in the carrier phase observations, as well as the effect of satellite orbit and fixed station errors, map into the positioning results with magnitudes that grow with increasing baseline length.**

The effect of external systematic biases on GPS phase data reductions can only be detected through their consistency as compared to other data. This is sometimes rather difficult, as comparing GPS results to existing terrestrial networks will show differences that may be due to errors in the terrestrial network. *A more appropriate method could be to compare repeated session "measurements" of GPS baselines at different times of the day, different times of the year for consistency.* However, care must be taken that other effects on baseline/session accuracy are isolated from these analyses (such as radically different baseline lengths). Note, systematic errors that effect all session results with a similar magnitude will not be detected using this technique.

(A) An Empirical Method of Determining Realistic Variances for GPS Baseline Solutions

The simplest method is to simply assign "reasonable" values for a and b, and to construct the s^2 quantities using eqn (9.4-1). These values are then used to modify the VCV matrix (using any of the methods described below), the adjustment is then re-run and the Variance Factor Test is re-applied. If the test still fails, then the values for a and/or b are changed, and the process repeated until the Variance Factor Test is finally passed. There are, however, several difficulties with this approach:

- Which quantity to vary? *a or b?*
- What if there were a variety of baseline lengths? *Varying the quantities could have unexpected effects.*
- What if a mixed set of GPS hardware was used (single or dual-frequency)?
- What if several types of field procedures were used (conventional static, "stop & go", "rapid static" techniques)?
- What if there was a mixture of ambiguity-free and ambiguity-fixed solutions?

In effect the values of a and b are functions of all of the above (and more)! Given these difficulties, a common method has been to estimate the values of s^2 directly (rather than its a and b components).

An alternative method is to estimate a scale factor for particular subsets of baselines sharing similar solution characteristics that may cause systematic differences in GPS accuracy, for example, according to whether they are produced using the same software, the baseline lengths are similar, etc. It must be emphasised that this is a more sophisticated method of estimating scale factors than simply using the aposteriori variance factor in an iterative process. This is discussed below.

(B) An Empirical Method of Determining VCV Scale Factors for GPS Baseline Solutions

The method makes use of repeat baseline results. The following procedure may be used (based on an approach suggested by JONES, 1995):

- Baselines which have been measured more than once are grouped according to baseline length, for example: 1-5km, 5-10km, 10-15km, 15-25km, 25-50km, >50km.
- For each baseline, associated with a pair of stations k-j, the mean values of the baseline component estimates are determined $B^m x_{kj}$, $B^m y_{kj}$, $B^m z_{kj}$, together with the associated mean variance terms $\sigma_m{}^2{}_{x_{kj}}$, $\sigma_m{}^2{}_{y_{kj}}$, $\sigma_m{}^2{}_{z_{kj}}$ (computed from applying the Law of Propagation of Variances to the individual variance terms $\sigma_i{}^2{}_{x_{kj}}$, $\sigma_i{}^2{}_{y_{kj}}$, $\sigma_i{}^2{}_{z_{kj}}$ from the diagonal elements of the VCV matrices).
- Compute the component differences for each individual baseline estimate i with respect to the mean values: $\Delta B^i x_{kj} = B^m x_{kj} - B^i x_{kj}$, $\Delta B^i y_{kj} = B^m y_{kj} - B^i y_{kj}$, $\Delta B^i z_{kj} = B^m z_{kj} - B^i z_{kj}$.
- Compute the corresponding variance factors for the component differences of the individual baseline estimate i with respect to the mean values (applying the Law of Propagation of Variances): $\sigma_i{}^2{}_{Dx_{kj}} = \sigma_i{}^2{}_{x_{kj}} + \sigma_m{}^2{}_{x_{kj}}$, $\sigma_i{}^2{}_{Dy_{kj}} = \sigma_i{}^2{}_{y_{kj}} + \sigma_m{}^2{}_{y_{kj}}$, $\sigma_i{}^2{}_{Dz_{kj}} = \sigma_i{}^2{}_{z_{kj}} + \sigma_m{}^2{}_{z_{kj}}$.
- The "standardised" differences for each baseline k-j (baseline estimate i) are computed by dividing the component differences by the formal error estimates (standard deviations):

$$DB^ix_{kj} = \frac{\Delta B^ix_{kj}}{\sigma_{iDx_{kj}}},\ DBy_{kj} = \frac{\Delta B^iy_{kj}}{\sigma_{iDx_{kj}}},\ DB^iz_{kj} = \frac{\Delta B^iz_{kj}}{\sigma_{iDx_{kj}}}.$$

- For each baseline group an *average* variance factor is computed:

$$s^2 = [\ \Sigma(\ DB^ix_{kj}\ )^2 + \Sigma(\ DB^ix_{kj}\ )^2 + \Sigma(\ DB^ix_{kj}\ )^2\ ]\ /\ (3.m) \qquad (9.4\text{-}2)$$

where m is the number of baselines in the group (the summation being taken over all baselines k-j, each observed a minimum of two times).

The size of the scale factors is dependent on the instrumentation, processing software and observing conditions. Examples of scale factors determined, as a function of baseline length, are (IBID, 1995):

Leica/SKI:

- 0-5km	--	400	(Tropical (1995) and Mediterranean (1992) climate)
- 5-10km	--	800	(Tropical (1995) and Mediterranean (1992) climate)
- 10-20km	--	800	(Mediterranean (1992) climate)

Trimble/GPSurvey:

- 0-2km	--	25	("rapid static": 5-20 minutes)
- 0-2km	--	49	(conventional static: 180 minutes)
- 2-4km	--	36	("rapid static": 5-20 minutes)
- 2-4km	--	49	(conventional static: 180 minutes)
- 4-8km	--	81	(conventional static: 180 minutes)
- 8-15km	--	121	(conventional static: 180 minutes)
- 15-30km	--	225	(conventional static: 180 minutes)

A similar procedure may be used for any group of baselines determined using a particular hardware/software system. Though care must be taken to ensure that they are truly representative of different observing environments.

Using Empirically Derived Variances to Alter the VCV Matrices

The use of empirically derived variance information to modify the GPS VCV matrices is now an accepted practice when combining individual baseline solutions, single session solutions or multi-session solutions.

Generally the values of a range from about 5-40mm and the values of b range from 2-10ppm, depending on the type quality of the initial GPS reduction. Some additional remarks regarding the way in which variances are derived from the values of a and b:

- The value of s^2 adopted for a certain baseline is the same for all three baseline components: Bx, By, Bz.

- The same a quantity may be assumed to apply to all three baseline components, or a different value used for each component (a_X, a_Y, a_Z), or the VCV_{XYZ} matrix may be transformed into the topocentric form VCV_{ENH} (see §11.1) and a different value assumed for the horizontal components (a_E, a_N) (arising from centring error) than for the vertical component (arising from height-of-antenna error) (a_H).

- The value of b may be assumed to relate to the baseline length and is to be distributed according to the baseline component lengths, in which case the variances are approximated by:

$$\begin{aligned} s_X{}^2 &= (a_X{}^2 + b.Bx)^2 \\ s_Y{}^2 &= (a_Y{}^2 + b.By)^2 \\ s_Z{}^2 &= (a_Z{}^2 + b.Bz)^2 \end{aligned} \qquad (9.4\text{-}3)$$

- The value of b may be assumed to relate to the baseline length and is to be distributed according to the estimated baseline component standard deviations, in which case the variances are approximated by:

$$\begin{aligned} s_X{}^2 &= (a_X{}^2 + b.\frac{\sigma_{Bx}}{\sigma}.L)^2 \\ s_Y{}^2 &= (a_Y{}^2 + b.\frac{\sigma_{By}}{\sigma}.L)^2 \\ s_Z{}^2 &= (a_Z{}^2 + b.\frac{\sigma_{Bz}}{\sigma}.L)^2 \end{aligned} \qquad (9.4\text{-}4)$$

  where σ_{Bx}, σ_{By}, σ_{Bz} are the component standard deviations and σ is the baseline length standard deviation.

- The VCV_{XYZ} matrix may be transformed into the topocentric form VCV_{ENH} and a different value of b assumed for the horizontal and vertical components (for example, to account for the fact that the height determination in GPS is weaker than the determination of the horizontal components):

$$\begin{aligned} s_E{}^2 &= (a_E{}^2 + b_E.L)^2 \\ s_N{}^2 &= (a_N{}^2 + b_N.L)^2 \\ s_H{}^2 &= (a_H{}^2 + b_H.L)^2 \end{aligned} \qquad (9.4\text{-}5)$$

  Variations based on eqns (9.4-3) and (9.4-4) are also possible.

Note that for short baselines, the influence of the constant component a is the strongest, while in the case of long baselines the length dependent component b has a greater influence than the constant part.

Although there are several strategies for computing the variances (see above), there are also several options for how they are used to modify the original VCV matrix:

(1) The computed VCV matrix of the initial baseline phase data reduction is ignored, and replaced by a diagonal VCV matrix containing only the empirically derived variance s^2. There are no covariance terms, hence *the correlations between baseline components have been changed.*

(2) Preserve the computed VCV matrix, and add the appropriate variance quantity s^2 to the diagonal elements of the matrix. Maintaining the original covariance terms means that *the correlations between baseline components have been changed.*

(3) Construct a VCV matrix using the new variance information s^2 and the original correlation information. The procedure would be:

- Compute the coefficients of correlation using the original VCV information:

$$r_{xy} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}, \; r_{xz} = \frac{\sigma_{XZ}}{\sigma_X\sigma_Z}, \; r_{yz} = \frac{\sigma_{YZ}}{\sigma_Y\sigma_Z} \tag{9.4-6}$$

- Determine new values for the variances s_X^2, s_Y^2, s_Z^2.
- Construct the new VCV matrix using eqn (9.1-7):

$$VCV^c = \begin{bmatrix} s_X^2 & r_{xy}s_Xs_Y & r_{xz}s_Xs_Z \\ \ldots & s_Y^2 & r_{yz}s_Xs_Z \\ \text{symp} & \ldots & s_Z^2 \end{bmatrix} \tag{9.4-7}$$

Option (3) is the preferred option.

Example of a Multi-Session Network Adjustment

The Molong GPS survey referred to earlier may be used as an example. Seven observation sessions, in which four receivers were used, were reduced independently (an example was given in §9.2). Although only Trimble GPS receivers were used, the observations were made over a two day period by the same field parties and all ambiguities were resolved, there were a number of factors that made deriving a multi-session network solution far from a straightforward process:

(1) Different single session solutions (Figure 9.3-3):

- Sessions A, B, C, and F were reduced as single baselines using the TRIMVEC™ software, and all the baselines (trivial and independent) were included in the secondary network adjustment.
- Sessions D and E were reduced together in a rigorous multi-station, multi-session solution using GAMIT, a sophisticated scientific program designed for GPS geodesy applications.
- Session G was reduced in multi-baseline mode using the TRIMMBP™ software.

(2) Varying length of baselines. The baselines varied in length from less than one kilometre to over 30 kilometres. The precisions derived from the various VCVs did not reflect any length dependence.

(3) Antenna station setup varied from being tripod mounted to using geodetic pillars. The precision of centring and/or antenna height measurement is expected to be uneven.

Using the various single baseline, multi-baseline and multi-station/session solution outputs, with their original computed VCVs, in a secondary multi-session network adjustment caused the Variance Factor Test to fail at the 95% confidence level (with 31 degrees of freedom). After some experimentation the stochastic information associated with the baseline observations was modified according to the following:

- The observations were divided into two groups: (1) the results of session A, B, C, F, G processing, and (2) session D and E processing.

- Stations occupied in sessions D and E were almost all existing geodetic pillars.
- The stochastic model for the observations was altered by adding to the diagonal elements of the computed VCV matrices the variance developed from eqn (9.4-5), in which a=10mm and b=5ppm for group (1) baselines; a=5mm and b=3ppm for group (2) baselines.

The Variance Factor Test was then successfully passed. The relative 2-D error ellipses connecting 6 of the 12 stations in the network are shown in Figure 9.1-3. The error figures are almost circular, and range in size from about 5 to 10mm. It should be pointed out that the magnitude of the semi-major axes of the error ellipses is dependent on the output VCV of the multi-session adjustment, and that this aposteriori VCV then is empirically altered leading to a new input VCV matrix for an iterated solution. However, the amount by which the VCV is changed is determined by the variance factor, which itself is influenced by the amount of redundancy in the overall network! Which brings up an important issue: **should trivial baselines be included in a network adjustment?** This is discussed below.

9.4.2 MULTI-BASELINE VCV MATRICES FROM SINGLE BASELINE REDUCTIONS

Trivial Baselines in a Network Adjustment?

In the above example, "quasi-independent" baselines from sessions A, B, C and F were included within the adjustment. There are a total of 32 baselines, of which only 20 can be classed as independent. If only the 20 independent baselines were included in the multi-session adjustment, and the iterative process of modifying the observation VCVs described above was used, it was found that the appropriate value of b that leads to a satisfactory variance factor is 5ppm for all baselines (the value of a was unchanged). The most noticeable effect is now in the size of the resulting error ellipses, some are up to 50% larger (Figure 9.4-1).

The main argument *against* incorporation of all baselines (independent and trivial) into a network adjustment is that the resulting solution statistics are "over-optimistic", that is the length of the axes of the error ellipses (or ellipsoids) are smaller than when only the independent baselines are used. Although there is at present by no means a total consensus on whether to include, or not to include, trivial baselines in any adjustment, there is strong support for including so-called "quasi-independent" baselines in network adjustment, for the following reasons:

(1) It can be argued that all baselines reduced separately (and particularly if an ambiguity-fixed solution is obtained) are in fact independent.

(2) Including all baselines from a session, if reduced separately, would (partly) account for the between baseline correlations that have been ignored at the single baseline reduction step.

(3) Increasing the number of observations, and hence the redundancies (or degrees of freedom), permits more reliable statistical testing.

(4) The use of all baselines, independent and quasi-independent, removes some of the arbitrariness involved in the selection of only independent baselines (different sets of independent baselines may lead to differing results). *The outcome is a unique solution.*

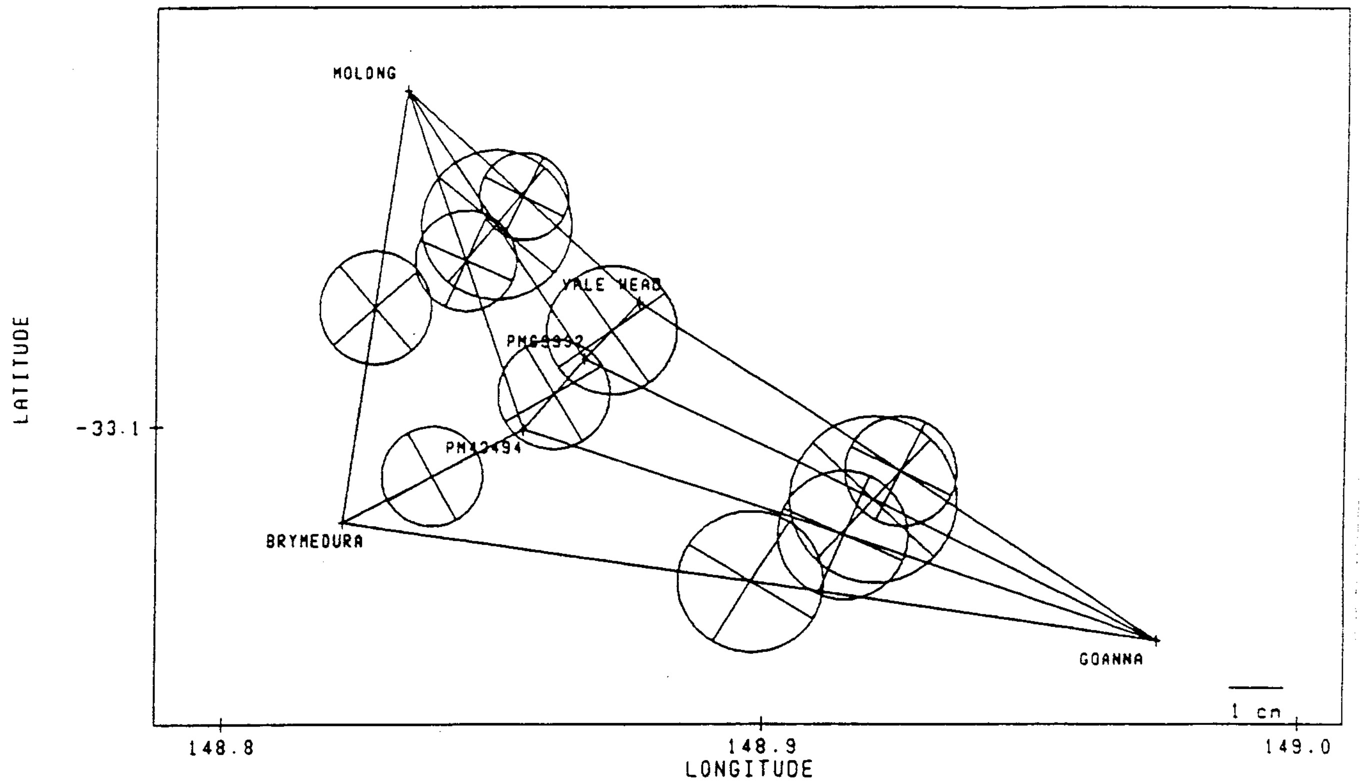


Figure 9.4-1. Example of relative 2-D error ellipses from a network adjustment involving only independent baselines (compare with Figure 9.1-3).

A Procedure for Incorporating Trivial Baselines into a Secondary Adjustment

CRAYMER & BECK (1992) have demonstrated that by including <u>all</u> baselines (reduced in single baseline mode), trivial and independent, in a network adjustment, such a secondary adjustment of baselines is *close to being equivalent to a rigorous simultaneous multi-baseline phase reduction in the case of a single session adjustment,* or a simultaneous multi-session phase reduction in the campaign network adjustment case. However, there are several conditions that have to be met:

- Include all single baseline results.
- Scale baseline VCVs by $\frac{n}{2}$, where n is the number of receivers operating during the session.
- The same ambiguities are resolved in single baselines as would have been for the multi-baseline processing.

This certainly confirms the intuitive arguments presented earlier that including all baseline somehow "reconstructs" the missing between-baseline correlation

information. The scaling of VCVs by the $\frac{n}{2}$ factor would overcome the problem of over-optimistic network VCVs.

A similar result has been verified by HAN & RIZOS (1995b), except that they suggest that it is the cofactor matrices that must be scaled by n/2, not the VCV matrices. Then each (scaled) co-factor matrix, for each baseline solution, is converted into the corresponding VCV matrix upon multiplication of a single session variance factor. This session variance factor is the mean of all the variance factors. The procedure for scaling the VCV matrices also includes the additional "standardisation" factor arising from the need to take into account the physical correlations between successive epochs of GPS observations (see below).

9.4.3 "STANDARDISATION" OF VCV MATRICES TO ACCOUNT FOR BETWEEN-EPOCH PHYSICAL CORRELATIONS

In §7.2 it was mentioned that in fact strong between-epoch physical correlations have been noticed in GPS observations. The higher the data rate, the stronger the correlation (eqn (7.2-14)). These correlations should be included in the input VCV matrix of the phase observations used in the Least Squares adjustment of double-differenced phase observations in a baseline reduction. However, no commercial software takes this correlation into account, hence the output VCV matrix of the baseline component and ambiguity parameters is incorrect. *The effect on the VCV matrix can be accounted for by a scale factor.*

The fact that neglecting between-epoch correlations simply leads to an incorrectly scaled VCV matrix means that there has been no need to hitherto tackle this problem. During secondary network adjustment, failure of the Variance Factor Test is usually addressed by scaling the VCV matrix factor and repeating the solution. In this way, the scale factor (the aposteriori variance factor) is *empirically* determined, and accounts for a number of effects including the neglected between-epoch physical correlation.

There would therefore have been no need to explicitly consider the between-epoch correlation except for changes in the way GPS surveys are carried out has made the issue more than of passing academic interest. *The scale factor is a function of the length of the observation session and the data rate.* In conventional GPS surveying the observation session lengths are relatively long (30-90 minutes), and the ratio of the longest observation session to the shortest session is generally only 2-3. Hence the variances of the estimated baseline components are reasonably compatible and the secondary network adjustment can accommodate these baselines. In the case of GPS "rapid static" and "stop & go" surveying (§5.5), the observation times are comparatively short (from less than 1 minute to 15 or so minutes), and hence the ratio of the longest observation session to the shortest session can be very large. As a result, the variance-covariance matrices for different baselines, determined using a variety of observation session lengths, will have very large differences in the magnitude of the VCV matrix elements, which will cause problems during network adjustment.

HAN & RIZOS (1995b) describes a procedure by which the baseline co-factor matrix may be scaled by an appropriate factor α that accounts for the neglected between-epoch correlations in the baseline reduction using phase observations:

$$\alpha = \frac{n.[1 + \chi(\tau)]}{n.[1 - \chi(\tau)] + 2\chi(\tau)} \qquad (9.4\text{-}8)$$

where $\chi(\tau)$ is the correlation coefficient for data collected at intervals of τ seconds, and n is the number of observations in the session.

The standardised cofactor matrix is converted to the standardised VCV matrix through multiplication by the standardised variance factor (a formula for which can be found in IBID, 1995b).

Chapter 10: GPS and Quality Control

10.1 GPS & QUALITY ISSUES

A Question of Quality

Though "quality" is nowadays a much abused buzzword, there is nevertheless an unprecedented interest in measuring, assuring, verifying and improving the quality of products and services. Hence there is a headlong rush by the manufacturing and service industry sectors to embrace such concepts as "customer oriented business", "Total Quality Management", "transparent operations", "commitment to quality", etc., each representing (it is widely assumed) some aspect of that elusive and ill-defined issue of **Quality**. *What does it all mean for the GPS surveyor?*

In previous chapters many concepts that are related to "quality" were discussed, including network redundancy, biases and errors, appropriate field procedures, baseline accuracy and precision, and many more. This chapter deals with *quality issues in GPS surveying* in order to highlight the role of **Quality Management** as an integral part of the GPS surveying process, and not simply a set of adhoc procedures which are applied in order to "check" if the results are OK. The language of Quality Management is replete with terms such as **Quality Control** (QC) and **Quality Assurance** (QA).

Although the two terms are often considered synonymous, a good working definition is that: (a) **QA** refers to the set of practices and procedures which are intended to maximise the chances that the product or service will satisfy the client's requirements, at a reasonable cost, while (b) **QC** refers to the procedures used to verify the level of quality achieved, and if it is inadequate, to detect the source of the problem and remedy it, if possible.

Quality Management (QM), in the context of GPS surveying, is concerned with *assuring* an agreed level of accuracy and reliability for the station coordinate results. The focus is therefore on *procedures* for defining, measuring and verifying quality, from the commencement of a project to the delivery of the results to the client. This is not to imply that past surveys were of "poor" quality. On the contrary, the set of survey practices which are now accepted as "standard" are the product of many years of hard-won experience. What is proposed is that in future the procedures themselves, and even the bases for the procedures, be: (a) documented (which procedures were followed? what were the outcomes?), (b) "de-mystified" (why was a certain procedure used? can it be justified?), and (c) premeditated (what was done to assure a quality result?). In this era of Total Quality Management it is no longer acceptable for a survey organisation to present such responses as:

> "I am a licensed/registered surveyor, all my work is of the requisite quality"

or

> "The client doesn't understand, he/she is only interested in the numbers and in making sure his/her bill is as low as possible"

or

> "If I start charging extra for this Quality Assurance stuff, he/she will take

his/her business to someone else"

or

"I paid good money to buy the best equipment, the results are certain to be good!"

So what is meant by Quality Management? QM is a management *system* which provides a framework for a consistent approach to managing <u>all</u> aspects of an organisation's operations. The intention is to "get it right first time, every time", rather than the historical quality approach of "let's check if we got it right -- if not let's fix it up". However, perhaps the defining characteristic of QM is that it promotes an organisational *culture* in which all quality procedures are subject to continual scrutiny and improvement, supported by a management system which encourages identified improvements to be put into practice.

In §10.1 procedures and checklists are presented to demonstrate how QA may be implement at every stage of a GPS survey. §10.2 deals with GPS "standards and specifications", which provide the basis by which "levels of service" may be defined. Although "rules-of-thumb" (and commonsense) generally suffice when it comes to GPS survey planning, modern survey techniques such as "rapid static" and "stop & go" (§5.5) present special problems because of the unusual deployment scenarios that are possible. For example, when three or more GPS receivers are used in a survey, will the use of two fixed base receivers result in an inherently stronger and more reliable network, than other configurations? In §10.3 a rigorous mathematical procedure for "measuring" the relative reliability of candidate network configurations is described. Finally, in §10.4, QC procedures for GPS surveying are presented and discussed.

10.1.1 QUALITY ASSURANCE PROCEDURES

Two important aspects of QA are highlighted here: (a) the importance of documentation and reporting, and (b) checklists that "track" the application of QA procedures from the beginning of a project to its completion.

Documentation and Reporting

In many countries geodetic control surveys are carried out by professional surveyors in private practice or government service. Appropriate reporting, documentation and archiving standards should be maintained in order to provide the client with a clear overview of all aspects of the survey "job". In some countries the results of geodetic control surveys are to be reported to the geodetic authority in a prescribed the format (in the U.S. this format is known as the "Blue Book"). The following comments are made:

- ❑ A written report should be prepared, and signed by the surveyors in responsible charge of the project. It should include as a minimum, the following data:
    - A narrative description of the project which summarises the project conditions, objectives, methodologies, QC/QA procedures, and conclusions.
    - A discussion of the observation plan, equipment used, satellite constellation status, and observables recorded.
    - A description of the data processing performed. Note the software used (version number, etc.) and the techniques employed (including ambiguity resolution), and error modelling.
    - Provide a summary and detailed analysis of the minimally constrained and the constrained Least Squares network adjustments performed. List the baseline

observations and parameters included in the adjustment. List the absolute and standardised residuals, the variance factor, and the relative error ellipse/ellipsoid information.

- Identify any data or baseline solutions excluded from the network with an explanation as to why it was rejected.
- Provide details of the transformation model used, or derived, and any GPS/geoid height information that was determined.
- Include a diagram of the project stations and control points at an appropriate scale. Descriptions for each of the monuments should be included, perhaps accompanied by photographs.

❑ Data files, including observations, computed baselines, network adjustments, and coordinates, if not submitted with the project report, should be archived for inspection and future analysis (perhaps in the SINEX format -- §9.3), as should all field sheets and other reconnaissance information.

Managing Quality: Tracking the GPS Surveying Process

Although the analogy cannot be carried too far, it is nevertheless useful to consider a GPS surveying "job" as in the case of any other product or service, and to therefore study the project *life cycle* :

(1) Customer request
(2) Feasibility study
(3) Analysis
(4) Design
(5) Production
(6) Delivery

For each of the above, the inputs and outputs (or outcomes) may be identified. Take for example a GPS survey project. <u>Customer Request</u> can be considered to be the process of defining the *Project Objectives* :

- Document the instructions.
- Determine if the required outputs can be: (a) produced, (b) verified, and (c) controlled.

The <u>Feasibility Study</u> phase is concerned with the *Project Definition* (§5.2):

- Identify the inputs, including: (a) reason for survey, (b) accuracy requirements, (c) completion date, (d) item to be delivered, (e) budget allocation, and (f) project authorisation/definition documentation.
- Assign responsibilities, such as: (a) project director, (b) technical manager, (c) design personnel, and (d) survey personnel.
- Generate definitive technical data for, for example: (a) procurement, (b) execution of the job, and (c) verification that products and processes conform to specifications.
- Prepare the (output) specifications, including: (a) stations & network geometry, (b) equipment & observation technique, (c) computation & adjustment, (d) project management, and (e) acceptance & rejection criteria.

The Analysis and Design phases are synonymous with *Mission Planning & Reconnaissance* (see §5.2 & §5.3):

- Assemble the inputs, including: (a) project definition documentation, (b) planning & design materials, (c) information regarding existing survey marks, and (d) liaison with public utilities, etc.
- Define the station selection criteria, for example: (a) suitability, (b) site access, (c) sky visibility, (d) multipath sources, (e) need for eccentric stations, and (f) azimuth mark.
- Perform the reconnaissance to: (a) review proposed use of existing stations, (b) check suitability of new stations, and (c) finalise station selection.
- Carry out the network design task, taking into account such factors as: (a) accuracy specifications, (b) connections to existing stations, (c) observation techniques, (d) number of station revisits, (e) independent baselines, (f) access & travel times, and (g) data processing considerations.
- Generate the necessary documentation, including: (a) map showing all stations & independent baselines, (b) sky plots, (c) observing schedule, and (d) reconnaissance forms, recovery & access sketches, photographs, etc.
- Prepare mission planning outputs, for example: (a) schedule of baselines to be observed each day, for each field party, (b) observing technique for each baseline (dependent on such things as baseline length, number of visible satellites, PDOP range, whether modern or conventional static GPS techniques used), (c) proposed start & end times, (d) numbering & naming conventions for stations & data files, and (e) personnel to be used.

The Production phase is concerned with *Data Acquisition & Primary Data Processing* (see §5.4, §5.5, chapters 7 & 8):

- Carry out the necessary onsite activities, including: (a) site checks & field documentation (verify site, etc.), (b) instrument setup & initialisation (antenna centring & height measurement, cable connections & power supply, enter appropriate receiver parameters, etc.), (c) instrument operation & data collection monitoring, and (d) data download, demount of hardware & travel to next site.
- Process baselines: (a) setup data processing software parameters, (b) data pre-processing, initial solutions & archiving, (c) process double-differenced data (dual-frequency? conventional static or modern technique? independent baselines & approximate coordinates?), and (d) apply QC procedures.

The Delivery stage is concerned with the *Network Adjustment & Result Transformation* (see chapters 9 & 11):

- Secondary network adjustment activities, including: (a) minimally constrained solutions, (b) check output, note inconsistencies & determine whether there is a need for reobservation, and then (c) refine adjustment.
- Project completion activities, for example: (a) generate full documentation, (b) highlight statistical information on accuracy & reliability, (c) constrain solution to existing control, (d) determine class & order of survey, (e) datum transformation, (f) orthometric height determination, and (g) assessment of total process & communicate with client.

What are presented in the remainder of this section are examples of checklists (provided by JONES, 1995), which may be customised by organisations to assist them with the process of applying and auditing their QA/QC procedures.

Project Definition Checklist (PD1)

Project Name ______________________________

Assignment of Personnel

Project Director ______________________________

Project Manager ______________________________

Design and Survey Personel **Selected** ____________________

Advised ____________________

Project Folder ____________________

Initial Project Documentation

Authorisation to Implement Project Obtained ____________________

Statement of Client Requirements Obtained ____________________

Completion Date Obtained ____________________

Budget Information Obtained ____________________

Technical Specifications and Implementation Schedule

Technical Specification Acquired/Completed ____________________

Implementation Schedule Prepared ____________________

Validate Project Proposal

Statement of User Requirements Checked ____________________

Technical Specifications Checked ____________________

Resource Requirements Checked ____________________

Implementation Schedule Checked ____________________

Amendments Completed ____________________

Mission Planning

Planning Information Checklist (MP1)

Project Name ____________________ **Project Manager** ________________________

Project Folder Established ________ **Reconnaissance Folder Established** ______

Technical Specification Obtained

Location of Survey ____

Purpose of Survey ____

Number of New Stations __________ **Instrumentation** __________________

Observation Method _____________ **Processing Software** ______________

Outputs Required ____

Name and Numbering Conventions Defined ____________________

Planning Materials - Maps

Topographic Maps _______________ **Index Maps** _____________________

Photocopies of Maps _____________

Planning Material - Reconnaissance Forms

Station Reconnaissance Forms __________________________

Reconnaissance Summary Forms __________________________

Planning Material - Existing Survey Marks

Search Area Defined _____________ **Existing Control** _______________________

Other BPN Marks _______________ **Other Authorities Marks** ________________

Planning Material - Meetings with Public Utilities

Utility 1 ____

Utility 2 ____

Mission Planning

Name and Number Conventions (MP2)

Project Name

Project Code

Station Number (Example)

Session ID (Example)

Path Name (Directory Structure) for SKI Project

Directory Structure for Backup Diskettes

Backup Diskette Numbers (Example)

Summary Log File Name (Example)

Mission Planning

Design and Reconnaissance Checklist (MP3)

Project Name____________________ **Project Manager**________________________

Preliminary Design

Plot Existing Survey Marks on Map ____________________

Check Plotting of Existing Survey Marks

- **Positions** ____________________
- **Numbers** ____________________
- **Colours** ____________________
- **Symbols** ____________________

Identify Sites for New Marks ____________________

Prepare Station Reconnaissance Forms ____________________

Prepare Reconnaissance Summary Forms ____________________

Reconnaissance

Reconnaissance of Existing Stations ____________________

Review of Existing Stations ____________________

Review Proposal for New Stations ____________________

Reconnaissance of New Stations ____________________

Final Station Selection ____________________

Complete Reconnaissance Documentation ____________________

Design Network Geometry

Initial Design ____________________

Higher Order Control Assessment ____________________

Final Design ____________________

<u>Mission Planning Documentation</u>

Network Design Map __________________

Reconnaissance Forms __________________

Sky Plots __________________

Observation Shedule __________________

<u>Mission Planning Design Review</u>

Compliance with Client Requirements __________________

Compliance with Technical Specifications __________________

Safety Considerations __________________

Legal Considerations __________________

Design Revision __________________

Measurement
Acquisition and Processing Checklist (AP1)

Project Name ____________________ **Project Manager** ________________________

Pre-Deployment

Personnel Selection

Mark Establishement Officers	____________________
Field Teams	____________________
Baseline Processing Roster	____________________

Establishment of New Survey Marks ____________________

Equipment Check

GPS Receivers	____________________
Supporting Field Equipment	____________________
Computers and Software	____________________
Backup Diskettes	____________________

Team Briefing ____________________

Baseline Measurement and Processing

Observations Completed	____________________
Reobservation Completed	____________________
Baseline Processing Completed	____________________

Adjustment

Free-Net Adjustment Completed	____________________
Fixed-Net Adjustment Completed	____________________

Post-Deployment

File Planning Documents and Field Notes ____________________

Check Equipment

GPS Receivers	____________________
Supporting Field Equipment	____________________
Computers and Data	____________________

Repairs Initiated ____________________

Delivery of Results

Delivery Procedure Checklist (RE1)

Project Name ____________________ **Project Manager** ________________________

Survey Documentation

Survey Mark Information ____________________

Field Pages ____________________

Baseline Reduction Records ____________________

Network Adjustment Records ____________________

Coordinate Listings ____________________

Digital Data Records ____________________

Written Project Report ____________________

Acceptance Inspections

Comparison with User Requirements ____________________

Comparison with Technical Specifications ____________________

Deficiencies Corrected ____________________

Delivery Administration

Results Prepared for Presentation ____________________

Date of Hand Over Meeting ____________________

Delivery Documentation Complete ____________________

Records Filed and Stored

- **Project Documentation** ____________________
- **Field Notes and Plans** ____________________
- **Computer Listings** ____________________
- **Coordinate Listings** ____________________
- **Digital Data** ____________________

10.2

GPS STANDARDS & SPECIFICATIONS

Accuracies considerably better than 1 ppm are currently achievable with GPS, following special field and office procedures. Even relative accuracies of several parts per million are achieved on a routine basis, which considerably exceed conventional standards for geodetic surveys using standard ground-based techniques. Nevertheless there is a need for the development of acceptable standards and specifications.

What does "standards" mean? What does "specifications" mean?

The <u>accuracy standards</u> are essentially those developed for conventional surveys, augmented with several categories of higher accuracy. <u>Specifications</u>, or <u>survey practices</u>, on the other hand, relate to network design and planning, GPS instrumentation to be used, and field and office procedures that promote efficiency (and reliability), and facilitate the classification of surveys. *They may take the form of strict requirements, recommendations or simply suggestions.*

The U.S. National Geodetic Survey was the first national geodetic organisation to prepare standards and specifications (S&S) for GPS surveys (FGCC, 1988). These S&S were be adopted in many other countries, including Australia, with minor modifications made to account for the different survey accuracy standards. (In Australia they are referred to as "standards & practices" -- see ICSM, 1994.) GPS field and technical experience has been drawn heavily in developing these specifications, and issues specific to the GPS technology have had to be addressed when framing the specifications:

- GPS provides three-dimensional information.
- Satellite geometry and ephemeris accuracy requirements need definition.
- Geoid slopes and accuracies need definition.
- Different measurement types and accuracies are available.
- Baseline, session, and network solutions are possible.
- GPS precisions are generally higher than the supposedly superior terrestrial control.
- GPS results are referred to a coordinate system that is geocentric, and is displaced in space relative to the origin of conventional geodetic datums.

However, must be emphasised that there are many aspects of the technology that are common to conventional geodetic techniques, including:

(1) The results are relative coordinates, as would be derived by conventional theodolite/EDM techniques.

(2) Networks can be built up in an analogous manner to geodetic traverses, and the quality control procedures are similar (no "hanging" traverses or "no check" lines!).

(3) The magnitude of the positional error between directly connected GPS stations is proportional to the distance separation of the stations.

(4) Redundancy can come from the number of occupations per stations, repeat baseline observations, measuring antenna height twice, etc.

A major difference between the U.S. S&S and the Australian S&P is that in the case of the former they are framed for GPS surveys alone, while the latter are general, and are intended to be valid for <u>all</u> survey technologies. Furthermore, the Australian S&P recognise the role of modern GPS techniques such as "rapid static" and "stop & go" (§5.5), whereas the U.S. S&S predate these new techniques.

The accuracy standards can be easily modified, however, the recommended practices have had to be conservative in order that they need not have to be changed often. Although attempts have been made in some countries to differentiate the practices according to a hierarchy of: <u>requirement</u>, <u>recommendation</u>, or <u>suggestion</u>; this seems unnecessarily restrictive and does not reflect the changing nature of the technology (and with it, its capabilities). Hence it is preferable to simply state the specifications without making value-judgements, although wherever possible it should be indicated how they have been arrived at (by conjecture, experience or scientific fact?), whether they should be considered "optional", and whether they are likely to remain valid for the foreseeable future, etc. **As a result, GPS specifications are likely to undergo continuous assessment and refinement.**

10.2.1 SURVEY ACCURACY STANDARDS

GPS accuracy standards are generally country specific.

In general, accuracy standards for horizontal coordinates are based on the ratio of the relative positional error of a pair of control stations to the horizontal separation of the points, at the 95% confidence level. As this ratio increases, the classification of the survey decreases. In the case of GPS this measure of positional accuracy may be extended to all three of the components (as in the U.S. S&S, while only the horizontal components are considered in the Australia S&P). Experience has shown that it is possible to measure relative position using GPS carrier phase data routinely to the 0.5-1cm + 1-2ppm level (§2.4 and §4.4). With careful observation procedures, and the appropriate data processing, precisions of the order of 0.01ppm have been achieved for scientific GPS surveys. This is 1000 times better than the existing "1st order" accuracy standard. *Hence the need for new survey categories.* This leads to the problem of how to measure the quality of a GPS survey, and the secondary concern of classification.

Tables 5.2-1 and 5.2-2 give the Australian and U.S. survey accuracy standards for GPS (and, in the case of Australia, for other types of horizontal surveys as well), expressed in terms of maximum allowable <u>base error</u> and <u>line-length error</u> for relative position.

In both the Australian and U.S. standards, since the previous survey standards were not adequate for classifying high precision GPS surveys, a two-tier system of classification has been adopted:

- ❑ A **general classification** for standard surveys, incorporating the present classifications for geodetic control, engineering, mapping and cadastral surveys. The results of general classification surveys are based on a constrained adjustment of GPS stations tied to the local network control. In the case of Australia, the long standing classifications "A", "B", "C", "D", "E" have been preserved.
- ❑ A **special classification** for high precision geodetic surveys such as, for example, crustal motion determination, land subsidence monitoring and precise engineering surveys. Special classification surveys are adjusted independently of the local network control and are therefore based only on the internal three-dimensional consistency of the GPS network. In the case of the U.S. new categories "AA", "A", "B"; and for Australia, new categories of "3A", "2A", have been established (Table 10.2-1).

To cater for these classifications, a new terminology was necessary to distinguish between the survey results:

☞ that were a function only of the field survey methods, reduction techniques and the results of a minimally constrained network adjustment, and those

☞ that were a function of the above AND the conformity of the new survey results with the existing network coordinate set after the transformation process required to convert results from the original datum (such as WGS84 in the case of GPS) to the local geodetic datum.

The U.S. GPS standards refer to the "geometric" classification based on the internal consistency of the survey, and the "NGRS" (National Geodetic Reference System) classification based on the fit to existing geodetic control. In a similar fashion, in Australia the **CLASS** of any survey is defined by the quality of its internal consistency, whether the survey was intended for general or special classification (for example, photogrammetric control purposes, or crustal motion monitoring), and the **ORDER** is a function of the class and the fit to the existing AGD (horizontal) and/or AHD (vertical) control.

How does the survey classification system function in Australia? The "Recommended Standards and Practices for Control Surveys" (ICSM, 1994) suggests survey techniques designed to achieve a specific CLASS of survey. If the survey methods or reduction techniques used are not commensurate with the desired class of survey, or if the network adjustment fails to achieve the desired class, the stations in the survey should be assigned the highest CLASS common to all three aspects. Therefore, the allocation of CLASS to a survey on the basis of the results of a **minimally constrained Least Squares network adjustment** (after the usual statistical testing for outliers, etc.) is strictly defined according to the test (IBID, 1994):

" whether the semi-major axis of each relative, one-sigma, standard error ellipse (for 2-D) or ellipsoid (for 3-D), is less than or equal to the length of the maximum allowable semi-major axis (r), in millimetres, using the following formula:

$$r = c.(d + 0.2) \qquad (10.2\text{-}1)$$

where c is an empirically derived factor represented by historically accepted precision for a particular standard of survey, and

d is the distance to any station in kilometres (minimum of 1km). "

The values of c assigned to various classes of survey are shown in Table 10.2-1. (The values of base error and line-length error shown in Table 5.2-1 are derived from this formula and the values of c given in Table 10.2-1.) Multiply the c values by 2.45 for the 95% confidence limits.

Experience has shown that the overall pattern of error propagation is not strictly proportional to distance (as implied by the formula above), but is a combination of instrumental and centring errors. A graph of the length of the maximum allowable semi-major axis against distance between any two stations is given in Figure 10.2-1, showing clearly the different error propagation characteristics for lines above and below one kilometre.

Table 10.2-1. Australian classification of horizontal control surveys.

CLASS	c (1-σ)	c (95%)	Typical applications
3A	1	2	Special high precision surveys
2A	3	8	High precision geodetic surveys
A	7.5	18	National and state geodetic surveys
B	15	35	Densification of geodetic survey
C	30	75	Survey coordination projects
D	50	125	Lower CLASS projects
E	100	250	Lower CLASS projects

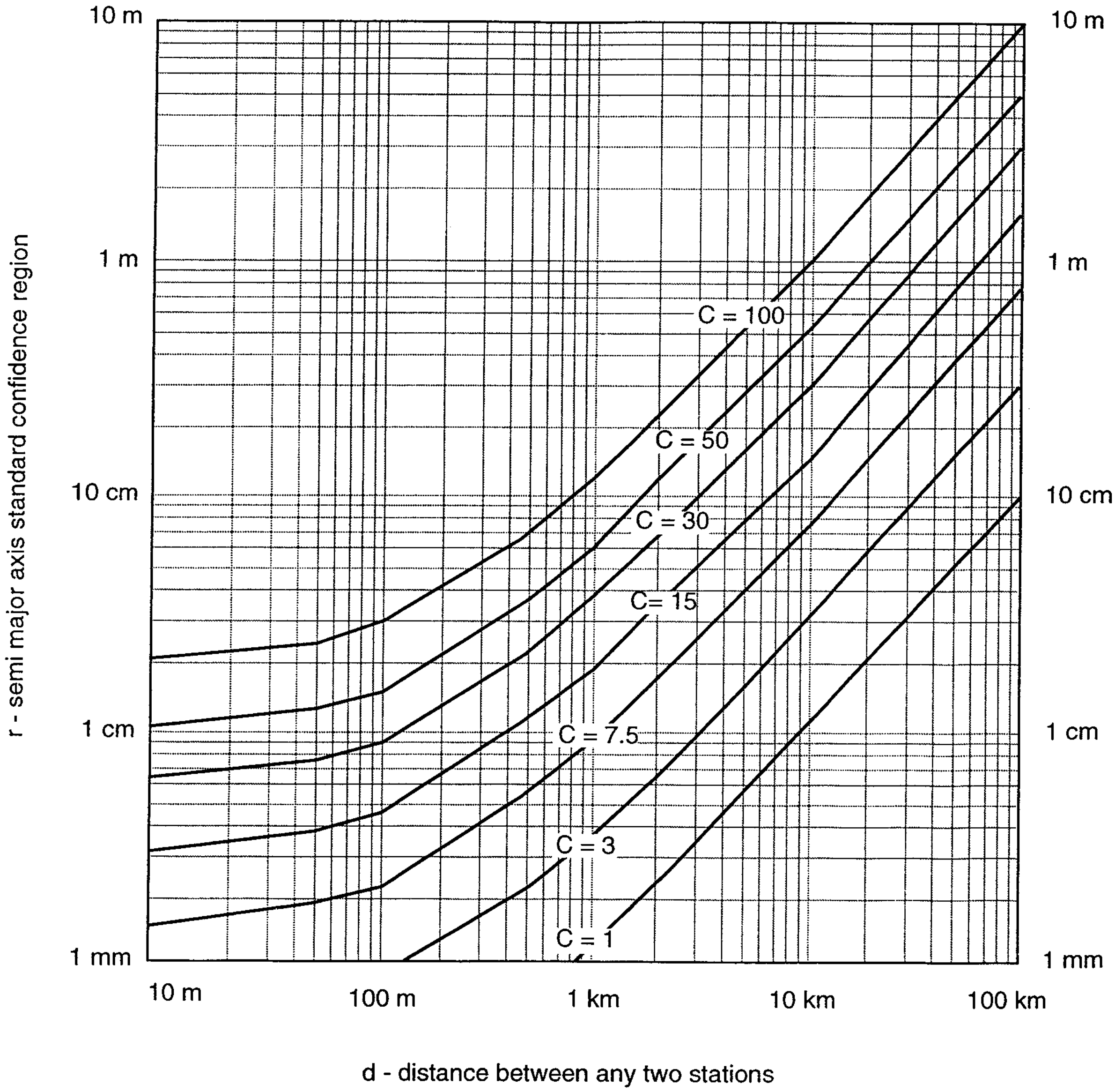


Figure 10.2-1. Class related values of "c" for 2-D surveys.
(ICSM, 1994)

Assigning the ORDER of a Survey

In Australia, the ORDER assigned to a survey network following constraint of that network to the existing control may be:

- ❑ not higher than the ORDER of the existing stations constraining that network, and
- ❑ not higher than the CLASS assigned to that survey.

The relationship between CLASS and ORDER is:

CLASS	ORDER
3A	Special high precision
2A	0
A	1
B	2
C	3
D	4
E	5

As in the case of the assignation of CLASS of survey, the allocation of ORDER is based on eqn (10.2-1), but applied to the output of the ***constrained* network adjustment**. As the concept of ORDER is based on the CLASS of the station, as well as the fit of the survey network to the existing control, the ORDER may be degraded by virtue of the quality of the surrounding control, or by the quality of the ties to that control.

Vertical Control Standards

In the case of heighting, the situation is a little more complicated. Some heighting techniques (for example, spirit levelling) propagate errors in proportion to the square root of the distance, while other techniques (for example, GPS and trigonometric levelling) propagate errors as a function of distance.

The preliminary Australian S&S propose the allocation of CLASS to <u>any</u> vertical survey on the basis of (ICSM, 1994):

> " whether the standard deviation of each height value (from a minimally constrained network adjustment) is less than or equal to the maximum allowable value (r) defined according to:
>
> $$r = c\sqrt{d} \qquad \text{for} \quad r > 0.25c \qquad (10.2\text{-}2)$$
>
> where r is the maximum allowable error in millimetres,
> c is an empirically derived factor, given below for each CLASS of survey, and
> d is the distance to any station in kilometres (> 1km). "

Values of c are assigned according to:

CLASS	c (one sigma)
A	4
B	8
C	12
D	50
E	100
F	250

The U.S. standards for conventional vertical surveys are unchanged (and are based on the usual square-root-distance classification), but the GPS standards recognise that the use of GPS heights can vary depending on whether:

☞ **Orthometric height** differences are required, then the accuracy of the GPS derived orthometric heights are a function of BOTH the accuracy of the GPS ellipsoidal heights and the accuracy of relative geoid heights.

☞ **GPS ellipsoid height** differences are being measured for the purpose of monitoring the change in height between stations, then it is not necessary to have any accurate information on the shape of the geoid.

The U.S. accuracy standards for GPS orthometric heighting are given in Table 10.2-2 (FGCC, 1988).

Table 10.2-2. U.S. orthometric height difference accuracy standards (95% C.I.).

ORDER-CLASS	Minimum elevation difference accuracy standard ppm	1:e	Minimum geoid height difference accuracy standard ppm	1:e
AA	2	1:500000	2	1:500000
A	2	1:500000	2	1:500000
B	5	1:200000	5	1:200000
1	15	1:67000	10	1:100000
2-I	20	1:50000	10	1:100000
2-II	50	1:20000	20	1:50000
3-I	100	1:10000	40	1:25000

Note that at the higher orders, the error is dominated by the accuracy of the relative geoid height values, whereas, for the lower orders the major source of error is in the ellipsoid height differences. A further discussion of "GPS levelling" is given in section 11.3.

10.2.2 SPECIFICATIONS FOR GPS SURVEYS

These are dependent on the accuracy standard which is to be achieved. They cover such aspects as:

- network design,
- instrumentation,
- field procedures and monumentation,
- office reduction procedures, and
- calibration and result validation procedures.

Furthermore, except for the last item, the above specifications hold equally for GPS "horizontal" and "vertical" surveys. No attempt is therefore made to optimise a GPS survey for height accuracy above that for horizontal accuracy. Each of these topics is discussed briefly below using the Australian S&P and U.S. S&S as examples. (For further details see ICSM, 1994, in the case of the Australian S&P, and FGCC, 1988, for the U.S. S&S.)

Network Design

This generally includes such factors as the minimum and maximum station spacing, minimum number of connections to existing horizontal and vertical control, minimum spacing between azimuth and station marks, length of observation sessions, type of GPS technique used, direct connection requirements, etc. No guidelines on network <u>shape</u> are usually given, as the location and distribution of points in a GPS survey are more likely to be influenced by the *intent* of the survey rather than by prescribed network "figures" that must be observed. On the other hand, more importance is placed on direct connections, multiple station occupancies, repeat observations, etc., for the purpose of aiding quality control.

Table 10.2-3 summarises the Australian network design specifications for various classes of GPS survey. Note that no recommendations are made as far as connection requirements to existing geodetic control (these would influence the ORDER of the GPS survey).

Instrumentation and Testing

This could include advice on the type of receiver, when to use dual-frequency receivers, clock stability standards, measurement interval (to ensure data from different receivers can be processed together), provision of a test network for regular calibration of instruments and software, etc. Obviously only GPS receivers that can track and record integrated carrier beat phase are appropriate. Unfortunately there is little further advice given. The Australian S&S mention nothing except to suggest dual-frequency instrumentation as a function of the class of survey and the baseline length (see "field procedure specifications").

The FGCS has tested GPS receivers (and the data reduction software) over a special calibration/test network in the Washington D.C. area (§4.4).

The Australian S&P mentions that "system testing" may be required after acquisition of new equipment, new software, or when trialing new procedures. The test consists of the measurement of a small network, as well as the ongoing analysis of production results using the instrumentation, software or procedures under scrutiny. It is not clear if it sufficient to furnish evidence of an overseas test, or one carried out by the GPS manufacturer.

Table 10.2-3. Australian network design specifications - internal GPS.

CLASS	3A	2A	A	B	C	D
c-values for the CLASS (Part A, 2.2.1	≤1	≤3	≤7.5	≤15	≤30	≤50
Technique						
Classic Static	✓	✓	✓	✓	✓	✓
Quick Static			✓	✓	✓	✓
Pseudo Kinematic				✓	✓	✓
Stop and Go				✓	✓	✓
Guide to minimum station spacing in kms (a)	5	1.5	0.5	0.1	NA	NA
Typical station spacing in kms (b)	100–500	10–100	0.5–10	0.1–5	>0.05	NA
Independent occupations per station (c) At least 3 × (% of total stations) (d) at least 2 × (% of total stations) (d)	50% 100%	40% 100%	20% 100%	10% 100%	— —	— —
Minimum common satellites	Four satellites, five or more an advantage					
Minimum PDOP required (e)	Less than ten (10)					
Minimum satellite elevation	Fifteen (15) degrees					
Data rate	Optional					
Minimum observation period (static method) (f)	180 min	120 min	60 min	30 min	30 min	30 min
Minimum independent baselines at each station	3	3	2	2	2	2

NA: Not applicable

Notes:

(a) Minimum station spacing is illustrated using a 5 mm noise level after adjustment. Below these minimum distances, special efforts are required to reduce the error budget. For a noise level of 10 mm these values are to be approximately doubled.

(b) These values relate to the using of conventional equipment and proprietary software.

(c) Independent occupations per station may be back to back, but the antenna should be re-setup for each occupation. Antenna heights are to be changed by at least 0.1 to 0.2 m unless set up on a pillar. The full specified minimum observation period should be observed with each occupation.

(d) For example for a **CLASS** A network aim for:

(i) 20% of stations are to be occupied at least three times;

(ii) 100% of stations are to be occupied at least twice.

(e) Less than 10 after resolution of ambiguities. See also **Section 2.6.4.**

(f) As a guide 30 minutes as a definitive minimum + about 2 minutes per kilometre.

Table 10.2-4 is the equivalent U.S. specifications, and note that recommendations with regards to connection to existing horizontal and vertical control.

Table 10.2-4. U.S. GPS network design, geometry and connection specifications.

Geometric relative positioning standards				
Group	AA	A	B	D
Order	AA	A	B	1,2-I&II,3
ppm	0.01	0.1	1.0	10,20,50,100
<u>Horizontal network control of NGRS[a], minimum number of stations:</u>				
When connections are to orders AA, A or B....	4	3	3	2
When connections are to order 1...............	na[b]	na[b]	na[b]	3
When connections are to orders 2 or 3........	na[b]	na[b]	na[b]	4
<u>Vertical network control of NGRS[a], minimum number of stations[c][d]</u>...........	5	5	5	4
<u>Continuous tracking stations (master or fiducials), minimum number of stations</u>....	6	5	3	op
<u>Station Spacing (km)</u>:				
Between ANY adjacent stations, not <u>less</u> than.	30	30	3	0.3
Between "existing network control" and CENTER of project:				
Not <u>more</u> than............................	10d	10d	7d	5d
50 percent not <u>less</u> than..................	$\sqrt{5d}$	$\sqrt{5d}$	$\sqrt{5d}$	$\sqrt{5d}$
Between "existing network control" located <u>outside</u> of project's outer boundary and the edge of the boundary, not <u>more</u> than...	300	300	100	50
<u>Location of network control</u> (relative to center of project); number of "quadrants", not less than...............................	4	4	4	3
<u>Direct connections are desirable</u>: Between ANY adjacent stations (new or old, GPS or non-GPS) located near or within project area, when spacing is <u>less</u> than (km).......	30	30	10	5

Legend:
- d - is the maximum distance in (km) between the center of the project area and any station of the project.
- NGRS - National Geodetic Reference System
- CL - Confidence level
- na - not applicable
- op - optional

5-9-88

NOTE: If it is not practical to plan a survey that is within the criteria, minor adjustments may be made provided that it is authorized by the agency requesting the survey.

Remarks: (a) Consult National Geodetic Survey officials whenever it is necessary to consider exceptions to these criteria, particularly, when the GPS survey project data are to be submitted to NGS for incorporation in the NGRS.

(b) If a survey with an accuracy standard of AA, A, or B is specified and one objective in the survey is to upgrade the existing network, then connections to a minimum of four stations are required or at least one station in each one-degree block with a minimum of four stations.

(c) First choice is vertical network control established and/or maintained by the National Geodetic Survey. When it is not possible to occupy the minimum number of NGRS points, non-NGRS control points may be used. This should be documented in the project report.

(d) If it is expected that the constrained adjustment for determination of the elevations within the project area will be based on more than one "bias group" (see discussion under section on Office procedures, Analysis and Adjustments) then the minimum number of stations specified is that which is required within the area for each "bias group." For example, if there two bias groups and ties required to four bench marks, then four bench marks will be incorporated within each area of the "bias group" for a total of 8 bench marks.

5-09-88

Field Procedure Specifications

Specifying such things as the number of receivers to observe simultaneously, length of observation span, number of satellites to be observed, cutoff elevation angle, measurement of meteorological parameters, etc.

Therefore, specifications for field procedures must include (beyond those existing for conventional techniques), as a function of desired accuracy:

- time per station occupation (if possible, as a function of satellite geometry),
- whether single or dual-frequency instrumentation is to be used,
- the GPS technique to be used,
- number of receivers deployed,
- satellite geometry,
- data rates and the number of satellites tracked,
- number of independent occupations per site,
- number of repeated baselines,
- cutoff elevation for tracking,
- quantity and quality of meteorological data to be collected, and
- antenna setup specifications, etc.

Tables 10.2-3 and 10.2-5 give the Australian observation requirements for GPS surveys (note that the modern GPS techniques are included), and Table 10.2-6 the equivalent U.S. specifications (only for conventional static GPS). In Table 10.2-5 an effort has been made to provide recommendations with regards to the GPS surveying technique that should be used as a function of antenna spacing. Table 10.2-3 makes suggestions concerning which GPS surveying technique should be used as a function of the desired ORDER of the survey. Table 10.2-6 (from the U.S. S&S is far more detailed, and perhaps contains outdated information).

Data Processing Specifications

Specifying the number of triangle and loop closures, the model for phase data reduction, observation weights, error analysis tests, data editing techniques, etc.

Specifications for processing data might include (as a function of accuracy desired):

- how the antenna offsets are applied,
- satellite elevation cutoff,
- ephemeris source and age,
- use of met data, or standard atmospheric models, for tropospheric delay reduction,
- measurement data quality and quantity,
- measurement data rejection criteria,
- measurement residual maximums,
- baseline or session adjustment specification, etc.,
- specifications for the storage and presentation of GPS results, and
- loop closure and repeat baseline analysis specifications.

As with GPS instrumentation, the only way of ensuring realistic results with confidence is to use data reduction software which has been tested on a three-dimensional calibration network, or with a standard test dataset.

Table 10.2-5. Australian observation requirements for GPS surveys.

	initial-isation	dual/ single frequency	common sats	continuous lock during travel	maximum spacing	PDOP
STATIC GPS METHODS						
classic-static	no	opt.	≥4	no	> 500 kms	*e
pseudo kinematic	no	opt.	≥4 *a	no, only at base	< 20 kms	*e
quick static methods	no	opt. *d	≥4	no	< 10 kms	*e
KINEMATIC GPS METHODS						
Continuous Kinematic	yes *b	opt. *c	5 preferred 4 possible	yes	< 20 kms	<10 *f
Stop and Go	yes	single	5 preferred 4 possible	yes	< 20 kms	<10 *f
"Ambiguity resolution on the fly" — kinematic	no	dual, also single	5 preferred 4 possible	preferred but not necessary	< 20 kms 7–10 kms best	

*a Four satellites required in both observation sessions, 5 or more an advantage.

*b Observe a known baseline (at beginning or end) and solve all ambiguities, or do an antenna swap or return to the starting point at the end of the survey.

*c Dual frequency receivers give an advantage.

*d Dual frequency P-code will enhance the speed of the solution.

*e Sufficiently changing geometry during a recording session assists in the determination of ambiguities and, where they have been resolved, PDOP should be low at some stage in the processed data.

*f The ambiguities are resolved through the initialisation and the PDOP should be low at some stage during each station occupation from that moment (refer to manufacturers specifications).

Table 10.2-6. U.S. field procedures for GPS surveys.

Geometric relative positioning standards (Group / Order / ppm)	AA / AA / 0.01	A / A / 0.1	B / B / 1.0	D / 1,2-I&II,3 / 10,20,50,100
Two frequency observations (1 and L2) required[a]: Daylight observations[b]	Y	Y	Y	op
Number of receivers observing simultaneously, not less than:	5	5	4	3
Satellite Observations: RDOP values during observing session (meters/cycle)[d] [TO BE ADDED IN FUTURE VERSION]				
Period of observing session (observing span), not less than (min):				
[4 or more simultaneous satellite observations][e]				
Triple difference processing[f]	na	na	240	60-120
Other processing techniques[g]: General requirements[h][i]	240	240	120	30-60
Continuous and simultaneous between all receivers, period not less than[i][j]	180	120	60	20-30
Data sampling rate - maximum time interval between observations (sec)	15	30	30	15-30
Minimum number of quadrants from which satellite signals are observed	4	4	3	3 or 2[k]
Maximum angle above horizon for obstructions[v] (degrees)	10	15	20	20-40
Independent occupations per station(l):				
Three or more (percent of all stations, not less than)	80	40	20	10
Two or more (percent of stations, not less than):				
New stations	100	80	50	30
Vertical control stations	100	100	100	100
Horizontal control stations	100	75	50	25
Two or more for each station of "station-pairs"(m)	Y	Y	Y	Y
Master or fiducial stations(n):				
Required, yes or no(o)	Y	Y	Y	op
If yes, minimum number	4	3	2	-
Repeat base line measurements, about equal number in N-S and E-W directions, minimum not less than (percent of total independently [nontrivial] determined base lines)	25	15	5	5
Loop closure, requirements when forming loops for post-analyses:				
Base lines from independent observing sessions, not less than	3	3	2	2
Base lines in each loop, total not more than	6	8	10	10
Loop length, generally not more than (Km)	2000	300	100	100
[NOTE: Also, see table 5]				
Loop closure (Continued):				
Base lines not meeting criteria for inclusion in any loop, not more than [percent of all independent nontrivial lines(p)]	0	5	20	30
Stations not meeting criteria for inclusion in any loop, not more than (percent of all stations)	0	5	10	15
Direct connections are required: Between ANY adjacent (NGRS and/or new GPS) stations (new or old, GPS or non-GPS) located near or within project area, when spacing is less than (Km)	30	10	5	3
Antenna setup:				
Number of antenna phase center height measurements per session, not less than	3(q)	3(q)	2	2
Independent plumb point check required(r)	Y	Y	Y	op

Geometric relative positioning standards — Group / Order / ppm	AA / AA / 0.01	A / A / 0.1	B / B / 1.0	D / 1,2-I&II,3 / 10,20,50,100
Photograph (closeup) and/or pencil rubbing required for each mark occupied.............	Y	Y	Y	Y
Meteorological observations:				
Per observing session, not less than.........	3(s)	3(s)	2(t)	2(t) or op
Sampling rate (measurement interval), not more than (min)...........................	30	30	60	60
Water vapor radiometer measurements required at selected stations?.......................	op	op	N	N
Frequency standard warm-up time (hr)(u):				
Crystal....................................	12	12	(u)	(u)
Atomic.....................................	1	1	(t)	(t)

LEGEND: nr - not required, na - not applicable, op - optional

REMARKS:

(a) If two-frequency observations can not be obtained, it is possible that an alternate method for estimating the ionospheric refraction correction would be acceptable, such as modeling the ionosphere using two-frequency data obtained from other sources.

Or, if observations are during darkness, single frequency observations may may be acceptable depending on the expected magnitude of the ionospheric refraction error.

(b) When spacing between any two stations occupied during an observing session is more than 50 km, two frequency observations may need to be considered for Accuracy Standards of Order 2 or higher.

(c) Multiple baseline processing techniques.

(d) Studies are underway to investigate the relationship of Geometric Dilution of Precision (GDOP) values to the accuracy of the base line determinations. Initial results of these studies indicate there is a possible correlation. It appears the best results may be achieved when the GDOP values are changing in value during the observing session.

(e) The number of satellites that are observed simultaneously cannot be less than the number specified for more than 25 percent of the specified period for each observing session.

(f) Absolute minimum criteria is 100 percent of specified period.

(g) "Other" includes processing carrier phase data using single, double, nondifferencing, or other comparable precise relative positioning processing techniques.

(h) The times for the observing span are conservative estimates to ensure the data quantity and quality will give results that will meet the desired accuracy standard.

(i)

(j) Absolute minimum criteria for the data collection observing span is that period specified for an observing session that includes continuous and simultaneous observations. Continuous observations are data collected that do not have any breaks involving all satellites; occasional breaks for individual satellites caused by obstructions are acceptable, however, these must be minimized. A set of observations for each measurement epoch is considered simultaneous when it includes data from at least 75 percent of the receivers participating in the observing session.

(k) Satellites should pass through quadrants diagonally opposite of each other

(l) Two or more independent occupations for the stations of a network are specified to help detect instrument and operator errors. Operator errors include those caused by antenna centering and height offset blunders.

When a station is occupied during two or more sessions, back to back, the antenna/tripod will be reset and replumbed between sessions to meet the criteria for an independent occupation. To separate biases caused by receiver and/or antenna equipment problems from operator induced blunders, a calibration test may need to be performed.

(m) Redundant occupations are required when pairs of intervisible stations are established to meet azimuth requirements, when the distance between the station pair is less than 2 km, and when the order is 2 or higher.

(n) Master or fiducial stations are those that are continuously monitored during a sequence of sessions, perhaps for the complete project. They could be sites with permanently tracking equipment in operation where the data are available for use in processing with data collected with the mobile units.

(o) If simultaneous observations are to be processed in the session or network for base line determinations while adjusting one or more components of the orbit, then two or more master stations shall be established.

(p) For each observing session there are r-1 independent base lines where r is the number of receivers collecting data simultaneously during a session, e.g. if there were 10 sessions and 4 receivers used in each session, 30 independent base lines would be observed.

(q) A measurement will be made both in meters and feet, at the beginning, mid-point, and end of each station occupation.

(r) To ensure the antenna was centered accurately with the optical plummet over the reference point on the marker, when specified, a heavy weight plumb bob will be used to check that the plumb point is within specifications.

(s) Measurements of station pressure (in millibars), relative humidity, and air temperature (in°C) will be recorded at the beginning, midpoint, and end depending on the period of the observing session.

(t) Report only unusual weather conditions, such as major storm fronts passing over the sites during the data collection period. This report will include station pressure, relative humidity, and air temperature.

(u) The amount of warm-up time required is very instrument dependent. It is very important to follow the manufacturer's specifications.

(v) An obstruction is any object that would effectively block the signal arriving from the satellite. These include buildings, trees, fences, humans, vehicles, etc.

A minimally constrained three-dimensional adjustment would reveal whether realistic observation weights were used and what the standard errors of the three-dimensional vector components are. A constrained adjustment using existing control coordinates accounting for scale, orientation, and geoid undulation relationships between the GPS satellite and the local geodetic control network datums would provided proper integration of the GPS relative positioning results (but obviously at a level corresponding to the existing control network accuracy).

The Australian S&P give no guidelines on office <u>checking</u> procedures (loop tests, etc.). Rather they give minimum data reduction procedures for all the common GPS surveying techniques (Table 10.2-7). (The meaning of terms such as "double- and triple-differencing", "ambiguity fixed and ambiguity float", etc., as they apply to GPS data processing were given in chapters 6 and 7.) On the other hand, the U.S. S&S make little mention of the data processing procedure to be used, and instead focuses on the checking procedures to be used. Table 10.2-8 summarises these office procedures.

Calibration and Result Validation: The Quality Control Issues

GPS surveying is a relatively new and complex technology which, in addition, is changing rapidly. It is therefore not realistic to draw up a strict list of instrumentation and procedures. With many different sets of equipment becoming available and expected refinements and changes to the system, the emphasis should be on validation of, rather than strict specification, of equipment, etc. *What can be done to ensure that a GPS survey has in fact been adequately carried out?*

❑ The GPS surveyor can be tested (equipment, expertise, procedures and software) on a test-network (and licensed for various grades of GPS survey?). *The issue of "legal* *traceability" of GPS surveys is being looked at closely in Australia.*

- ❑ Multiple occupations of stations permit checks to be made on station coordinate results on a session-by-session basis. *The amount by which the results deviate from each other and the proportion of stations with multiple occupancies can be defined in the S&S.*
- ❑ Ditto for repeat baselines, in which the same pairs of stations are occupied more than once. *The amount by which the results deviate from each other and the proportion of multiple occupancies can be defined in the S&S.*
- ❑ Loop miscloses provide reliability validation information only if baselines comprising the loop come from at least two independent observing sessions. *The acceptable misclose can be defined in the S&S, but this is not a statistical test!*
- ❑ The relative accuracies derived from the minimally constrained solution should be checked against the standards defined for the class of GPS survey in question.
- ❑ The baseline results derived initially from the phase data adjustment can be compared with the derived results from the final network (constrained and unconstrained) solution. *The acceptable discrepancy can be defined in the S&S.*
- ❑ A study of the measurement residuals after adjustment should be made. *Ultimately, if there is sufficient redundancy present, this is the best evidence of the quality of a GPS survey.*

Table 10.2-7. Recommended GPS reduction procedures (ICSM, 1994).

CLASS	3A	2A	A	B	C	D	E
c-values (one sigma)	≤1	≤3	≤7.5	≤15	≤30	≤50	≤100
<8 km	D*, DD, FX	D*, DD, FX	S, DD, FX	S, DD, FX	S, DD, FX	S, DD, FT	S, DD, FT
8–25 kms	D, DD, FX	D, DD, FX	D, DD, FX	D, DD, FX	S, DD, FX	S, DD, FT	S, DD, FT
25–50 kms	D, DD, FX(25)–FT(50)	D, DD, FX(25)–FT(50)	D, DD, FX–FT	D, DD, FX–FT	D, DD, FX–FT	D, T	D, T
50–90 kms	D, DD, FT	DD or T**, D, FT	DD or T**, D, FT,	DD or T**, D, FT,	DD or T**, D, FT	D, T, NCP	D, T, NCP
>90 kms	D, T	D, T	D, T	D, T	D, T	D, T, NCP	D, T, NCP

Where: S = single frequency
D = dual frequency
DD = double differences
FX = ambiguity fixed solution
FT = ambiguity float solution, with repaired cycle slips
T = triple difference solution with sufficient observation length, allowing change of geometry.
NCP = Narrow correlation, C/A code or Pseudorange methods, e.g. DGPS
* L1 solution, from a dual frequency receiver, in order to enable ambiguity resolution by widelaning.
** Double difference preferred, triple difference solution increasingly acceptable the longer the distance, if the observation length allows sufficient geometry change.

Table 10.2-8. U.S. office procedures used for classifying GPS surveys.

Geometric relative positioning standards — Order: / ppm :	AA 0.01	A 0.1	B 1.0	1 10	2-I 20	2-II 50	3 100
Ephemerides:							
Orbit accuracy, minimum (ppm)............	0.008	0.05	0.5	5	10	25	50
Precise ephemerides required?............	Y*	Y*	Y	op	op	N	N
Loop closure analyses(b) - When forming loops, the following are minimum criteria:							
Base lines in loop from independent observations not less than............	4	3	2	2	2	2	2
Base lines in each loop, total not more than..................................	6	6	6	10	10	15	15
Loop length, not more than (Km)..........	2000	2000	300	200	200	200	200
Base lines not meeting criteria for inclusion in any loop, not more than (percent of all independent lines)....	0	0	5	20	30	30	30
In any component (X,Y,Z), "maximum" misclosure not to exceed (cm).........	10	10	15	25	30	50	100
In any component (X,Y,Z), "maximum" misclosure, in terms of loop length, not to exceed (ppm)....................	0.2	0.2	1.25	12.5	25	60	125
In any component (X,Y,Z), "average" misclosure, in terms of loop length, not to exceed (ppm)....................	0.09	0.09	0.9	8	16	40	80
Repeat base line differences:							
Base line length, not more than (Km).....	2000	2000	500	250	250	100	50
In any component (X,Y,Z), "maximum" not to exceed (ppm)....................	0.01	0.1	1.0	10	20	50	100
Minimally constrained adjustment analyses:							
(Criteria is being developed and will appear in an updated version of this document)							

REMARKS:

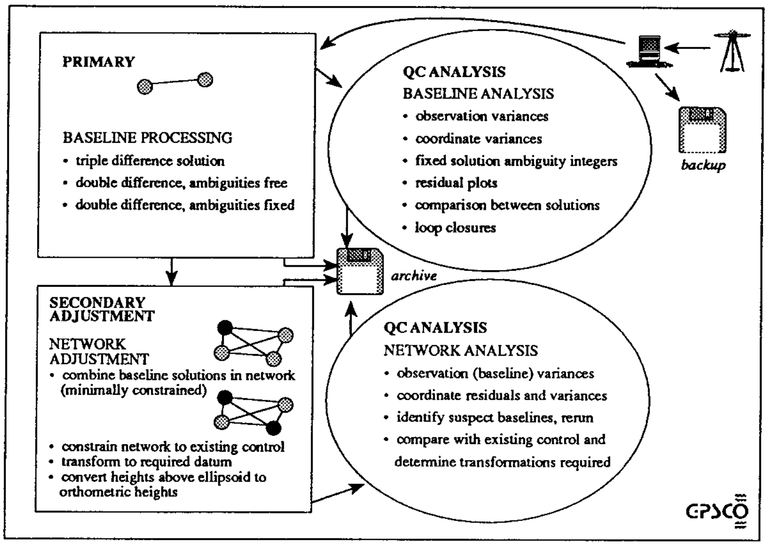
(a) The precise ephemerides is presently limited to an accuracy of about 1 ppm. By late 1989, it is expected the accuracy will improve to about 0.1 ppm. It is unlikely orbital coordinate accuracies of 0.01 ppm will be achieved in the near future. Thus to achieve precisions approaching 0.01 ppm, it will be necessary to collect data simultaneously with continuous trackers or fiducial stations. (see criteria for field procedures, table 5.) Then the all data is processed in a session or network solution mode where the initial orbital coordinates are adjusted while solvinmg for the base lines. In this method of processing the carrier phase data, the coordinates at the continuous trackers are held fixed.

(b) Between any combination of stations, it must be possible to form a loop through three or more stations which never passes through the same station more than once.

10.3

QUALITY CONTROL PROCEDURES FOR GPS NETWORKS

In relation to GPS surveying, "QUALITY CONTROL" refers to those procedures or indicators that provide a measure of the quality (<u>accuracy</u> and <u>reliability</u>) of a GPS survey. Quality Control therefore has as much to do with detecting observational errors, as in interpreting the primary and secondary Least Squares adjustments peculiar to GPS surveying. (It is assumed that there is no problem with the software or the functional models underlying the adjustments.)



There are three categories of quality control measures that can be applied sequentially:

- Quality control measures related to evidence gleaned from the station logs, as well as that obtained from other sources (for example, relating to the GPS system health).
- Those quality control measures that relate to the <u>internal consistency of the GPS-only network adjustment</u> (the so-called "minimally constrained" solution). *Only when the quality of this solution is assured is the third phase of quality assessment commenced.*
- Those quality control measures that <u>incorporate external position information</u> from geodetic control stations or other previously coordinated points.

For the moment the focus will be on those quality control issues related to the field operations and the minimally constrained GPS network solution. With regard to the latter, it is possible, in addition, to distinguish between:

- Those procedures and quality indicators that are relevant for single session (perhaps even single baseline) results. These involve **DATA QUALITY and SUFFICIENCY** issues.
- Those relevant for multi-session solutions. These test the **CONSISTENCY and RELIABILITY of the COORDINATE RESULTS**.

10.3.1 SINGLE-SESSION TESTS

Prior to Processing

There are a number of checks that can be made before any data is processed. These can be carried out in the field, or immediately after the observation session's, or day's, data is transferred to the field (or local computation) office. They include the following:

- Was the height of antenna measured? Was it checked (by double measurement, or by another person)? Do the values appear reasonable? Was the correct phase centre measured to?
- If the antenna was set up on an eccentric station, were the appropriate measurements made to connect the surveyed point to the ground mark?
- Does the station log indicate any instrument or power problems? Were the correct satellites (as noted in the log) observed? Were the observation times as planned?
- Were all field procedures correctly carried out? Is there any evidence to suggest that the data is in any way unreliable?
- Was the correct site occupied? (Check photo evidence, pencil rubbing of ground mark, etc.)
- Check system status via electronic bulletin board or information service (§3.4).

Baseline Processing

The primary source of quality indicators is the GPS data solution itself (including pre-processing for cycle slips, the triple-difference solution, etc.). We can differentiate between *preliminary* processing for data validation purposes and the *final* GPS data processing. The strategy and options used for the final processing are often based on the outcomes of the preliminary solutions.

The following are some of the quality information that can be gleaned during the preliminary GPS data reduction phase:

- Check the integrity of the data after download, and during reformatting (there may be internal checks for this). Backup copies of the data should be made and verified (see §5.4).
- Scan data on a station-by-station basis, to determine the distribution of data to satellites (including the number of obstructions, breaks in data, cycle slips, etc.), the length of the observation session, the satellites tracked (and their health status, etc.). If the common tracking for particular baselines does not satisfy certain criteria (length of session, number of satellites, etc.), a search should be made for station pairs that

are best in this regard. *The outcome is a station set suitable for processing, as well as initial indications of the best receiver pairings for baseline definition or processing.*

- Triple-difference solutions, baseline-by-baseline, will provide good apriori coordinates for the final double-difference solution. They can be performed before or after data pre-processing for cycle slip detection and repair. *If the data has been "cleaned" prior to the triple-difference solution, any data outliers (data with large residuals) may be uncorrected cycle slips still in the data.* The data pre-processing procedures should then be rerun.
- If all indications are that the data collection process was trouble-free and the data quality has been validated, the double-difference solution can be attempted.

The above remarks are applied to processing of data collected for conventional static GPS surveys. Because of the short observation sessions and entirely different processing steps followed in the case of high productivity survey techniques such as "rapid static", "stop & go", etc. (§5.5), quality control procedures for data processing are minimal. *Instead, the network processing step provides the basis for quality control.*

With regards to conventional baseline processing, a number of steps can be taken to satisfy the analyst that the data is indeed sound and the results reliable. Some information useful for quality checks:

- Type of solution --> triple-, ambiguity-free or ambiguity-fixed.
- The estimate of height difference is of the order of 2-3 times less accurate than the horizontal components.
- Input and output coordinates (solve-for and fixed), in various systems, for example Cartesian, geodetic, topocentric. Mean position from single-point pseudo-range solutions.
- Echo of receiver (serial numbers, etc.) and station information (site name, antenna height, etc.).
- Estimated standard deviation of coordinate components (from variance-covariance matrix).
- The correlation matrix or VCV matrix of the coordinate parameters.
- Optional estimate of quality of satellite geometry, for example indicators such as "RDOP", etc.
- Tracking data acquired, logging times at individual sites, tracking channels used, satellites tracked, signal quality flags, etc.
- Number of observations used in solution, as well as those rejected, sampling rate used, data edit criteria, etc.
- Summary of ephemeris information, health warning flags in Navigation Message, etc.
- Any data reduction performed, for example the tropospheric bias model used.
- Indicator of fit of observations to final solution (the residuals) via some indicator such as the "rms tracking".
- Possibly the results of statistical tests on the residuals.
- If ambiguity-fixed solution attempted, summary of number of ambiguities resolved.
- Possibly a summary recommendation as to the quality of the solution.

In general, some of the above information is examined in detail, and an assessment made as to whether they are "reasonable" or "acceptable". If they are "reasonable", that is:

(a) they are what is expected from past experience, or
(b) they match the manufacturers' specifications, or
(c) they pass the requirements set down by the standards and practices for this class of survey, or

(d) they satisfy requirements set out in the original contract,

the first phase of quality control assessment has been completed.

It must be emphasised that there are no "hard-and-fast" rules. Rarely does GPS reduction software give the analyst access to the observation residuals, and other quantities that would allow extensive statistical tests to be applied. It is therefore important for the analyst to be aware of certain characteristics of double-difference solutions (§7.3 and §8.1) that could be considered "rules-of-thumb". Some of these are:

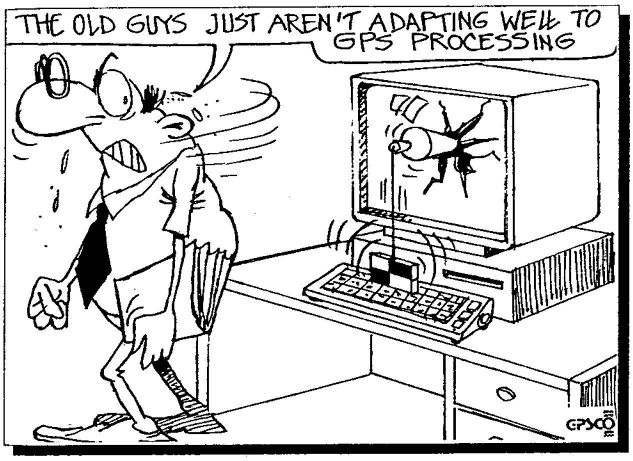
- The chances of successful ambiguity resolution is essentially a function of baseline length, number of satellites tracked, and length of observation session.
- The "rms of tracking" increases with increasing baseline length.
- The coordinate standard deviations are lower for an ambiguity-fixed solution than for an ambiguity-free solution.
- If there are doubts concerning the quality of the ambiguity-fixed solution, it is preferable to accept the ambiguity-free solution in its place.
- If the ambiguity-free solution indicates high "rms tracking" values, for example because the baselines are > 50km, the triple-difference solution may be preferable.
- Improvement in the modelling of biases (for example, through the use of dual-frequency instrumentation), or sophistication of solution (rigorous adjustment, additional parameters, etc.) leads to better and more reliable results for the same length baseline.

A quality assessment can be made on the basis of the session solution, either derived from single baseline or multi-baseline processing. Some recomputation of the session solution may confirm any suspicions concerning <u>data quality</u>, and even permit the source to be pinpointed. No checks can be made on the <u>consistency of the results</u>, as that requires examination of the multi-session solution. However, where a problem related to data integrity has been identified (or is suspected) in a session double-difference solution, additional "trouble-shooting" can be carried out:

- For example, <u>all baselines can be processed separately</u>. It may indicate which of the independent baselines are "best" (those for which the ambiguities were resolved, or the greatest period of common tracking of satellites occurs). *These baselines could then be explicitly identified for the final adjustment.*
- Baselines may be <u>processed in the two directions</u> (fixing first one end of the baseline and solving for the other, and then reversing the process). This could, for example, check the entry of such parameters as the antenna height, especially if the processing were carried out by different analysts.
- The <u>processing of the baselines can be carried out twice</u>, by different analysts, preferably using different processing options, to obtain an appreciation of the sensitivity of the GPS solutions to operator subjectivity. Although most processing software has default options to cater for the majority of GPS jobs, check processing using different analysts and options can partially guard against institutional "mind-set" taking hold with regards to GPS data processing.

The outcome is either a satisfactory session solution or an unsatisfactory session solution which cannot be "salvaged" by any amount of re-processing, and which would have to be scrapped. If the quality testing were being carried out as the GPS survey was progressing, it is generally a relatively

simple matter to reoccupy the stations and repeat the observations.



10.3.2 MULTI-SESSION TESTS

With regards to multi-session testing, the following comments can be made:

- The primary source of information to gauge the ultimate quality of the minimally constrained (multi-session) network is obtained from the network adjustment program. In particular, certain statistical tests can be applied to the network results, as much of the data that the analyst needs is often provided. (Many network adjustment programs will carry out a battery of statistical tests as a matter of course.)
- Multi-session testing is capable of (a) confirming the existence of data quality problems affecting the session results, AND (b) identifying any anomalous result arising from inconsistent or incorrect station occupation, antenna height measurements, etc.
- The level of quality control information in relation to both (a) and (b) is, however, directly proportional to the number of redundant connections (between sessions) in the GPS network.
- The number and type of redundant connections is often defined by recommended standards and specification (§10.2), contract obligations, etc.
- Greater benefit will be obtained if some (or all) sites are occupied more than once in order to satisfy the requirement for redundancy, rather than having more than one baseline in a session terminating at a site.
- The greater the number of redundancies, the better the chance of pinpointing the problem through a process of elimination (or "trouble-shooting").
- The quality testing may also be influenced by the nature of the individual session solutions. The "detectability" of type (a) and (b) problems is to some extent dependent on whether processing is by the single baseline or multi-baseline mode.
- Remedial action can be: (i) leave out a suspect station, or (ii) possible correction of type (b) problem, or (iii) reobserve all or some of the session stations.

Some statistical information derived from secondary adjustments that should be evaluated (though not an exhaustive list!) include:

- Network variance factor and the degrees of freedom. *The VF can be influenced through the modification of the baseline VCV matrices as described in §9.4.*
- The RMS, minimum and maximum residual values, and the standard deviation of the absolute observation residuals.
- The standardised residuals (absolute residuals divided by their propagated standard error) can be compared against the Chi-square test (tests appropriateness of apriori errors) and the Tau criterion (for data outlier detection), see §9.1.
- Aposteriori errors at the 95% (2 sigma) confidence level for the adjusted station coordinates and for the relative positions for all station combinations. *Error ellipses/ellipsoids provide a useful evaluation of the station solution confidence: nearly circular and uniformly small error figures are an indication of a well-conditioned network, while irregularly shaped or unusually large error figures indicate problems with solutions or weakness in network design.*
- Relative confidence between station combinations can be evaluated according to the claimed geometric accuracy standard. *The propagated relative error (at the 95% confidence level) should be less than the maximum allowable for the accuracy standard sought.*
- Prior to carrying out a constrained adjustment the existing control coordinates should be examined and transformation parameters determined. The rms and standard deviation of the computed transformation parameters, and the coordinates, should be investigated.
- Any significant changes between the statistics of the minimally constrained adjustment and the constrained adjustment should be flagged and investigated.

The Nature and Detectability of Observational Errors

To master multi-session testing, an understanding of the nature of the observational errors and how they propagate through session and, ultimately, network solutions is necessary.

There are two classes of errors:

(1) Those that **cannot be detected because they do not propagate into baselines containing redundant stations.**

(2) Those that **are detectable because they propagate into the coordinates of redundant stations.**

These statements need amplification, and should be compared with the notions of "quality control" as practised in conventional geodetic networks:

- Data quality problems will propagate into other station coordinates during a session (and ultimately through the entire network) if the station at which the problem has occurred is connected directly to other stations (as in a traverse). The effect of the error on other stations "downstream" is unpredictable, as is the case of an error in the early part of a traverse. *If the station were part of a "spur" (or radiation in conventional survey practice), then the problem will be undetectable.*
- Each instrument setup in a traverse (from which backsights and foresight observations are made) is analogous to a GPS session. *It is the sites common to two sessions that provide the link in the "GPS traverse".*

- In a similar way, any problem with a station setup is detectable "downstream" only if the station is occupied in at least two independent sessions, and if one was "correct" and the other was "incorrect". If the same mistake (for example, misidentification of station) is made both times, the error is *undetected* (until perhaps some years later when this station is connected into another survey network). This is equivalent to the foresight station in the previous setup not being the same as the instrument station in the current setup. Hence an inconsistency is introduced, and although the measurement process is satisfactory, all subsequent station coordinates are corrupted.
- If the station is occupied only once (and hence the mistake occurs once) then, as in the case of the same mistake being made twice, the error lies dormant. This is equivalent to a single radiation from the main traverse line. The radiated point cannot be checked.

Redundancy

"Redundancy", in relation to a multi-session network solution, is provided by the multiple occupation of a station (for two or more sessions) *over and above that required to transfer datum from one session to another.* Thus if more than one of the stations in a session has been previously occupied, then there is a redundancy (equal to the number of stations occupied more than once, minus one).

The following comments may be made regarding to GPS network redundancy:

- They provide independent "pathways" joining all the network stations (Figure 10.3-1). Investigation of the closes around different pathways can assist in pinpointing the source of inconsistent session results.

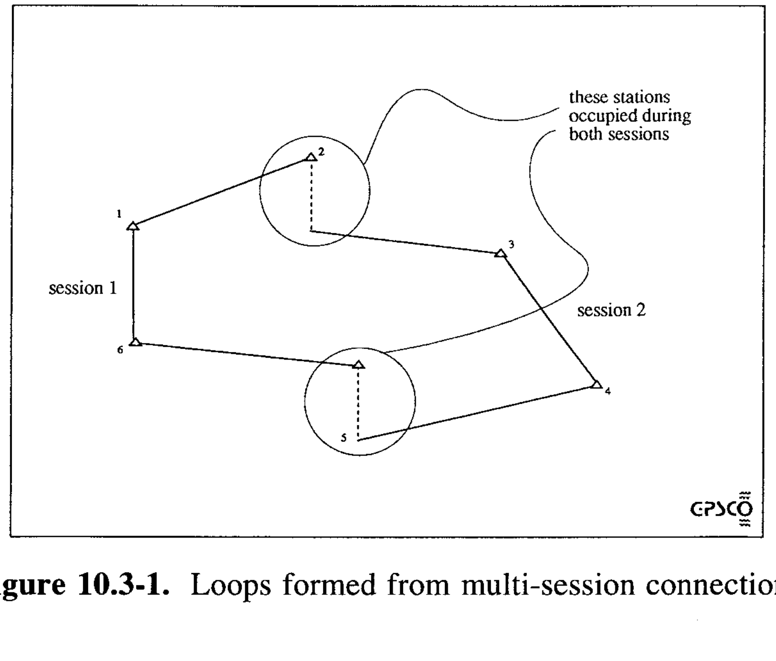


Figure 10.3-1. Loops formed from multi-session connections.

- The multiple occupation of a pair of stations in the same sessions leads to repeat baseline results (Figure 10.3-2), which can provide a measure of GPS repeatability.

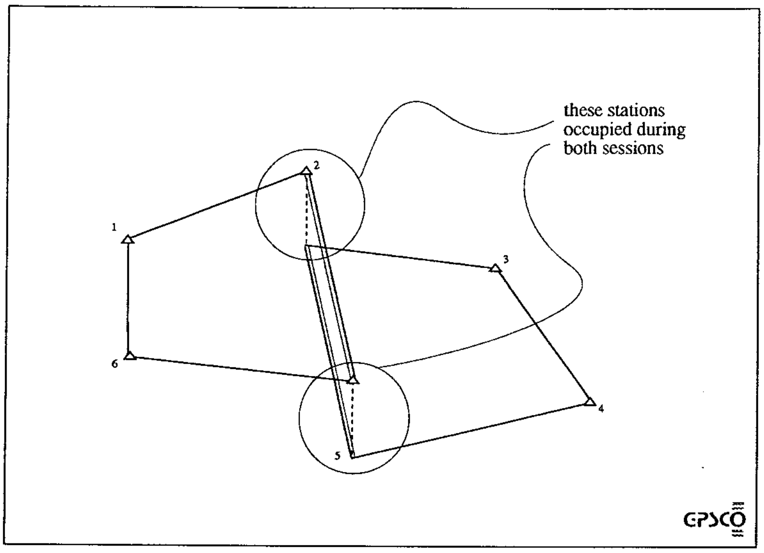


Figure 10.3-2. Repeated baselines.

- The amount of redundancy that is required (multiple stations or baselines) is often explicitly indicated in the standards and specifications for GPS surveys (§10.2). They may in fact be integral to the classification of the GPS survey.

- The provision of adequate redundancy is therefore an integral part of the network design process (§5.2).

Loop Closure Tests

Quasi-independent "loops" can be constructed by linking stations from different sessions (Figure 10.3-1). They are useful whether the multi-session solution is performed with the aid of network software, or the vectors extracted manually from several adjustments. They may serve two purposes:

(1) A sampling of loops can provide verification of the overall consistency of the GPS network results. This may be required in the U.S. to satisfy the authorities as to the "class" of GPS survey performed (Table 10.2-8). In this case, such parameters as the maximum and average allowable misclose may need to be explicitly stated. Note, however, that if the multi-session solution is being accumulated within a geodetic network program, the final summary residuals and station coordinate standard deviations provide the same information. *Hence Australian geodetic authorities do not require an explicit loop closure test to be performed.*

(2) Judicious selection of loops can assist with problem diagnosis. If a problem has been detected by first carrying out a GPS (multi-session) network solution and studying the residuals, etc., the aim may be to see if it is possible to salvage part of a session solution by eliminating the problem station from the final adjustment. Generally, however, such tests are carried out to find out "where" the problem is, and "why". *The most likely outcome is to reobserve if there is insufficient redundancy.*

Standards & Specifications: Quality Control Guidelines

The basic criteria for survey classification is the relative accuracy of neighbouring stations. These are given in Tables 5.2-1 (for Australian recommended standards and practices) and 5.2-2 (for the U.S. standards and specifications), and are the values which must not be exceeded by the semi-major axes of the relative error ellipses (or ellipsoids). The relative error is composed of two parts: a constant part (a) and a length-dependent part (b):

$$e = a + b.L \qquad \text{(for Australia)}$$

$$e = \sqrt{a^2 + (b.L)^2} \qquad \text{(for the USA)}$$

where L is the interstation distance in kilometres, e and a are in millimetres, and b is in parts per million.

The CLASS of the GPS survey is dependent on whether the semi-major axes are all below the error level defined by the appropriate standards & specifications (§10.2). An example of the application of this is given below, in which the magnitudes of the semi-major and semi-minor axes of the relative 2-D error ellipses, and their orientation, have been calculated for some baselines from the Molong (N.S.W.) GPS survey (§9.3). The error ellipses have been plotted in Figure 9.1-2. Note that the last column shows, for the selected baseline, the "class" label that can be applied, according to the survey specifications given in Table 5.2-1.

However, even these values cannot be considered "objective" as they are influenced by the network redundancy. The same network, but compiled using only the independent baseline solutions, can have error ellipses 50% larger (see Figure 9.4-1, compared with Figure 9.1-3). Hence **the classification of the GPS survey can be manipulated!**

RELATIVE HORIZONTAL ERROR ELLIPSES & VERTICAL UNCERTAINTY

AT STATION / TO STATION	DISTANCE / AZIMUTH	HORIZONTAL AXES LENGTH (m)	AZIM (deg)	VERTICAL STD.DEV. (m)	CLASS
PM43494	1798.188	.0064	115.1	.0065	2A
PM69992	50.1	.0064	205.1		
PM43494	3690.144	.0058	51.7	.0060	2A
BRYMEDURA	242.6	.0058	141.7		
PM43494	7273.225	.0065	118.4	.0067	3A
MOLONG	109.1	.0065	208.4		
PM43494	11773.821	.0111	113.7	.0113	3A
GOANNA	288.6	.0109	203.7		
PM69992	1499.537	.0070	123.3	.0071	A
VALE HEAD	46.4	.0069	213.3		
PM69992	6310.905	.0079	132.5	.0080	2A
MOLONG	123.5	.0079	222.5		
PM69992	11494.322	.0121	117.4	.0123	2A
GOANNA	296.4	.0119	207.4		
BRYMEDURA	8982.113	.0060	62.3	.0062	3A
MOLONG	81.6	.0060	152.3		
BRYMEDURA	14278.616	.0113	116.1	.0115	3A
GOANNA	278.3	.0111	206.1		
VALE HEAD	5909.096	.0059	120.2	.0061	3A
MOLONG	132.4	.0059	210.2		
VALE HEAD	11380.181	.0100	115.8	.0102	3A
GOANNA	303.4	.0098	205.8		

10.3.3 REMEDIAL ACTION: SUMMARY AND SOME GUIDELINES

There are a number of remedial actions that can be taken with regard to quality issues:

- Remove the suspect station(s) from the session, and recompute the multi-session adjustment. This is possible only if there are sufficient redundancies (multiple occupancies) so that the network does not suffer by the removal of some of the baselines.
- Correct the problem. This is only feasible in the case of a wrong station occupation or an incorrect antenna height measurement. By using the station logs it may be possible to salvage the necessary information (eccentric station offsets or correct antenna height) without having to reobserve a session.
- Reobserve some or all of the stations in the session, particularly if the entire session is poor due to, for example, data insufficiency (and the ambiguities were therefore not resolved).

The effectiveness of multi-session quality control is enhanced if multiple occupations and care is especially given to the measurement of antenna height. Having different field parties (using different instrumentation) occupy the same station maximises the chance that any station error (such as groundmark misidentification) may be restricted to one setup, and hence is detectable in the multi-session solution. Having different observers note the height of antenna maximises the chance that the correct height is measured (at least once)!

Ensuring that quality control is being carried as the survey is progressing makes it feasible to reobserve sessions with minimum disruption (and cost)!

Data Quality Issues

DATA QUALITY PROBLEMS affect, in the first instance, the quality of a single session solution, and can hence be largely identified at this level. There is, in general, no satisfactory remedy but to discard the data and reobserve some or all the stations in the session. In summary, the factors affecting data quality, and their characteristics, are:

(1) Data quality factors (insufficient epochs, satellites, etc.) affecting a baseline, but is an isolated incident due to poor field procedures.
Detect: *Preliminary processing and session solution quality indicators.*
Remedy: *Reobserve and take greater care in future.*

(2) Data quality characteristic that is site-dependent, for example multipath.
Detect: *Preliminary processing and session solution quality indicators.*
Remedy: *Improve site conditions, or change site location and reobserve. (Could have been prevented by better reconnaissance?)*

(3) Data quality factor which is due to an instrumental problem.
Detect: *Preliminary processing and session solution quality indicators.*
Remedy: *Replace receiver and reobserve affected stations. (Regular calibration of equipment is advised.)*

(4) Data quality factor which affects all sites in a session, for example ionospheric problems or bad Navigation Message.
Detect: *Preliminary processing and session solution quality indicators.*
Remedy: *Reobserve entire session.*

Result Inconsistency Issues

RESULT INCONSISTENCY PROBLEMS affect the quality of a multi-session GPS network solution. They can be detected at this level only if sufficient redundancy is designed into the network, and only if the problem occurs at a station with multiple occupancies. In addition, the degree of redundancy will determine whether remedial action other than merely reobserving some or all stations affected can be taken. In summary, the factors affecting the internal consistency of a network solution, and their characteristics, are:

(1) Incorrect station identification or antenna height error, being an isolated incident due to poor field procedures.

Detect: *Multi-session adjustment residuals, or loop closures. Detectable only if occurred at site with multiple occupancies.*

Remedy: *Drop affected station occupation from multi-session solution, or if not possible, reobserve and take greater care in future!*

(2) Systematic effect at stations which is due to an instrumental problem, for example antenna incorrectly fitted so that electrical phase centre not where it is indicated on housing.

Detect: *Multi-session adjustment residuals, or loop closures involving the same receiver. Detectable only if different receivers are deployed at multiple occupations.*

Remedy: *Replace receiver and reobserve affected stations. (Regular calibration of equipment is advised.)*

If such problems occur at sites occupied only once (that is, no redundancies involving these sites) they are undetectable by any procedure.

Chapter 11: Result Transformation and Presentation

11.1 TRANSFORMING GPS SURVEYING RESULTS: COORDINATE SYSTEMS & DATUMS

The results of a GPS survey, the minimally constrained 3-D network, are generally of limited use to the client as he/she usually requires the coordinates of points to be given in relation to a previously defined geodetic datum. This datum may be:

- one previously established by GPS, or
- a conventional "local" geodetic datum, usually comprising two distinct systems, one for HORIZONTAL position and the other for the HEIGHT component.

Both are essentially handled in the same way, that is, it cannot be assumed that the earlier GPS survey is on the same datum as the present survey, hence it cannot be integrated in the same manner as would individual GPS session solutions into a campaign solution, as described in chapter 9. *As the most common problem is to relate GPS survey results to prior terrestrial surveys, this chapter discusses the procedures for transforming GPS results between coordinate systems in order to facilitate the integration of the results into conventional survey datums.* The procedures are, however, equally applicable to the combination of two GPS surveys, relating the GPS datum to other space datums, and even relating two terrestrial networks.

There are several distinct issues involved:

☞ The MATHEMATICAL RELATIONS between the various commonly used coordinate systems: Cartesian, ellipsoidal, topocentric and projection.

☞ Definition of the TRANSFORMATION between the 3-D coordinate systems involved: that implied by the GPS network on the one hand, and that defined by the local control stations on the other.

☞ GPS HEIGHTING: as it relates to both the transformation process and the direct determination of elevation.

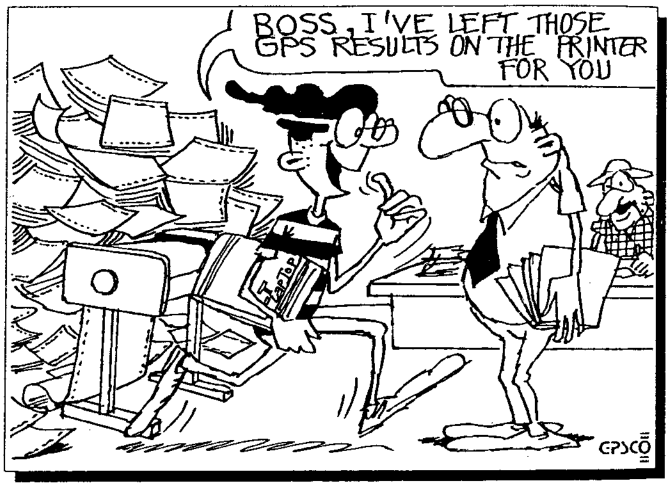
☞ NETWORK VALIDATION and CONSTRAINT: involving the manipulation of the GPS network to make it consistent with the local datum through the incorporation of local control station information.

These topics are dealt with in this and the following chapter. GPS transformation is dealt with both in this section and in §11.2, as it can be carried out in two ways:

(1) Using **published transformation parameters**.

(2) Through the **determination of the appropriate transformation parameters**.

The transformation models are presented in this section, and the transformation options in Australia are given in §11.2. Where the transformation parameters between the GPS datum and the local geodetic datum are to be determined, stations whose coordinates are known in the local geodetic datum must also be occupied during the GPS survey.



11.1.1 GEODETIC DATUMS

The GPS Datum

The commonly used space-based datums are founded on the geocentric (earth-centred) 3-D Cartesian systems defined by a global ensemble of U.S. military tracking stations (specifically those of the Defense Mapping Agency) that have been established over the decades since the start of the Space Age. As far as GPS is concerned this datum is:

(1) maintained by the assigned Cartesian coordinates of the handful of tracking stations making up the Control Segment (§2.2),

(2) transferred to the (time-varying) satellite coordinates broadcast to users in the Navigation Message during the orbit determination process, carried out by the system controllers at the Master Control Station, and

(3) ultimately is realised in a local region by the network coordinates that result from a GPS survey.

The official "GPS datum" is the so-called World Geodetic System 84 (WGS84), nominally defined as having:

- The origin at the earth's centre of mass.
- The Z-axis aligned parallel to the direction of the Conventional Terrestrial Pole (CTP) for polar motion, as defined by the International Earth Rotation Service (formerly the BIH).
- The X-axis being the intersection of the WGS84 Reference Meridian Plane and the

plane of the CTP Equator (the Reference Meridian being parallel to the Zero Meridian defined by the IERS).

- The Y-axis completes a right-handed, earth-centred, earth-fixed (ECEF) orthogonal coordinate system, measured in the plane of the CTP Equator, 90° east of the X-axis.

This global geometric datum requires no reference ellipsoid or geoid model to be specified.

Ultimately, how well the GPS survey results relate to WGS84 is dependent on the quality of the Broadcast Ephemerides in the Navigation Message, and the quality of the Datum Station coordinates within the network. The accuracy of the GPS orbit information (related to WGS84 system) is at the 5-50m level, although it possibly could get worse under Selective Availability (§2.4).

Because GPS surveying techniques only provide relative position, and hence are insensitive to absolute coordinates (that is, those relative to the WGS84 / geocentre origin), the coordinates of the datum station in effect defines the origin of the GPS survey in what is in reality a "near-WGS84" system. That is:

- the origin is defined by the vector of the datum station coordinates (if based on a pseudo-range point position it may be up to 100 metres in error).
- the scale of the system may vary depending upon the accuracy of modelling the phase observations (for example, single frequency results may be biased by the neglected ionospheric effects), hence different GPS survey networks may have different implicit scales.
- the orientation of the coordinate axes implicit in the GPS network results in relation to the WGS84 system although probably comparatively well defined, there may nevertheless be a deviation of a few arcseconds.

Hence, how well the Datum Station in the GPS network is related to WGS84 will influence how well the GPS results are referred to the WGS84 system. The implication therefore is that: (a) there may be a suitable set of published transformation parameters that will precisely relate the GPS results to a local datum, or (b) the transformation parameters relating the GPS survey to a local datum may have to be determined from the survey data itself.

The Conventional Geodetic Datum

In general, conventional geodetic datums are "locally defined". The label "local" may refer to an arbitrary datum of convenience, having relevance only in a very localised area, or it may be the more familiar "national" geodetic datum. There are in general two distinct components:

- HORIZONTAL datum system based on curvilinear (ellipsoidal) coordinates defined in relation to some "reference ellipsoid", and
- an independent VERTICAL datum based on a local geoid definition (at one or more tide gauge, or "vertical datum" stations).

The datum is in general *non-geocentric*, the origin being the centre of the reference ellipsoid associated with the datum. However, the datum origin plays no part in defining the vertical datum. In relation to the horizontal datum:

- The datum is, in the first instance, defined (origin and curvature) by the selection of a best-fitting ellipsoid to the geoid over region of the datum.

- It is realised "on the ground" by a single Datum Station, and the network of control stations established from and linked to this station.
- The Z-axis coincides with the semi-minor axis of the reference ellipsoid, the X-axis passes through the point λ=0°, ϕ=0°, and the Y-axis completes the right-handed triad system.

In the Australian context, the datums are:

(1) The **Australian Geodetic Datum** (AGD) for horizontal control comprising several thousand geodetic control stations distributed across Australia (Figure 11.1-1). In particular the AGD is realised by the AGD66 coordinate set in the states of New South Wales, Victoria, Tasmania and the Northern Territory; and the AGD84 coordinate set in the remaining states of Australia (ALLMAN & VEENSTRA, 1984). The origin station is the Johnston Trigonometrical Station, with the *assigned* coordinates (NMC, 1986):

ϕ = 25°56'54.5515"S
λ = 133°12'30.0771"E
h = 571.2m (ellipsoidal height)

and the associated reference ellipsoid is the Australian National Spheroid, defined by the parameters:

semi-major axis	6378160m
inverse of ellipsoidal flattening	298.25

(2) The **Australian Height Datum** (AHD) for vertical control is realised by several thousand level benchmarks across Australia (Figure 11.1-2). However, the zero height datum surface is not coincident with any unique equipotential surface of the earth's gravity field, but is the surface of mean sea level as defined by 30 tide gauge sites around the coast of Australia (ROELSE et al, 1971).

The Australian datum is in the process of being redefined (§12.1). The new datum is known as the **Geocentric Datum of Australia** (GDA) (MANNING & HARVEY, 1994). As the name implies, this datum is "geocentric" and hence very close to the WGS84 datum. However, what is even more remarkable is that the basic framework of the GDA is provided by a super-precise GPS survey linking together selected, evenly-distributed geodetic control stations (at an approximate spacing of 500km) with permanent GPS tracking stations around the world. Such a datum redefinition has already occurred in the U.S. with respect to the North American Datum 1983 (NAD83).

Despite the different bases of the coordinate datums (whether GPS or conventional), all geodetic datums may be related to one another, or with any of the global space datums (including WGS84), by a number of transformation models. Associated with the transformation process are the conversion procedures relating ellipsoidal, topocentric and Cartesian coordinates (and any variance-covariance information), the projection coordinate systems and the issue of height systems.

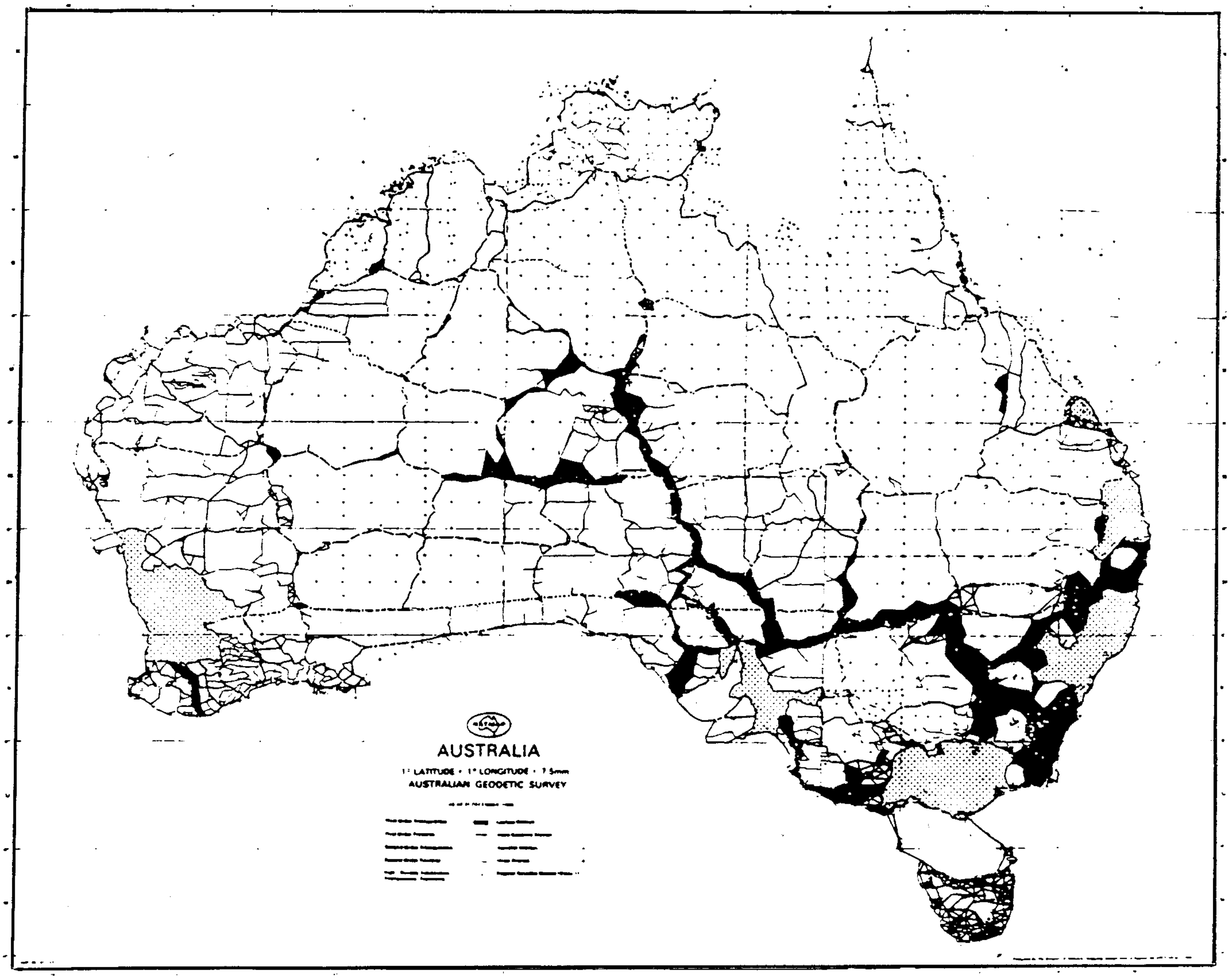
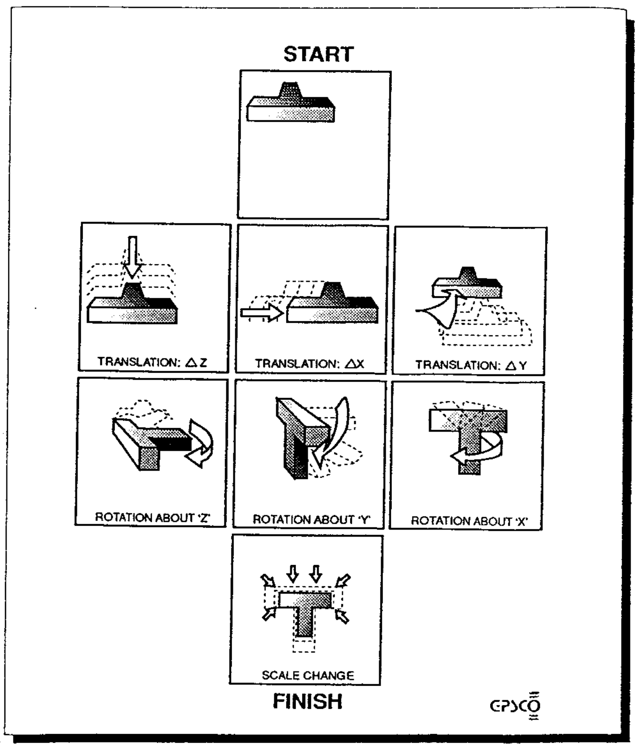


Figure 11.1-1. Layout of the Australian Geodetic Survey.



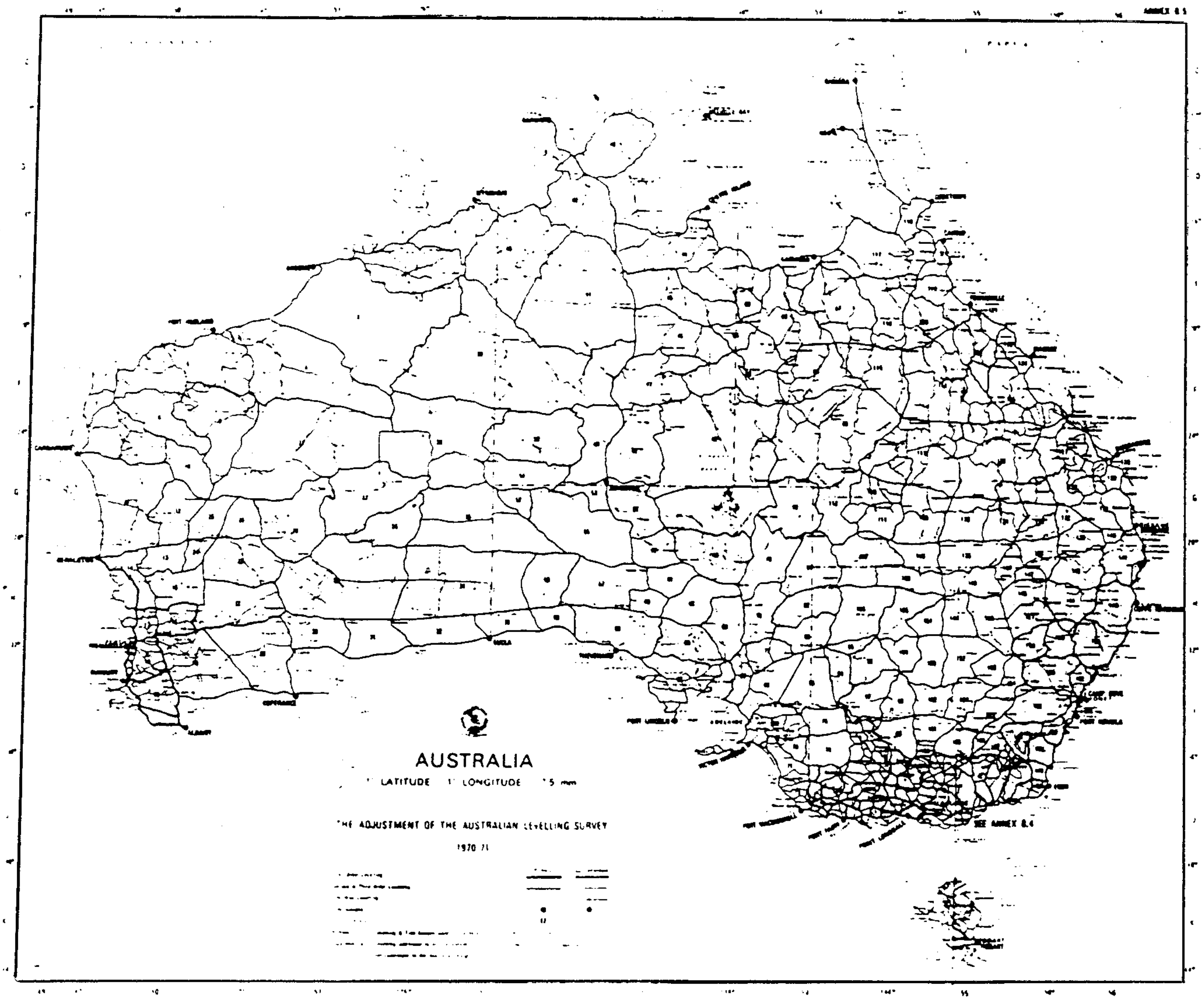


Figure 11.1-2. Lines of the Australian Levelling Survey.

11.1.2 TRANSFORMATION MODELS AND PROCEDURES

There are a number of ways of defining the relationship between one reference system and another. (In subsequent discussions the reference system coordinate sets will simply be referred to as "networks", and although implying they are distinct collections of points, they are in fact often the same physical points but for which two or more sets of coordinates are available.) The choice of the most appropriate network transformation model is influenced by such factors as:

- Whether the model is to be applied to a small area, or over a large region.
- Whether one (or both) networks have significant distortions.
- Whether the networks are three-dimensional in nature, 2-D or even 1-D.
- The accuracy required.
- Whether the transformation parameters are available, or must be determined.

The most general of the transformations is the **affine transformation**. An affine transformation transforms straight lines to straight lines and parallel lines remain parallel. Generally the size, shape, position, and orientation of lines in a network will change. The scale factor depends on the orientation of the line but not on its position within the net. Hence the lengths of all lines in a certain direction are multiplied by the same scalar. Alternatively it is possible to define a **projection transformation** where the scale factor is also a function of position.

A transformation in which the scale factor is the same in all directions is called a **similarity transformation**, and is by far the most widely used of the transformation models. A similarity transformation preserves shape, so angles will not change, but the lengths of lines and the position of points may change. An **orthogonal transformation** is a similarity transformation in which the scale factor is unity. In this case the angles and distances within the network will not change, but the positions of points do change on transformation. The 3-D similarity transformation model relating coordinates of points in the $X_BY_BZ_B$ network to coordinates in the $X_AY_AZ_A$ network is:

$$\begin{pmatrix} X_B \\ Y_B \\ Z_B \end{pmatrix} = s \, . \, \mathbf{R} \, . \begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \qquad (11.1\text{-}1)$$

where s is the scale factor and $\mathbf{R}$ is a 3x3 orthogonal rotation matrix (eqn (11.1-4)). Note that there are seven parameters: three rotation angles, three translation components and one scale factor. The translation terms T_x, T_y, T_z are the coordinates of the origin of the $X_AY_AZ_A$ net in the frame of the $X_BY_BZ_B$ net.

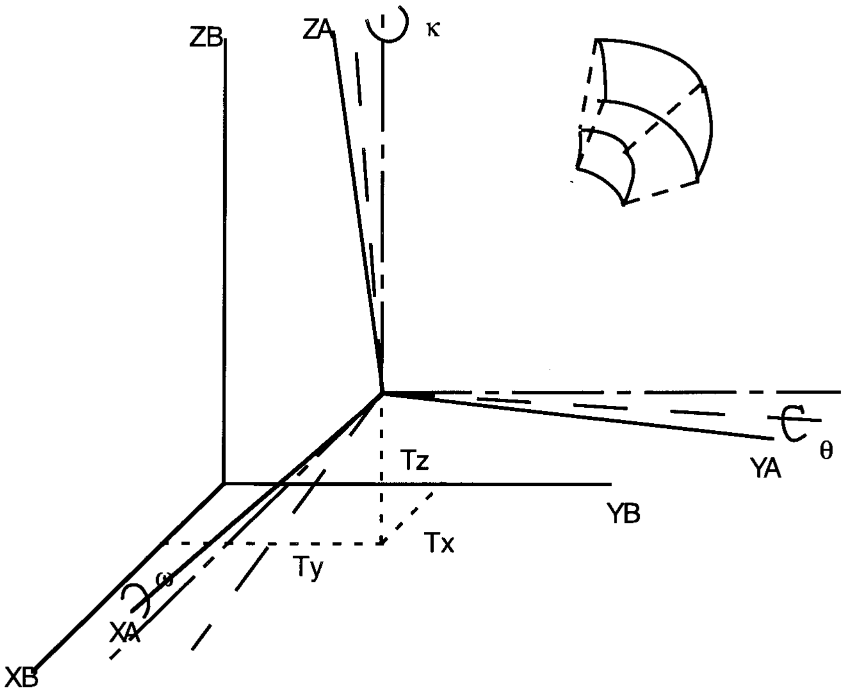


Figure 11.1-3. The seven parameter 3-D similarity transformation model.

It is presumptuous to assume that similarity transformations, rather than affine or projection transformations, correctly describe the differences between any two datums, or coordinate sets, but there's no denying their popularity. **Why are similarity transformations so popular?** Some of the reasons are:

- A comparatively small number of parameters are involved, the minimum necessary to account for differences between 3-D reference systems: between their two origins, and the differences in the orientation and lengths of the respective coordinate axes.
- It is a simple model that can be implemented easily in software (there are several computational procedures to chose from).
- If the seven parameters must be determined from common stations in the two networks, it is preferable to estimate a small number of parameters as there are generally only a limited number of common points.
- The model is adequate for relating two networks which are homogeneous (that is, no local distortion in scale or orientation).

There are, however, circumstances where even the seven parameter similarity transformation model is too elaborate, and a number of simpler models may be adequate:

- The simplest procedure is not to use an algorithmic model at all. An example would be a contour map of the differences between the coordinate sets for the common points of two networks. Such a procedure is adequate for converting the coordinates of the non-common points from one reference system to the another, but no mathematical relationship is explicitly established between the reference systems.
- Another option is to merely use a subset of the similarity transformation parameters. Because geodetic networks normally cover only small portions of the earth's surface, some transformation parameters are highly correlated. Often the rotation parameters are

insignificant and/or poorly determined. Hence a transformation model based on, say, a simple three parameter block shift (no scale or rotation parameters) can be used.

- If the networks are relatively small in extent, of the order of less than say 100km width and breadth, a local 2-D transformation may be defined. By converting the Cartesian or ellipsoidal coordinates to topocentric quantities, the three axes are oriented (east, north, and normal to the horizontal plane) in such a way as to solve the original 3-D transformation problem through separate horizontal and vertical procedures. For example, the horizontal transformation may involve two block shift parameters (along the east, north axes), a rotation about the normal axis, and a scale factor.

The two-dimensional transformation is described in HOFMANN-WELLENHOF et al (1994).

The similarity transformation model may be considered a suitable compromise between elaborate models such as the affine or projection transformations on the one hand, and crude limited parameter models on the other. *When used wisely, and with an appreciation of its shortcomings, the similarity transformation is ideal for relating 3-D GPS networks to other GPS or terrestrial networks.* In particular, using a similarity transformation on a large network may distort local scale and orientation. Therefore an important consideration is the magnitude of local distortions in scale and orientation. In this regard it should be noted that:

- Similarity transformations will tend to smooth out local distortions.
- It may be more appropriate to divide a region into smaller zones, each with their own set of transformation parameters.
- The results of the transformation may be unreliable outside the area spanned by the common points used to derive the transformation parameters.

The issue of network distortions, and the common strategy of determining *"tailored" transformation parameters for small areas*, requires that a distinction be made between:

(1) Universal or global model parameters, used, for example, to relate two homogeneous satellite datums and derived from globally distributed network data. The transformation parameters in Table 1.2-2 are examples of such global models.

(2) National parameters, intended for use on a datum-wide basis. They are often derived by the national geodetic authority and typically relate a satellite datum to the national geodetic datum, for example WGS84 and AGD84 (see later discussions on Australian options).

(3) Local parameters are determined for any given area, for example, by a surveyor so that he/she may relate his/her survey work to the surrounding network. This is often the case for GPS surveys.

The Similarity Transformation Model

In the geodetic context, the general transformation model in eqn (11.1-1) is often referred to as the **Bursa-Wolf model**. When this model is invoked for small networks, the rotation parameters are highly correlated with the translation parameters. (The reader can convince themselves of this by considering, for example, a rotation about the Z-axis of a network on the Greenwich meridian; the effect is almost indistinguishable from a translation of the network along the Y-axis.) An alternative formulation that avoids this correlation "problem" is the **Molodensky-Badekas model** (HARVEY, 1986):

$$\begin{pmatrix} X_B \\ Y_B \\ Z_B \end{pmatrix} = \begin{pmatrix} X_m \\ Y_m \\ Z_m \end{pmatrix} + s \,.\, \mathbf{R} \,.\, \begin{pmatrix} X_A - X_m \\ Y_A - Y_m \\ Z_A - Z_m \end{pmatrix} + \begin{pmatrix} T_x' \\ T_y' \\ T_z' \end{pmatrix} \qquad (11.1\text{-}2)$$

where $X_m = \Sigma X_{Ai}/n$, $Y_m = \Sigma Y_{Ai}/n$, $Z_m = \Sigma Z_{Ai}/n$ are the coordinates of the centroid of the network. Alternatively X_m, Y_m, Z_m may be selected to be the coordinates of one of the points in network A. Although the translation parameters are different, the rotation matrix and the scale factor are the same as for the Bursa-Wolf model.

Rotation Matrices

The rotation matrices about the X-, Y-, and Z-axes are:

$$\mathbf{R_z}(\kappa) = \begin{pmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{R_y}(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \quad \mathbf{R_x}(\omega) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{pmatrix} \qquad (11.1\text{-}3)$$

The most common combined rotation matrix is: $\mathbf{R} = \mathbf{R_z}(\kappa).\mathbf{R_y}(\theta).\mathbf{R_x}(\omega)$, leading to:

$$\mathbf{R} = \begin{pmatrix} \cos\kappa\cos\theta & \cos\kappa\sin\theta\sin\omega + \sin\kappa\cos\omega & \sin\kappa\sin\omega - \cos\kappa\sin\theta\cos\omega \\ -\sin\kappa\cos\theta & \cos\kappa\cos\omega - \sin\kappa\sin\theta\sin\omega & \sin\kappa\sin\theta\cos\omega + \cos\kappa\sin\omega \\ \sin\theta & -\cos\theta\sin\omega & \cos\theta\cos\omega \end{pmatrix} \qquad (11.1\text{-}4)$$

For small rotations this matrix may be approximated by:

$$\mathbf{R} \approx \begin{pmatrix} 1 & \kappa & -\theta \\ -\kappa & 1 & \omega \\ \theta & -\omega & 1 \end{pmatrix} \qquad (11.1\text{-}5)$$

where ω, θ, and κ are the rotation angles in radians about the X-, Y-, and Z-axes respectively. The small angles assumption is usually valid for rotation angles up to 10". The rotation angles depend on the baseline vectors (that is, the relative positions) and not on the absolute coordinates. Thus it does not matter where the origin of coordinates is because the estimated rotation angles will be the same. For this reason the same rotation matrix is used in the Bursa-Wolf model as in the Molodensky-Badekas model.

Scale factor

A scale factor can be visualised as follows. Imagine a network drawn on the surface of an inflatable sphere. As the sphere is inflated, the points of the network spread apart from each other, and from the centre of the sphere (Figure 11.1-3). The inflation of the sphere is equivalent to the application of a scale factor greater than unity.

Multiplication of a set of rectangular Cartesian coordinates by a scale factor is identical to multiplying the corresponding baseline lengths by the same scale factor. Hence the scale factor can be determined from either the 3-D site coordinates or from the baseline lengths. Thus, as with the rotation angles, the origin of the coordinates has no affect on the results.

In the case of ellipsoidal coordinates the longitude is not affected by a scale change but the geodetic latitude does change slightly. For example, a 1ppm scale change will change geodetic latitudes by less than about 0.0007" (2cm). However, the effect on ellipsoidal height is significant. For example, a 1ppm scale change will produce a change in height of about 6.4 metres.

Ellipsoidal Transformations

Eqns (11.1-1) and (11.1-2) relate sets of Cartesian coordinates. There are also formulae that relate two sets of *ellipsoidal* coordinates. In addition to accounting for the differences in scale, origin and orientation between the two datums, they must also relate the results on different ellipsoids (due to possible differences in the size and shape of the ellipsoids). These formulae are collectively referred to as the **Molodensky Formulae**. They are particularly useful for 2-D transformations as only the effect on latitude and longitude need be considered (the height transformation is defined by eqn (11.3-8)). The *abridged* Molodensky Formulae (accounting for only differences in origin and the size/shape of the reference ellipsoids), giving the corrections (in arcseconds) to the ellipsoidal coordinates of Datum 1, to transform them to Datum 2, are:

$$\Delta\phi = (-\Delta X.\sin\phi.\cos\lambda - \Delta Y.\sin\phi.\sin\lambda + \Delta Z.\cos\phi + (a.\Delta f + \Delta a.f).\sin 2\phi) \,/\, (R_M.\sin(1\ \text{second}))$$

$$\Delta\lambda = (-\Delta X.\sin\lambda + \Delta Y.\cos\lambda) \,/\, (R_N.\cos\phi.\sin(1\ \text{second})) \qquad (11.1\text{-}6)$$

where:

a	is the semi-major axis of the Datum 1 ellipsoid
b	is the semi-minor axis of the Datum 1 ellipsoid
$\Delta X, \Delta Y, \Delta Z$	are the offsets of the Datum 1 ellipsoid relative to the Datum 2 ellipsoid origin
Δa	is the change in semi-major axis (from Datum 1 ellipsoid to Datum 2 ellipsoid)
Δf	is the change in flattening (from Datum 1 ellipsoid to Datum 2 ellipsoid)

$$R_N = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$ where e is the eccentricity of the Datum 1 ellipsoid

$$R_M = \frac{a(1-e^2)}{(1-e^2\sin^2\phi)^{1.5}}$$

An example of the application of these corrections is for the definition of the approximate ellipsoidal transformation from WGS84 to the Cape Datum (South Africa). The corrections have been computed on a grid and available to users in the form of contour maps (Figure 11.1-4). In the case of the transformation from the Australian datum AGD66 to WGS84, the values of $\Delta X, \Delta Y, \Delta Z$ are -133, -48, 148 metres, respectively; and from AGD84 to WGS84, the values of $\Delta X, \Delta Y, \Delta Z$ are -134, -48, 149 metres, respectively. The values of $\Delta a, \Delta f$ are, in both cases, -23m and -0.000000081204, respectively.

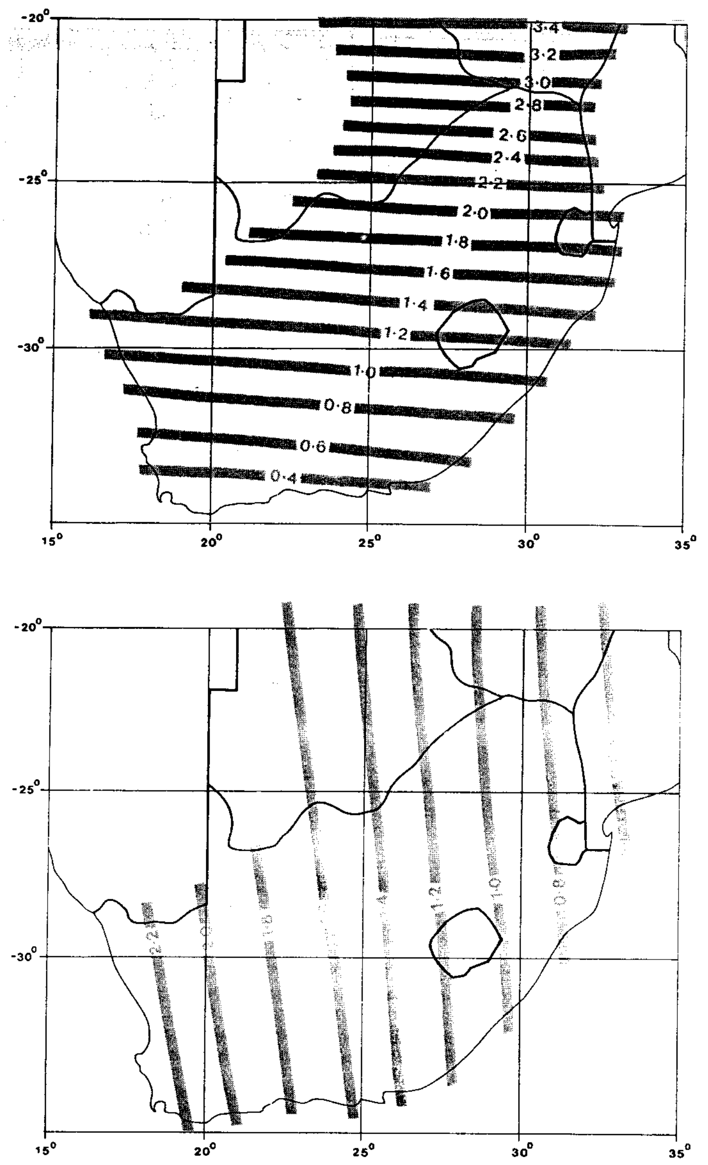


Figure 11.1-4. Corrections to latitude (top) and longitude (bottom) from WGS84 to Cape Datum - South Africa (in arcseconds), based on the abridged Molodensky transformation formulae.

VCV of Transformed Coordinates

When known parameters are used to transform coordinates the variance-covariance (VCV) matrix of these coordinates should also be transformed. The VCV of transformed coordinates can be determined by applying the Law of Propagation of Variances. The following assumes small rotation angles, but similar equations could be derived for the full **R** matrix if required. Assuming the following transformation model (eqn (11.1-1)):

$$\begin{pmatrix} X_B \\ Y_B \\ Z_B \end{pmatrix} = s \cdot \begin{pmatrix} 1 & \kappa & -\theta \\ -\kappa & 1 & \omega \\ \theta & -\omega & 1 \end{pmatrix} \cdot \begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \tag{11.1-7}$$

The appropriate transformed VCV is:

$$\mathbf{VCV_{XYZ'B'}} = \mathbf{J} \cdot \begin{pmatrix} \mathbf{VCV_{XYZ'A'}} & \mathbf{0} \\ \mathbf{0} & \mathbf{VCV_P} \end{pmatrix} \cdot \mathbf{J^T} \tag{11.1-8}$$

where:

$$\mathbf{J} = \begin{pmatrix} s & s\kappa & -s\theta & X_A+\kappa Y_A-\theta Z_A & 0 & -sZ_A & sY_A & 1 & 0 & 0 \\ -s\kappa & s & s\omega & Y_A-\kappa X_A+\omega Z_A & sZ_A & 0 & -sX_A & 0 & 1 & 0 \\ s\theta & -s\omega & s & \theta X_A-\omega Y_A+Z_A & -sY_A & sX_A & 0 & 0 & 0 & 1 \end{pmatrix}$$

In $\mathbf{VCV_P}$ (the 7x7 VCV matrix of the parameters) the transformation parameters are in the order: s, ω, θ, κ, T_x, T_y, T_z. $\mathbf{VCV_{XYZ'A'}}$ and $\mathbf{VCV_{XYZ'B'}}$ are the 3x3 VCV matrices of the coordinates of the point being transformed.

Similarly, the VCV matrix of a <u>network</u> of transformed coordinates X_{B1},Y_{B1},Z_{B1} to X_{Bn},Y_{Bn},Z_{Bn} can be determined as follows:

$$\mathbf{VCV_{XYZ'Bn'}} = \mathbf{J} \begin{pmatrix} \mathbf{VCV_{XYZ'An'}} & \mathbf{0} \\ \mathbf{0} & \mathbf{VCV_P} \end{pmatrix} \mathbf{J^T} \tag{11.1-9}$$

where:

$$\mathbf{J} = \begin{pmatrix} \mathbf{F} & \mathbf{0} & \mathbf{0} & \cdots\cdots & \mathbf{G_1} \\ \mathbf{0} & \mathbf{F} & \mathbf{0} & \cdots\cdots & \mathbf{G_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{F} & \cdots\cdots & \mathbf{G_3} \\ & & & \cdots\cdots & \\ \cdot\cdot & \cdots & \cdot & \mathbf{F} & \mathbf{G_n} \end{pmatrix} \tag{11.1-9}$$

and

$$\mathbf{F} = \begin{pmatrix} s & s\kappa & -s\theta \\ -s\kappa & s & s\omega \\ s\theta & -s\omega & s \end{pmatrix} \text{ and } \mathbf{G_i} = \begin{pmatrix} X_{Ai}+\kappa Y_{Ai}-\theta Z_{Ai} & 0 & -sZ_{Ai} & sY_{Ai} & 1 & 0 & 0 \\ Y_{Ai}-\kappa X_{Ai}+\omega Z_{Ai} & sZ_{Ai} & 0 & -sX_{Ai} & 0 & 1 & 0 \\ \theta X_{Ai}-\omega Y_{Ai}+Z_{Ai} & -sY_{Ai} & sX_{Ai} & 0 & 0 & 0 & 1 \end{pmatrix}$$

$\mathbf{VCV_{XYZ'An'}}$ and $\mathbf{VCV_{XYZ'Bn'}}$ are 3nx3n matrices.

Thus the VCV matrix of the transformed coordinates $\mathbf{VCV_{XYZ'Bn'}}$ is a combination of the VCV matrices of the original coordinates $\mathbf{VCV_{XYZ'An'}}$ and the transformation parameters used $\mathbf{VCV_P}$.

11.1.3 CONVERSION OF COORDINATE QUANTITIES

Ellipsoidal to Cartesian

Some transformation models require the Cartesian coordinates (X,Y,Z) and their VCV matrix. However, frequently, in the case of the terrestrial network, only ellipsoidal coordinates and VCV are available and it is necessary to convert them to the Cartesian frame (Figure 11.1-5). The Cartesian coordinates can be calculated from the well known relations:

$$\begin{aligned} X &= (\nu + h)\cos\phi\cos\lambda \\ Y &= (\nu + h)\cos\phi\sin\lambda \\ Z &= \{ (1 - e^2)\nu + h \}\sin\phi \end{aligned} \tag{11.1-10}$$

with $\nu = \dfrac{a}{\sqrt{1 - e^2\sin^2\phi}}$ and $e^2 = 2f - f^2$, where a is the semi-major axis of the reference ellipsoid and f is the inverse flattening of the ellipsoid.

The converted VCV matrix can be calculated from:

$$\mathbf{VCV_{XYZ}} = \mathbf{J} \,.\, \mathbf{VCV_{\phi\lambda h}} \,.\, \mathbf{J}^T \tag{11.1-11}$$

where $\mathbf{J}$ is the Jacobian matrix defined by:

$$\mathbf{J} = \begin{pmatrix} \dfrac{\nu e^2\sin\phi\cos^2\phi\cos\lambda}{1-e^2\sin^2\phi} - (\nu+h)\cos\lambda\sin\phi & -(\nu+h)\cos\phi\sin\lambda & \cos\phi\cos\lambda \\ \dfrac{\nu e^2\sin\phi\cos^2\phi\sin\lambda}{1-e^2\sin^2\phi} - (\nu+h)\sin\lambda\sin\phi & (\nu+h)\cos\phi\cos\lambda & \cos\phi\sin\lambda \\ \dfrac{\{\nu e^2\sin^2\phi\cos\phi + \nu\cos\phi\}(1-e^2)}{1-e^2\sin^2\phi} + h\cos\phi & 0 & \sin\phi \end{pmatrix} \tag{11.1-12}$$

The Jacobian matrix given above is used to transform the VCV matrix for a single point. For conversion of the VCV matrix for more than one point the full Jacobian matrix is given by:

$$\mathbf{J} = \begin{pmatrix} J_1 & 0 & 0 & 0 & . & . \\ 0 & J_2 & 0 & 0 & . & . \\ 0 & 0 & J_3 & 0 & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & J_n \end{pmatrix} \tag{11.1-13}$$

Cartesian to Ellipsoidal

Similarly, the output adjusted GPS coordinates and the VCV matrix of transformation models is often expressed in the Cartesian formulation. The corresponding ellipsoidal coordinates can be calculated by several methods, one of which is:

$$\phi = \arctan\left\{\frac{(Z + e^2\nu\sin\phi)}{R}\right\} \qquad \textit{[iterate]}$$

$$\lambda = \arctan\left(\frac{Y}{X}\right) \tag{11.1-14}$$

$$h = \frac{R}{\cos\phi} - \nu$$

where $R = \sqrt{X^2 + Y^2}$.

The VCV matrix of these ellipsoidal coordinates can be calculated from:

$$\mathbf{VCV_{\phi\lambda h} = J \;.\; VCV_{XYZ} \;.\; J^T} \tag{11.1-15}$$

where the Jacobian matrix $\mathbf{J}$ is given by:

$$\mathbf{J} = \begin{pmatrix} A & \frac{Y.A}{X} & B \\ \frac{-Y}{R^2} & \frac{X}{R^2} & 0 \\ \frac{X}{R\cos\phi} + A.C & \frac{Y}{R\cos\phi} + \frac{Y.A.C}{X} & B.C \end{pmatrix} \tag{11.1-16}$$

where $A \approx \dfrac{X\tan\phi}{R^2(e^2 - \sec^2\phi)}$, $B \approx \dfrac{1}{R\sec^2\phi - e^2\nu\cos\phi}$, $C = \dfrac{R\sin\phi}{\cos^2\phi} - \dfrac{\nu e^2\sin\phi\cos\phi}{(1 - e^2\sin^2\phi)}$

The conversion of the VCV matrix of more than one point can be done in the same manner as indicated in eqn (11.1-13).

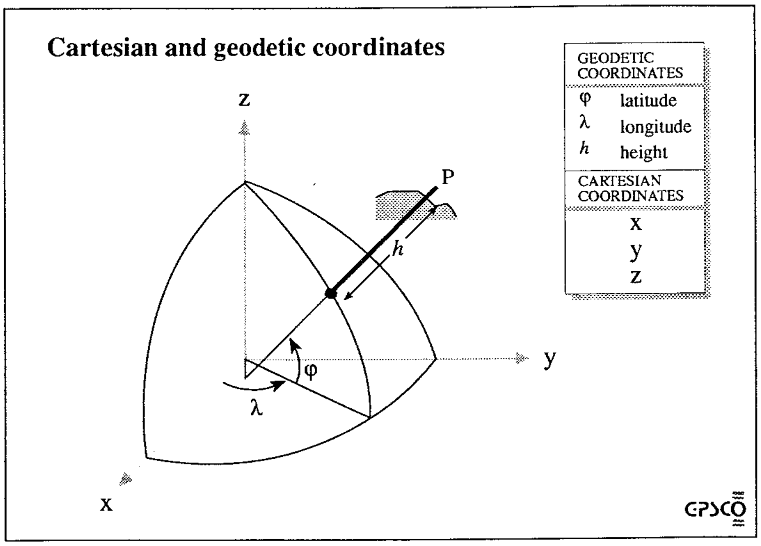


Figure 11.1-5. The Cartesian and ellipsoidal coordinate systems.

Cartesian to Topocentric

An alternative formulation for two-dimensional coordinates is a local topocentric system. The horizontal plane is tangential at a point with Cartesian coordinates (X_o, Y_o, Z_o). The (E,N,H) components form a right-handed system. The Cartesian coordinates can be converted to local topocentric components using the relation (SEEBER, 1993):

$$\begin{pmatrix} E_A \\ N_A \\ H_A \end{pmatrix} = \mathbf{T} \cdot \left(\begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} - \begin{pmatrix} X_o \\ Y_o \\ Z_o \end{pmatrix} \right) \tag{11.1-16}$$

where the rotation matrix $\mathbf{T}$ is given by:

$$\mathbf{T} = \begin{pmatrix} -\sin\lambda & \cos\lambda & 0 \\ -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi \\ \cos\phi\cos\lambda & \cos\phi\sin\lambda & \sin\phi \end{pmatrix} \tag{11.1-17}$$

The converted VCV matrix can be calculated from:

$$\mathbf{VCV_{ENH}} = \mathbf{T} \cdot \mathbf{VCV_{XYZ}} \cdot \mathbf{T}^T \tag{11.1-18}$$

This relation is very useful for converting GPS results to a form that can be used to compute 2-D relative error ellipses. In Australia, the semi-major axes of the relative error ellipses can be tested in order to check if the GPS survey has satisfied the requirements for a certain ORDER of GPS survey (§10.2).

Topocentric to Cartesian

The inverse transformation is (IBID, 1993):

$$\begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} = \mathbf{T} \cdot \begin{pmatrix} E_A \\ N_A \\ H_A \end{pmatrix} + \begin{pmatrix} X_o \\ Y_o \\ Z_o \end{pmatrix} \tag{11.1-19}$$

where the rotation matrix $\mathbf{T}$ is given by:

$$\mathbf{T} = \begin{pmatrix} -\sin\lambda & -\sin\phi\cos\lambda & \cos\phi\cos\lambda \\ \cos\lambda & -\sin\phi\sin\lambda & \cos\phi\sin\lambda \\ 0 & \cos\phi & \sin\phi \end{pmatrix} \tag{11.1-20}$$

The converted VCV matrix can then be calculated from:

$$\mathbf{VCV_{XYZ}} = \mathbf{T} \cdot \mathbf{VCV_{ENH}} \cdot \mathbf{T}^T \tag{11.1-21}$$

Projection Coordinates

Often ellipsoidal coordinates ϕ, λ are converted to plane projection coordinates x, y. There are many **map projections** which strive to somehow accommodate the complex curvature of the ellipsoid on a flat map sheet without too much distortion (MATHER, 1978). In general, conformal projections are used in surveying and mapping. *Conformality* means that an angle on the ellipsoid caused by a pair of geodesic lines is preserved. However, the geodesics themselves are curved lines after projection. Common *classes* of conformal projections are (HOFMANN-WELLENHOF et al, 1994):

(1) Conical projection: formed by considering a cone tangential to the ellipsoid at some *standard parallel of latitude*. After the cone is flattened out, the meridians of longitude are straight lines converging to an apex, which is also the centre of circles representing the projected parallels. The *Lambert projection* is one example of this type of projection.

(2) Azimuthal projection: is a special case of the conical projection where the apex is the pole and the cone degenerates to a plane tangential at the pole. The pole is therefore the centre of circles representing the parallels and of the straight lines representing the meridians. One example of this projection is the *stereographic projection.*

(3) Cylindrical projection: is a special case of the conical projection where the apex is moved to infinity so that the cone becomes a cylinder which is tangential to the equator. After the cylinder is unrolled, the equator is mapped without distortion. The *Universal Transverse Mercator (UTM) projection* is one of the most widely used of all projection systems. In the transverse position the cylinder is tangential to a *standard meridian.*

In the conventional Transverse Mercator (TM) projection the standard meridian (or "central meridian") is mapped without distortion as it is the line of tangency of the spherical approximation of the ellipsoid with the cylinder. The central meridian is the y-axis (north direction) of the projection, while the x-axis is the mapping of the equator. The ellipsoid is partitioned into 120 zones of 3° longitude, each with the central meridian in the centre of the zone. As a point moves away from the central meridian the distortions grow larger as the scale factor increases. A modification of the TM is the UTM. Firstly, the ellipsoid is partitioned into 60 zones, each 6° longitude in width. Secondly, the scale factor at the central meridian is 0.9996 to reduce the large distortions in the fringes of the zone.

Detailed formulae defining the conformal mappings for the various projection systems can be found in standard texts on maps and projections (for example, MATHER, 1978).

11.1.4 VERTICAL DATUMS AND HEIGHT SYSTEMS

What is meant by a "physical" height system? To illustrate this concept, and others below, consider the basic operation of spirit levelling. The concept is simple enough over short distances. The difference in height between two points P and Q is determined by setting up a levelling instrument between the two points and noting the readings on linearly graduated staves set vertically at the points. The instrument's line of sight is tangential to that level surface of the earth's gravity field passing through the instrument's horizontal axis. The height difference between P and Q is therefore the difference in readings made on the staves. This process is extended over an area so that if the height of one point is given, the height differences thereby obtained are used to form a network of heights.

A closer study of this operation reveals an assumption that is not strictly correct. If the points P and Q are so far apart that the above procedure has to be applied repeatedly, then the sum of the levelled height differences between P and Q *will not necessarily be equal to the difference in height of the respective level surfaces through P and Q* (HEISKANEN & MORITZ, 1967). This is due to the non-parallelism of the level surfaces of the earth's gravity field (that is, they tend to converge towards the poles), as illustrated in Figure 11.1-6. The solution to this problem is to convert the linear staff increments dn to gravity potential differences dW:

$$dW = -g\,dn \tag{11.1-22}$$

where g is observed gravity.

However, a height system such as this is only based on height differences. To provide for a reference surface, geodesists make use of a datum level surface with gravity potential W_o, such that the potential of a point P is uniquely defined as:

$$\Delta W = W_P - W_o = -\int_{geoid}^{P} g\;dn \tag{11.1-23}$$

To convert the units from gravity potential to metric ones, these potential values (actually potential differences from the datum) are scaled by the mean value of gravity along the vertical to the geoid/ellipsoid between the point P and the datum surface to give the orthometric height:

$$H_P = \frac{\Delta W_P}{\bar{g}} \tag{11.1-24}$$

Gravity therefore plays a central role in the levelling operation:

- The spirit bubble (or compensator mechanism in modern levels) is aligned to the local level surface, and in effect defines the local horizon.
- Gravity information is required to convert staff readings into orthometric height differences (eqn (11.1-24)).
- The datum surface to which the accumulated orthometric height differences are ultimately referred is the GEOID, a level surface of the earth's gravity field with potential W_o.

Another distinctive feature of the levelling process is that, in a sense, it makes no assumptions concerning the shape of the earth, or even the geoid. The field operations, the data reductions and the final elevation value assigned to a point are all free of the influence of the reference ellipsoid.

The reference ellipsoid is a mathematical approximation of the real earth, and is central to most other geodetic techniques. Although measurements by, for example, theodolite or EDM, are made in the real world, the observations are very quickly "reduced" to the artificial ellipsoid surface. The subsequent calculations and the coordinate values themselves are expressed in this "artificial" reference system. The same holds true for GPS surveys.

From these two viewpoints: the physical (or measurement-based) and the mathematical (or computational-based), there are *two definitions of height* (Figure 11.1-7):

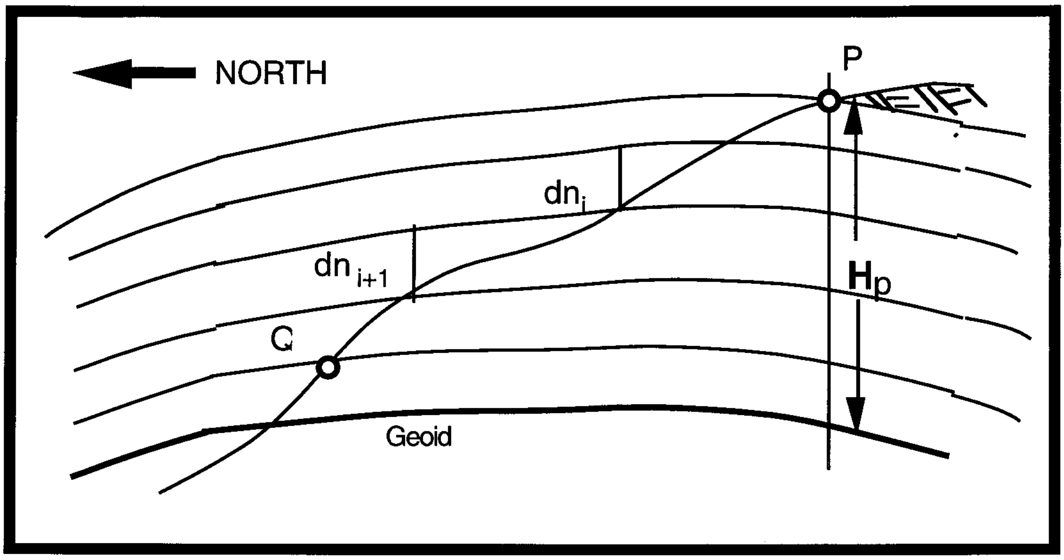


Figure 11.1-6. Level staff increments and the non-parallelism of the earth's gravity field.

- Heights referred to the geoid, with the height system being based on **orthometric heights** (H), defined as follows (MITCHELL, 1990):

  "The orthometric height of a point is the distance of the point from the geoid, measured along the curved vertical, with the positive direction upwards from the geoid."

- Heights referred to the ellipsoid, with the height system being based on **ellipsoidal heights** (h), defined as follows (IBID, 1990):

  "The ellipsoidal height of a point is the geometrical distance of the point from the reference ellipsoid, as measured along the normal to the ellipsoid."

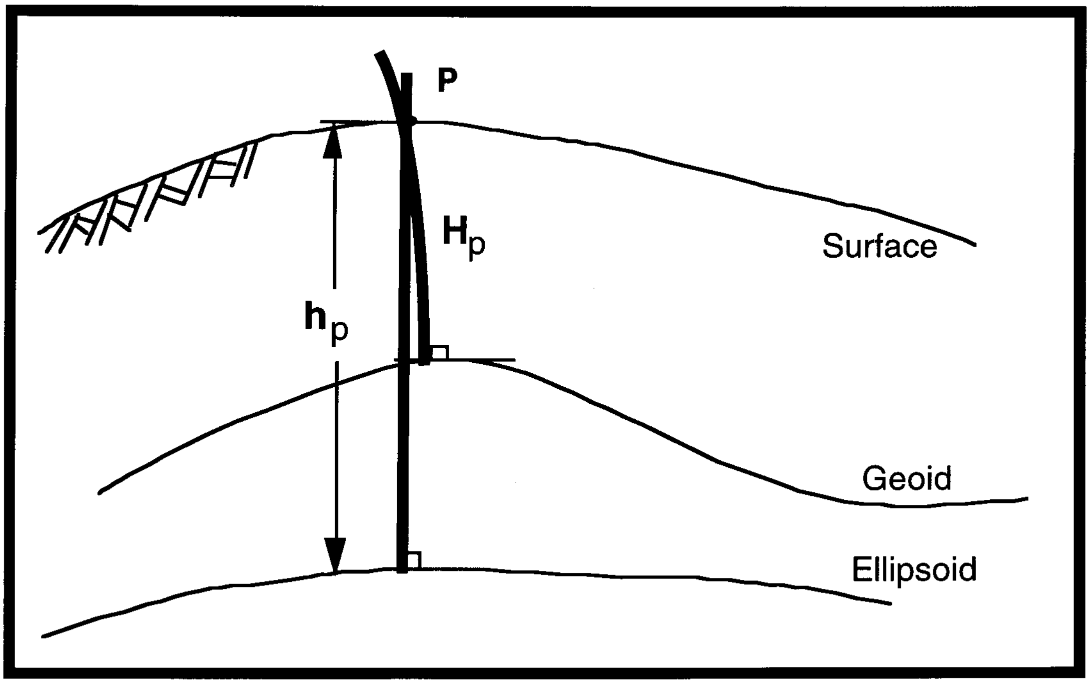


Figure 11.1-7. The geoid, the ellipsoid and the surface of the earth.

11.2
GPS NETWORK TRANSFORMATIONS

Geodetic datums permit station coordinates to be related within a rigorous framework. The coordinate sets may be stated as either Cartesian components (X,Y,Z), or ellipsoidal values (ϕ,λ,h), and can be converted from one coordinate system to the other using closed formulae such as those given in §11.1. In either case there is an implied origin, as well as three graduated axes.

An adjusted survey network on a specific datum will result in a unique coordinate set for all stations in the network. A readjustment of the same network, but with different constraints or an updated data set, will result in another coordinate set. *Nevertheless the datum remains unchanged.* The AGD66 and AGD84 coordinate sets are examples of this, both are realisations of the Australian Geodetic Datum. However, the differences in the two coordinate sets can be accommodated within a transformation model.

In the case of GPS results, the datum is nominally that of WGS84. There is therefore often the need to transform the GPS results into another datum, such as that implied by the local terrestrial control coordinate set. Hence, there are two important roles for transformation models in the context of GPS surveying:

- **To relate two datums**, for example the GPS satellite datum to a local geodetic datum.
- **To relate two coordinate sets**, either two GPS networks or two local geodetic networks, both nominally on the same datum.

To maintain generality in the following discussions it will be assumed that the transformation models are used to relate reference systems, and only where it is important will the distinction be made between transforming datums and transforming coordinate sets.

There are a number of questions relating to this transformation process:

☞ What is the exact functional model that relates one reference system to another?

☞ Are the parameters of the model available?

☞ Must the precision of the transformation parameters be taken into account?

☞ If transformation parameters had to be determined, what software is available and what operational procedures should be followed?

☞ Are the transformation parameters so determined "universal", that is, are they applicable to other networks?

☞ Can the transformation process alter the relative disposition of points in the network, hence in effect being indistinguishable from another network "adjustment"?

Useful references for further reading on the topic of transformation models are HARVEY (1986) and HARVEY (1994), with valuable hints on procedures for the reliable estimation of transformation parameters. STEED (1990) is a good reference for transformation procedures appropriate for Australian surveys.

In §11.1, the transformation model based on the similarity transformation was described, realisable in the two forms: the **Bursa-Wolf** model, and the **Molodensky-Badekas** model. Where the transformation parameters are available, these can be applied to transform a set of Cartesian coordinates in one datum to those of another datum (using eqns (11.1-1) and (11.1-2)). In this section, the procedures that can be used to empirically determine transformation parameters are discussed.

11.2.1 DETERMINING TRANSFORMATION PARAMETERS

The determination of transformation parameters is the product of a Least Squares adjustment in which the coordinates of points in both networks are considered "observations". There are a number of considerations, some unique to the problem of determining transformation parameters, others shared by all Least Squares adjustments, which should be mentioned:

- ❑ As the observations are the coordinates of the common points, they will be adjusted as a by-product of the estimation process. *Hence the process of determining transformation parameters is a form of constrained network adjustment.* The amount of constraint is controlled by the apriori VCV information. For example, if a point is to be held fixed, its coordinates can be given a very small variance, say $(0.1\text{mm})^2$, and set all the covariances involving this station equal to zero.
- ❑ If Bayesian Least Squares (§7.1 and HARVEY, 1994) is applied, the transformation parameters are treated as quasi-observables, and the apriori VCV can be assigned. Some parameters may therefore be held fixed or estimated as free parameters.
- ❑ In any Least Squares adjustment the correct stochastic model (VCV matrix) must be available. As always there is a problem of discerning between precision and accuracy. (Further comments on this are made below.)
- ❑ There are a number of analysis issues that need to be highlighted, such as the impact of network geometry, the effect of distortions in the networks, checking the results, and the application of statistical tests to verify the quality of the adjustment. (These are also discussed below.)

VCV Matrix of Coordinate "Observations"

The VCV matrix for the coordinates in both networks reflects the precisions of the individual coordinate components, as well as the correlations between the different coordinate components. In the case where one network is GPS-determined, and the other is a conventionally defined terrestrial network, there are several unique problems.

VCV of Terrestrial Networks

In practice the input VCV of the coordinate set used as data in the determination of the transformation parameters is full because the (X,Y,Z) coordinates of the points are all correlated. However, rarely does a surveyor have access to the VCV information of the terrestrial control points (if the points are part of the national geodetic network containing say 5000 stations, the VCV information of interest is contained within the full 15000x15000 VCV matrix of parameters produced by the 3-D adjustment of all the geodetic observations!). *Hence the input VCV is often assumed to be diagonal, with very approximate variances on the diagonals.*

As input to a 3-D transformation parameter adjustment, only the horizontal ground coordinates from a conventional 2-D ellipsoidal adjustment are available and these must be combined with ellipsoidal heights, determined from sum of the orthometric height (from levelling) and geoid-ellipsoid separation (Figure 11.1-6). It is generally assumed that these heights are not correlated with the horizontal coordinate components. The ellipsoidal coordinates and the VCV matrix must be converted to the Cartesian equivalents using eqns (11.1-10) to (11.1-13).

Finally, in both GPS and terrestrial network adjustments the coordinates of one point are conventionally assigned zero variance (the so-called "minimally constrained solution"). For terrestrial networks, only a portion of the ground net is used as input for the transformation parameter determination. Apart from the problem of assembling the VCV matrix, the subnet may be a considerable distance from the fixed datum point (for example, the Johnston origin station in the case of the AGD) and hence the variances of coordinates of the points within the subnet will usually be very large. Formulae are available for converting VCV matrices so that they have minimum trace, or zero variances for some point(s) within the selected portion of the network (HARVEY, 1986).

VCV of GPS Networks

Compared to the problems listed above: (a) obtaining the VCV information, (b) constructing a 3-D VCV, and (c) modifying the VCV to reflect a more convenient datum station; the VCV for a GPS network is generally easily extracted and manipulated. The VCV is already in 3-D form, and the fixed station is usually within the network to be transformed (and is often selected to be one of the common points from which the transformation parameters are derived). As usual, there is the problem that the elements of the VCV matrix of the GPS coordinates represent only precision estimates. However, if the GPS network is the result of a multi-session secondary network adjustment (§9.3), the VCV matrix may already have been scaled or modified in some way, so that the precisions are more realistic (that is, closer to what may be considered the *accuracy* of the network).

The Solution Procedure

The minimum number of observations required to solve for the parameters of a 3-D similarity transformation is seven (§11.1). This condition could be satisfied by having three common points, each with their three coordinate components known in each net, leaving a redundancy of two. However it is desirable to have as many common stations as possible so as to ensure a reasonable "degree of freedom" for the Least Squares determination of the transformation parameters. At the same time, the analyst may wish to leave some common points out of the solution, to be used later as "check points".

There are a number of network adjustment programs that will estimate the parameters of a similarity transformation as a by-product of the integration of one network with another. There are also many commercial GPS software packages that have the capability of determining datum transformation parameters.

Functional Model

For the **Bursa-Wolf** formulation (eqn (11.1-1) the functional model is:

$$\mathbf{0} = \mathbf{F} = (1 + \delta s)\ .\ \mathbf{R}\ .\begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} - \begin{pmatrix} X_B \\ Y_B \\ Z_B \end{pmatrix} \tag{11.2-1}$$

Note that the scale parameter has been redefined to represent the change in scale between the two networks (expressed in "parts per million"). Each of the model equations involve a mixture of parameters and more than one observable. Hence neither the "parametric" nor the "condition" method of Least Squares adjustment is suitable (§7.1). Rather, the "combined" method of Least Squares is used (see HARVEY, 1994, for details). A similar functional model can be defined for the **Molodensky-Badekas** form (eqn (11.1-2)). The partial derivatives for the design matrix are given in IBID (1994).

Network Geometry

The geometrical formulation of a similarity transformation represents a continuum. Replacing this continuum by a discrete set of common points may lead to *errors of interpolation*. Moreover, in the presence of observational and computational errors, the accuracy of the estimated parameters may vary considerably. This depends on the spatial distribution of the points used. Some guidelines are:

- Points should not be distributed in a collinear fashion because components of rotations about axes parallel to the line of points cannot be determined.
- For a stable solution it is important that the points are spatially well distributed. For example, a network with an uneven geographical spread of points will cause the solution to be biased towards the areas of higher point density. Points in areas of lower density will be disadvantaged and probably result in large corrections to their coordinates.
- The common points should be mostly around the perimeter of the region in which the transformation parameters are to be used.
- If distortions in scale and orientation are present in the geodetic network, then it may be better to solve for sets of local parameters rather than determining a single set of transformation parameters for the whole network.

Parameterisation

Although there are seven parameters in a 3-D similarity transformation model, there may be a choice between estimating more than seven parameters, or less.

If a network does contain distortions it is possible to add parameters representing systematic errors to the basic similarity transformation model. However, if the model contains too many parameters, the adjustment may lead to a poorly conditioned system of equations. Many parameters will usually fit the data better (that is, produce smaller residuals) than a few parameters, but the estimates of the parameters may not be reliable. Furthermore, the degrees of freedom of the adjustment will be reduced and statistical testing will be less effective. A careful investigation of the network should be made before using additional parameters to "soak up", for example, scale distortions due to ground height errors.

On the other hand, solving less than the full set of seven parameters can sometimes be justified. In general, all seven parameters are estimated and then tested using statistical tests to determine if a subset of parameters is significantly different to zero. Then another solution may be obtained in which these parameters are held fixed. It often happens that the scale factor is insignificant, or that some of the rotation angles are not separable from the translation parameter estimates.

The Output

The results of a transformation adjustment are a set of parameters, and two sets of adjusted coordinates.

The Adjusted Coordinates

The adjusted coordinates of network A have the same weighted scale, orientation and location as their apriori coordinates. Similarly, the adjusted coordinates of network B have the same weighted scale, orientation and location as their apriori coordinates. Thus, the overall scale, location and orientation of each network remains unchanged, only the individual coordinates, angles and lengths change. In other words, the movements of points within a network are small and the overall characteristics of the nets do not change.

Statistical Testing

There are a number of statistical tests that can be applied (see HARVEY, 1994), among which are:

- Variance factor test.
- Tests of parameters, to determine if they are statistically different to zero.
- Tests on the residuals, for example to see if they are normally distributed, or to search for outliers.
- Tests for regional or local distortions.

The usefulness and reliability of such tests grows with an increase in the "degrees of freedom" of the adjustment (that is, by including more common points). Some tests are based on leaving out certain points, and using these to check the transformation parameter model.

11.2.2 TRANSFORMATION OPTIONS IN AUSTRALIA

The transformation options for converting coordinates from one well defined datum or coordinate system to another are summarised in Figure 11.2-1 (STEED, 1990). There are a variety of transformation models; some are mathematically defined, others have been determined empirically (from an analysis of common points); some are global transformation models, others have only national or even local relevance (for example, interpolation of difference contours); some relate satellite datums, others different coordinate sets of the Australian Geodetic Datum. The transformation of particular relevance to GPS surveying is that relating AGD84 (the latest coordinate set of the AGD) and the GPS datum WGS84. Note there are two options:

(1) Using Higgin's formula: AGD84 <--> WGS84, or

(2) Using the path: AGD84 <--> NSWC9Z2 <--> WGS72 <--> WGS84.

In fact, the parameters of HIGGINS (1984) are derived empirically from a combination of the three other transformation models in Table 11.2-1 (DMA, 1987; SEPPELIN, 1974; ALLMAN & VEENSTRA, 1984), and were not computed from observed WGS84 positions. Because GPS is a relative surveying technique, it is in fact very difficult to obtain a national set of GPS coordinates that are on the WGS84 datum (in an absolute sense). Until this is done a set of transformation parameters cannot be empirically derived to model the conversion from WGS84 to AGD84 (and visa versa), and replace the Higgins parameters. (The new Geocentric Datum of Australia is defined to be within one metre of WGS84 -- §12.1). In the meantime, Table 11.2-1 summarises most of the transformation options relevant for Australian surveys.

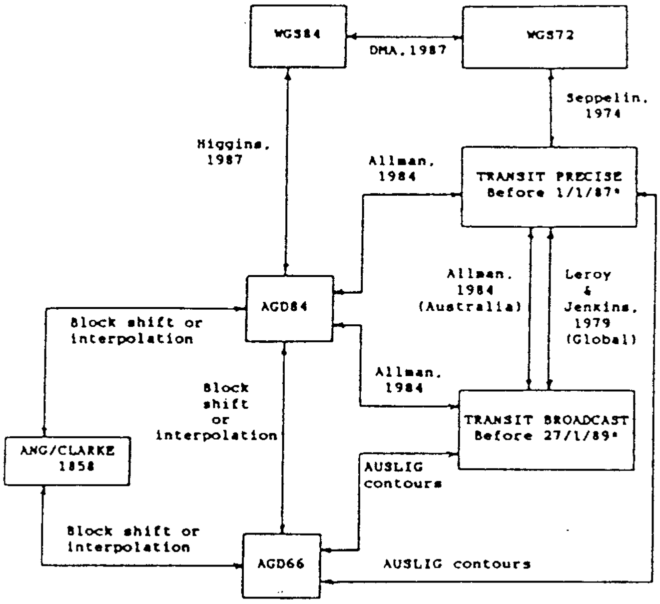


Figure 11.2-1. Australian transformation options.
(STEED, 1990)

However, there is one transformation which is commonly sought, but for which no readily available transformation model is available: AGD84 <--> AGD66. The AGD66 was derived from a Least Squares adjustment of the Australian geodetic network performed in 1966 and was used until the new adjustment was performed in 1984. Because of the limitations of the adjustment model used, and the systematic errors present in some of the observations, the AGD66 coordinate set suffers significant distortions. Distortions which were clearly visible when the AGD84 coordinate set was compared with the earlier set (see ALLMAN & VEENSTRA, 1984). The transformation model AGD66 <--> AGD84 is in fact defined by the maps of displacement vectors given in IBID (1984).

The AGD66 coordinate set is still the official basis for the geodetic control network in several states of Australia, including New South Wales, Victoria, Tasmania and Northern Territory. An approximate transformation model AGD66 <--> WGS84 has been derived by the Land Information Centre (LIC) (Table 11.2-1). Note that in Table 11.2-1 the full VCV information is not available, only the estimated standard deviations of the parameters are given (shown in brackets).

Table 11.2-1. Transformation parameters of relevance to Australian surveys.

Source	HIGGINS (1987)	A & V[a] (1984)	DMA (1987)	SEPPELIN (1974)	LIC93[b]
FROM	WGS84	NSWC9Z2	WGS84	WGS72	WGS84
TO	AGD84	AGD84	WGS72	NSWC9Z2	AGD66
T_x (m)	116.00 (2.3)	116.00 (1.2)	0.00	0.00	137.98
T_y (m)	50.47 (2.3)	50.47 (2.3)	0.00	0.00	42.58
T_z (m)	-141.69 (2.5)	-137.19 (1.5)	-4.50	0.00	-162.28
ω (")	0.23 (0.04)	0.23 (0.04)	0.00	0.00	-0.015
θ (")	0.39 (0.04)	0.39 (0.04)	0.00	0.00	-0.591
κ (")	0.34 (0.04)	-0.47 (0.04)	0.554	0.26	0.298
δs (ppm)	-0.098 (0.07)	-0.699 (0.07)	-0.226	0.827	1.017

[a] ALLMAN & VEENSTRA (1984)
[b] For New South Wales only! WGS84 only approximate at one metre level

Transforming GPS Results In Australia

In summary, in Australia, to transform the results of a GPS survey into coordinate information referred to a local geodetic datum we may either:

(1) Use published transformation parameters (Table 11.2-1), for example the Higgins parameters for WGS84 <--> AGD84, or the LIC93 parameters for WGS84 <--> AGD66(NSW).

(2) Determine the transformation parameters as an integral part of the network adjustment process.

With regards to the first option, it is of little use in relating GPS results to the local control stations because:

- It is assumed that there are no significant distortions in the local network, and that the coordinates of the control stations are related to the geodetic datum without error.
- It is assumed that the GPS results are related, with high accuracy, to the WGS84 system.

Neither of these conditions are usually fulfilled. In particular, the GPS network coordinates, while related to one another to a high precision, are only weakly related to the WGS84 system. For example, the uncertainty in the absolute coordinates of the Datum Station has little effect on the relative precision of the GPS network (see "Reference Station Bias" in §6.2). Hence any (absolute) coordinate error will affect all stations in the network, and the previously derived transformation parameters will have little relevance in relating any GPS network to the geodetic datum.

It is therefore preferable that in order to maintain the high precision of a GPS network, any combination of the GPS results with coordinate information related to the local datum must ensure that the transformation parameters are also estimated, in some way, as part of the process of GPS result integration. In addition, the full VCV matrix of the parameters can be obtained, and used to determine the VCV of the coordinates after transformation. An example is given below.

Example of Local Transformation Parameter Determination

Let us consider the example of the Molong (N.S.W.) survey referred to earlier in §9.3 (Table 9.3-1 and Figure 9.3-3). Seven of the GPS stations had coordinates in the local geodetic datum (although of varying quality). The seven transformation parameters for the Bursa-Wolf model (eqn (11.1-1)) had to be estimated. The following information was available:

- The GPS coordinates in the (nominal) WGS84 reference frame (Net A), as well as the full VCV matrix (with the Canobolas station assumed fixed), determined from the minimally constrained network solution described in §9.3.
- The local coordinate information (Net B) is in the form of latitude and longitude on the one hand, and levelled height (to which the estimated geoid height has been added) on the other. The VCV information is incomplete -- only the standard deviations could be guessed at. The latitude/longitude for one station (Vale Head) was poorly known and therefore assigned an uncertainty of about 10m, while all other stations were assumed to have horizontal coordinates good to about 10cm. The height data consisted of some spirit levelled values (for Vale Head, Canobolas, Yuranigh and Goanna), assigned an uncertainty of 10cm, some trig levelled values (Brymedura and Molong), assigned an uncertainty of 60cm, and one value (Nandillyan) which was uncertain and assigned an uncertainty of 2m.

NET A DATA DESCRIPTION - GPS BASELINE ADJUSTMENT

APRIORI XYZ COORDINATES OF NET A:

X	Y	Z	SITE
-4580212.441	2765441.532	-3461437.257	VALE HEAD
-4575501.983	2768568.678	-3465321.921	BRYMEDURA
-4571934.531	2749030.854	-3486696.817	CANOBOLAS
-4578429.387	2765618.487	-3463675.206	YURANIGH
-4584249.057	2761558.635	-3459326.514	NANDILLYAN
-4581684.119	2755868.691	-3467412.500	GOANNA
-4580243.023	2770142.456	-3457857.061	MOLONG

APRIORI VARIANCE COVARIANCE MATRIX OF NET A (in units m^2):

```
 0.21378D-09
-0.11873D-10  0.14965D-09
 0.55969D-11 -0.65800D-12  0.16153D-09
 0.21049D-09 -0.11141D-10  0.53647D-11  0.23403D-09
-0.11165D-10  0.14661D-09 -0.58324D-12 -0.14601D-10  0.16510D-09
 0.53510D-11 -0.58559D-12  0.15748D-09  0.60871D-11 -0.98200D-12  0.18059D-09
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.10000D-13
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.10000D-13
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.10000D-13
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.15550D-08
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
```

```
 0.00000D+00  0.00000D+00  0.00000D+00 -0.17629D-10  0.12310D-08
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.11130D-10 -0.57692D-11  0.78562D-09
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.13502D-08
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
-0.85759D-11  0.64985D-09
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.12025D-10 -0.96851D-12  0.97232D-09
 0.14277D-09 -0.31618D-11  0.26458D-11  0.14059D-09 -0.27848D-11  0.25106D-11
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.16097D-09
-0.30014D-11  0.94285D-10 -0.93602D-12 -0.26076D-11  0.92381D-10 -0.87944D-12
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00 -0.60572D-11  0.11084D-09
 0.27421D-11 -0.93462D-12  0.90401D-10  0.26160D-11 -0.88224D-12  0.88133D-10
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.38935D-11 -0.10319D-11  0.11000D-09
 0.20826D-09 -0.10790D-10  0.52217D-11  0.21768D-09 -0.12067D-10  0.55519D-11
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.13911D-09 -0.24321D-11  0.25392D-11
 0.23598D-09
-0.10751D-10  0.14515D-09 -0.57300D-12 -0.12004D-10  0.15244D-09 -0.67474D-12
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00 -0.25541D-11  0.91466D-10 -0.87159D-12
-0.14662D-10  0.16782D-09
 0.51873D-11 -0.58537D-12  0.15406D-09  0.55327D-11 -0.68538D-12  0.16295D-09
 0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00  0.00000D+00
 0.00000D+00  0.00000D+00  0.00000D+00  0.24238D-11 -0.86924D-12  0.86218D-10
 0.60159D-11 -0.78991D-12  0.18071D-09
```

NET B DATA DESCRIPTION - AUST GROUND DATA

ELLIPSOID = ANS/AGD FLAT = 1/298.250000 RADIUS = 6378.160km

APRIORI COORDINATES OF NET B:

LATITUDE			LONGITUDE			ELLIPSOIDAL HEIGHT	SITE
-33	4	42.62580	148	52	33.95750	598.638	VALE HEAD
-33	7	11.02890	148	49	16.79740	698.100	BRYMEDURA
-33	20	45.08590	148	58	51.70580	1416.140	CANOBOLAS
-33	6	8.90700	148	51	52.57720	618.621	YURANIGH
-33	3	19.52200	148	56	2.49100	664.600	NANDILLYAN ECC-2
-33	8	30.32230	148	58	19.45680	779.470	GOANNA
-33	2	21.73000	148	49	59.50000	706.900	MOLONG

APRIORI VARIANCE COVARIANCE MATRIX OF NET B (units km^2 & $rads^2$):

Diagonal matrix with variances:

LATITUDE	LONGITUDE	ELLIPSOIDAL HEIGHT	SITE
0.1929D-09	0.1929D-09	0.1D-07	VALE HEAD
0.1929D-15	0.1929D-15	.36D-06	BRYMEDURA
0.1929D-15	0.1929D-15	0.1D-07	CANOBOLAS
0.1929D-15	0.1929D-15	0.1D-07	YURANIGH
0.1929D-15	0.1929D-15	0.4D-04	NANDILLYAN ECC-2
0.1929D-15	0.1929D-15	0.1D-07	GOANNA
0.1929D-15	0.1929D-15	.36D-06	MOLONG

THE ADJUSTED PARAMETERS

PARAMETER	ADJUSTED VALUE	SIGMA
SCALE FACTOR (ppm)	-4.084 ±	2.798
ROTN ABT X AXIS (secs)	2.633 ±	1.081
ROTN ABT Y AXIS (secs)	-2.207 ±	1.896
ROTN ABT Z AXIS (secs)	-3.884 ±	2.676
TRANS ALONG X AXIS (m)	194.089 ±	67.096
TRANS ALONG Y AXIS (m)	192.728 ±	75.054
TRANS ALONG Z AXIS (m)	-187.048 ±	36.058

CORRELATION & STANDARD DEVIATION MATRIX:

1)	0.28E-05						
2)	-0.03	0.00					
3)	0.01	-0.63	0.00				
4)	-0.01	-0.79	0.89	0.00			
5)	0.19	0.72	-0.95	-0.96	0.07		
6)	-0.10	0.87	-0.86	-0.98	0.91	0.08	
7)	0.27	-0.35	0.92	0.72	-0.77	-0.68	0.04

ADJUSTED XYZ COORDINATES OF NET A:

X	COR	Y	COR	Z	COR
-4580212.440±0.015	0.001	2765441.533±0.012	0.001	-3461437.255±0.013	0.002
-4575501.982±0.015	0.001	2768568.680±0.013	0.002	-3465321.919±0.013	0.002
-4571934.531±0.000	0.000	2749030.854±0.000	0.000	-3486696.817±0.000	0.000
-4578429.395±0.039	-0.008	2765618.491±0.035	0.004	-3463675.209±0.028	-0.003
-4584249.050±0.036	0.007	2761558.627±0.025	-0.008	-3459326.529±0.030	-0.015
-4581684.119±0.013	0.000	2755868.692±0.010	0.001	-3467412.498±0.010	0.002
-4580243.022±0.015	0.001	2770142.457±0.013	0.001	-3457857.059±0.013	0.002

ADJUSTED COORDINATES OF NET B:

LATITUDE	COR	LONGITUDE	COR	ELLIPSOIDAL HEIGHT	COR
-33 4 42.62889±0.00149	-0.00309	148 52 33.96126±0.00150	0.00376	598.766±0.069	0.128
-33 7 11.02962±0.00150	-0.00072	148 49 16.80218±0.00158	0.00478	698.372±0.132	0.272
-33 20 45.08564±0.00252	0.00026	148 58 51.70513±0.00261	-0.00067	1416.148±0.099	-0.008
-33 6 8.90916±0.00176	-0.00216	148 51 52.57924±0.00202	0.00204	618.548±0.076	-0.073
-33 3 19.51725±0.00171	0.00475	148 56 2.48942±0.00178	-0.00158	663.165±0.111	-1.435
-33 8 30.32707±0.00140	-0.00477	148 58 19.45627±0.00147	-0.00053	779.438±0.097	-0.032
-33 2 21.72952±0.00179	0.00048	148 49 59.49800±0.00183	-0.00200	705.543±0.099	-1.357

The transformation parameter determination was carried out using a computer program developed at the School of Geomatic Engineering, University of New South Wales (HARVEY, 1994). The example shown above is a REAL one, not a simulation, and therefore a number of aspects of the solution should be commented on:

- The transformation parameters are highly correlated, and their values are very sensitive to the network geometry and quality. The full VCV matrix should be used when propagating the network A uncertainties to network B (eqn (11.1-8)).
- Following the comment above, the transformation parameters bear no relation to the national values (for the "Higgins" or "LIC93" values in Table 11.2-1), and are not valid for use beyond the area of this network.
- The coordinates of network A have received less correction (typically a few centimetres), than some components of network B. (There appears to be some problems with the heights of the points in network B.)
- The adjusted coordinates of network A can now be adopted as fixed stations in a subsequent constrained adjustment of the GPS baselines (before finally transforming the new network results for the non-common stations), see §12.1.

11.3
GPS HEIGHTING

"...the geoid is that level surface of the earth's gravity field coincident with Mean Sea Level *in an average sense...*"

The italicised phrase in the above formal definition of the geoid emphasises the fundamental tension between the notion of the geoid as a real, but elusive surface, which plays a central role in classical geodesy on the one hand, and the utilitarian definition as a datum for heights required by practising surveyors on the other.

The geometric relationship between the reference ellipsoid -- the datum for ellipsoidal heights -- and the geoid is illustrated in Figure 11.3-1, and defined by the simple equation:

$$N = h - H \qquad (11.3\text{-}1)$$

where N is the geoid-ellipsoid separation, or simply the geoid height, h is the ellipsoidal height and H is the orthometric height.

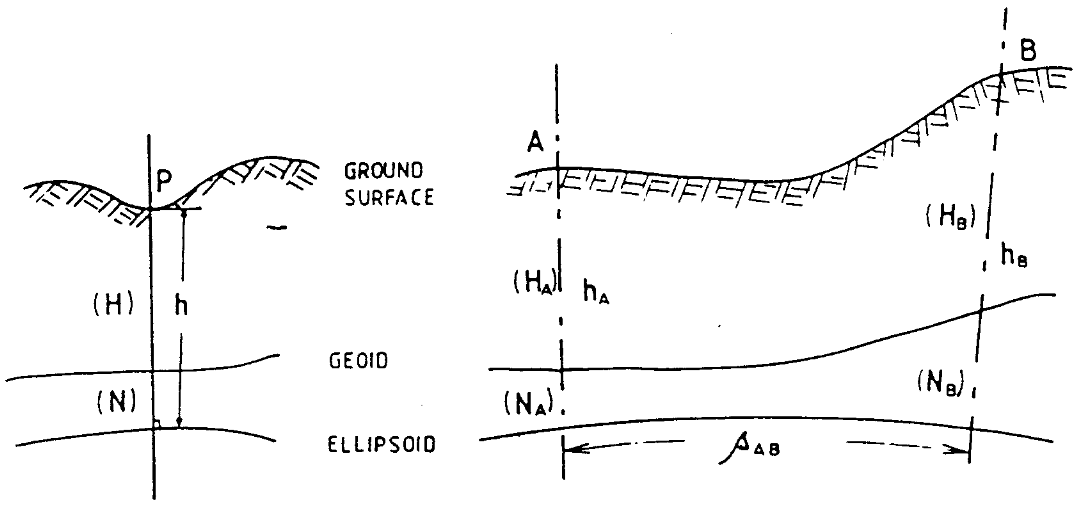


Figure 11.3-1. Relationships between ellipsoidal, orthometric and geoidal height for single point and relative heighting.

Geoid height is also a function of:

- the size and shape of the ellipsoid, and
- location and orientation of the ellipsoid in space,

to which it is being referred (Figure 11.3-2). This illustrated, for a small network in Sydney, in Table 11.3-1. The same geoid model (based on the OSU89A geopotential model) is "mapped" on both the WGS84 geocentric ellipsoid, and the Australian National Spheroid in relation to the Australian Geodetic Datum (for a non-geocentric location for the ellipsoid centre).

Absolute geoid heights are not very accurate ...

Relative geoid heights can be accurate at the few ppm level.

ELLIPSOID:

- Mathematical definition
- Simple geometrical surface described by a few parameters
- Cannot be "sensed" by instruments

GEOID:

- Physical (gravitational) definition
- Complicated surface described by infinite number of parameters
- Can be "sensed" by geodetic instruments

Table 11.3-1. The Sydney GPS network, OSU89A geoid on different ellipsoid and datum.

Name	Latitude	Longitude	N_{WGS84}	N_{AGD}
B402	S33°54'56.567"	E151°13'52.399"	22.454	14.459
BOTY	S33°58'23.355"	E151°14'10.566"	22.217	14.376
CETS	S33°55'05.557"	E151°13'57.743"	22.443	14.452
CPDH	S33°55'06.198"	E151°14'03.196"	22.440	14.448
G106	S33°55'13.377"	E151°13'56.767"	22.435	14.451
GLEB	S33°52'15.439"	E151°11'05.280"	22.679	14.624
KYMA	S33°56'45.707"	E151°09'39.054"	22.448	14.640
LAKE	S33°56'18.011"	E151°12'40.436"	22.398	14.496
MBTS	S33°56'28.596"	E151°15'53.198"	22.297	14.325
MMCH	S33°51'33.954"	E151°13'18.439"	22.679	14.539
PETE	S33°54'36.572"	E151°10'31.635"	22.551	14.621
Q744	S33°54'02.521"	E151°15'03.756"	22.485	14.419
Q855	S33°54'01.506"	E151°14'56.739"	22.489	14.425
ROSE	S33°54'49.779"	E151°12'18.046"	22.498	14.535
UNSW	S33°55'08.060"	E151°13'52.148"	22.442	14.456
UU33	S33°55'02.962"	E151°13'40.084"	22.452	14.467
WBTS	S33°51'10.704"	E151°17'09.094"	22.618	14.367

N_{WGS84} mapped on ellipsoid $a = 6378137.0$; $f^{-1} = 298.257223563$

N_{AGD} mapped on ellipsoid $a = 6378160.0$; $f^{-1} = 298.25$

Coordinates of AGD ellipsoid centre relative to WGS84 ellipsoid centre are assumed to be: X = 116m, Y = 50m, Z = -142m

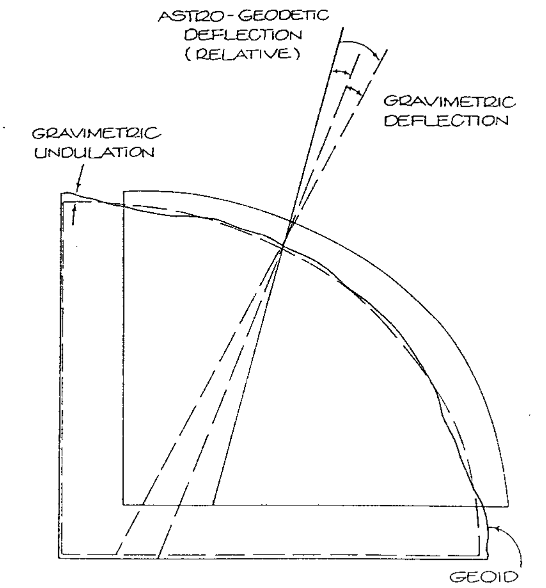


Figure 11.3-2. Geoid height in relation to different ellipsoids and datums.

The Role Of Geoid Height Information

Geoid height information plays two crucial roles in GPS result transformation:

(1) To convert the external (non-GPS) control information into the mathematically equivalent reference system to which the GPS results refer. The control information is generally in the form of ϕ, λ (horizontal components) and H (the orthometric height), and the necessary steps are:
  - compute the ellipsoidal height ($h = H + N$), and
  - then the ϕ, λ, h system can be converted into Cartesian components (§11.1).

The transformation parameters may then be determined between the two Cartesian systems, given at least three stations with coordinates in both the GPS system, and the local geodetic system in which the results are required.

(2) The GPS-derived, and transformed (to the local geodetic system), 3-D Cartesian results can then be converted to the ϕ, λ, h system. However, to obtain orthometric height information the geoid height needs to be subtracted. This is referred to as **GPS Levelling**.

There are several comments that need to be made at this stage:

- The geoid height information is required to be given in terms of the ellipsoid implicit in the local geodetic datum: the reference ellipsoid. *That is, it must be non-geocentric.*
- The geoid height accuracy requirements vary considerably. For the transformation they are only modest, but for GPS levelling it depends on the orthometric height accuracy sought (usually similar to levelling standards -- see §10.2).
- For survey accuracy GPS heighting, relative geoid height between ends of GPS baselines is required, as illustrated in Figure 11.3-1.

11.3.1 GEOID HEIGHT DETERMINATION

The techniques for geoid height determination have been discussed in many reports and papers (references presenting an Australian perspective include HOLLOWAY, 1988; KEARSLEY, 1988; MITCHELL, 1990). The basic geoid height determination techniques are:

(1) Astro-geodetic method: directly related to the local geodetic network because the necessary data (the deflections of the vertical) are the astronomic coordinates on the one hand, and the geodetic coordinates of the same stations on the other. *Profiles of geoid heights are calculated between these astro-geodetic stations.*

(2) Geopotential models of the earth's gravity field: derived from a combination of satellite and terrestrial data. The geoid height is one quantity that can be determined from a mathematical formula whose input is the ϕ, λ, h of the point, and the coefficients of the geopotential model. *Very high order models (with up to 360^2 coefficients) are available.*

(3) Use of surface gravity in techniques such as Stokes' Integral: requires gravity data in the vicinity of point at which the geoid height is to be evaluated. *This is a severe restriction in some parts of the world, but the technique is potentially the most accurate of all the geoid determination procedures.*

(4) Geometric or interpolation methods: a local representation of the geoid is obtained only at points which have both levelled (orthometric) heights and heights derived using GPS (ellipsoidal). *Geoid heights at other points are found by interpolation.*

The Astro-Geodetic Method

The following are some characteristics, advantages and disadvantages of this method of geoid height determination:

- ❑ Geoid height information obtained is *relative to the local ellipsoid.* This is indeed the only geoid height determination technique that results in values of N that relate to the reference ellipsoid implicit to the local geodetic datum. Figure 11.3-5 shows the astro-geodetic geoid for Australia.

- ❑ The basic data are the deflection of the vertical: the difference between the computed ellipsoidal coordinates from a 2-D geodetic adjustment and the corresponding astronomically-determined coordinates. Such stations are known as astro-geodetic stations.

- ❑ Suitable only in areas where astronomic coordinates are available. Most developed countries have halted such observations many years ago. Many other countries do not even have such astro-geodetic networks.

- ❑ Geoid height information is essentially the slope of the geoid between astro-geodetic stations (generally 20-100km apart). Profiles of the geoid are determined through chains of astro-geodetic stations.

- ❑ Information generally available in the form of maps of geoid contours drawn from a network of interlocking geoid height profiles. The user needs to interpolate to any other point.

- ❑ Variable accuracy dependent upon:
  - distances between astro-geodetic stations,
  - distances between profiles,
  - smoothness of geoid, and
  - accuracy of astro observations (of the order of 0.1-1arcsec).
- ❑ Relative geoid height accuracy (between astro-geodetic stations) of the order of 0.1-0.5m.
- ❑ Relative geoid height accuracy overall probably only a few metres.

<u>Summary:</u> Limited application in the future, including for GPS heighting.

Geopotential Models

The geoid height is defined by:

$$N_{SHM} = \frac{GM}{R\gamma} \sum_{n=2}^{n_{max}} \left(\frac{a_e}{R}\right)^n \sum_{m=0}^{n} P_{nm}(\sin\phi)[C^*_{nm}\cos m\lambda + S_{nm}\sin m\lambda] \qquad (11.3\text{-}2)$$

where:

ϕ, λ	are the geocentric latitude and longitude of the point of computation,
R	is the geocentric radius to the point of computation on the ellipsoid,
γ	is normal gravity on the ellipsoid,
a_e	is the radius of sphere to which coefficients refer,
G	is the Gravitational Constant,
M	is the mass of the earth,
P_{nm}	is the associated Legendre function of the degree **n** and order **m**,
C^*_{nm}, S_{nm}	are the spherical harmonic coefficients of the geopotential, of degree **n** and order **m**. C^*_{nm} are corrected for the model potential of the reference ellipsoid to which the geoid heights refer, and
n_{max}	is maximum degree and order of the spherical harmonic expansion.

The following are some characteristics, advantages and disadvantages of the use of this method:

- ❑ Geoid height information that is obtained is relative to a geocentric ellipsoid, with parameters fixed for the particular model (modern spherical harmonic models adopt the WGS84 parameters for the size and flattening of the ellipsoid).
- ❑ The basic "data" are the <u>spherical harmonic</u>, or <u>geopotential</u>, coefficients of the model.
- ❑ Applicable globally, on land, at sea, and in near-earth orbit.
- ❑ Geoid height information is derived using a mathematical algorithm at the point requested. Maps are available, or computer software can be used (particularly desirable if different models are to be used).
- ❑ There are different models, independently generated by research organisations in the U.S. and Europe. The most commonly used are the Ohio State University models

OSU89A & B (RAPP & PAVLIS, 1990) and OSU91A & B, and the NASA-DMA model GEM-96. Figure 11.3-6 illustrates the OSU89A geoid model over the Australian continent.

- Variable accuracy, dependent upon:
  - amount of local gravity input into models,
  - amount and quality of satellite tracking data,
  - smoothness of geoid, and
  - maximum degree of model.
- *Absolute geoid height accuracy is of the order of a few metres.* The relative geoid height accuracy is better because the presence of biases (or long wavelength errors) at the points of computation will largely cancel in differential geoid height computations. Experience in Australia, the U.S. and Europe, indicates that the latest geopotential coefficient models are capable of relative geoid height accuracies of the order of 5 parts per million or better.
- Resolution is dependent on the maximum degree of the geopotential model. Highest degree (n_{max}) = 360, hence geoid features with half wavelengths down to approximately 25km can be represented by such a model.

Summary: Very convenient, but may be limited by accuracy and resolution in some areas.

Gravimetric (Stokes' Integral) Methods

The basic method involves the integration of terrestrial gravity observations using Stokes' Integral:

$$N_{Stoke} = \frac{R}{4\pi\gamma} \int \int f(\psi)\, \Delta g\, d\sigma \qquad (11.3\text{-}3)$$

where:

R	is the mean radius of the earth,
γ	is normal gravity on the sphere,
$f(\psi)$	is the Stokes' function (HEISKANEN & MORITZ, 1967),
Δg	is the gravity anomaly (= observed gravity reduced to the geoid minus normal gravity at the corresponding point on the ellipsoid) for the surface element $d\sigma$, and
ψ	is the angular radius, measured from the geocentre, between the computation point and the point at which the gravity anomaly is located.

Note that, in principle, the integration is carried out over the entire globe. The requirement to have gravity anomalies over the entire globe had limited the use of the gravimetric techniques until the advent of the Space Age, and in particular the availability of global gravity field models such as those afforded by spherical harmonic coefficients. Nowadays the gravimetric technique is actually a combination of Stokes' Integral and spherical harmonic models (SCHWARZ & SIDERIS, 1993), such that:

$$N = N_L + N_S \qquad (11.3\text{-}4)$$

where N_L is given by eqn (11.3-2), and N_S is the short wavelength information derived from the integration of surface gravity, but derived from a modification of eqn (11.3-3):

$$N_S = \frac{R}{4\pi\gamma} \int_0^{\psi_o} \int_0^{2\pi} f(\psi)\, \Delta g'\, d\alpha\, d\psi \tag{11.3-5}$$

in which the integration is carried out only for a limited "cap", centred at the computation point, with radius ψ_o. Usually this is of the order of a degree ($\approx$110km) or so, and the gravity anomaly is actually a residual quantity:

$$\Delta g' = \Delta g + \Delta g_L \tag{11.3-6}$$

The quantity Δg_L is derived from the spherical harmonic model:

$$\Delta g_L = \frac{GM}{R^2} \sum_{n=2}^{n_{max}} \left(\frac{a_e}{R}\right)^n (n-1) \sum_{m=0}^{n} P_{nm}(\sin\phi)[C^*_{nm}\cos m\lambda + S_{nm}\sin m\lambda] \tag{11.3-7}$$

There are a number of software packages that use the above formulation to derive geoid height. A description of the package developed at the School of Geomatic Engineering, University of New South Wales, can be found in such references as HOLLOWAY (1988) and KEARSLEY (1988).

The following are some characteristics, advantages and disadvantages of this method:

- ❑ Geoid height information is obtained relative to a geocentric ellipsoid.
- ❑ The basic data is terrestrial gravity observations in the vicinity of the computation point. The computation is essentially a "weighted mean" of gravity anomalies (appropriately scaled by Stokes' Function).
- ❑ Suitable only in areas with good local gravity coverage (as in most developed countries).
- ❑ Information is sometimes provided in the form of maps of geoid contours. Otherwise needs to be computed, but requires sophisticated software and detailed (and large) gravity data sets. The software and/or the gravity data may not be generally available.
- ❑ Variable accuracy, dependent upon:
  - quality and density of gravity coverage, and
  - sophistication of software.
- ❑ Relative geoid height (between two points up to several 100km apart) accuracy is higher than absolute values. Tests in Australia, U.S. and Europe indicate that relative geoid height can be determined at the few parts per million level.

Summary: Highest accuracy possible, but most inconvenient of the GPS heighting techniques to use by non-experts. Must be precomputed.

Geometric or Interpolation Method

This method relies on the direct evaluation of N from eqn (11.3-1), where h is provided by the GPS survey and H is determined from conventional levelling. If there are a number of points in the network for which N can be estimated in this way, N at other points can be obtained from an interpolation technique such as plane surface (and other surface) fitting, collocation, cubic splines, etc.

The following are some characteristics, advantages and disadvantages of this method:

- ❑ Simple to use, not requiring sophisticated software. No outside assistance is required and hence the GPS surveyor is in complete control of the computations.
- ❑ Geoid height information can be obtained relative to a local (non-geocentric) ellipsoid.
- ❑ Suitable where there are sufficient stations having both orthometric and GPS heights, and at the appropriate density to model the form of the geoid.
- ❑ Because no assumption is made about the datum for either system of heights, it is particularly suitable where the local height system may not be rigorously defined. For example, because the Australian Height Datum is not a true orthometric system, this method is still applicable as the "geoid" height is not determined by independent means, but from the (possibly flawed) data itself.
- ❑ The geoid height results can be represented as a contour map, or a set of coefficients for the functional model used for the interpolating surface.
- ❑ Variable accuracy, dependent upon:
    - quality and density of common station coverage, and
    - smoothness of geoid (for example, linear interpolation over distances even as low as 25km can cause 10cm errors in parts of Australia).
- ❑ Principal problem is the requirement to (GPS) survey levelled benchmarks (that is, additional stations). At least three are required if using the plane surface fitting option. Hence if GPS heighting is required, this aspect of additional fieldwork has to be borne in mind during the GPS survey planning process.
- ❑ Further, the results cannot be used beyond the perimeter of the GPS network.

<ins>*Summary:*</ins> Variable accuracy, but simple to use, and hence a commonly applied technique for medium accuracy applications.

11.3.2 GEOID HEIGHT APPLICATIONS

The various uses of geoid height information, the appropriate geoid height determination technique(s) and the accuracy required are summarised in Table 11.3-2.

Table 11.3-2. GPS applications of geoid height data.

N used for	"Order" for N: specification	Possible means of evaluating N
1. Geophysical exploration; reconnaissance surveys	Low: $\sigma_N \approx \pm 5$-10m	Low order ($n_{max} = 36$) geopotential models
2. Transforming between geodetic datums	3rd order: $\sigma_N \approx \pm 1$-2m	High order ($n_{max} = 180$ to 360) geopotential models
3. Control surveys for large scale mapping; engineering projects	2nd order: $\sigma_{\Delta N} \approx \pm 10$-$20$cm over 20 km (5-10 ppm)	a) Surface fitting N, as measured by GPS and levelling b) Detailed gravimetric evaluation
4. Height control (between 2nd to 3rd order) from GPS; special projects	1st order: $\sigma_{\Delta N} \approx \pm 20$-$30$cm over 110 km (2-3ppm)	Detailed gravimetric evaluations using surface gravity data

Although four categories of user have been identified in the above table, there are in fact two distinct user applications:

- the **use of geoid heights in the network transformation process**, and
- **converting GPS heights to levelled heights**.

These are discussed below.

Geoid Heights and the Transformation Process

Three types of coordinate transformation can be identified:

(1) Cartesian (X,Y,Z) <--> Geodetic (ϕ,λ,h)

(2) Geodetic (ϕ_1,λ_1,h_1) <--> Geodetic (ϕ_2,λ_2,h_2)

(3) Cartesian (X_1,Y_1,Z_1) <--> Cartesian (X_2,Y_2,Z_2)

Figure 11.3-3 summarises the various transformation options involving the coordinate conversions referred to above. In the context of GPS surveying Datum 1 can be identified as the GPS datum and Datum 2 as the local geodetic datum. GPS provides data directly in the Cartesian system on a global geocentric datum (for example, WGS84). Data can be easily transformed into the geodetic coordinate system (ϕ,λ,h) using eqn (11.1-14) without the need for any geoid height information. Geoid height mapped on the global datum ellipsoid must be used to transform to (ϕ,λ,H). On the other hand, conventional geodetic networks such as the AGD and AHD provide the coordinate triad (ϕ,λ,H), and in order to relate them to either (ϕ,λ,h) or (X,Y,Z) requires geoid height mapped on the local datum ellipsoid. Note that to perform a similarity transformation, the two sets of coordinates must be Cartesian. Hence a prerequisite for determining, or applying, a set of parameters to transform from Datum 1 to Datum 2 (and visa versa) is the conversion of local (ϕ,λ,H) to (X,Y,Z). *How can this be done?* Only the astro-geodetic and the geometric technique of geoid determination is capable of directly mapping the geoid on the local datum reference ellipsoid.

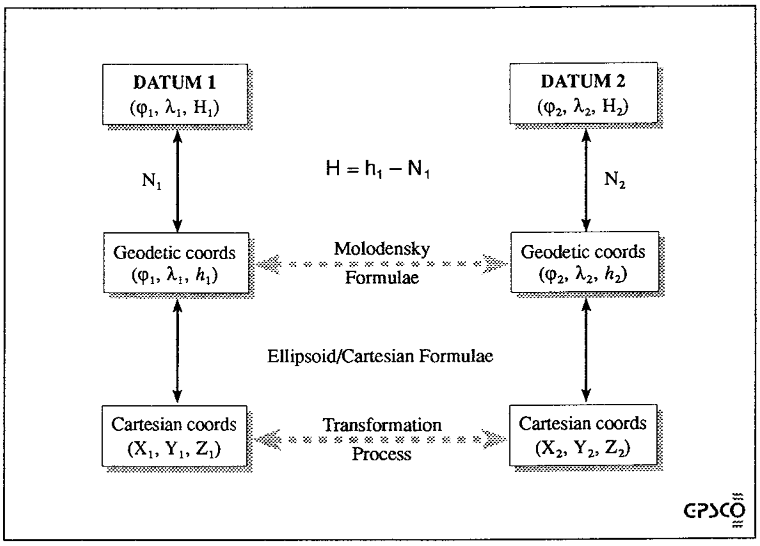


Figure 11.3-3. Coordinate transformation options.

In Table 11.3-2 it is suggested that a geoid model provided by spherical harmonics is adequate for this purpose. The procedure would therefore be:

(1) Use eqn (11.3-2) to evaluate geoid height N_1 on Datum 1 (that is, on the WGS84 ellipsoid).

(2) Assume $h_1 = N_1$ and transform (ϕ_1,λ_1,h_1) to (ϕ_2,λ_2,h_2) using the Molodensky formula for transforming ellipsoid height (it is not necessary to transform the latitude and longitude components):

$$\Delta h = \Delta X.\cos\phi.\cos\lambda + \Delta Y.\cos\phi.\sin\lambda + \Delta Z.\sin\phi - \Delta a.\frac{a}{R_N} + \Delta f.\frac{b}{a}.R_N.\sin^2\phi \tag{11.3-8}$$

where: a is the semi-major axis of the WGS84 ellipsoid,
b is the semi-minor axis of the WGS84 ellipsoid, and
ΔX, ΔY, ΔZ are the offsets of the WGS84 ellipsoid relative to the

local ellipsoid origin.
In the case of the AGD, the values are 116.0m, 50.47m and -141.69m (Table 11.2-1).

Δa is the difference between the semi-major axis of the WGS84 ellipsoid and the local ellipsoid, and hence in the case of the AGD $\Delta a = 23$m.

Δf is the difference between the flattening of the WGS84 ellipsoid and the local ellipsoid, and in the case of the AGD $\Delta f = 0.812045^{-7}$.

$$R_N = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$ where e is the eccentricity of the WGS84 ellipsoid.

(3) Apply Δh to N_1 to obtain N_2, the quantity required to convert H_2 to h_2.

This is the procedure used to determine the quantities in Table 11.3-1.

Under what circumstances can a knowledge of geoid height be dispensed with entirely? This condition is satisfied if there are only three common points used to determine the transformation parameters, as the geoid slope (relative to the an ellipsoid surface) can be accommodated by some of the transformation parameters.

GPS Levelling

GPS levelling has the following characteristics:

- ❑ Involves relative geoid heights.
- ❑ Is independent of any reference system, as it involves determining the separation of two physical surfaces: the topographic surface and the geoid. Requires geoid heights on a reference ellipsoid to which the GPS heights are also referred (either a local ellipsoid if GPS heighting process applied after GPS transformation to local datum, or a global ellipsoid if GPS results are not first transformed).
- ❑ GPS levelling can be considered a by-product of the GPS survey process. GPS heighting may therefore be a more economical heighting technique than standard levelling procedures.
- ❑ The resulting relative orthometric height accuracy is dependent on the accuracy of both the (relative) GPS-derived ellipsoid heights and the (relative) geoid heights. The accuracy can be expressed in terms of the length of the baseline, that is, in "parts per million".
- ❑ Relative (ellipsoidal) heighting by GPS survey is generally a factor of 2-3 times less precise than the horizontal components. Residual atmospheric biases are the source of the greatest uncertainty.
- ❑ GPS levelling involves additional calculations for geoid height determination, of which only the *geometric method* is relatively simple.
- ❑ Software can be purchased for computing geoid heights from geopotential models. Different models can then be used.

❑ Geoid heights from gravimetry, in Australia, can be obtained from the Australian Survey and Land Information Group (AUSLIG) (this is referred to as AUSGEOID93, see STEED & HOLTZNAGEL, 1994). In Canada, the geoid data is precomputed on a grid, and a program is available for purchase to interpolate to any point.

GPS Levelling Accuracy

GPS levelling is more economical than standard levelling techniques, but the accuracy is generally lower. The allowable misclose for spirit levelling is:

- 1st order: $4\sqrt{k}$ (mm) or 4cm in 100km 2-way levelling (0.4ppm), or 1.2cm in 10km (1.2ppm).
- 2nd order: $8\sqrt{k}$ (mm) or 8cm in 100km (0.8ppm), or 2.5cm in 10km (2.5ppm).
- 3rd order: $12\sqrt{k}$ (mm) or 12cm in 100km (1.2ppm), or 3.8cm in 10km (3.8ppm).

Note that the accuracy of spirit levelling is a function of the $\sqrt{\ }$ of the distance, and not just of the distance itself (as in the case of GPS accuracy). This is illustrated in Figure 11.3-4.

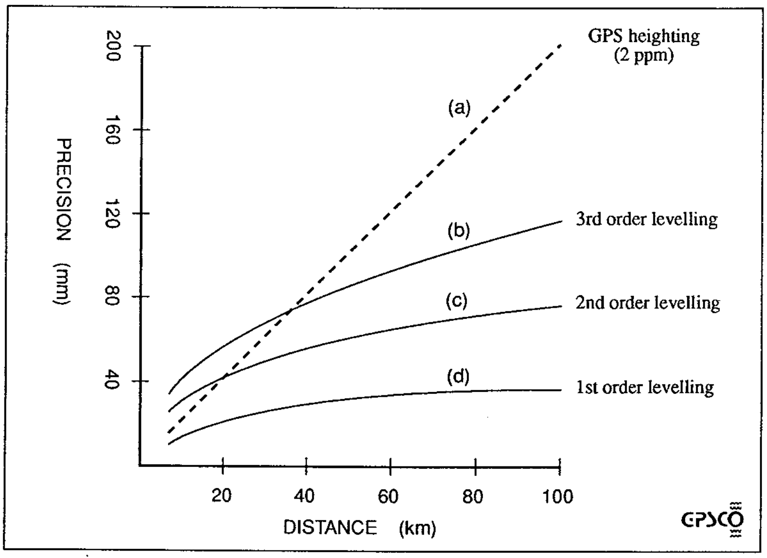


Figure 11.3-4. Comparison of accuracies of ellipsoidal height differences from GPS and from spirit levelling. (a) GPS heighting (2ppm), (b) 3rd order levelling, (c) 2nd order levelling, (d) 1st order levelling.
(MITCHELL, 1990)

The following comments can be made concerning GPS levelling accuracy:

❑ GPS accuracies appear to be of the order of 2-10ppm. At best, GPS levelling can provide 3rd order levelling standard for short distances (<20-30km), even if the geoid height computation is error-free.

❑ Geopotential models can provide relative geoid height data to an accuracy of the order of 3-10ppm.

- ❑ Geoid heights determined from gravimetry, in Australia, are considered to be accurate at the 2-6ppm level.
- ❑ The accuracy of geoid heights derived from simple interpolation techniques varies considerably, from several ppm to 10 or more ppm (Figure 11.3-8).
- ❑ GPS orthometric heights obtained from gravimetrically-derived geoid heights can be used to independently check (for blunders in) spirit levelled data. This is particularly feasible in Australia where the AHD is only of third order (MITCHELL, 1990).

11.3.3 GPS AND THE GEOID IN AUSTRALIA

The Australian Height Datum

The Australian Height Datum (AHD) was implemented on mainland Australia in 1971, and in Tasmania in 1979 (ROELSE et al, 1971; MITCHELL, 1990). This was the first attempt to define a continent-wide levelling datum. Prior to this, Australia had a multitude of levelling datums, each of which was adopted by a government utility to suit its own purpose in a region. These datums were usually based on a local tide gauge (HOLLOWAY, 1988).

The levelling data used in the AHD was largely of third order standard, observed during the previous decade and a half. In addition, some earlier first and second order levelling was included, as was a significant amount of one-way levelling in Queensland and the Northern Territory. Hence the standard of levelling throughout the network was not uniform. A total of almost 100,000 km of levelling in 261 loops was adjusted (ROELSE et al, 1971). The estimated internal precision of the AHD is about 8mm per $\sqrt{k}$, where k is the length of the level run in kilometres. That is, a height difference between points 100km apart should have a precision of the order of a decimetre.

In the 1971 adjustment, AHD was tied to Mean Sea Level (MSL) at 30 tide gauges distributed fairly uniformly around the Australian mainland coast. The MSL at each of these tide gauges was derived from several years of readings. These 30 MSL estimates acted as a constraint in the levelling adjustment by being assigned a "fixed" height of 0.000 metres. In Tasmania, the AHD is connected to MSL at two tide gauges, Burnie and Hobart. The AHD (Tasmania) is independent of AHD (Mainland), with no connection ever being made between them.

Is the AHD an orthometric height system? The answer is NO, for several reasons, the most important of which are:

- Observed gravity was not used to convert the staff readings to geopotential differences (eqn (11.1-22)), and to convert heights in the geopotential system to orthometric heights (eqn (11.1-24)). Rather, model values of gravity were used (HOLLOWAY, 1988).
- The assumption that MSL, as defined by the 30 mainland tide gauges (and two Tasmanian gauges) was coincident with the geoid. The effect of this neglect of Sea Surface Topography becomes obvious when the differences between the AHD (Mainland) and a "free" adjustment of the same levelling data holding only one tide gauge height fixed is inspected (Figure 6.2, MITCHELL, 1990). There is a significant distortion of the network indicating that it was inappropriate to assign an orthometric height of zero to all MSL estimates.

The second point is of particular concern for GPS heighting, as eqn (11.3-1) is no longer strictly valid as far as the AHD is concerned (refer to the early discussion regarding the insensitivity of the geometric method of geoid modelling to weaknesses in the height datum definition). Hence an understanding of the nature and magnitude of errors in the AHD is important if GPS is to used for heighting. This issue was studied by analysts appointed by the GPS User Group and their report (IBID, 1990) details current approaches to GPS heighting in Australia. In the section below is summarised the status of geoid height determination in Australia.

Geoid Studies

Astro-Geodetically Determined Geoids

The first such determination was carried out in 1967, using 600 astro-geodetic deflections of the vertical. In 1971 a new geoid was prepared, relative to the AGD, but based on almost double the number of astro-geodetic data (Figure 11.3-5). In classical geodetic practice the preparation of an astro-geodetic geoid was a necessary part of the process of datum definition known as the "development method", by which the assigned ellipsoidal coordinates of the datum origin station (in Australia's case, the Johnston Trig datum station) were altered in such a way as to ensure an astro-geodetic geoid that minimised the absolute values of the geoid undulations across the datum. The value of N at the Johnston station for a geoid that satisfied this "best-fit" condition should have been -6m. However, due to the legislated value of h at Johnston Origin of 571.2m on the one hand, and the levelled height H of 566.3m on the other, the implied geoid height is in fact +4.9m. Hence the geoid in Figure 11.3-5 has a datum shift of +10.9m relative to a "best-fit" geoid. Such a geoid map can be used for converting AHD heights to ellipsoid heights on the AGD, and hence transforming to the Cartesian formulation.

Geoids from Geopotential Models

A high degree spherical harmonic model such as the OSU89 (RAPP & PAVLIS, 1990) or OSU91 provides a very convenient representation of the geoid (Figure 11.3-6). Testing of the OSU89 model, and others, has been carried out and described in, for example, KEARSLEY & GOVIND (1991). One test has been to compare the gravity anomaly implied by OSU89A (eqn (11.3-7)) with the observed values. From an admittedly small sample of data, it appears that in areas where the rms of the difference between the observed and model gravity anomalies is less than 10mGal, the OSU89A model recovers relative geoid height with a precision of at least 5ppm (IBID, 1991).

Table 11.3-3 shows the difference between the OSU89A and OSU91A geoid height values for a small Sydney network. Note that although there is an approximately 0.25m difference, there is also a small grade and hence the relative geoid heights between points in the network will be slightly different (though only at the few centimetre level), depending upon which geoid model is used.

Although the evaluated series (eqn (11.3-2)) gives N relative to a geocentric ellipsoid, the procedure described earlier can be used to transform the geoid height so that it may refer to a local (non-geocentric) ellipsoid such as that implicit for the AGD, as in the case of Figure 11.3-5. A computer program, developed at the School of Geomatic Engineering, University of New South Wales, permits the geoid height implied by the OSU91 geopotential model to be obtain on any reference ellipsoid, through the application of the transformation formula at eqn (11.3-8).

Table 11.3-3. The Sydney GPS network, OSU89A and OSU91A geoid models.

Name	Latitude	Longitude	N_{OSU89A}	N_{OSU91A}	Δ
B402	S33°54'56.567"	E151°13'52.399"	22.454	22.710	*.256*
BOTY	S33°58'23.355"	E151°14'10.566"	22.217	22.432	*.215*
CETS	S33°55'05.557"	E151°13'57.743"	22.443	22.696	*.253*
CPDH	S33°55'06.198"	E151°14'03.196"	22.440	22.692	*.252*
G106	S33°55'13.377"	E151°13'56.767"	22.435	22.687	*.252*
GLEB	S33°52'15.439"	E151°11'05.280"	22.679	22.988	*.309*
KYMA	S33°56'45.707"	E151°09'39.054"	22.448	22.725	*.277*
LAKE	S33°56'18.011"	E151°12'40.436"	22.398	22.654	*.256*
MBTS	S33°56'28.596"	E151°15'53.198"	22.297	22.511	*.214*
MMCH	S33°51'33.954"	E151°13'18.439"	22.679	22.972	*.293*
PETE	S33°54'36.572"	E151°10'31.635"	22.551	22.846	*.295*
Q744	S33°54'02.521"	E151°15'03.756"	22.485	22.735	*.250*
Q855	S33°54'01.506"	E151°14'56.739"	22.489	22.740	*.251*
ROSE	S33°54'49.779"	E151°12'18.046"	22.498	22.774	*.276*
UNSW	S33°55'08.060"	E151°13'52.148"	22.442	22.697	*.255*
UU33	S33°55'02.962"	E151°13'40.084"	22.452	22.710	*.258*
WBTS	S33°51'10.704"	E151°17'09.094"	22.618	22.876	*.258*

Δ: $N_{OSU91A} - N_{OSU89A}$

Gravimetric Geoids

The first gravimetrically-determined geoid was computed in 1969. Subsequent to that, several refinements to the computational procedures have occurred over the last two decades, but it is only in the last few years has there been a revival of interest in an Australia-wide gravimetric geoid. The task of computing a national geoid (to support GPS activities) was formally assumed in July, 1989, by Australia's national geodetic agency: the Geodetic Services Section of the Australian Surveying and Land Information Group (KEARSLEY & GOVIND, 1991). AUSLIG has completed the task at a resolution of 0.1° using a suite of programs largely developed in the School of Geomatic Engineering, University of New South Wales (HOLLOWAY, 1988; KEARSLEY, 1988). The computation is based on the combined gravimetric and geopotential method (based on eqn (11.3-4). The geoid model, known as AUSGEOID93 (STEED & HOLTZNAGEL, 1994), is available in electronic form, either as point values or grids, for map sheets at several scales.

Geometric Methods

By their nature these geoid models are local, computed by the GPS surveyor, and have relevance only in the area of the GPS survey. HOLLOWAY (1988) describes several tests using this method of GPS heighting. In addition, the algorithm for obtaining the plane surface fit to the geoid "spot" heights is also described. Figure 11.3-7 is an example of the geoid height map that can be derived using such an interpolation method, in this case for a GPS network in South Australia. There were 45 stations in this network that had both GPS and AHD heights. This is far more than would normally be available. *Hence the contour map of the geoid shows considerable detail.* (It must be emphasised however that the map is not an independent determination of the geoid separation in this area as it also contains errors in the AHD.)

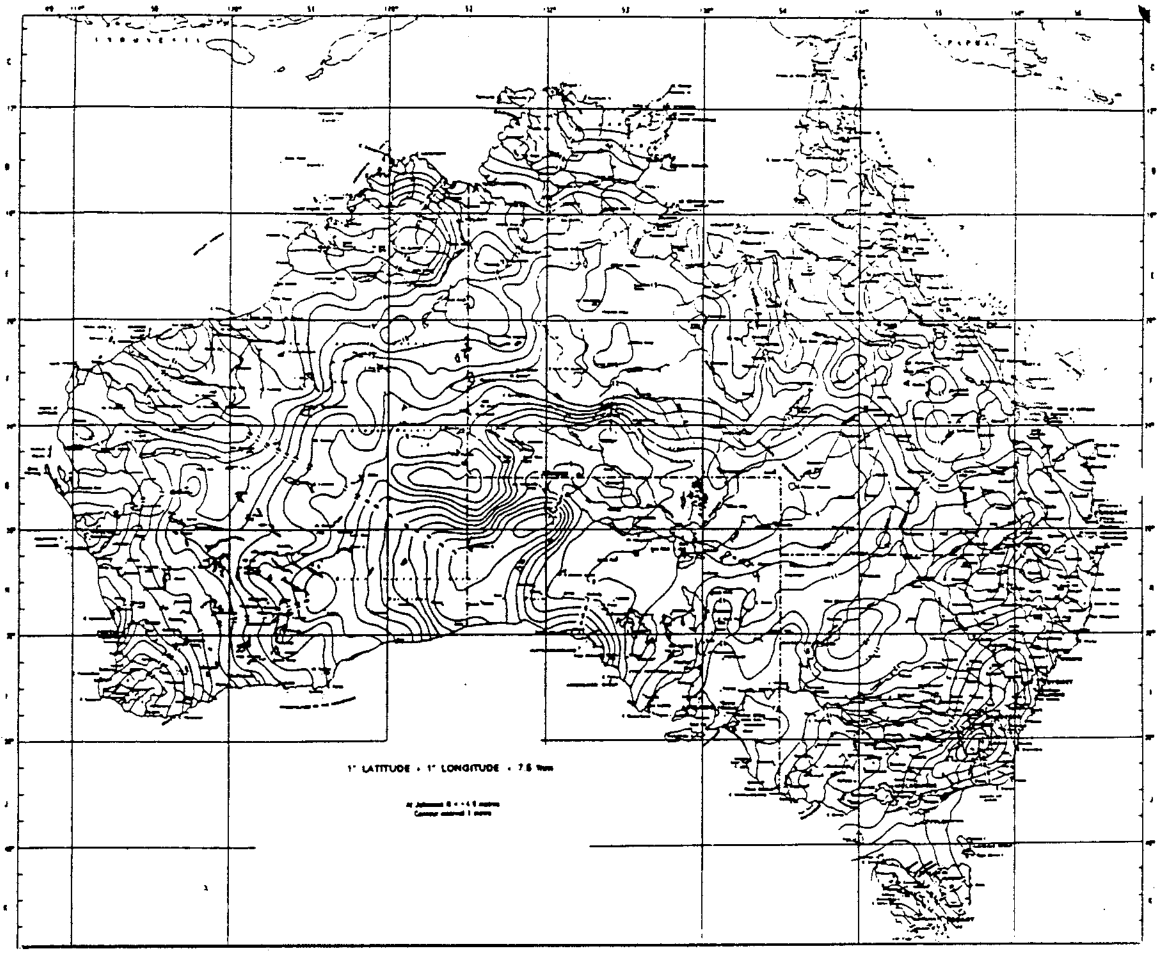


Figure 11.3-5. Astro-geodetic geoid of Australia (mapped relative to the AGD).

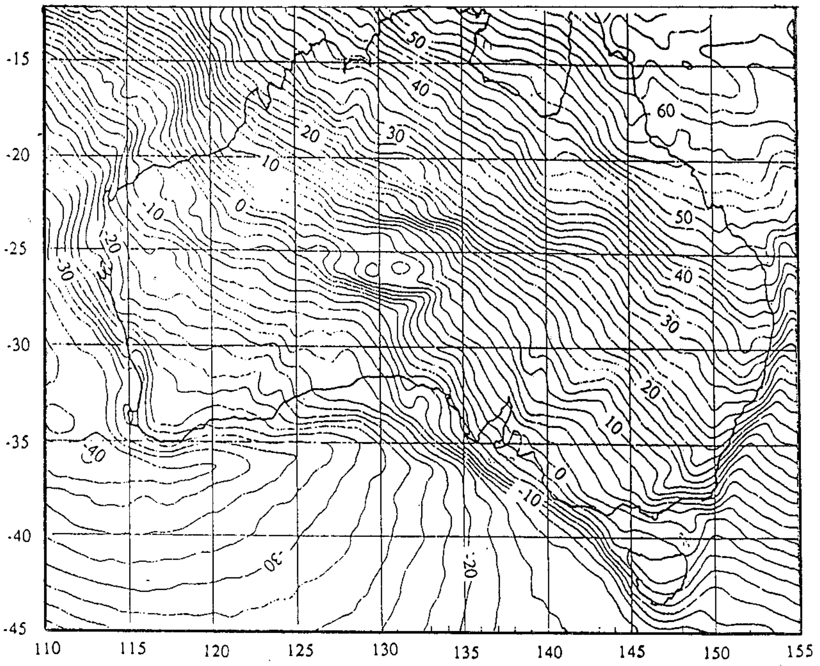


Figure 11.3-6. Geoid of Australia from OSU89A (relative to the WGS84 ellipsoid).

A compendium of relative geoid comparison results between the AUSGEOID93 gravimetric geoid and geometrically-derived geoids is given by STEED & HOLTZNAGEL (1994). Figure 11.3-8 illustrates the results in terms of parts per million, and in millimetre miscloses in AHD height, for a variety of tests involving Western Australian and New South Wales GPS results. Apart from some poor comparisons for short lines, Figure 11.3-8 implies that 5ppm is achievable from geometric methods for orthometric height determination using GPS. Table 11.3-4 shows the comparison for the small Sydney GPS network referred to earlier.

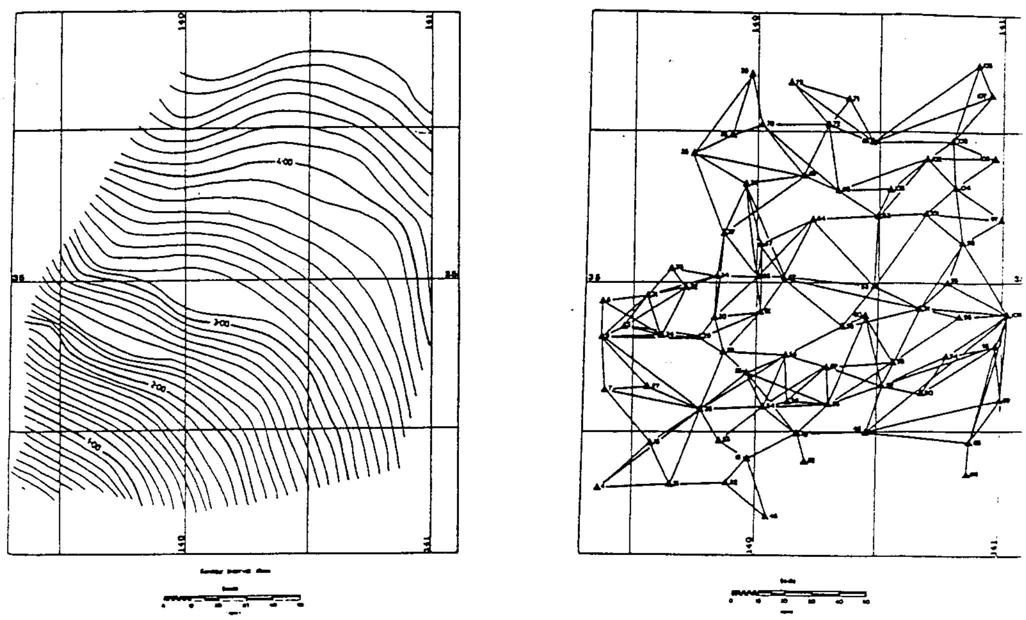


Figure 11.3-7. Geoid map derived from interpolating spot values of GPS height minus levelled height for a network in South Australia. *(HOLLOWAY, 1988)*

Example of GPS-Derived Orthometric Height in the Sydney GPS Net

A variety of geoid height information was available: OSU89, AUSGEOID93 and AHD minus GPS-derived ellipsoidal height. Table 11.3-4 summarises the results of the comparisons. Note that the datum for the GPS network is the Mather pillar at the University of New South Wales. Any error in the height of this fixed station will influence (in an absolute sense) all the derived GPS heights.

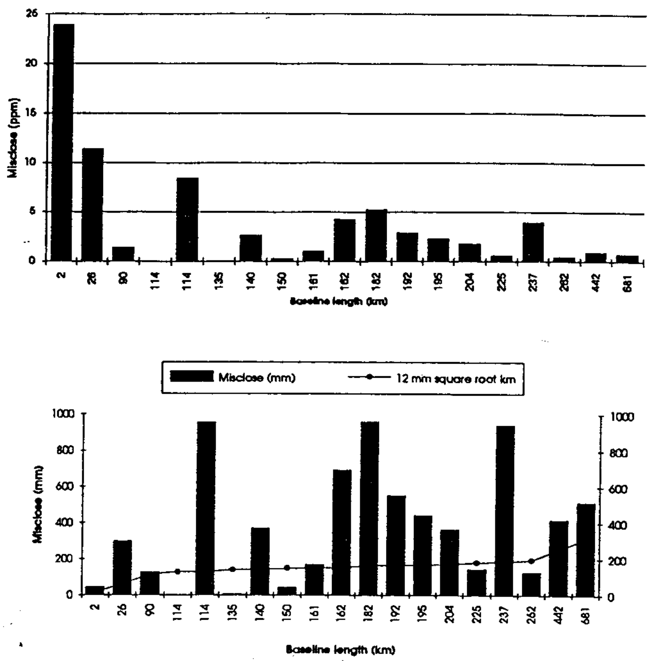


Figure 11.3-8. Difference in AHD minus GPS derived difference in AHD using AUSGEOID93 geoid height values. *(STEED & HOLTZNAGEL, 1994)*

Table 11.3-4. The Sydney GPS network, a comparison of geoid models.

Name	N_{OSU89A}	N_{GRAV}	Δ1	h_{GPS}	H	N_{GPS}	Δ2
B402	22.454	22.498	*.044*	59.008	36.410	22.598	-*.100*
BOTY	22.217	22.269	*.052*	41.407	19.003	22.404	-*.135*
CETS	22.443	22.482	*.039*	109.199	86.605	22.594	-*.112*
CPDH	22.440	22.486	*.046*	90.837			
G106	22.435	22.470	*.035*	81.777	59.075	22.702	-*.232*
GLEB	22.679	22.646	-*.033*	24.225	1.45*	22.77*	-*.12**
KYMA	22.448	22.412	-*.036*	24.336	1.812	22.524	-*.112*
LAKE	22.398	22.406	*.008*	59.685	37.15*	22.53*	-*.12**
MBTS	22.297	22.300	*.003*	47.437	24.962	22.475	-*.175*
MMCH	22.679	22.705	*.026*	33.206	10.375	22.831	-*.126*
PETE	22.551	22.531	-*.020*	29.272	6.604	22.668	-*.137*
Q744	22.485	22.521	*.036*	99.863	77.206	22.657	-*.156*
Q855	22.489	22.519	*.030*	78.809	56.15*	22.66*	-*.14**
ROSE	22.498	22.520	*.022*	82.964			
UNSW	22.442	22.480	*.038*	88.192	65.579	22.613	-*.133*
UU33	22.452	22.505	*.053*	50.332	27.623	22.709	-*.202*
WBTS	22.618	22.652	*.034*	108.506	85.748	22.758	-*.106*

Δ1: $N_{GRAV} - N_{OSU89A}$

Δ2: $N_{GRAV} - N_{GPS}$
$N_{GPS} = h_{GPS} - H$

N_{GRAV} computed from OSU91A + observed gravity data (AUSGEOID93)

* implies accuracy of orthometric height is weak

Chapter 12: Datums and the Future of GPS

12.1 CONSTRAINING GPS NETWORKS

The results of a GPS survey are a set of 3-D coordinates, nominally in the WGS84 datum. *Rarely are these results in the form that is useful to the client or sponsor of the survey.* The results must therefore be converted into more "useful" quantities. This could require for example:

☞ Separation of the results into horizontal and vertical components (the client may only be interested in the horizontal components) -- §11.1.

☞ Transformation to a local geodetic datum -- §11.1 and §11.2.

☞ Reduction of the GPS heights to the same system as used for geodetic levelling -- §11.3.

☞ A further adjustment, or constraint, of the GPS network to "fit" into a previously defined network.

This list is not intended to be complete, or to imply that all of the above will necessarily be required. It merely serves to illustrate the relatively complex post-GPS adjustment operations that may need to be performed in order to make the GPS network results truly useful. Many of these operations require that external control information be used.

By "external information" is meant the position information of some GPS points (either 3-D coordinates, horizontal components only, or even just levelled heights), that have been obtained from:

(a) other GPS surveys, or

(b) conventional surveys, or

(c) official geodetic databanks.

There are a number of "external" adjustments that are pertinent to GPS surveying. Although the most common is of the GPS / local control adjustment, it is by no means the only type. For example, GPS surveys may involve "primary" points and "secondary" points. A greater effort may have been invested in coordinating the primary stations, with a higher percentage of baseline redundancies, longer observation sessions, etc. One adjustment strategy would be to carry out a minimally constrained multi-session adjustment for the primary stations, as described in §9.3. To subsequently incorporate the session solutions for the secondary stations into a multi-session campaign solution in which the primary stations are held fixed.

However, to illustrate some of the steps and issues involved in combining external coordinate information with current GPS results the more common example of the *integration* of a GPS survey on the one hand, and local geodetic control information on the other will be considered.

The introduction of external geodetic control information into an already adjusted (though "minimally constrained") GPS network generally requires that:

- ❑ The VCV information of the GPS-only network should correctly reflect the true errors of the multi-session adjustment (§9.4).
- ❑ Geodetic datum coordinates of the local control stations (occupied during the GPS survey) are somehow introduced into the GPS-only adjustment.
- ❑ The combined adjustment must account for the differences in the datums, through some transformation modelling (§11.2).
- ❑ The appropriate weights between the different observations (GPS relative station coordinates on the one hand, and the externally defined coordinates of the control stations on the other), and the degree of constraint to be introduced into the combined GPS / external data adjustment, will need to be carefully specified.
- ❑ The combined adjustment should take into account the *incomplete* height information available concerning control stations. Usually only orthometric height is available, however the required ellipsoidal height contains geoid height information as well (§11.3).
- ❑ As with any network adjustment, an examination of the combined network solution should be made to ensure that there are no inconsistencies which may indicate a problem with the GPS survey <u>or</u> the control station information. *This is an issue of Quality Control.*

The final "external" adjustment will usually (though not always) result in the 3-D coordinates of all the GPS stations, on the local geodetic datum, together with their uncertainties; as well as the local transformation parameters relating the GPS survey datum to the local geodetic datum.

12.1.1 THE BASIS FOR THE EXTERNAL ADJUSTMENT

Defining the External Adjustment

A network adjustment concerned with combining the GPS-only network solution with local geodetic control information must confront the following issues:

(1) Specification of the <u>degree of constraint to be applied</u>. The GPS-only solution usually involves only one fixed (datum) station. The options now available are, for example:

- If this Datum Station is, in addition, a local control station, the <u>minimally constrained</u> nature of the original GPS solution can be preserved after the adjustment and transformation of the network coordinates to the local geodetic datum. The final coordinate values will be a combination of the GPS-derived, and the prior geodetic values.

- On the other hand, all control stations can be held fixed in the final combined adjustment. This would be appropriate if the aim of the GPS survey was *geodetic control densification.* The danger is, however, that any errors in the control stations will lead to a distortion of the relatively high quality GPS-only results to fit the framework defined by the control system.
- Options between these two extremes are possible, in which some of the control stations may be held "almost" fixed (by appropriate choice of apriori VCV information), or no station is held absolutely fixed and all are allowed to "float".

(2) The completeness of the transformation model adopted. If a minimally constrained solution is sought, the three shift (translation) parameters are simply the differences between the GPS and the local geodetic coordinates of the Datum Station and need not be directly estimated. In other cases, the shift parameters are estimated as a "weighted mean" of the differences between the two sets of coordinates available for the control stations. The scale and three rotation parameters are often estimated in addition.

(3) The geoid heights at the control stations being specified in order to ensure the local geodetic coordinates are in fact directly comparable to the 3-D GPS results. For small area networks, or where only three control stations are involved, it is usually possible to neglect geoid heights, as the geoid can be assumed smooth and the estimated transformation parameters will absorb this effect.

The second issue was discussed in detail in §11.2, while the use of geoid height (external) information was discussed in §11.3. A further examination of the transition from a minimally constrained network solution to one that is constrained to some degree by the external control information must be made, and the following will need to be considered:

- The nature of the GPS results on the one hand, and the external information on the other. Both sets of coordinates are in fact the "observations" which must be input to a further network adjustment.
- How to alter the constraint of an adjustment via the observation VCV information.
- How to affect such an adjustment using commonly available network adjustment software.

The "Observations" for an External Adjustment

A GPS-only solution is generally minimally constrained, but includes redundancies through the multiple occupation of some of the network stations. The multi-session adjustment seeks to determine the "best-fit" of the GPS baseline data. **The incorporation of additional coordinate data on some of the stations has the effect of introducing further redundancies into the network solution.** ***Hence the adjustment changes.***

This additional information is usually in terms of horizontal coordinate components (latitude and longitude) and height, and is extracted, for example, from the relevant control station database maintained by the responsible geodetic authority. (Note that if an eccentric station were occupied, it is these coordinates that are required!)

At this stage it is worth considering how the two sets of observations can be reconciled. Some ideas include:

- To avoid the use of datum transformations completely, the external coordinate information may be converted into (largely) datumless quasi-observations such as baseline components, and the adjustment proceeds in the standard manner, as in the example of a GPS multi-session adjustment based on the original (GPS) datum station.

- The transformation parameter determination procedure based on the Least Squares analysis of the coordinates of the common points in fact provides the external adjustment sought, but only for the common points. The remaining GPS stations can be transformed using the solved-for transformation parameters and then combined with the external information.
- The external adjustment and the determination of the transformation parameters can be carried out simultaneously.

The quality of the data also needs to be specified in the adjustment. In the case of GPS, the VCV information reflects the internal consistency of the GPS-only network (see below). The precision of the local control station coordinates will also influence the quality of the final combined adjustment, especially if they are used to constrain the adjustment. Obviously the precision assumed for the local control can affect the quality classification the survey receives (§10.2), and hence particular care needs to be exercised. **The problem is that the VCV matrix information from the original geodetic datum adjustment is not usually available**.

From experience, apart from the great difficulty in assigning realistic VCV information to the local control coordinates mentioned above, the major causes of problems in combined network adjustments are errors in the coordinates of the control stations, or the misidentification of the station mark to which they refer. The issue of "Quality Control", as discussed in §10.4 for GPS-only networks, is therefore just as important for any combined external adjustments.

It should be mentioned that measured EDM distances, or theodolite observations, etc., can *theoretically* be included in the external adjustment, and will have the same impact as external control information. However, this option is rarely exercised, but may be more commonly used in future as geodetic network readjustments are attempted using all available data: GPS results and historical non-GPS data.

Scaling the GPS-only VCV Matrices

The VCV matrices output by the GPS data reduction software (for example, the single session solutions) are generally over-optimistic. They indicate that the relative positioning accuracies are of the order of a few parts per million (ppm). The VCV matrix of the estimated parameters is principally influenced by the baseline-satellite geometry and the observation precision. Although they may be a true indication of the internal precision (random error effect), they give no information on the (external) systematic errors, in particular the satellite orbit and atmospheric refraction uncertainties.

Some indication of the (variable) systematic errors can be obtained from the multi-session network adjustment (§9.3). It is not unusual to find that the final (relative) station coordinate uncertainties are of the order of 10ppm or worse, after modifying the GPS VCV matrices using any of the strategies described in §9.4.

As with the construction of the minimally constrained (GPS-only) multi-session solution, the external network adjustment itself can be used to provide a clue as to the appropriateness of the stochastic information. If the VCV matrices from the GPS adjustment solution are input, without alteration, into a network adjustment program such as GEOLAB™, the corrections to the station coordinates in the secondary network adjustment may be too large in comparison to the apriori accuracy of the GPS baselines or stations. The variance factor is often significantly greater than unity. The adjustment may then be re-run, but only when the aposteriori variance factor is used to **scale** the input VCV matrices, ensuring that the variance factor test passes after the second iteration adjustment.

Note that although **the baseline "observation" VCV matrices during the multi-session adjustment phase may have been modified separately through the use of length dependent variances in the VCV matrix, when this minimally constrained network is further adjusted** (after the incorporation of external control data), **the total VCV matrix** (of dimension 3nx3n, where n is the number of stations in the GPS-only network) **is usually multiplied by a single factor**. (This reflects the fact that often baseline "observations" are input into GPS-only multi-session adjustments, and their VCV matrices can be easily modified using length dependent factors, whereas the output of a minimally constrained adjustment is a set of station coordinates, and their VCV matrix is more easily modified using a scale factor rather than a length dependent factor.)

GEOLAB™: A One-Step External Adjustment

A one-step combined adjustment is carried out within a program such as GEOLAB™ by:

- Specifying the GPS-only baseline input (with unaltered VCVs, or re-scaled in order to pass the variance factor test), and identifying this input group with a transformation parameter set.
- Input the control station coordinates in the form of horizontal coordinate components (latitude and longitude), and ellipsoid height values, with appropriate uncertainties (which will be converted into the VCV matrix associated with these "observations"). Geoid height values at the stations, where available, can be input as separate observations. The control data may be identified with a datum different from the GPS data.
- Specifying the apriori transformation model, and identifying the parameters to be adjusted.
- Nominate the datum station (or stations) whose coordinates are to be held fixed. The choice influences the transformation parameter results as well as the network coordinate results.

GEOLAB™ is an example of the total network adjustment tool. It is widely used by GPS surveyors in Australia and other countries. The manual gives many hints on how to tackle a variety of survey adjustment problems. In the case of GPS, there are a number of useful features, such as the ability of the program to scan the output files of many of the commonly used GPS reduction packages (such as TRIMVEC™, GPSURVEY™, SKI™, GPPS™, PRISM™, etc.) and extract the relevant information in order to construct the GEOLAB™ input files.

As with any adjustment the results must be analysed using statistical tests, and the like, in order to verify that the adjustment results are reliable, and are not biased by outlier observations, inappropriate VCVs, errors in station identification, and a host of other potential problems. In fact, the label "one-step adjustment" is really a misnomer, as several trial adjustments are often performed in order to identify problem data, or to determine the "best" empirical scaling factor for the input VCVs, or the appropriate transformation model to use (generally a choice of whether to solve for all seven of the similarity transformation parameters, or just a subset), etc.

12.1.2 THE CONSTRAINED NETWORK SOLUTION

Analysing External Adjustment Results

As is the case with any Least Squares adjustment, there are a number of indicators that should be checked. Analysing adjustment output is a topic of its own. *Although there may be a number of well documented procedures that may be followed, it must be emphasised that quality control analysis of adjustment results is as much an "art" as a "science".* The challenge is to use the indicators available to (a) determine whether the adjustment was successful, but if not, to (b) interpret the telltale indicators as to the possible source of the error or distortion in the adjustment. The former should be routine, while the latter more correctly belongs to the set of "trouble-shooting" procedures.

Some of the indicators are statistical in nature, for example many analysts look immediately to the <u>variance factor</u>. But there are, in addition, the results of various statistical tests on the residuals and on the estimated parameters, the correlations between parameters, etc., to be noted. Some indicators are deterministic, for example, the "root-mean-square" of the residuals, the largest residual, etc. But there are a large number of indicators which cannot be so easily classified. They belong to a "fuzzy" group of indicators that generally are very application specific. For example, in chapters 7 and 8, several "rules-of-thumb" for double-differenced phase solutions were quoted. In the case of secondary network adjustments (either minimally constrained or incorporating external control information), drafting such "rules-of-thumb" is more difficult. The following is merely a list of some ingredients of a "good adjustment":

- The network must have "symmetry": well spaced stations within the network, well spaced control points and a "well shaped network" (for example, not linear in form).
- "Sufficient" degrees of freedom in the adjustment.
- Control station coordinates of approximately similar accuracy (for example, mixing spirit and trigonometrically levelled heights can lead to problems if the VCV information is not accurately defined).
- The observations should be relatively homogeneous (mixing results from different eras, different field procedures, or from different software, may lead to problems).
- Careful trouble-shooting should have been exercised at all earlier phases of the adjustment, by following the strategy of "adjust a little, test a little".

There are a number of further points that can be made with reference to the combined adjustment solution, although they are not elevated here to the status of "rules-of-thumb":

- <u>Any large residuals or corrections to the coordinates usually indicate a problem with the control information</u>. Following the quality control tests made on the GPS-only solution, this solution can be assumed to be largely free of error. The cause of error may be incorrect control coordinates in the local datum.
- <u>If the large residuals are generally restricted to the vertical components, this may indicate a height error at the control stations</u>. A common error is to input the available levelled (or orthometric) height, rather than the ellipsoidal height. Even approximate values of the geoid height, as obtained from a spherical harmonic model (§11.3), can often appreciably improve an external adjustment.
- <u>The aposteriori variance factor should be examined</u>. If it is greater than unity this indicates either that the uncertainties assigned to the control information are too low,

or the GPS-only observation precisions are too optimistic. Either of the VCVs may be scaled to obtain a more "balanced" solution.

- The resulting transformation parameters should not be large (>0.5ppm in scale, >0.5" in rotations, but beware of the strong correlations between rotations and translations) as this may indicate a possible problem with the control information. Large uncertainties in the parameter estimates may also be evidence for this. Some trouble-shooting may isolate the problem, but only if there is sufficient redundancy in the external adjustment (more than the minimum number of common points in the GPS and local network).
- The statistical indicators should be studied first. If there is a hint of a problem, generally the next step is to scan the residuals. It may or may not be possible to immediately pinpoint the source of the problem. If it is not obvious (and it is not uncommon to have more than one error, the effect being for the errors to interact and to cloud the sources of the problems), certain trouble-shooting procedures such as investigating loop closures (§10.3) may have to be resorted to.

Solution Output and Survey Classification Issues

As with the primary GPS reductions described in chapter 7, and the secondary (GPS-only) network adjustments discussed in chapter 9, "quality control" is an important issue (chapter 10). There are a number of strategies that can be used to verify the quality of the adjustment, some specific to the external adjustment, others common to all types of Least Squares operations. One of the biggest differences is the scope for finally estimating the GPS accuracy, if the external ground coordinates are of the adequate quality. This also impinges directly on the issue of survey accuracy assurance, as discussed in §10.2. As a reminder, there are two levels of GPS survey result accreditation in Australia and the U.S.:

(1) Those that are a function only of the field survey methods, reduction techniques and the results of minimally constrained network adjustment -- *the CLASS*.

(2) Those that are a function of the above **AND** the conformity of the new survey results with the existing network coordinate set after the transformation process required to convert results from the original datum (WGS84 in the case of GPS) to the local geodetic datum -- *the ORDER*.

If the external control is superior to that expected from the GPS survey procedures used (which in Australia define the CLASS of survey -- level (1)), then a distortion of the GPS-only net to fit the external control *does not degrade the quality of the GPS results*. The resulting GPS survey ORDER will not be greater than the CLASS (level (2)).

However, if the surrounding control information is of a poorer quality than that expected from the GPS survey procedures used, then the external adjustment will lead to a GPS survey ORDER classification that is lower than the CLASS. That is, the GPS survey will be classified as having an ORDER no higher then the external control, no matter how precise (or how high the CLASS) of the survey.

12.1.3 GPS AND THE FUTURE OF GEODETIC NETWORKS

The challenge for geodetic authorities is to promote the use of GPS in the most effective manner, but to still maintain the basic fabric of control network methodologies, such as:

- The notion of a "hierarchy" of control network points with high precision, or so-called "first order", points established at a lower density than the "second" or "third" control points. *This is consistent with the basic survey practice of "working from the whole-to-the-part".* The issues therefore are:
  - Given that GPS surveying is able to deliver relative accuracies of the order of several parts per million (better than the old "first" order standards), should the same multi-level primary, secondary, tertiary network system be maintained?
  - Efforts should be made to redefine or renovate the geodetic network to a level of accuracy at least an order of magnitude better than that possible using the "standard" GPS surveying methods.
- That control points should be located where they are needed, hence:
  - The practice of even geographic distribution of points may need to be changed. For example, where the population is sparse and no resource development needs must be met by the geodetic network, a close and even distribution of geodetic marks may not be necessary.
  - The control points should be located where they are most accessible, such as in the valleys rather than on hilltops. However, a degree of station intervisibility must still be assured.
- A geodetic control network is an asset that must be maintained and upgraded by the geodetic authority, hence:
  - Policies must be established that ensure continuity with existing practice of station marking, maintenance, documentation.
  - The geodetic authority must develop and enforce survey methodologies and practices that will ensure the continued maintenance, as well as the extension or densification, of the geodetic network to some specified level of accuracy.
  - Given that the cost of surveying a station using GPS may be significantly less than the cost of establishing a permanent mark, there may be justification for a much sparser network of high quality GPS points.
  - What are the respective roles of the federal geodetic authorities vis a vis state lands departments?
- The definition of a geodetic network in terms of classical notions of a non-geocentric datum based on a single local origin station, with the reference ellipsoid oriented in such a manner as to ensure a "best fit" of the geoid over the area of interest, is outmoded because:
  - Users of GPS technology in point-positioning (or navigation) mode want to easily relate their position information to the datum (and the maps based on that datum). *Although point-positioning accuracies are 100m (95% of the time), datums such as the AGD are several hundred metres offset from the geocentre!*
  - GPS survey users can avoid the need for coordinate transformations if the local geodetic datum is "close" to WGS84.

- ❑ To make use of ALL geodetic data in the establishment of the geodetic network, hence some of the issues are:
  - How will GPS baseline results be used to strengthen the network?
  - What network adjustment strategies should be used when combining GPS and historical geodetic data (terrestrial measurements of distance, angles, levelling, etc., and space-based observations such as TRANSIT and SLR coordinates, and VLBI baselines)?
  - How is the network recomputation (or **renovation**) to be carried out in the context of datum **redefinition**?

With the progressive improvement in GPS "geodesy" methodologies, relative accuracies of the order of 0.01ppm can now be achieved. These accuracies are necessary, for example, for scientific studies of crustal motion. Some of the most important legacies of such studies are:

- The establishment of an international network of permanent GPS tracking stations supporting the production of high precision orbits by the International GPS Service for Geodynamics (IGS). *This fundamental network makes it possible for geodetic authorities to carry out GPS surveys to a higher accuracy than possible previously by providing a direct connection to a superior network as well as satellite ephemeris information with accuracies at least an order of magnitude higher than that available from the Navigation Message.*
- In addition to precision orbits, the IGS stations contribute to the definition and maintenance of the global high precision geocentric datums known as the ITRS (International Terrestrial Reference System -- §2.1). *These are "close" to WGS84 (on average within a metre), but of higher internal consistency.*
- The notion of four-dimensional geodesy, whereby station coordinates expressed in the ITRS must be accompanied with an epoch identifier. *Making connections to this network therefore gives rise to issues of precise datum definition that were never previously considered when geodetic datums were defined in isolation on a local or regional basis.*

The IGS and ITRS therefore provide for the first time a convenient means of densifying high precision, so-called "zero order" networks to the local, regional or continental level (§12.2). These "zero order" networks can be expected to have internal accuracies at the 0.1ppm level or better, and hence are ideal for providing the "backbone" for the new high precision datums for the 21st century.

The IGS and the New Geodetic Datum for Australia

The IGS and ITRS have provided the impetus for a renovation and redefinition of the AGD (MORGAN et al, 1996). By datum **renovation** is meant the strengthening of the geodetic network in Australia through the establishment of a superior GPS-only network at an approximate 100-500km spacing, with the highest internal accuracy (far higher than that possible through the use of commercial GPS surveying techniques). Datum **redefinition** is achieved by tying the GPS-only network to several "fixed" stations of the IGS/ITRS network, and hence effectively making (a) the datum geocentric, and (b) for all intents and purposes coincident with the GPS satellite datum WGS84.

The new datum is known as the Geocentric Datum of Australia 1994 (GDA94). The New Zealand datum is also undergoing a redefinition to make it compatible with GDA94.

To understand how the AGD is being transformed into the GDA it is important to recognise the roles of the various GPS networks (MANNING & HARVEY, 1994):

(1) There are three stations in Australia that are a subset of the **IGS core tracking network** (Figure 6.2-4). These are Yaragadee, Tidbinbilla and Hobart (Figure 12.1-1), and their coordinates can be considered "fixed" in the ITRS. They therefore provide the fundamental link to the global geocentric datum.

(2) The **Australian Regional GPS Network** (ARGN) consists of up 15 permanently operating GPS receivers, 8 located on the continent (Figure 12.1-1) with additional sites on Cocos Island, Macquarie Island, in New Zealand, and in Antarctica. They are operated by the Australian Survey and Land Information Group (AUSLIG). These will allow continuous connection to the ITRS to be maintained, as well as support Australian initiatives such as a real-time GPS Integrity Monitoring Network (for GPS navigation users), special GPS campaigns for crustal motion studies and for the monitoring of the stability (in height!) of the high precision tide gauge network around Australia, as well as for any GPS surveys by state or federal geodetic authorities.

(3) The **Australian Fiducial Network** (AFN) is a subset of the ARGN, consisting of the eight continental stations.

(4) The **Australian National Network** (ANN) is a network of GPS points surveyed by AUSLIG, together with state geodetic authorities, in order to provide an approximately 500km density of "zero order" geodetic stations across Australia (Figure 12.1-2). These points were surveyed during several observation campaigns in 1992, 1993 and 1994. The data was reduced using the GAMIT scientific software in a rigorous multi-baseline / multi-session phase adjustment. To achieve the 0.1ppm relative accuracies the GPS orbits had to be adjusted as well as local tropospheric parameters (see §6.2), and only the IGS core stations were constrained to their apriori ITRS coordinates.

(5) The ANN forms the "backbone" for each state's geodetic network. **Each state was responsible for supplying all geodetic observations (traditional terrestrial measurements such as theodolite directions, distances, etc., as well as GPS baselines) to AUSLIG. The ANN sites were kept fixed and all the geodetic observations then adjusted into the GDA in a secondary network adjustment** (see, for example, KINLYSIDE, 1992). There are about 8,000 stations in the GDA network (of various accuracy orders), and over 70,000 observations. Of these observations over 10,000 are GPS baselines or coordinates (from multi-baseline solutions)! For example, in the states of New South Wales and Victoria there are many baselines making up the primary 100km GPS-only network (observed during the last 8 years -- Figure 12.1-3). Lower order geodetic stations will then be adjusted using different adjustment strategies, including the application of simple block shift transformations to local sets of tertiary coordinates.

The Geocentric Datum of Australia is a bold step in the evolution of the geodetic datum from one that is locally defined to one truly global in context, and from one adequate for survey technologies at the 10ppm to one able to support GPS surveys at accuracies better than 1ppm. The GDA is both a timely response to the challenge of accommodating GPS surveying technology, and is absolutely reliant on it for it's definition and high internal accuracy. The reader is referred to MORGAN et al (1996) for a full account of the procedures used to establish the GDA.

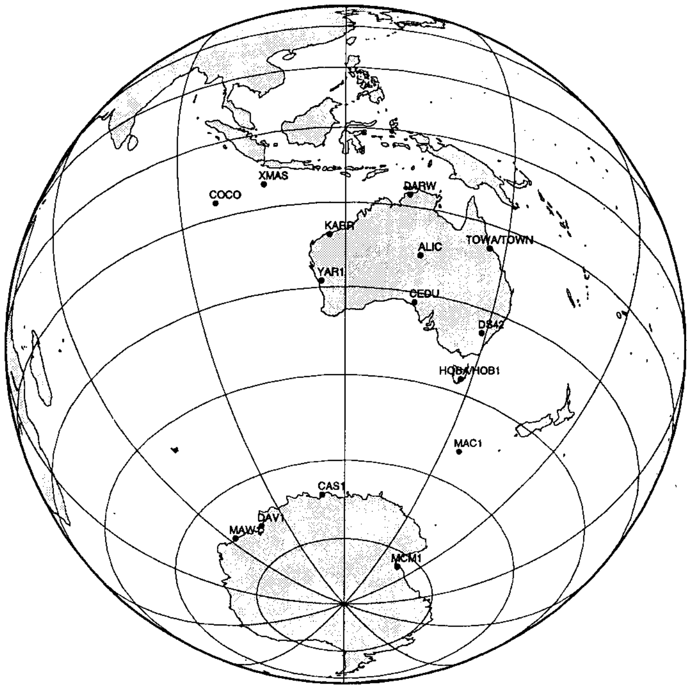


Figure 12.1-1. The Australian Regional GPS Network.

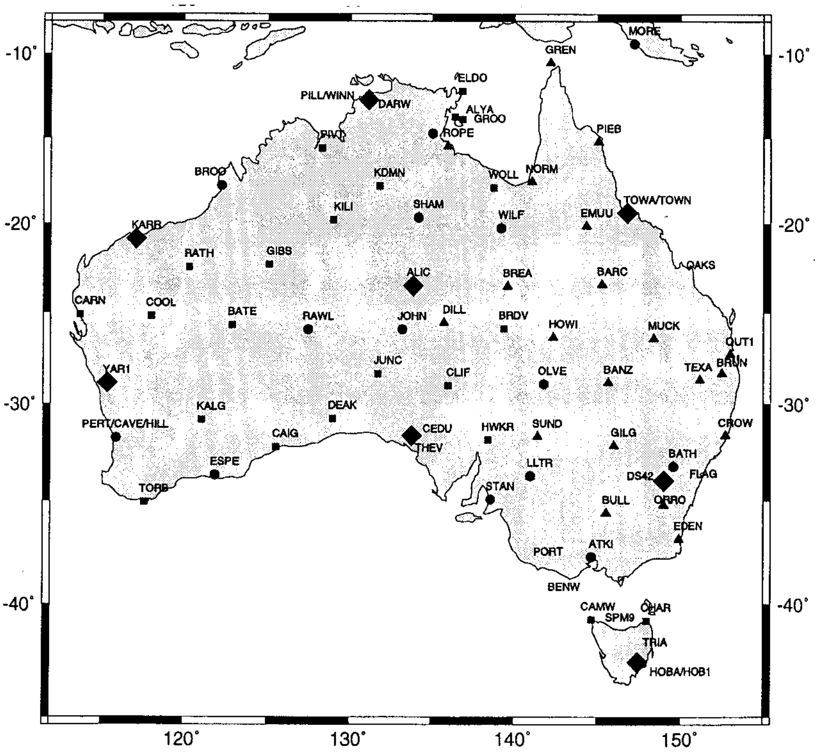


Figure 12.1-2. The Australian National Network.

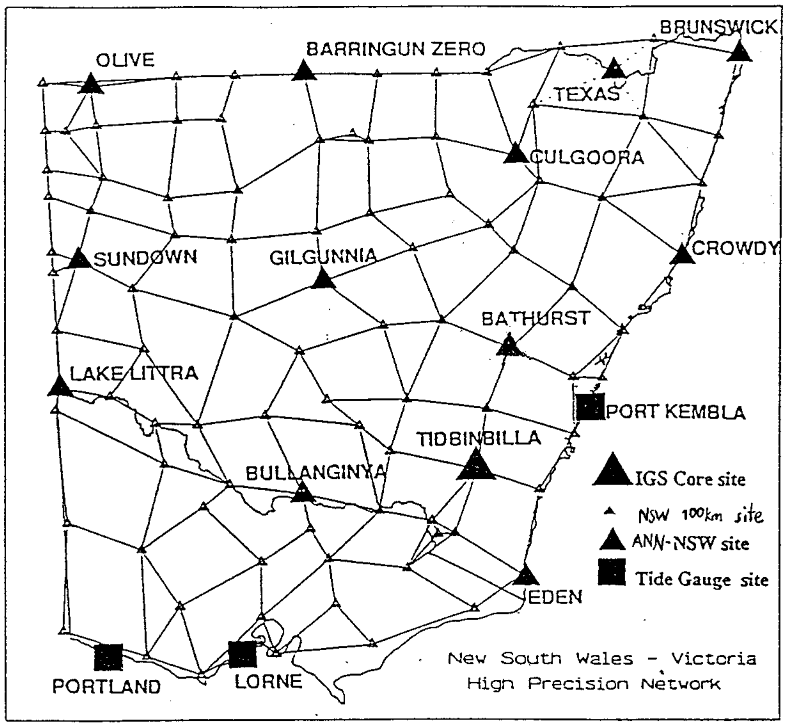


Figure 12.1-3. The NSW and Victorian GPS State Networks.

12.2
THE GLOBAL GPS INFRASTRUCTURE

12.2.1 THE INTERNATIONAL GPS SERVICE FOR GEODYNAMICS

The history and development of the International GPS Service for Geodynamics (IGS) demonstrates the unique capability of international geodetic groups and agencies to work successfully together for a common goal.

The primary motivation for the establishment of the IGS was the recognition in 1989 that the most demanding GPS users, the community of scientists involved in the measurement of crustal motion, were experiencing difficulties interpreting the results of GPS field campaigns because there were: (a) no global reference system, and (b) no GPS satellite orbit ephemerides *accurate enough*. The relative positioning accuracy requirement for such users is of the order of 0.01 parts per million, about 100 times the accuracy of standard GPS surveying (§2.4). As discussed in §6.2, in order to satisfy high accuracy users, several biases have to be accounted for. Two of these biases are the "satellite ephemeris bias" and the "reference station bias". The IGS was therefore established to address these two biases.

A description of the IGS, its functions and products, the background history and present status, can be found in a series of articles in ZUMBERGE et al (1995). Much of the following information is taken from this source. Another useful information source is the IGS Web page (see §3.4).

Background

At the International Association of Geodesy (IAG) General Meeting in 1989, the idea of a global cooperative permanent GPS tracking network, and associated data processing service, to support the geodynamic community was proposed. On February 1, 1991, a "Call for Participation" was issued and more than 100 scientific organisations (including universities) and governmental survey agencies announced their willingness to participate either as an "observatory" (part of the IGS tracking network), or as an analysis/data centre. At the 20th General Assembly of the IUGG in Vienna, in August 1991, several workshops and test campaigns were organised.

The "1992 IGS Test Campaign", June 21 to September 22, 1992, focussed on the routine determination of high accuracy orbits. "Epoch'92" was a two week concentrated campaign in the middle of the three month IGS Campaign for the purpose of densifying the network of tracking stations. Following the success of this test campaign, the IGS Pilot Service was launched in order to bridge the gap between the 1992 IGS Test Campaign and the start of the official service. *The official IGS service started on January 1, 1994.*

Why Global GPS Networks?

A globally distributed network of GPS tracking receivers provides a comprehensive and robust source of tracking data which may be processed to yield: (a) precise GPS satellite orbits, and (b) a set of accurate station coordinates that function as a realisation of a global datum.

The first operational GPS tracking network was established by the U.S. Department of Defense in 1980. By 1985, the network comprised ten stations (Figure 12.2-1), a combination of U.S. Air Force and Defense Mapping Agency stations which produced data for precise post-processed ephemerides.

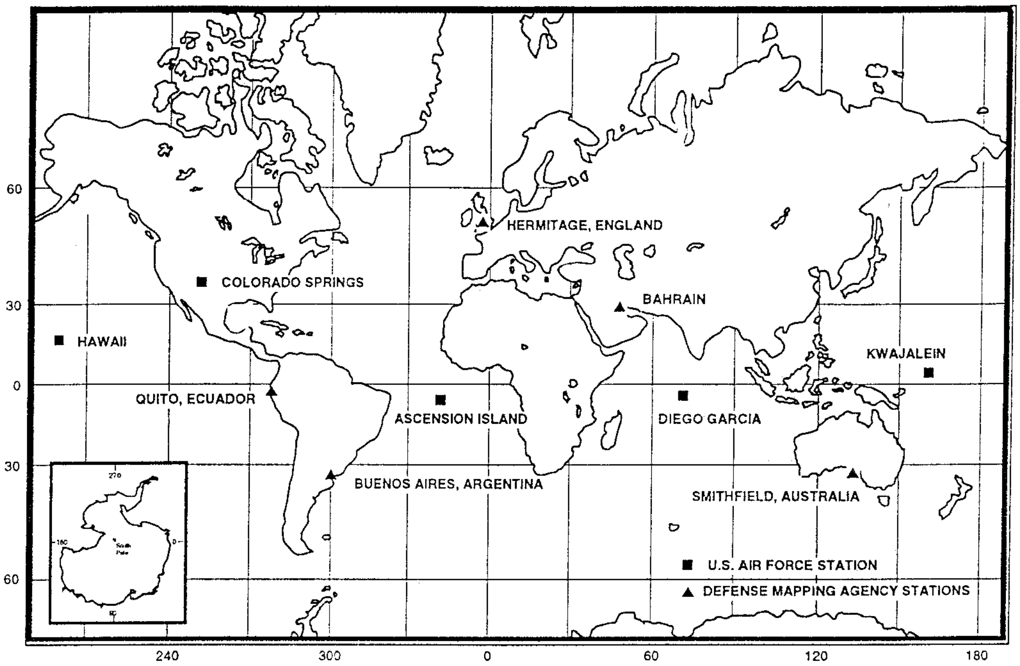


Figure 12.2-1. U.S. Air Force and Defense Mapping Agency GPS tracking network.

While the ephemerides produced by the U.S. DoD ground network were used by certain groups in the early 1980's, certain groups of civilian users began to look at the GPS technology as a cheaper, more mobile system than both the Satellite Laser Ranging (SLR) and Very Long Baseline Interferometry (VLBI) techniques, to monitor crustal motion and other ultra precise applications. Through the 1980's there was an increasing demand for GPS receivers *and experimental support.* Regional field campaigns began to mushroom, and more and more international groups started to collaborate. By the late 1980's it was realised that a global distribution of tracking stations is essential for such applications.

The conclusion was that a continuous, standardised, precise tracking network was a preferable option, compared to the costly exercise of deploying receivers to remote locations for a matter of just a few weeks -- the average length of campaign experiments. The Cooperative International GPS Network (CIGNET) was an important early activity promoted by the U.S. National Geodetic Survey. The 1989 network shown in Figure 12.2-2 was soon augmented by other international partners, to form the core of the initial IGS Network.

1991 was a crucial year for the development of the global tracking network with the announcement of the "Call for Participation" in the "International GPS Service for Geodynamics". The IGS successfully demonstrated the service during the three month "1992 IGS Test Campaign", June 21 to September 22, 1992, with data from tracking stations being accessed by the seven Analysis Centres within three days of data capture. Precise orbits were made available electronically on the Internet to users within two to three weeks. Figure 12.2-3 shows the tracking network configuration during the IGS demonstration campaign in 1992.

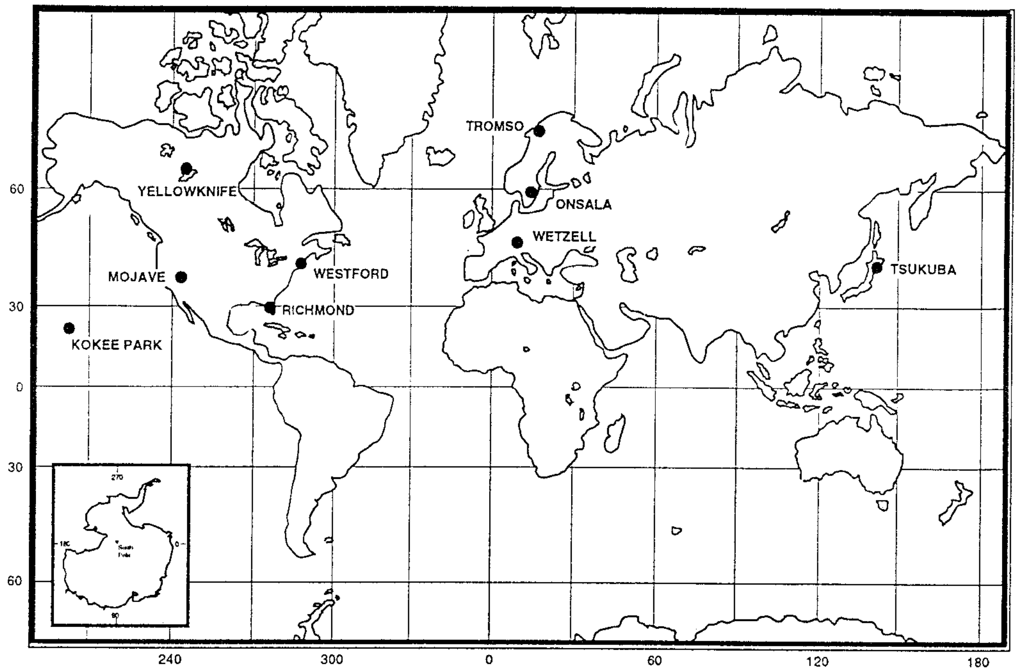


Figure 12.2-2. Cooperative International GPS Network (CIGNET) in 1989.

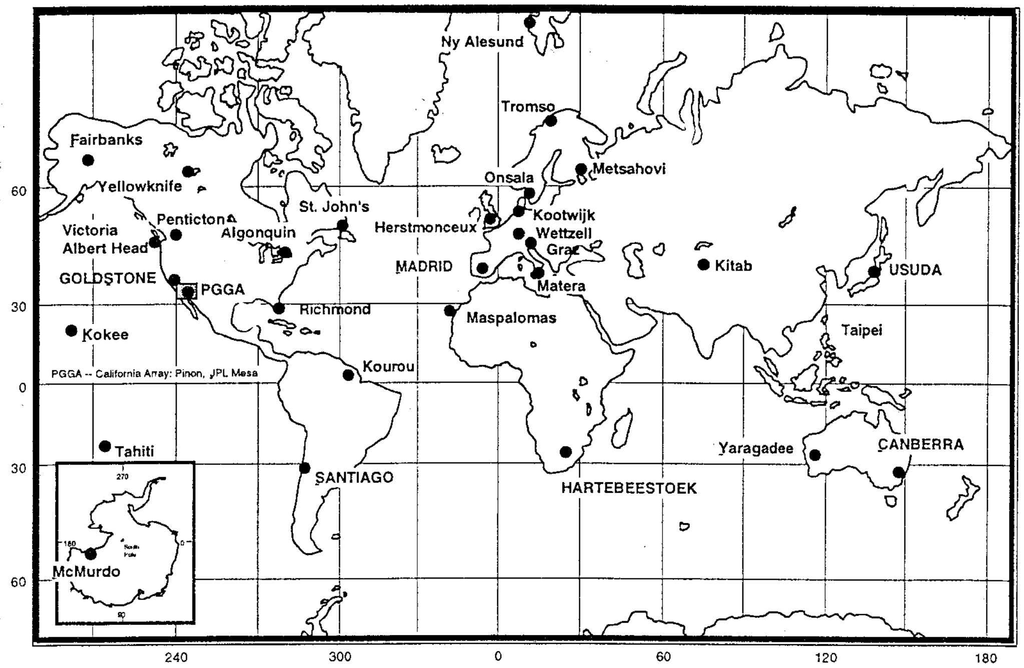


Figure 12.2-3. Tracking network configuration during the IGS demonstration campaign, June 21 to September 22, 1992.

The IGS Organisation

The IGS is a member of the Federation of Astronomical and Geophysical Data Analysis Services (FAGS), and operates in close cooperation with the International Earth Rotation Services (IERS). The primary objective of the IGS is to provide a service to support, through GPS data products, geodetic and geophysical research activities. The secondary objective is to support a range of activities performed by users, such as those developing and maintaining national geodetic datums. Hence, the IGS also develops the necessary standards and specifications (not too dissimilar to those developed for standard GPS surveys, and described in §10.2), and encourages international adherence to these standards and its conventions.

The IGS collects, archives, and distributes GPS observation datasets of sufficient accuracy to satisfy the requirements of a wide range of applications. These datasets are used by the IGS to generate the following data products:

(1) High accuracy GPS satellite ephemerides (the "post-processed" orbits).
(2) Earth rotation parameters (a by-product of the orbit determination process).
(3) Coordinates and velocities of the IGS tracking stations.
(4) GPS satellite and tracking station clock information.
(5) Ionospheric information.

The accuracies of these products are sufficient to support current scientific objectives including:

- Realisation and improvement of the International Terrestrial Reference System (ITRS).
- Monitoring deformations of the solid earth crust (mainly the horizontal motion attributable to plate tectonics, as well as local and regional fault motion).
- Monitoring earth rotation.
- Monitoring variations in the liquid earth, particularly through monitoring the vertical component of the tracking stations (sea level, ice sheets, etc.).
- Scientific satellite orbit determination.
- Ionosphere and troposphere monitoring.

The IGS accomplishes its mission through the following *organisation* :

- Networks of tracking stations.
- Data Centres.
- Analysis Centres.
- Analysis Coordinator.
- The Central Bureau.
- The Governing Board.

The GPS networks consists of about 30-40 "core" tracking stations, and a large number (150-200) "fiducial" tracking stations. The core stations provide continuous GPS tracking for the primary purposes of computing satellite ephemerides, monitoring the ITRS and determining the earth rotation parameters. The fiducial stations may be occupied continuously or intermittently (and then repeatedly at regular intervals) for the purpose of extending the ITRS to all parts of the globe. *Three stations of the Australian Regional Network are part of the core GPS network, while the remainder can be considered to belong to the fiducial GPS network.*

The data centres fall into three categories: operational, regional and global. The operational data centres are in direct contact with the tracking stations, downloading the data on a timely basis (generally daily), reformatting the data files to the RINEX format (§7.3), maintenance of a local data archive, and the electronic transmission of the data files to a regional or global data centre. The regional data centres collect tracking data from several operational data centres, maintain a local data archive, and transmit these data to the global data centres. *The Australian Survey and Land Information Group (AUSLIG), in Canberra, is one of the regional data centres.* The global data centres are the main interfaces to the analysis centres and to the outside user community, maintaining archives of the raw tracking data as well as the IGS data products. The three global data centres are:

- Crustal Dynamics Data Information System, NASA Goddard Space Flight Center, Greenbelt, USA.
- Institut Geographique National (IGN), Paris, France.
- Scripps Institution of Oceanography, University of California, San Diego, USA.

The analysis centres receive and process tracking data to produce the IGS products. The analysis centres are committed to generate products, without interruption, and to deliver them to the global data centres, to the IERS, and to other designated bodies. The main products are the satellite ephemerides (on a weekly basis), and station coordinates (on a quarterly basis). The seven analysis centres are:

- CODE Astronomical Institut-University of Bern, Switzerland.
- European Space Operations Center, European Space Agency, Germany.
- FLINN Analysis Center, Jet Propulsion Laboratory, USA.
- GeoForschungsZentrum (GFZ), Potsdam, Germany.
- Geosciences Research Lab, National Oceanic and Atmospheric Administration (NOAA), USA.
- Natural Resources Canada.
- Scripps Institution of Oceanography, University of California, San Diego, USA.

The Analysis Coordinator monitors the analysis centres' activities to ensure that the IGS objectives are carried out. Specific tasks include quality control and performance evaluation through the inter-comparison of the products of the different analysis centres. The Analysis Coordinator is also responsible for the combination of the analysis centre products into a single IGS product. The current IGS Analysis Coordinator is Jan Kouba, Natural Resources Canada.

The Central Bureau is responsible for the general coordination and management of the IGS. The primary functions of the Central Bureau are: (a) to facilitate communications (through the maintenance and operation of the Central Bureau Information System -- see §3.4); (b) coordinate day-to-day IGS activities; (c) coordinate the establishment of IGS standards and promote compliance with the standards; (d) monitor quality assurance of the data and products; and (e) maintain documentation, organise reports, meetings, and workshops. The IGS Central Bureau is located at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

The IGS Governing Board is the international body which exercises oversight and control over the activities of the service. There are 15 members of the Governing Board.

Current IGS Network

The configuration of the IGS network at the end of 1994 is indicated in Figure 12.2-4 (ZUMBERGE et al, 1995). A comparison of Figures 12.2-3 and 12.2-4 indicates a rapid growth of the network over the past years, with the network nearly doubling in size each year! The specifications for the site, the receiver and the data communications procedures are carefully defined by the IGS. By the end of 1995, more than 100 permanently operating stations were contributing data to the IGS. Of this, perhaps between 30 and 40 strategically located stations can be considered "core" sites, being processed on a regular basis by three or more Analysis Centres. Many of these sites have been established by a small number of agencies. However, quite a number of the sites (particularly the new ones) have been established by governments to serve multiple functions, such as commercial base stations, fundamental datum stations, Integrity Monitoring stations, or stations generating "differential" corrections for real-time DGPS navigation users.

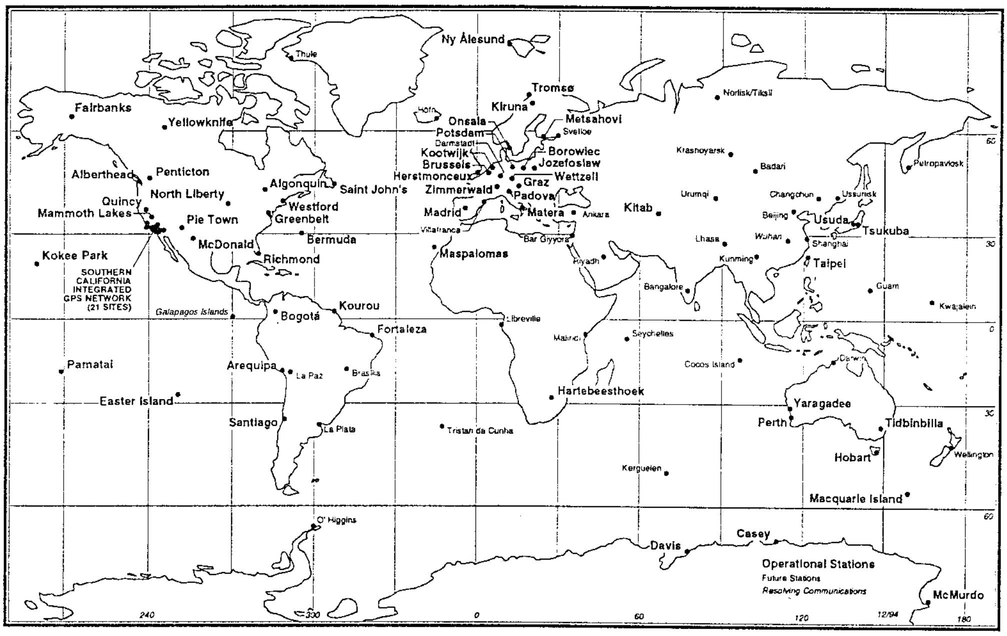


Figure 12.2-4. IGS tracking network at the close of 1994.

The future growth of the IGS global network is aimed at "filling in the gaps" in coverage. These additional sites will contribute nothing to improved orbit products (the present network is more than adequate for this purpose), but represents a densification of the ITRS. The IGS is developing the logistical ability and the techniques to include many hundreds of well distributed GPS stations for the purpose of determining station coordinates and velocities as part of the ITRS. This densification will ensure that most GPS users will be within about 1000km of a precise reference point on which to precisely link their local or regional survey (or datum).

12.3
CURRENT TRENDS IN GPS

Predicting the future is always a risky enterprise! GPS is undergoing tremendous growth, with many product innovations being made. Furthermore, technology advances are leading resulting in organisational and institutional changes which will have a profound impact on the "positioning professions", as well as on the community in general. *Perhaps the most obvious trend therefore is the ever expanding range of GPS applications.* In this section, summary remarks will be made with regards to:

- The GPS **technology** -- satellite developments, user hardware and software.
- The **institutional issues** relating to GPS system control, national and international infrastructure, datum definition and maintenance.
- The expanding range of GPS **applications** and the impact on society.

12.3.1 TRENDS IN GPS TECHNOLOGY

GPS Satellites

- The U.S. government has given an assurance that GPS will be freely available until 2004 --> *what happens then?*
- In 1996 the first of the Block IIR satellites will be launched --> *these will have inter-satellite ranging and onboard ephemeris processing capabilities.*
- The design of the Block IIF satellites is still being finalised, and they will be launched from 2001 onwards --> *transmission of a possible 3rd (civilian) frequency.*
- Increased opportunities of the transmission of GPS-like signals from other (mainly communications) satellites --> *increased availability and reliability for navigation applications.*
- Integration of GPS system into more complex Wide Area Augmentation System (WAAS) --> *particularly for air navigation applications.*

GPS Instrumentation

- Reduction in size, power consumption and cost of receivers --> *especially for navigation.*
- All digital processing, increased reliability, faster sampling rates (<10Hz), lower noise observations, multipath-resistance.
- "Middle tier" of GPS receivers based on low-cost navigation instruments, to which have been grafted carrier phase capabilities --> *targeted at GIS users.*
- Increased capabilities of survey receivers --> *dual-frequency measurement, increased internal memory, greater automation of operation, more optimised for kinematic positioning.*

- Automatic cycle slip detection and repair.
- High precision (P-code) pseudo-range without knowledge of code (under AS).
- High precision C/A code pseudo-range using "narrow-correlator" technology.
- Universal data output in standard "RINEX" format for processing in 3rd party software.
- Greater computing power within receiver.
- Real-time operation will be the norm for all applications (surveying & navigation).
- Combined GLONASS-GPS instrumentation is now available, and more may be developed.
- Greater integration of GPS technology into complex survey and navigation systems.
- Growth in products based on OEM receivers.
- Better antennas, overall decrease in susceptibility to multipath reflections.

GPS Software

- Improved user-friendliness, more automatic processing capabilities --> *windows™ based.*
- "On-the-Fly" ambiguity resolution is refined --> *possibly single-epoch carrier phase positioning.*
- Increasing variety of processing strategies provided for in the software package(s).
- Explosion of software for kinematic surveying / precise navigation / GIS.
- Increased "targeting" of software for special applications.
- Increased demand for quality assurance measures --> *particularly for navigation applications, but also "rapid GPS surveying", etc.*
- Adoption of standard "receiver independent output format" for baseline results --> *for example the SINEX format.*
- Increasing proportion of software that is not instrument-specific.
- Increasing use of precise IGS orbit products.
- Increased range of and improved secondary software --> *for example, network adjustments, GIS input, etc.*

12.3.2 TRENDS IN GPS INSTITUTIONAL ISSUES

GPS Control

- Increasing "tug-a-war" between civilian and military users --> *will the U.S. military abandon GPS in favour of a new military-only system?*
- Abandon "Selective Availability" possibly by the end of the century --> *reduce the need for GPS base stations for transmitting DGPS corrections.*
- Bypass "Anti-Spoofing" through transmission of additional (civilian) frequency on Block

IIF GPS satellites --> *make "on-the-fly" ambiguity resolution more certain.*

- Integration of GLONASS satellites to improve availability and reliability for critical navigation users --> *not likely to be of much use to survey users.*
- Increased concern about GPS control, and plans developed for civilian-only Global Navigation Satellite System (GNSS) --> *will it be good enough for positioning at survey accuracies?*

GPS Datum Issues

- More countries will redefine their national geodetic datums to be coincident (at some level of accuracy) with ITRS --> *a geocentric datum.*
- Convergence of definition of WGS84 with ITRS --> *already happening.*
- GPS is unchallenged as the geodetic tool for establishing, maintaining or renovating national control networks.
- In addition to the traditional levels of national control network (first order/class, second order/class, etc.), there will additional high accuracy control points established using "GPS geodesy" techniques --> *these will be the "backbones" of new datums.*
- With the increased efficiency of GPS for establishing control point coordinates, there will be a trend to reduce the number of physical control marks --> *easier to establish marks when and where they are needed.*
- Where permanent control marks are established, they will be increasingly located where they are most needed not on inaccessible hilltops --> *intervisibility of stations not necessary for GPS, but some is necessary to define azimuth to support conventional ground survey techniques.*
- A range of GPS hardware can be used, and the RINEX data reduced using a preferred software package --> *including the "GPS geodesy" software.*
- Real-time operation of GPS base stations will permit real-time GPS control surveys.

GPS Infrastructure Issues

- Increased blurring of GPS infrastructure --> *for example, permanent GPS control marks will also support real-time DGPS, Integrity Monitoring, etc.*
- Increased acceptance of the IGS service --> *precise orbit information and access to the International Terrestrial Reference System*
- Increased network of permanently operating GPS receivers tied to ITRS --> *continuous "active" control station connection to ITRS.*
- Issues such as GPS system testing, accreditation of GPS surveyors, procedures for GPS cadastral surveys, legal traceability of GPS results, etc., will become increasingly important.
- GPS positioning will not only be the prerogative of the "positioning professions" such as Navigation and Surveying.
- Increasing "internationalisation" of GPS --> *leading to more integration of GPS datums such as has occurred in the case of Australia and New Zealand.*

12.3.3 TRENDS IN GPS APPLICATIONS

General Comments

- A blurring of the distinctions between GPS navigation, surveying and geodesy.
- The number of civilian users will continue to grow, compared to the number of military users --> *the ratio is 9:1 and will increase.*
- GPS will facilitate the "position information society" --> *where location will be a common and useful piece of information during everyday activities.*
- Quality Assurance will be an increasingly important issue.

GPS Navigation

- Expanding applications as the cost of hardware drops.
- GPS will be the primary navigation tool for air, land and sea travel.
- Largest market will be for land vehicles --> *many applications, from autonomous navigation to "fleet management".*
- GPS will likely become the most insignificant component of larger and more complex system --> *for example, "fleet monitoring", "Intelligent Transportation System", "Future Air Navigation System", etc.*
- The elimination of Selective Availability will ensure absolute positioning accuracy at the 10-20 metre level.
- DGPS use will expand, however, requiring the establishment of transmitting base stations, both terrestrial and satellite-based.
- GPS and Map Display technology will converge into new navigation products --> *for example, ITS, Electronic Chart Systems, etc.*
- Satellite communications will play a large role in GPS and DGPS applications --> *particularly for sea and air applications.*
- Wide-Area DGPS will be refined, and encroach on Local-Area DGPS.
- Increasing use of low-cost handheld receivers used for geo-coding (data capture) in support of GIS applications.
- Increasing integration of GPS with other navigation sensors --> *for example, GPS and Dead Reckoning systems.*

GPS Surveying

- GPS will be a commonly used technique for all surveys, especially for distances greater than about 5km.
- GPS surveying applications will expand into the engineering and cadastral areas --> *changes in Surveyors' law may be required.*
- Increased use of "non-conventional" positioning techniques --> *for example, new methodologies using GPS survey receivers, but also increased use of low-cost handheld receivers for submetre accuracy applications.*

- Base station operation will be increasingly supported --> *post-processing as well as real-time mode.*
- GPS deformation surveys --> *engineering structures, earthquake and volcanic zones, etc.*
- Mixing of GPS survey receivers will no longer be unusual.
- More and more applications will be addressed in real-time.
- Precision airborne GPS will be used extensively for remote sensing, airborne geophysics and photogrammetry.
- Refinement of one-epoch carrier phase positioning --> *no longer a distinction between static and kinematic surveying.*

GPS Levelling

- A viable tool for 3rd or 4th order levelling.
- A viable tool for checking conventional levelling networks.
- Increased range of geoid computation packages and/or geoid map products will be available.
- Research effort directed towards attaining 0.01ppm height accuracies in GPS geodesy --> *for monitoring sea level rise, subsidence of land, etc.*

GPS Geodesy

- Accuracies at the few 0.01ppm level will become routine.
- GPS geodesy will play the dominant role in geodynamic studies.
- GPS will be used for other non-positioning applications such as ionosphere monitoring and determination of tropospheric conditions --> *GPS "meteorology"!*
- GPS geodesy will be performed using permanent GPS networks, as well as the traditional "campaign-style" survey.
- GPS geodesy will play the dominant role in geodynamic studies.
- The IGS will support all precise GPS surveys.
- Improvements in instrumentation, reference system definition (ITRS), global tracking (IGS), observation modelling and network design.
- All of these will require greater partnership between academics, researchers, government survey organisations, instrument manufacturers --> *GPS geodesy will be less the prerogative of a select band of experts, and the government agencies will take on more of the role.*
- More frequent GPS campaigns for geodynamic studies, at a lower cost and on a local area basis --> *with IGS there is no need for orbit computation!*
- Software refinements to make data reduction easier --> *data processing is the biggest bottleneck at present.*
- Permanent arrays or networks of precision GPS receivers operating 24hr/day --> *on global, continental, national and local scales for different applications.*

REFERENCES

ABIDIN, H.Z., 1993. Computational and geometrical aspects of on-the-fly ambiguity resolution. Dept. of Geodesy & Geomatic Eng., University of New Brunswick, Canada, *Tech. rept. no.164*, 290pp.

ABIDIN, H.Z., 1994. On-the-fly ambiguity resolution. ***GPS World***, 5(4), 40-50.

ALLMAN, J.S. & VEENSTRA, C., 1984. Geodetic Model of Australia. *Div. of National Mapping Tech. rept.no.33*, Canberra.

ASHJAEE, J., 1993. An analysis of Y-code tracking techniques and associated technologies. ***Geodetical Info Magazine***, 7(7), 26-30.

BESER, J. & PARKINSON, B.W., 1982. The application of NAVSTAR differential GPS in the civilian community. ***Navigation***, 29(2), 107-136.

BEUTLER, G., BAUERSIMA, I., GURTNER, W. & ROTHACHER, M., 1987. Correlations between simultaneous GPS double difference carrier phase observations in the multistation mode: implementation considerations and first experiences. ***Manuscripta Geodaetica***, 12, 40-44.

BEUTLER, G., BAUERSIMA, I., GURTNER, W., ROTHACHER, M., SCHILDKNECHT, T. & GEIGER, A., 1989. Atmospheric refraction and other important biases in GPS carrier phase observations. In monograph 12, *"Atmospheric Effects on Geodetic Space Measurements"*, F.K.Brunner (Ed.), School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 15-44.

BOCK, Y., GOUREVITCH, S.A., COUNSELMAN, C.C., KING, R.W. & ABBOT, R.I., 1986. Interferometric analysis of GPS phase observations. ***Manuscripta Geodaetica***, 11, 282-288.

BOEY, S. & HILL, C., 1995. Can GPS measurements be legally used for cadastral surveying? ***Aust. Surveyor***, 40(2), 101-111.

BOUCHER, C. & ALTAMIMI, Z., 1996. International Terrestrial Reference Frame. ***GPS World***, 7(9), 71-74.

BRUNNER, F.K. & WELSCH, W.M., 1993. Effect of the troposphere on GPS measurements. ***GPS World***, 4(1), 42-51.

CASPARY, W.F., 1987. **Concepts of Network and Deformation Analysis**. Monograph 11, School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 183pp.

CHEN, D., 1993. Fast Ambiguity Search Filter (FASF): a novel concept for GPS ambiguity resolution. Proc. 6th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, 22-24 September, 1993, 781-787.

CLARKE, B., 1994. **Aviator's Guide to GPS**. McGraw-Hill, Inc., 235pp.

CRAYMER, M.R. & BECK, N., 1992. Session versus single-baseline GPS processing. Proc. 5th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, 16-18 September, 1992, 995-1004.

CROSS, P.A., 1983. **Advanced Least Squares Applied to Position Fixing**. Working Papers, North East London Polytechnic, Dept. of Surveying, 205pp.

D.M.A., 1991. Department of Defense World Geodetic System 1984. *Technical Report 8350.2*, 2nd ed. U.S. Defense Mapping Agency.

DIXON, K., 1995. Global reference frames with time. ***Surveying World***, 3(6), 22-24.

EL-RABBANY, A.E-S., 1994. The effect of physical correlations on the ambiguity resolution and accuracy estimation in GPS differential positioning. Dept. of Geodesy & Geomatic Eng., University of New Brunswick, Canada, *Tech. rept. no.170*, 161pp.

ERICKSON, C., 1992. An analysis of ambiguity resolution techniques for rapid static GPS surveys using single frequency data. Proc. 5th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, 16-18 September, 1992, 453-462.

F.G.C.C., 1988. Geometric geodetic accuracy standards and specifications for using GPS relative positioning techniques (version 5). *Pub. of the U.S. Federal Geodetic Control Committee*, 50pp.

FREI, E. & BEUTLER, G., 1990. Rapid static positioning based on the Fast Ambiguity Resolution Approach "FARA": theory and first results. ***Manuscripta Geodaetica***, 15, 325-356.

GEORGIADOU, Y. & DOUCET, K.D., 1990. The issue of Selective Availability. ***GPS World***, 1(5), 53-56.

GRANT, D.B., RIZOS, C. & STOLZ, A., 1990. Dealing with GPS biases: some theoretical and software considerations. In Unisurv S-38, *"Contributions to GPS Studies"*, C. Rizos (Ed.), School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 204pp.

GURTNER, W., 1994. RINEX - the receiver-independent exchange format. ***GPS World***, 5(7), 48-52.

HAN, S., 1995. Ambiguity resolution techniques using integer least squares estimation for rapid static or kinematic positioning. Proc. 2nd Australian "Satellite Navigation Technology Symposium", Brisbane, Australia, 26-28 June, 1995.

HAN, S. & RIZOS, C., 1995a. A suggested procedure for on-the-fly ambiguity resolution for long range GPS kinematic positioning. Proc. 4th Int. Conf. "Differential Satellite Navigation Systems DSNS95", Bergen, Norway, 24-28 April, 1995.

HAN, S. & RIZOS, C., 1995b. Selection and scaling of simultaneous baselines for GPS network adjustment, or correct procedures for processing trivial baselines. ***Geomatics Research Australasia***, 63, December, 1995, 51-66.

HAN, S. & RIZOS, C., 1996. Validation and rejection criteria for integer least-squares estimation. Survey Review, Vol.33, No.260, 375-382.

HARVEY, B.R., 1986. Transformation of 3D coordinates. ***Aust. Surveyor***, 33(2), 105-125.

HARVEY, B.R., 1994. **Practical Least Squares and Statistics for Surveyors.** Monograph 13, School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 2nd ed., 319pp.

HATCH, R.R., 1982. The synergism of GPS code and carrier measurements. Proc. 3rd Int. Symp. on "Satellite Doppler Positioning", New Mexico, 8-12 February, 1982, 1213-1231.

HATCH, R.R., 1990. Instantaneous ambiguity resolution. Proc. "Kinematic Systems in Geodesy, Surveying & Remote Sensing", IAG Symposium 107, Springer-Verlag Vienna New York, 299-308.

HEISKANEN, W.A. & MORITZ, H., 1967. **Physical Geodesy**. W.H. Freeman & Co., San Francisco, 364pp.

HELLWIG, H., 1979. Microwave time and frequency standards. ***Radio Science***, 14(4), 561-572.

HIGGINS, M., 1987. Transformation from WGS84 to AGD84, an interim solution. *Dept. of Geographic Information memo*, Queensland.

HOFMANN-WELLENHOF, B., LICHTENEGGER, H. & COLLINS, J., 1994. **GPS Theory and Practice**. Springer-Verlag, Vienna New York, 3rd ed., 355pp.

HOLLOWAY, R.D., 1988. **The Integration of GPS Heights into the Australian Height Datum**. UNISURV S-33, School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 151pp.

HUSTI, G.J., SLUITER, P.G. & ZOMERDIJK, J., 1994. A comparison of geodetic receivers under A/S conditions. Proc. 7th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, 20-23 September, 1994.

I.C.S.M., 1994. **Standards and Practices for Control Surveys**. *Australian Inter-Governmental Committee on Surveying and Mapping, vers.1.2, pub.no.1*.

IVANOV, N.E. & SALISTCHEV, V., 1991. GLONASS and GPS: prospects for a partnership. ***GPS World***, 2(4), 36-40.

JONES, A., 1995. Personal communication.

JONES, R.H. & TRYON, P.V., 1987. Continuous time series models for unequally spaced data applied to modeling atomic clocks. ***SIAM J. Sci. Stat. Comput.***, 8(1), 71-81.

KAPLAN, E. (Ed.), 1996. **Understanding GPS: Principles & Applications**. Artech House Publishers, Boston London, 554pp.
KEARSLEY, A.H.W., 1988. The determination of the geoid-ellipsoid separation for GPS levelling. ***Aust. Surveyor***, 34(1), 11-18.
KEARSLEY, A.H.W. & GOVIND, R., 1991. Geoid evaluation in Australia, a status report. ***Aust. Surveyor***, 36(1), 30-40.
KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A. & COLLINS, J., 1987. **Surveying with GPS**. Dümmler, 128pp.
KINLYSIDE, D. A., 1992. Densification surveys in New South Wales - coping with distortions. Unisurv S-42, School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 209pp.
KLEUSBERG, A., 1990. Comparing GPS and GLONASS. ***GPS World***, 1(6), 52-54.
KLOBUCHAR, J.A., 1991. Ionospheric effects on GPS. ***GPS World***, 2(4), 48-51.
LACHAPELLE, G., 1995. Far reaching recommendations on GPS made by high level U.S. study. ***GIM journal***, August, 1995, 48-49.
LACHAPELLE, G. & HENRIKSEN, J., 1995. GPS under cover: the effect of foliage on vehicular navigation ***GPS World***, 6(3), 26-35.
LANDAU, H. & EULER, H.J., 1992. On-the-fly ambiguity resolution for precise differential positioning. Proc. 5th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, 16-18 September, 1992, 607-613.
LANGLEY, R.B., 1991a. The GPS receiver - an introduction. ***GPS World***, 2(1), 50-53.
LANGLEY, R.B., 1991b. The orbits of GPS satellites. ***GPS World***, 2(3), 50-53.
LANGLEY, R.B., 1991c. The mathematics of GPS. ***GPS World***, 2(7), 45-50.
LANGLEY, R.B., 1991d. Time, clocks, and GPS. ***GPS World***, 2(10), 38-42.
LANGLEY, R.B., 1993. The GPS observables. ***GPS World***, 4(4), 52-59.
LINDLOHR, W. & WELLS, D.E., 1985. GPS design using undifferenced carrier beat phase observations. ***Manuscripta Geodaetica***, 10, 255-295.
MANNING, J. & HARVEY, B., 1994. Status of the Australian geocentric datum. ***Aust. Surveyor***, 39(1), 28-33.
MATHER, R.S., 1978. **The Theory and Geodetic Use of Some Common Projections.** Monograph 1, School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 2nd ed., 125pp.
MERMINOD, B. & RIZOS, C., 1988. A surveyor's perspective to Kalman filtering. ***Aust. J. Geod. Photo. Surv.***, 49, 1-38.
MERMINOD, B., GRANT, D.B. & RIZOS, C., 1990. Planning GPS surveys - using appropriate precision indicators. ***Can. Inst. Surv. & Mapping J.***, 44(3), 113-129.
MITCHELL, H.L., 1990. GPS heighting and the A.H.D. Rept. by the GPS Heighting Study Group, *Australian GPS Users Group*, 77pp.
MORGAN, P., BOCK, Y., COLEMAN, R., FENG, P., GARRARD, D., JOHNSTON, G., LUTON, G., McDOWALL, B., PEARSE, M., RIZOS, C., & TIESLER, R., 1996. A zero order GPS network for the Australian region. Unisurv S-46, School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 187pp.
NATO, 1991. NAVSTAR GPS user equipment. *Public report released by US Air Force Space Systems Division*, NAVSTAR-GPS Joint Program Office.
NAVSTAR, 1993. Global Positioning System Standard Positioning Service signal specification. *Dept. of Defense rept.*, November, 1993, 46pp + appendices.
N.M.C., 1986. The Australian Geodetic Datum Technical Manual. *National Mapping Council Special Pub.no.10*, 65pp+Appendix.
PARKINSON, B.W., 1983. NAVSTAR: Global Positioning System -- ten years later. ***IEEE vol.71(10)***, 1177-1186.
PARKINSON, B.W., 1994. GPS eyewitness: the early years. ***GPS World***, 5(9), 32-45.
PARKINSON, B.W., STANSELL, T., BEARD, R. & GROMOV, K., 1995. A history of satellite navigation. ***Navigation***, 42(1), 109-164.
RAPP, R.H. & PAVLIS, N.K., 1990. The development and analysis of geopotential coefficient models to spherical harmonic degree 360. ***J. Geophys. Res.***, 95(B13), 21885-21911.

RIZOS, C. & GRANT, D.B., 1990. Time and the Global Positioning System. In Unisurv S-38, *"Contributions to GPS Studies"*, C. Rizos (Ed.), School of Geomatic Engineering (formerly Surveying), The University of New South Wales, 204pp.

ROELSE, A., GRANGER, H.W. & GRAHAM, J.W., 1971. The adjustment of the Australian Levelling Survey -- 1970-1971. *Div. of National Mapping, Tech. rept. no.12*, Canberra.

SANDLIN, A., McDONALD, K. & DONAHUE, A., 1995. Selective Availability: to be or not to be? ***GPS World***, 6(9), 44-51.

SANTERRE, R. & BOULIANNE, M., 1995. New tools for urban GPS surveyors. ***GPS World***, 6(2), 49-54.

SCHWARZ, K.P. & SIDERIS, M.G., 1993. Heights and GPS. ***GPS World***, 4(2), 50-56.

SEEBER, G., 1993. **Satellite Geodesy: Foundations, Methods & Applications**. Walter de Gruyter, Berlin New York, 531pp.

SEPPELIN, T.O., 1974. The Department of Defense World Geodetic System 1972. Proc. Int. Symp. "Problems Related to the Redefinition of the North American Geodetic Networks", Fredericton, New Brunswick, Canada.

SLUITER, P.G. & HAAGMANS, M.E.E., 1995. Comparative test between geodetic Y-code GPS receivers (part 2). ***GIM journal***, August, 1995, 35-41.

SNAY, R.A., 1986. Network design strategies applicable to GPS surveys using three or four receivers. ***Bulletin Geodesique***, 60(1), 37-50.

SOLER, T. & HOTHEM, L.D., 1989. Important parameters used in geodetic transformations. ***J.Surv.Engin.***, 115(4), 414-417.

SPILKER Jr., J.J., 1980. GPS signal structure and performance characteristics. In: *Global Positioning System*, papers published in ***Navigation***, reprinted by the U.S. Inst. of Navigation, vol.1, 29-54.

STEED, J., 1990. A practical approach to transformation between commonly used reference systems. ***Aust. Surveyor,*** 35(3), 248-264.

STEED, J. & HOLTZNAGEL, S.,1994. AHD heights from GPS using AUSGEOID93. ***Aust. Surveyor***, 39(1), 21-27.

TALBOT, N.C., 1987. GPS signal structure and receiver operation. In proc. R.M.I.T. Centenary GPS Conf., 24-26 August, 1987.

TEUNISSEN, P.J.G., 1994. A new method for fast carrier phase ambiguity estimation. Proc. IEEE "Position Location and Navigation Symposium PLANS94", Las Vegas, 11-15 April, 1994, 562-573.

TIBERIUS, C. & DE JONGE, P., 1995. Introduction to GPS surveying (part 2). ***GIM journal***, August, 1995, 71-77.

TORGE, W., 1980. **Geodesy**. Walter de Gruyter, Berlin New York, 254pp.

UNGUENDOLI, M., 1990. A rational approach to the use of a large number of GPS receivers. ***Bulletin Geodesique***, 64(4), 303-312.

VAN DIERENDONCK, A.J., RUSSELL, S.S., KOPTIZKE, E.R. & BIRNBAUM, M., 1980. The GPS navigation message. In: *Global Positioning System*, papers published in ***Navigation***, reprinted by the U.S. Inst. of Navigation, vol.1, 55-73.

VAN DIERENDONCK, A.J., 1995. Understanding GPS receiver terminology: a tutorial. ***GPS World***, 6(1), 34-44.

WANNINGER, L., 1993. Effects of the equatorial ionosphere on GPS. ***GPS World***, 4(7), 48-54.

WELLS, D.E., BECK, N., DELIKARAOGLOU, D., KLEUSBERG, A., KRAKIWSKY, E.J., LACHAPELKLE, G., LANGLEY, R.B., NAKIBOGLU, M., SCHWARZ, K.P., TRANQUILLA, J.M. & VANICEK, P., 1987. **Guide to GPS Positioning**. Canadian GPS Associates, Fredericton, New Brunswick, Canada, 2nd. ed., 600pp.

ZUMBERGE, J.F., LIU, R. & NEILAN, R.E., 1995. International GPS Service for Geodynamics 1994 Annual Report. September, 1995, IGS Central Bureau, 329pp.

Publications from

THE SCHOOL OF GEOMATIC ENGINEERING

(Formerly School of Surveying)

THE UNIVERSITY OF NEW SOUTH WALES

All prices include postage by surface mail. Air mail rates on application. (Effective January 1997)

To order, write to Publications Officer, School of Geomatic Engineering
The University of New South Wales, Sydney 2052, AUSTRALIA

NOTE: ALL ORDERS MUST BE PREPAID

UNISURV REPORTS - S SERIES

S8 - S20	Price (including postage) :		$10.00
S29 onwards	Price (including postage) :	Individuals	$25.00
		Institutions	$30.00

S8 A. Stolz, "Three-D Cartesian co-ordinates of part of the Australian geodetic network by the use of local astronomic vector systems", Unisurv Rep. S8, 182 pp, 1972.

S10 A.J. Robinson, "Study of zero error & ground swing of the model MRA101 tellurometer", Unisurv Rep. S10, 200 pp, 1973.

S12. G.J.F. Holden, "An evaluation of orthophotography in an integrated mapping system", Unisurv Rep. S12, 232 pp, 1974.

S14. Edward G. Anderson, "The Effect of Topography on Solutions of Stokes` Problem", Unisurv Rep. S14, 252 pp, 1976.

S16. K. Bretreger, "Earth Tide Effects on Geodetic Observations", Unisurv S16, 173 pp, 1978.

S17. C. Rizos, "The role of the gravity field in sea surface topography studies", Unisurv S17, 299 pp, 1980.

S18. B.C. Forster, "Some measures of urban residential quality from LANDSAT multi-spectral data", Unisurv S18, 223 pp, 1981.

S19. Richard Coleman, "A Geodetic Basis for recovering Ocean Dynamic Information from Satellite Altimetry", Unisurv S19,332 pp, 1981.

S20. Douglas R. Larden, "Monitoring the Earth's Rotation by Lunar Laser Ranging", Unisurv Report S20, 280 pp, 1982.

S29 Gary S Chisholm, "Integration of GPS into hydrographic survey operations", Unisurv S29, 190 pp, 1987.

S30. Gary Alan Jeffress, "An investigation of Doppler satellite positioning multi-station software", Unisurv S30, 118 pp, 1987.

S31. Jahja Soetandi, "A model for a cadastral land information system for Indonesia", Unisurv S31, 168 pp, 1988.

S33. R. D. Holloway, "The integration of GPS heights into the Australian Height Datum", Unisurv S33, 151 pp.,1988.

S34. Robin C. Mullin, "Data update in a Land Information Network", Unisurv S34, 168 pp. 1988.

S35. Bertrand Merminod, "The use of Kalman filters in GPS Navigation", Unisurv S35, 203 pp., 1989.

S36. Andrew R. Marshall, "Network design and optimisation in close range Photogrammetry", Unisurv S36, 249 pp., 1989.

S37. Wattana Jaroondhampinij, "A model of Computerised parcel-based Land Information System for the Department of Lands, Thailand," Unisurv S37, 281 pp., 1989.

S38. C. Rizos (Ed.), D.B. Grant, A. Stolz, B. Merminod, C.C. Mazur "Contributions to GPS Studies", Unisurv S38, 204 pp., 1990.

S39. C. Bosloper, "Multipath and GPS short periodic components of the time variation of the differential dispersive delay", Unisurv S39, 214 pp., 1990.

S40. John Michael Nolan, "Development of a Navigational System utilizing the Global Positioning System in a real time, differential mode", Unisurv S40, 163 pp., 1990.

S41. Roderick T. Macleod, "The resolution of Mean Sea Level anomalies along the NSW coastline using the Global Positioning System", 278 pp., 1990.

S42. Douglas A. Kinlyside, "Densification Surveys in New South Wales - coping with distortions", 209 pp., 1992.

S43. A. H. W. Kearsley (ed.), Z. Ahmad, B. R. Harvey and A. Kasenda, "Contributions to Geoid Evaluations and GPS Heighting", 209 pp., 1993.

S44. Paul Tregoning, "GPS Measurements in the Australian and Indonesian Regions (1989-1993)", 134 + xiii pp, 1996.

S45. Wan-Xuan Fu, "A study of GPS and other navigation systems for high precision navigation and attitude determinations", 332pp, 1996.

S46. Peter Morgan et al, "A zero order GPS network for the Australia region", 187 + xii pp, 1996.

MONOGRAPHS

Prices include postage by surface mail

		Price
M1.	R.S. Mather, "The theory and geodetic use of some common projections", (2nd edition), 125 pp., 1978.	$15.00
M2.	R.S. Mather, "The analysis of the earth's gravity field", 172 pp., 1971.	$8.00
M3.	G.G. Bennett, "Tables for prediction of daylight stars", 24 pp.,1974.	$5.00
M4.	G.G. Bennett, J.G. Freislich & M. Maughan, "Star prediction tables for the fixing of position", 200 pp., 1974.	$8.00
M8.	A.H.W. Kearsley, "Geodetic Surveying", 96 pp, (revised) 1988.	$12.00
M11.	W.F. Caspary, "Concepts of Network and Deformation Analysis", 183 pp., 1988.	$25.00
M12.	F.K. Brunner, "Atmospheric Effects on Geodetic Space Measurements", 110 pp., 1988.	$16.00
M13.	Bruce R. Harvey, "Practical Least Squares and Statistics for Surveyors", (2nd edition), 319 pp., 1994.	$30.00
M14.	Ewan G. Masters & John R. Pollard (Ed.), "Land Information Management", 269 pp., 1991. (Proceedings LIM Conference, July 1991).	$20.00
M15/1	Ewan G. Masters & John R. Pollard (Ed.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol 1" 295 pp., 1993 (Proceedings of LIM & GIS Conference, July 1993).	$30.00
M15/2	Ewan G. Masters & John R. Pollard (Ed.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol 2" 376 pp., 1993 (Proceedings of Advanced Remote Sensing Conference, July 1993).	$30.00
M16.	A. Stolz, "An Introduction to Geodesy", 112 pp., 1994.	$20.00
M17	Chris Rizos, "Principles and Practice of GPS Surveying", 565 pp., 1997.	$50.00