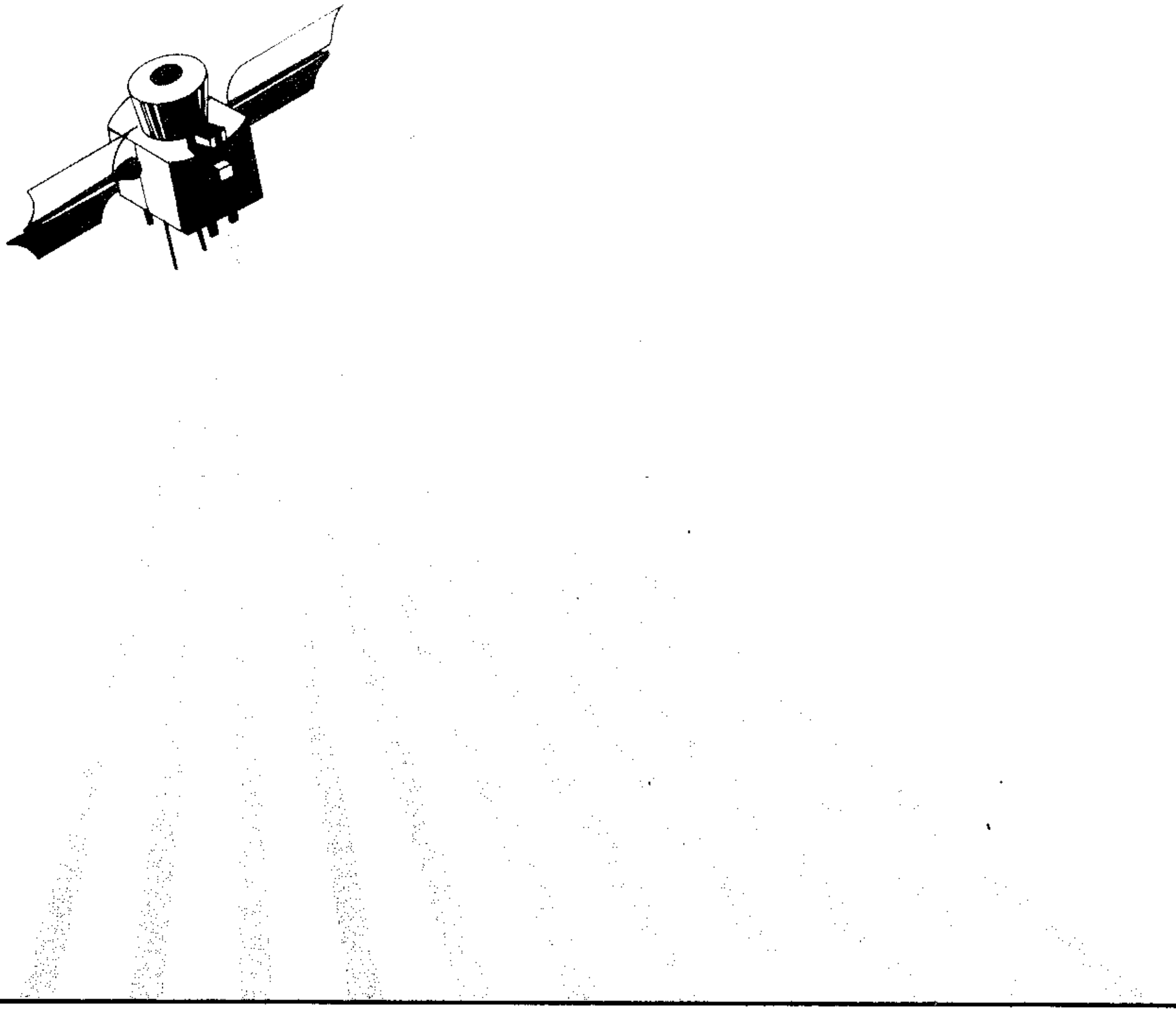


SURVEYING WITH GPS

R.W. King E.G. Masters C. Rizos A. Stolz J. Collins

SURVEYING WITH GPS

R.W. KING, E.G. MASTERS, C. RIZOS,
A. STOLZ AND J. COLLINS

MONOGRAPH 9
SCHOOL OF SURVEYING
THE UNIVERSITY OF NEW SOUTH WALES
KENSINGTON, N.S.W. AUSTRALIA

Copyright 1985 by
School of Surveying
The University of New South Wales
P.O. Box 1, Kensington,
N.S.W., 2033, Australia.

First published; November, 1985.

National Library of Australia
Card Number and ISBN
0-85839-042-6

FOREWORD

This monograph grew out of a set of lecture notes which we prepared for a workshop on the Global Positioning System (GPS) held at the University of New South Wales. The interest with which this workshop was received convinced us that the surveying community would benefit from, and indeed was in need of, a reference book that thoroughly described the principles of GPS satellite surveying. Our pragmatically oriented and detailed presentation will, we hope, make this book a useful reference.

Since the background of the readership will vary from the more practically oriented to the theoretically inclined, it is always difficult to balance the mathematical and non-mathematical approach to topics. We have tried to balance both of these aspects in this book and hope that the one does not detract from the other. Chapter 1 is a simple introduction to the subject and its basis. A review of existing instrumentation appears in Chapter 2. We recognise that any discussion on GPS receiver technology will age quickly. However, the book would have been incomplete without such a section. Chapter 3 describes the procedurès of surveying with GPS: survey planning, survey execution and post-survey procedures are highlighted. This is followed by an assessment of the impact of GPS on the entire spectrum of surveying activities ranging from surveys for geophysical prospecting to those for monitoring crustal movement. The information presented in Chapter 4 provides the surveyor with a basic understanding of the geodetic reference systems used in satellite surveying and the transformation relationships that exist between them. The principles of positioning with GPS are dealt with in Chapter 5. Least squares adjustment procedures are reviewed and the observation equations for the various GPS observables are presented together with the partial derivatives required for improving station positions, clock parameters and the like. By combining the material presented in Chapters 4 and 5, the motivated reader will be able to develop his own GPS processing software. Chapter 6 deals with vertical surveying using GPS. Important aspects of GPS orbit determination are presented in Chapter 7.

It is difficult to acknowledge properly all those responsible for the work presented here. However, we would like to thank particularly Yehuda Bock, David Close, Chuck Counselman, Rod Eckels, Don Grant and Bruce Harvey who read the manuscript and made helpful suggestions for its improvement. Ewan Masters and Chris Rizos are supported under the Australian Research Grants Scheme. Robert King was supported by the United States Air Force Geophysics Laboratory, through contract F19628-82-K-0002 with the Massachusetts Institute of Technology.

E. G. Masters
C. Rizos
A. Stolz
Kensington, N.S.W.,
AUSTRALIA

R. W. King
Cambridge, Mass., U.S.A.

J.Collins
Rockville, Md., U.S.A.

November, 1985.

CONTENTS

		Page
	Foreword	i
	Acronyms	iv
1	The Global Positioning System	1
	1.1 Preamble	1
	1.2 GPS Measurement Types	3
	1.3 Measurement Differencing	4
	1.4 GPS Orbit Considerations	8
	1.5 Suggested Reading List	9
2	GPS Receiver Technology	11
	2.1 GPS Signal Structure	11
	2.2 GPS Measurement Modes	12
	2.3 GPS Instrumentation	17
3	Surveying with GPS	24
	3.1 Introduction	24
	3.2 Survey Planning	25
	3.3 Survey Execution	27
	3.4 Post-Survey Processing	29
	3.5 Impact of GPS on Surveying	33
4	Time, Earth Rotation, Reference Systems and Geodetic Datums	40
	4.1 Definitions of Time	40
	4.2 Celestial and Terrestrial Reference Systems	43
	4.3 Relating the Conventional Terrestrial System to Geodetic Datums	50
5	GPS Data Processing	55
	5.1 GPS Phase Observables	55
	5.2 Analysis of GPS Observations	64
	5.3 Error Analysis	79
	5.4 Dual-Frequency Observations	82

6 Height Determination Using GPS 86

6.1 Introduction 86
6.2 Geoid Computation Techniques 87
6.3 Remarks on "Absolute" Geoid Determination 94
6.4 Relative Heights from GPS 95

7 Satellite Orbital Motion 98

7.1 Motion of a Near-Earth Orbiting Satellite 98
7.2 Motion of GPS Satellites 102
7.3 Computation of Satellite Orbits 109
7.4 GPS Ephemerides 115

References 121

ACRONYMS

AGD	Australian Geodetic Datum
BIH	Bureau International de l'Heure
CIO	Conventional International Origin
CCRS	Conventional Celestial Reference System
CTRS	Conventional Terrestrial Reference System
DMA	Defense Mapping Agency (USA)
EDM	Electronic Distance Measurement
GAST	Greenwich Apparent Sidereal Time
GMST	Greenwich Mean Sidereal Time
GPS	Global Positioning System
GPST	Global Positioning System Time
IAG	International Association of Geodesy
IAU	International Astronomical Union
JD	Julian Date
LO	Local Oscillator
NASA	National Aeronautical and Space Administration
NGS	National Geodetic Survey (USA)
NSWC	Naval Surface Weapons Center (USA)
SLR	Satellite Laser Ranging
TAI	International Atomic Time
TDB	Barycentric Dynamical Time
TDT	Terrestrial Dynamical Time
UT	Universal Time
UTC	Coordinated Universal Time
VLBI	Very Long Baseline Interferometry
WGS	World Geodetic System
WSSR	Weighted Sum of Squared Residual

1. THE GLOBAL POSITIONING SYSTEM

1.1 PREAMBLE

The Global Positioning System (GPS) is revolutionising surveying technology. Like its predecessor, the TRANSIT Doppler system, GPS shifts the scene of surveying operations from ground-to-ground measurements to ground-to-sky, with obvious implications: intervisibility of marks is no longer a criterion for their location; operations are possible in nearly all kinds of weather and can be performed during day or night; and the skills required to utilise the technology are different both in field operations and data processing. But GPS is not merely a replacement for TRANSIT. The simultaneous visibility of multiple satellites allows effective cancellation of the major sources of error in satellite observations, with the result that with GPS, relative positioning accuracies of one part per million (ppm) or better over distances from one kilometre to thousands of kilometres are possible. **This means that GPS can compete with terrestrial techniques over short distances, and can achieve more accurate results in less time than TRANSIT observations over longer distances.**

GPS was designed primarily as a navigation system, to satisfy both military and civilian needs for real-time positioning. This positioning is accomplished through the use of coded information, essentially clever timing signals, transmitted by the satellites. Each GPS satellite transmits a unique signal on two L-band frequencies: L1 at 1575.42 MHz and L2 at 1227.60 MHz (equivalent to wavelengths of approximately 19 and 24 cm, respectively). The satellite signals consist of the L-band carrier waves modulated with a "Standard" or S code (formerly called the C/A code), a "Precise" or P code and a Navigation Message containing, amongst other things, the coordinates of the satellites as functions of time -- the "Broadcast Ephemerides". The S code which is intended mainly for civilian use, yields a range measurement precision of about 10 m. The navigation service provided by this code is referred to as the Standard Positioning Service (SPS). The P code is intended for military and selected civilian use only and yields a measurement precision of about 1 m. The navigation service provided by the P code is therefore referred to as the Precise Positioning Service (PPS). Although both codes can be used for surveying, a more accurate method is to measure the phase of the carrier signal. For this reason, we will not discuss the detailed characteristics of the codes in this monograph. For details we refer the reader to Janiczek (1980).

There are currently eight usable satellites in orbit (see Table 1.1). These are the experimental, "Block 1" satellites, which will be progressively replaced as the "Block 2", operational satellites are placed into orbit beginning in 1986. By 1989 the system should be complete, with 18 satellites in six orbital planes -- at about 20 200 km altitude -- (Figure 1.1), allowing for simultaneous visibility of at least four satellites at any time of day almost anywhere in the world. The present constellation of satellites is configured to provide the most favourable geometry for testing the system over North America.

Table 1.1 Satellite launch dates and other data.

Navstar I.D.	Sat/Vehicle I.D.	Launch Date	Clock Type
1	4	22-02-1978	Quartz
3	6	7-10-1978	Rubidium
4	8	11-12-1978	Rubidium
6	9	26-04-1980	Cesium/Rubidium
8	11	14-07-1983	Rubidium
9	13	13-06-1984	Cesium
10	12	10-09-1984	Cesium
11	3	21-10-1985	Cesium

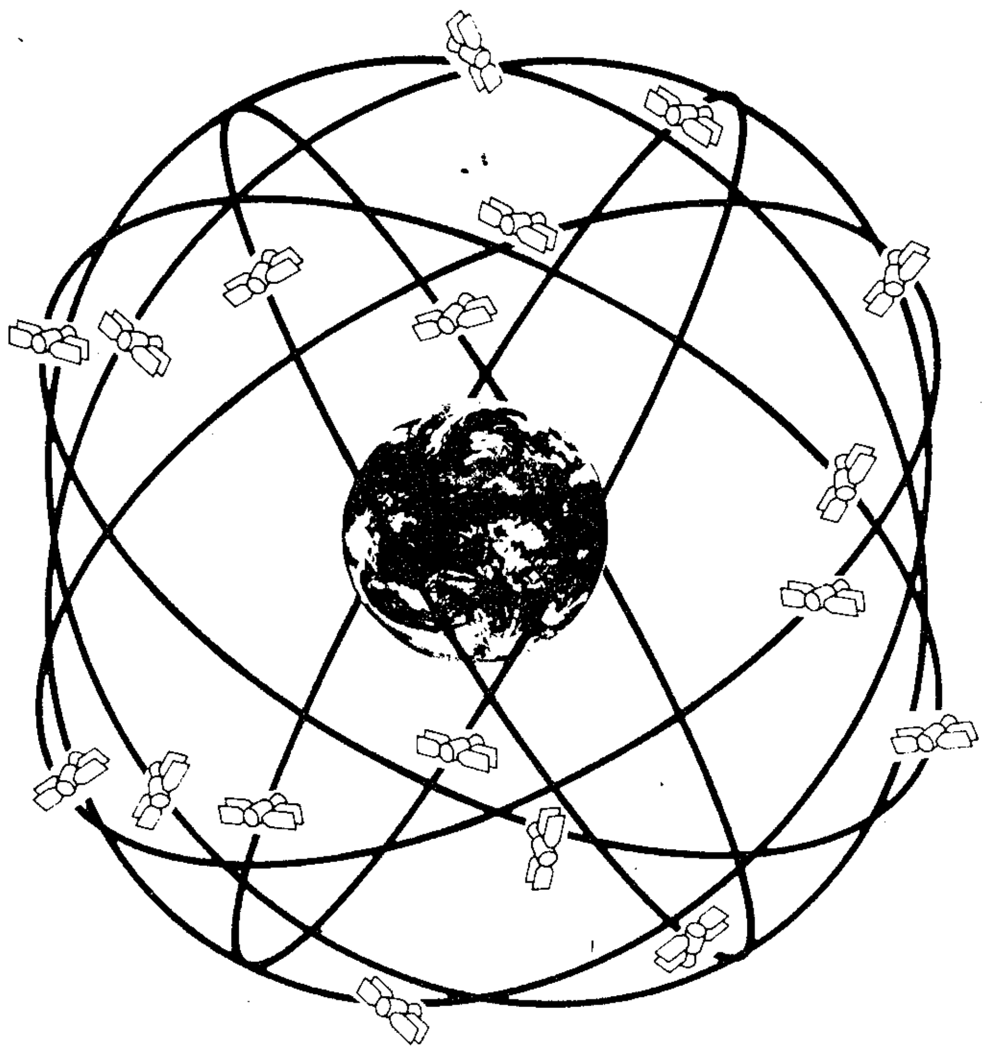


Figure 1.1 The GPS Satellite Constellation: 18 satellites, 6 orbital planes, 55° inclination, 20 200 km altitude, 12 hour orbits. (Courtesy Wild Heerbrugg)

As it happens, the observation geometry is equally favourable in Australia, and it is possible now to obtain surveying accuracies equal to those obtainable when the system is fully configured, but only for about six hours per day. At the time of writing (November 1985), the period of maximum mutual visibility of the satellites in eastern Australia is between 6 pm and midnight local time. This period regresses by 4 minutes per day (or 2 hours per month), returning to the same times a year from now. This period of useful visibility will increase as additional satellites are launched from late 1985.

As with TRANSIT, much higher accuracies are obtained in relative positioning from observations made simultaneously at two observing stations. Consequently, unless otherwise indicated, all discussion concerning data acquisition and processing will assume a two-receiver configuration. This is often referred to as the differential mode. The position differences so determined constitute the baseline vector or simply the baseline between the points occupied by two receivers.

All satellite positioning systems provide ground coordinates of a receiver (or the baseline vector between a pair of receivers) in an earth-centred coordinate system. The orientation of the system is determined by the tabulated coordinates or ephemerides of the GPS satellites. In order to relate coordinates determined by GPS surveying to the local geodetic datum a transformation relationship needs to be established. The GPS coordinate system and the nature of datum transformations is discussed in Chapter 4.

The following factors influence the final positioning accuracy obtainable with GPS:

(a) The precision of the measurement and the receiver-satellite geometry.
(b) The measurement processing technique adopted.
(c) The accuracy with which atmospheric and ionospheric effects can be modelled.
(d) The accuracy of the satellite ephemerides.

Each of these factors is discussed briefly in the next three sections and in more detail in Section 5.3. Section 1.5 contains a general reading list on selected aspects of GPS.

1.2 GPS MEASUREMENT TYPES

GPS measurements can be made using either the carrier signal or the codes. Code measurements are called pseudo-ranges and can be based on either the P code or the S code. Knowledge of the properties of each of these types of measurements is necessary for understanding and evaluating GPS instruments.

Pseudo-ranges are the simplest to visualize geometrically, as they are essentially a measurement of distance contaminated by clock errors. Throughout this monograph, we use the terms clock, frequency standard and oscillator to denote the same thing, namely, a device for precisely measuring a time interval. When four satellites are observed simultaneously, it is possible to determine the three-dimensional position of the ground receiver, and the receiver clock offset, at a single epoch. This is simply **resection by distance**, in surveying terminology, with the satellites serving as the control stations. As with the resection technique, the precision is a function of the **geometry** of the receiver in relation to the four visible satellites. The best geometry would be when the satellites are in each of the four quadrants and each at an elevation angle of 40°-70° above the horizon. However, pseudo-range measurements are not nearly as precise as phase measurements of the carrier wave itself. In order to achieve position accuracies of 10 m from P code measurements or 100 m from S code measurements (adequate for navigation), it was only necessary to design a code structure which allowed metre level measurement precision. Moreover, the more precise P code will likely be encrypted, and may therefore not be available for non-military use, when the system becomes fully operational in 1989. An additional impediment to accurate pseudo-ranging arises from multipath effects; that is, the tendency of some fraction of the satellite signal to reach the receiver antenna via reflection off the ground or other surfaces. The size and signature of multipath effects depend on antenna design and height of the antenna above ground but probably cannot be reduced below a few decimetres with practical configurations.

Carrier phase **measurements are more precise than the pseudo-ranges and not as vulnerable to multipath effects.** The wavelength of the stronger of the two L-band carrier signals, L1, is 19 cm, so even rough interpolation of phase gives centimetre level precision. From the technique of Electronic Distance Measurement (EDM), we know that phase measurements are ambiguous, and unless one can determine the absolute range difference at the initial epoch, phase measurements give only the changes in range ("integrated Doppler") over the observing period. However, the absolute range difference can often be easily determined, as we discuss in Chapter 5.

Carrier phase can be determined from the code-modulated signal either by using the code or other techniques (see Section 2.2). The L1 signal, which has both P code and S code modulation, can thus be tracked with S or P code receivers or with codeless receivers. The L2 signal, useful for removing ionospheric effects for very precise applications (< 2 ppm for relative positioning), has no S code modulation, so that receivers for these applications must either have P code capability or operate without code.

It is also possible to track the phase of the 10.23 MHz P code transition signal or P code sub-carrier without knowledge of the codes (see Section 2.2). The long wavelength (approximately 30 m) of this signal compared with the L-band carrier allows relatively easy resolution of the integer-cycle ambiguity, producing in effect a pseudo-range measurement. However, the long wavelength makes the measurements more susceptible to multipath effects, roughly to the same degree as pseudo-range measurements. Figure 1.2 illustrates the distinction between coded pseudo-range and phase measurements.

1.3 MEASUREMENT DIFFERENCING

Like TRANSIT, the dominant source of error in a single GPS measurement or series of measurements between a satellite and a ground station is **the unpredictable behaviour of the time and frequency standards serving as reference for the transmitter and receiver.** Even though the GPS satellites carry atomic frequency (Cesium or Rubidium) standards, the instability of these standards would still limit positioning to the several metre level were it not for the possibility of eliminating their effect through <u>signal differencing</u>. Differences may be taken between satellites, between stations, between epochs, or combinations of any of these. It can be implemented either by simple subtraction of signals or by more elaborate procedures and estimation schemes (see Section 5.2.6), but the effect is the same: the unpredictable and dominant errors from the ground and satellite "clocks" are eliminated. We examine the effect and application of each below.

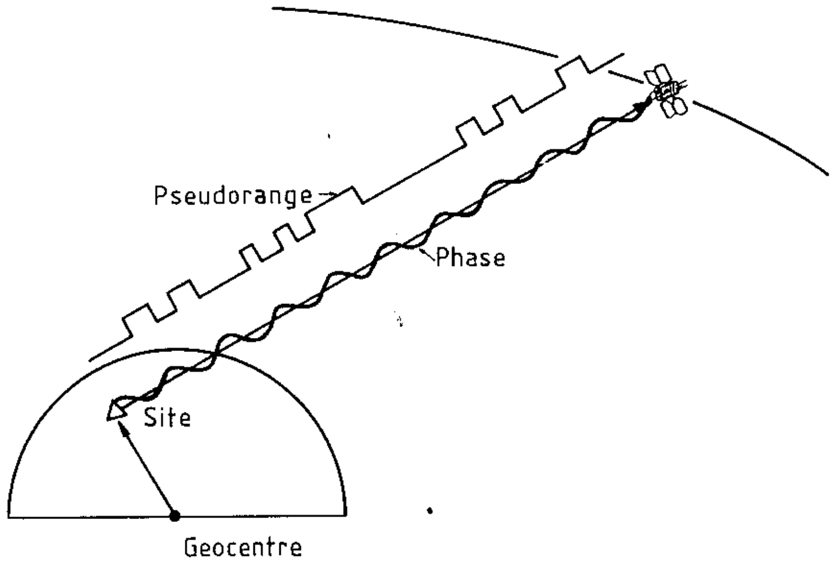


Figure 1.2 Pseudo-range and Phase Measurements. (Courtesy Wild Heerbrugg)

For a single satellite, differencing pseudo-ranges or phases of signals received simultaneously at each of two ground stations eliminates the effects of bias or instabilities in the satellite clock. This measurement is commonly called the <u>between-stations-difference</u>, or <u>single-difference</u> observable and is the basis of the differential mode of GPS receiver deployment. If the ground stations have crystal oscillators (normally the case for field instruments), then single differences are not very useful since noise from the receiver oscillators will dominate the measurement whether or not the effects of the satellite clocks have been cancelled. However, if the receivers are connected to atomic frequency standards, as may be the case for fixed tracking stations used for determining the orbits of the GPS satellites, the single-difference observable is very useful.

For a single station, differencing simultaneous observations received from each of two satellites (the <u>between-satellites-difference</u>) eliminates mainly the effect of bias or instabilities in the receiver clock. Since the satellites carry atomic clocks, pseudo-range or phase observations from several satellites can be used to point-position a receiver which carries only a crystal oscillator. Instabilities in the satellite frequency standards, as well as errors in the satellite orbits, normally limit the accuracy of these point positioning results to the several metre level, even after several hours of tracking.

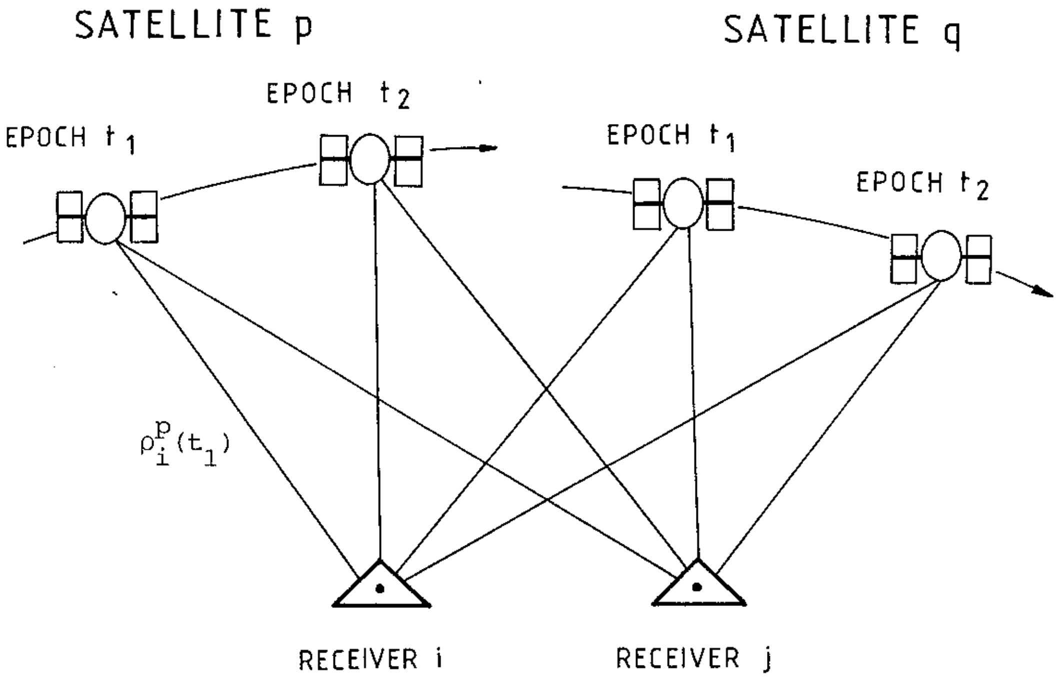
Double-differencing, between satellites and between stations, cancels the effect of instabilities in both the satellite and receiver oscillators. This is the observable used in almost all GPS surveying. Because double differencing eliminates the unknown initial phases of the satellite and receiver oscillators, it is the only phase observable in which the integer-cycle ambiguity can be resolved (see Sections 5.2.4 and 5.4.2).

Double-differences can, in turn, be differenced between epochs, that is, the double differences between the same satellites and the same stations at two consecutive time epochs are differenced. This is sometimes referred to as triple differencing. In this "observable" the cycle ambiguities are eliminated, making it very useful for quick, though lower accuracy, solutions.

Figure 1.3 illustrates differenced measurements between receivers, between satellites and between time epochs. The precise mathematical models for the various difference observables are given in Section 5.1.

Positions of a roving receiver can be determined relative to a fixed master station. This idea can be expanded to incorporate permanent GPS stations -- the base station concept. Even if the distance between the two receivers is very large, this technique of "point-positioning by relative-positioning" is more accurate than using the undifferenced, one-way observations. For example, at a separation of 1000 km, the 1-2 ppm accuracy usually attainable in relative positioning implies a point position accuracy of 1 m. This is a factor of 3-10 better than can be obtained today with single-receiver point positioning, and much better than may be possible in the future if the stability of the satellites' oscillators is intentionally degraded. This is also a promising technique for increasing the accuracy of S code point-positioning. The differential positioning of one roving S code navigation receiver relative to a fixed base receiver, is possible through the use of baseline vector updates derived by monitoring the measured three-dimensional coordinates of the base receiver. This technique may be used to satisfy many lower accuracy survey requirements.

In relative positioning over fairly short distances (10-50 km), differencing signals between stations also reduces errors due to unmodelled tropospheric and ionospheric refraction effects. To the extent that the receivers are observing the satellites through the same propagation medium, these effects will cancel. The magnitude of tropospheric and ionospheric effects under varying conditions is discussed in Section 5.3.2. However, a useful rule-of-thumb is that for single-frequency observations, the ionosphere will usually dominate and will produce relative positioning errors at the level of 0.5 ppm (at night) to 2ppm (in the daytime).



BETWEEN-RECEIVER	BETWEEN-SATELLITE	BETWEEN-EPOCH
$[\rho_i^p(t_1)-\rho_j^p(t_1)]$	$[\rho_i^p(t_1)-\rho_i^q(t_1)]$	$[\rho_i^p(t_2)-\rho_i^p(t_1)]$
$[\rho_i^q(t_1)-\rho_j^q(t_1)]$	$[\rho_i^p(t_1)-\rho_i^q(t_1)]$	$[\rho_j^p(t_2)-\rho_j^p(t_1)]$
$[\rho_i^p(t_2)-\rho_j^p(t_2)]$	$[\rho_j^p(t_2)-\rho_j^q(t_2)]$	$[\rho_i^q(t_2)-\rho_i^q)t_1)]$
$[\rho_i^q(t_2)-\rho_j^q(t_2)]$	$[\rho_j^p(t_2)-\rho_j^q(t_2)]$	$[\rho_j^q(t_2)-\rho_j^q(t_1)]$

Figure 1.3 GPS Phase Measurement Differencing. (Adapted from Wells 1985)

1.4 GPS ORBIT CONSIDERATIONS

An ephemeris is a list of coordinates defining the orbital position of a satellite at various times. All GPS measurement processing techniques require the input of ephemerides for the time span of the observations in order to determine a ground receiver's position, either absolute in the point positioning mode, or relative when deployed in a differential mode.

For point positioning, errors in the GPS ephemerides introduce errors in position of roughly the same magnitude as the ephemeris errors. The position errors may be reduced by a factor of 2-3 by averaging when multiple satellites are observed for several hours. In relative positioning however, a number of error sources, including ephemeris error, produce nearly equal shifts in the estimated positions of both stations and the effect of these errors on relative positioning is significantly reduced. An adequate rule-of-thumb for estimating the effect of orbit error on the baseline between two receivers is that **a given error in the satellite ephemeris introduces an error in relative positioning (baseline) reduced by the ratio of the baseline length to the satellite's altitude.** For example, a 20 m orbit error will result in a baseline error of 1 ppm, or 1 cm in 10 km, while a 200 m orbit error results in an error of 10 cm in a 10 km baseline, and so on. An important limitation on the precision of GPS surveys is the accuracy of the ephemerides available for data reduction.

Ephemerides can generally be classed either as post-processed or predicted. The post-processed ephemeris is determined after observations are made of the satellite and is therefore an estimate of the satellite's position in the period of the observations. Tracking data are acquired from a number of fixed stations and processed in order to obtain the orbit that best fits the data in a least squares sense The limitation on the ephemeris accuracy obtainable is due to such factors as the quality and coverage (in space and time) of the tracking data and the models (geometric and dynamic) used in the orbit computation procedure (Section 7.3.4). Orbits with accuracies at the metre level may be possible with special processing techniques and a dedicated tracking campaign. Post-processed ephemerides accurate at the 10-50 m level are available at present from the U.S. Department of Defense and at least one U.S. commercial surveying firm. It is likely that in the future a number of government and commercial organisations will compute post-processed ephemerides.

The "Broadcast Ephemerides" which are obtained by extrapolating a post-processed orbit for a few days into the future, allow real-time positioning using pseudo-ranges. The accuracy of these ephemerides is presently estimated to be at the 20-100 m level. Although the GPS Broadcast Ephemerides are intended to support air, land and sea navigation, their use in the differential mode provides baseline accuracies of the order of a few ppm (a few cm in a 10 km baseline). Only GPS receivers with a built-in code/navigation message deciphering ability will be able to utilise these ephemerides directly. While observations are made, the Navigation Message (Section 7.4.2) is interrogated and the ephemerides recorded for later use in the data reduction phase. It is expected that the quality of the Broadcast Ephemerides available to the general user will be degraded when GPS enters its operational phase in 1989. Consequently it may be necessary for surveyors to use post-processed ephemerides by the end of the decade.

1.5 SUGGESTED READING LIST

Below we have compiled a short list of important references. For the motivated reader keen to learn more about GPS, the Proceedings of the First International Symposium on Precise Positioning with GPS (Goad 1985) provides the most complete source of up-to-date information on GPS instrumentation, processing software and GPS surveying in general. Janiczek (1980) contains several articles describing the GPS ground and space segments, the signal structure and the use of GPS for navigation purposes. Wells (1985) has standardised technical terminology that has grown around GPS and has clarified present definitions of the receiver deployment modes, GPS measurement techniques and signal processing procedures. Bock et al (1985) present the methodology for processing GPS measurements for surveying applications and describe a GPS survey carried out in the Eifel region of West Germany in great detail. The Eifel survey was carried out to densify the German geodetic control net and to date represents the best documented survey performed with GPS. Collins (1984) describes the types of surveys that have been carried out by a commercial surveying firm using the Macrometer GPS receiver. Rizos et al (1984) assess the likely impact of GPS on surveying, geodesy and geodynamics, with particular reference to Australia. Engelis et al (1985) discuss the determination of orthometric heights using GPS, with particular reference to the Eifel network. Bossler (1984) discusses the possible impact of modern technologies such as GPS on traditional geodetic control networks and questions the future role of a monumented control network.

BOCK, Y., ABBOT, R.I., COUNSELMAN III, C.C., GOUREVITCH, A. & KING, R.W., 1985, "Establishment of Three-dimensional Geodetic Control by Interferometry with the Global Positioning System", J.Geophys.Res., 90, p.7689-7703.

BOSSLER, J.D., 1984, "The Impact of VLBI and GPS on Geodesy", EOS, Trans.Am.Geophys.U., 64, p.569-570.

COLLINS, J., 1984, "GPS Satellite Surveying", ACSM Bulletin, 88, p.23-26.

ENGELIS, T., RAPP, R.H. & BOCK, Y., 1985, "Measuring Orthometric Height Differences with GPS and Gravity Data", Manuscripta Geodaetica, 10, p.187-194.

GOAD, C.C. (convenor), 1985, "Proceedings of the First International Symposium on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985", U.S. Dept of Commerce, National Geodetic Information Center, NOAA, Rockville, Md. 20852, May 1985, 946pp.

JANICZEK, P.M. (ed.), 1980, "Global Positioning System", papers published in Navigation, reprinted by the Institute of Navigation, Washington, D.C.

RIZOS, C., STOLZ, A. & MASTERS, E.G., 1984, "Surveying and Geodesy in Australia with GPS", Aust.Surveyor, 32, p.202-225.

WELLS, D., 1985, "Recommended GPS Terminology", in proc. First Int. Symp. on Precise Positioning with GPS -- Positioning with GPS-1985, U.S. Dept. of Commerce, National Geodetic Information Center, NOAA, Rockville Md. 20852, May 1985, p.903-923.

2. GPS RECEIVER TECHNOLOGY

2.1 GPS SIGNAL STRUCTURE

Each GPS satellite transmits a unique navigational signal centred on two L-band frequencies: L1 at 1575.42 MHz (equivalent to a wavelength of approximately 19 cm) and L2 at 1227 60 MHz (equivalent to a wavelength of approximately 24 cm). The satellite signals consist of the L-band carrier wave modulated with the P code or the S code and the Navigation Message -- see Figure 2.1. The primary function of the P and S codes is to permit the signal transit time from satellite to receiver to be determined. The transit time when multiplied by the velocity of light gives the range to the satellite.

At present the L1 carrier is modulated with both P and S codes, whereas the L2 carrier is modulated only with the P code. Both carriers contain the Navigation Message, a data stream 30 seconds long, broadcast continuously, containing satellite ephemerides (see Section 7.4.2), clock parameters, and general system status.

P code ranging on the two frequencies allows the ionospheric refraction correction to be determined. The absence of a S code on L2 is, of course, intentional and is one of the accuracy limitations imposed on non-authorised (that is, non-P code) users of the system. Authorised users are the military and those working in the national interest of the U.S. and its allies. Once GPS becomes fully operational, the P code technology will be tightly controlled and any receivers using the P code will need to be "secure" and operated only within special guidelines.

Although both codes have the characteristics of random noise, they are in fact **binary codes** generated by a mathematical algorithm and are therefore referred to as "pseudo-random noise". The codes are accurate time marks that permit the receiver's internal processor to compute the time of transmission of the satellite signal. The signal transit time is essentially the phase shift between identical code sequences (P or S) generated by frequency oscillators in the satellite and in the user's receiver. All the satellite clocks are synchronised to the GPS time system (Section 4.1.2).

If the receiver were equipped with a high precision clock, synchronised perfectly with the GPS time system, it would measure the "true" range. By ranging simultaneously to three satellites the user's position would be defined by the intersection of three spheres of known radii, centred at each satellite whose coordinates are provided by the Navigation Message. Normally receivers will be equipped with crystal clocks that do not necessarily keep the same time as the more stable satellite clocks. Consequently this range is contaminated by the receiver clock error. This quantity is referred to as "pseudo-range" and the user is required to track four satellites and solve four equations in the four unknowns: the three-dimensional position components and the receiver-clock offset.

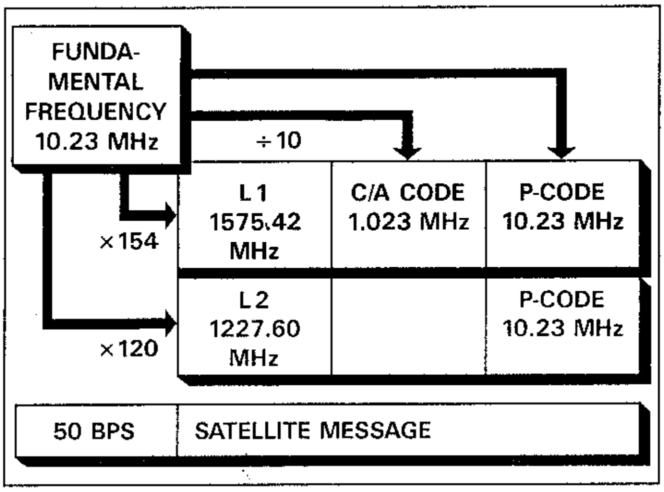


Figure 2.1 GPS Satellite Signal.
(Courtesy Wild Heerbrugg)

What is the nature of the codes? The P code is generated at the GPS clock frequency of 10.23 MHz. One element in the code sequence therefore corresponds to a time interval of about 100 nanoseconds, which is approximately equivalent to 30 m in range. The resolution can be improved to the submetre level by interpolation. The P code does not repeat itself for 267 days. This 267 day sequence is subdivided in such a manner that each satellite is assigned a unique 1 week portion of the code that does not overlap with that assigned to any other satellite.

By comparison the S code is not as complex. It is a code sequence of frequency 1.023 MHz, corresponding to a range resolution of the order of 300 m (which can also be improved upon by interpolation). The S code repeats itself every millisecond, which implies that range measurements have an ambiguity of integer multiples of 300 km. For P-coded receivers, the standard observing procedure is initially to "lock-on" to the S code, decode the Navigation Message and use the "handover" synchronisation data contained therein to switch from the S code to the P code for precise pseudo-range measurement. All this is accomplished in real-time. We refer the reader to Spilker (1978) for details.

2.2 GPS MEASUREMENT MODES

There are three basic types of measurements that can be made on the GPS signals. Where possible we follow the definitions and descriptions given by Wells (1985).

Pseudo-range is the time shift required to align (correlate) a replica of the GPS code generated in the receiver with the incoming GPS code, scaled into distance by the speed of light. Pseudo-range can be obtained either from the S code or the more precise P code. Figure 2.2 illustrates the principle of S coded pseudo-range measurements. GPS navigation receivers use this measurement type to perform real-time positioning. Pseudo-range is biased by the time offset between the satellite and receiver clocks.

Carrier phase, or more precisely, "carrier beat phase", is the difference between the phase of the incoming Doppler-shifted satellite carrier signal and the phase of an oscillator signal of nominally constant frequency generated in the receiver. The measurement can be obtained by reconstructing the carrier signal after removing the codes from the incoming signal, either by correlation with the codes or effectively squaring the signal (see discussion below). The carrier phase is usually sampled at epochs determined by the receiver clock. The carrier phase changes according to the continuously integrated (accumulated) Doppler shift of the incoming signal and is biased by the unknown offset between the satellite and receiver reference oscillators (see Section 5.1.1). The carrier phase measurement is like a very precise pseudo-range but with an integer number of carrier cycles (wavelengths) missing -- the so-called integer-cycle ambiguity. The integer-cycle ambiguity can be thought of as the (unknown) number of whole carrier wavelengths between receiver and satellite, while the phase measurement is the fraction of a cycle remaining in the receiver-satellite range.

P code transition phase is the difference between the phase of the incoming Doppler-shifted L1 or L2 P code sub-carrier and the phase of a reference oscillator in the receiver. This measurement can be made with or without the use of codes. It is less precise than carrier phase, but because of the longer wavelength (30 m versus 19 cm) its integer-cycle ambiguity is more easily resolved.

Instruments that require a knowledge of the codes -- for any of the measurement modes -- are referred to here as code-correlating instruments. These are by far the most common of the GPS receivers as they include all the navigation receivers. The codeless receivers were specifically developed for high precision surveying applications. Only two commercial receivers are based on codeless principles -- the MACROMETER V-1000 (R)[1] (and its successor, the MACROMETER II (TM)[1]) and the GPS Land Surveyor (TM)[2] Model 1991 (and its successor, the Model 2002).

Compared with a code resolution of 30 m for the P code and 300 m for the S code, the carrier cycle permits measurement resolutions which are better than 20 cm. Interpolation yields an equivalent linear measurement of subcentimetre precision for carrier phase, compared to decimetres for P code and metres for S code measurements. However, all measurement types contain errors due to the propagation delay of the atmosphere and are affected by clock offsets and clock instabilities in both the satellites and receivers. Usually, the clock instability

---

[1] MACROMETER is a registered trademark and MACROMETER II is a trademark of Aero Service Division, Western Geophysical Company of America.

[2] GPS Land Surveyor is a trademark of GPS Services, Inc., USA.

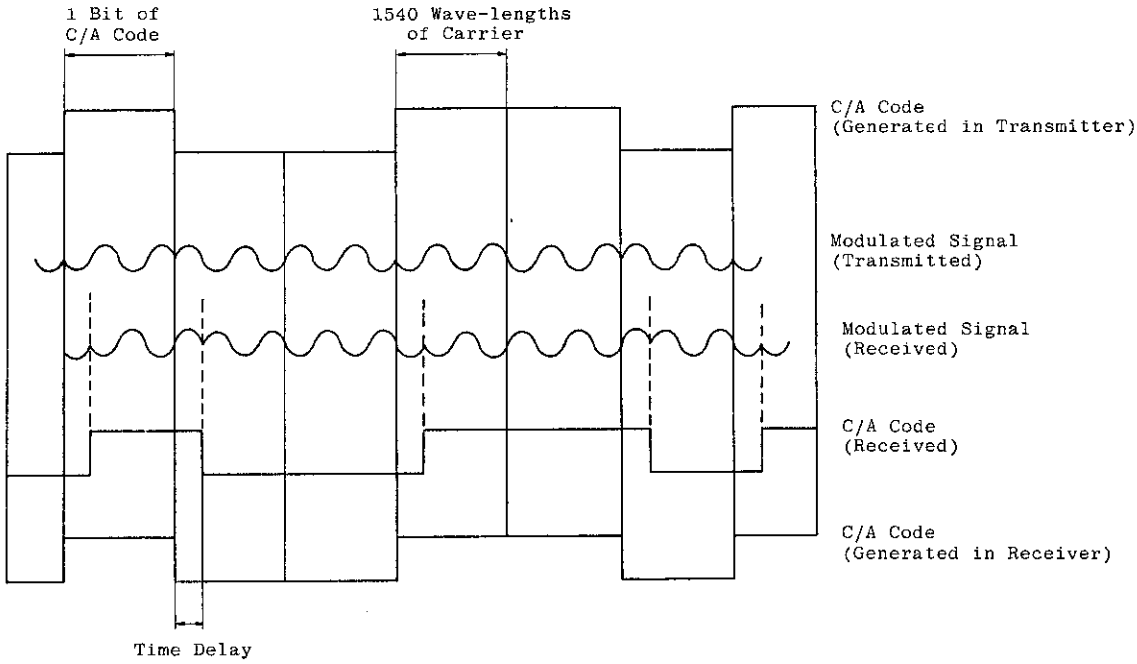


Figure 2.2 Principles of S coded Pseudo-range Measurement.

problem is overcome by differencing the observations between satellites and receivers (see Section 1.3 & 5.1). S code pseudo-ranging receivers give the least accurate measurements. Carrier phase receivers which use both the L1 and L2 frequencies to remove ionospheric effects will give the most accurate measurements. Hence, they are likely to be used for crustal movement surveys or where the highest accuracy is sought. The majority of surveying receivers measure the carrier phase on the single L1 frequency.

A GPS receiver generally has one or more channels. A channel consists of the hardware and software required to track the signal from one satellite at one of the two carrier frequencies. Receivers can be <u>multi-channel</u> and thus contain a number of channels equal to the number of satellite signals which are to be tracked simultaneously (for example the Macrometer V-1000 has six channels) or <u>multiplexing</u> where one channel is quickly sequenced through a number of satellite signals (for example, the Texas Instruments TI 4100). Some of the new generation receivers are both multi-channel and are capable of multiplexing (the Magnavox/Wild WM 101 has five channels and can track 9 satellites).

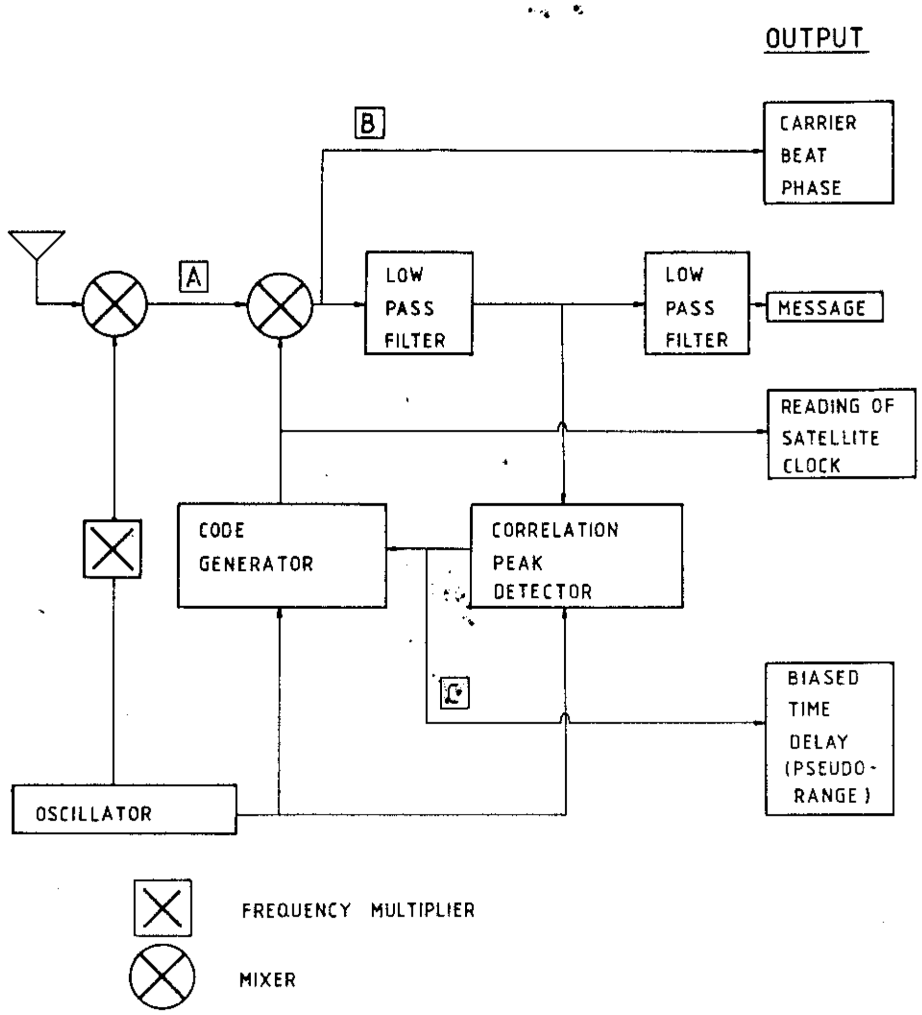


Figure 2.3 GPS Correlation-type Channel.
(Adapted from Wells 1985)

A correlation-type channel (Figure 2.3) uses a delay lock loop to maintain alignment (correlation peak) between the replica of the GPS code generated in the receiver, and the incoming code. There are many advanced techniques now used for the correlation process. The TI 4100 is an example of a code-correlating geodetic receiver. A number of other receivers which also use code-correlation techniques are under various stages of development and testing (see Section 2.3.1).

A squaring-type channel multiplies the received signal by itself to obtain a second harmonic of the carrier which does not contain the code modulation. The squaring concept is easily demonstrated by squaring a standard cosine signal:

$$y = A\cos(\omega t + \phi) \tag{3.1}$$

to obtain

$$y^2 = A^2\cos^2(\omega t + \phi) = A^2[1 + \cos(2\omega t + 2\phi)]/2 \tag{3.2}$$

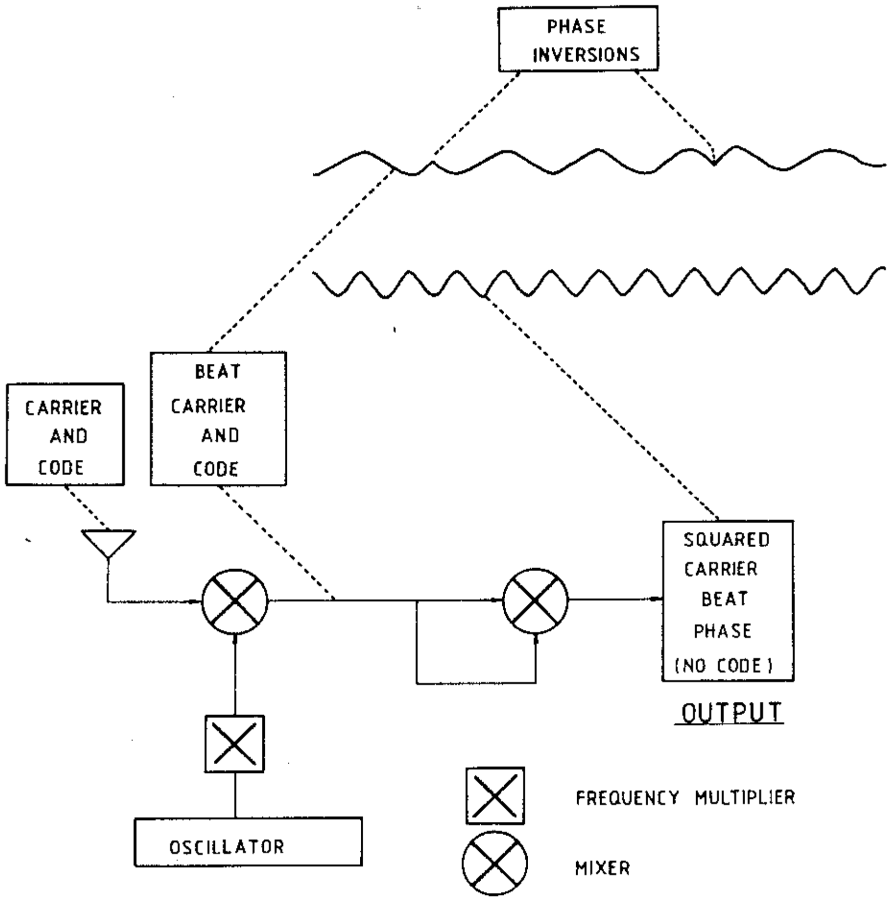


Figure 2.4 GPS Squaring-type Channel.
(Adapted from Wells 1985)

where A is the sequence of +1's or -1's generated by the code, ω is the frequency of the carrier, and ϕ is its phase. In practice, this simple conceptual description of the squaring process is implemented by one of several proprietary techniques which have been developed. The Macrometer V-1000 (and the later dual-frequency model) employs this technique to recover the demodulated carrier signal.

An alternative to the squaring process, which also does not require detailed knowledge of the codes, is the SERIES technique (MacDoran 1983, Buennagel et al 1984), in which the spectra of modulations of the GPS signals are compressed into the audio band and the Doppler shift of the P code modulation transitions is measured. The GPS Land Surveyor 1991 (and the later 2002 model) is the only commercially available geodetic receiver that employs this technique. P code ranging is also performed to assist in the cycle ambiguity resolution (only 30 m for the P code modulation).

2.3 GPS INSTRUMENTATION

Any discussion on GPS receiver technology will date quickly as new instruments are introduced, product lines discontinued and receiver specifications change. Thus this monograph cannot hope to do justice to the fast moving area of GPS instrument development. We do, however, include a review of existing GPS instrumentation:

(a) to provide a "historical" perspective in which to view the development of commercial receivers.
(b) to compare the coming generation of receivers with those that pioneered GPS satellite surveying.
(c) to demonstrate that many of the operational characteristics of the first generation receivers are similar to those of the newly developed receivers.

2.3.1 GEODETIC RECEIVERS

At the time of writing, three GPS geodetic receivers have been independently tested and used on a commercial basis, and can therefore be described as operational. What follows is a discussion of the salient points of each of these instruments.

MACROMETER V-1000

The first GPS receiver designed and built for the general surveying market was the MACROMETER V-1000, developed by Macrometrics Inc., formerly a subsidiary of Steinbrecher Corporation of Woburn, Massachusetts (USA). The Aero Service Division of Western Geophysical Company of America, Houston, Texas, acquired the Macrometrics company from Steinbrecher in March 1984.

In January 1983, the U.S. Federal Geodetic Control Commission (FGCC) carried out an evaluation of the MACROMETER V-1000 (Hothem & Fronczek 1983), comparing relative positioning results for station separations from less than a kilometre to over 40 km to those of ground surveys. Baseline accuracies of a few mm on the shorter baselines, and 1 ppm or better on the longer lines were demonstrated by the FGCC test. Since then the MACROMETER V-1000 has been in continuous use by both private and government organisations. Besides Aero Service, organisations that have acquired MACROMETER units include Geo/Hydro Inc., Rockville, Maryland (USA), U.S. National Geodetic Survey (NGS) and the Texas State Department of Highways and Transportation. During the first two years of operation thousands of control points have been established -- see Collins (1984) for a list of surveys carried out by a private surveying firm using these instruments.

A number of surveys performed using the V-1000 have been particularly well documented. In the Eifel region of West Germany, a 34-station network was surveyed in order to densify the German geodetic network. The average baseline length was about 10 km and the final accuracies were about 1-2 ppm (Bock et al. 1985). Similar surveys have been carried out in the Inn Valley of Germany (Heister et al. 1985) and Manitoba, Canada (McArthur et al. 1985). A 200-station network in Alberta, Canada was successfully surveyed to an accuracy of 5 ppm using helicopters to transport the receivers between monuments (Leeman et al. 1985). Urban/suburban control densification was achieved by surveys in Montgomery County, Pennsylvania (near Philadelphia) (Cuppels and Collins 1985); Collins & Leick 1985) and in Sainte-Foy, Quebec (Moreau et al. 1985). Use of the V-1000 in a high precision engineering survey has been described by Ruland & Leick (1985).

Essentially, the MACROMETER V-1000 comprises three units: a portable receiver/recorder with power supply, an omni-directional antenna connected to a horizontal ground plane and the P-1000 processor. The P-1000 is a microcomputer containing the software for carrying out the various tasks associated with GPS surveying; for example, preparing the "A-files" needed by the V-1000 to acquire the satellite signals and for processing the recorded data. A description of the field operation and data reduction procedure for the V-1000 is given in Chapter 3 (the reader is also referred to Hothem et al 1984).

As mentioned previously, the fundamental measurement made by the MACROMETER is the difference between the phase of the L1 carrier signal transmitted by the satellite and the phase of the reference signal generated within the receiver. An important marketing feature of the instrument is that it can track the phases of the carrier signals without requiring a knowledge of the codes. The observation time required to achieve 1 ppm accuracy for baselines less than 100 km is of the order of 2-3 hours. Similar accuracy with observation times of 15-30 minutes have been claimed in initial tests with the MACROMETER II, the recently unveiled dual-frequency model (Ladd et al 1985).

Texas Instruments TI 4100

The Texas Instruments TI 4100 was developed by Texas Instruments with funding from three U.S. government agencies -- the Defense Mapping Agency, National Geodetic Survey and Geological Survey -- hence the TI 4100 is also referred to in the literature as the Tri-Agency Receiver. The TI 4100 was designed to satisfy all positioning requirements from navigation (using pseudo-ranges) to high precision surveys (using carrier phase measurement at both L-band frequencies). An important feature of the TI 4100 is that P code pseudo-ranging is performed simultaneously with carrier phase measurement (can be used to assist in the ambiguity resolution and to estimate the offset of the receiver clock from GPS time -- see Section 5.2.3). The TI 4100 receiver consists of three separate units: the receiver/processor assembly, an omni-directional antenna and preamplifier, and a data recorder.

The TI 4100 has been subjected to many tests by the military, in a variety of environments, in order to evaluate its suitability as a general-purpose geodetic receiver (for example, Henson et al 1985). Many of these tests have concentrated on positioning (both static and dynamic) using pseudo-ranges. Surveys have been carried out by the U.S. National Geodetic Survey, by Nortech Inc. of Calgary, Canada, and by Geophysical Services Inc. of Dallas, Texas, a subsidiary of Texas Instruments. Delikaraogou et al (1985) describe use of the TI 4100 to establish 3-dimensional geodetic control with a five-station network near Ottawa, Canada, and Cannon et al (1985) report on the measurement of a 1700 km baseline in North America.

Comparisons of the TI 4100 and the Macrometer V-1000 have been conducted by NGS (Goad et al 1985), the Geodetic Survey of Canada (McArthur et al 1985), and the University of New Brunswick (Kleusberg et al 1985). The results of these comparisons indicated no significant difference in accuracy, at the level of 1-2 ppm, for baseline lengths from about 10 km to over 1000 km.

The proposed encryption of the P code by the U.S. Department of Defense has made prospective buyers cautious. Nevertheless several organisations have purchased the TI 4100 in order to take advantage now of the freely available P code facility.

GPS Land Surveyor

The GPS Land Surveyor 1991 was the first geodetic receiver developed by ISTAC Inc., Pasadena, California and is distributed by Geo/Hydro Inc., a surveying firm located in Rockville, Maryland (USA). It is an outgrowth of the SERIES technology, originally conceived and subsequently developed by MacDoran and his co-workers at the Jet Propulsion Laboratory, Pasadena, California (MacDoran et al 1982). SERIES was intended for crustal motion studies in which repeated centimetre level accuracy measurements of distances between widely spaced points on the earth's crust are required. However, the original instrument configuration proved too cumbersome for general surveying (SERIES requires a campervan to house it). The GPS Land Surveyor is a far more compact, more efficient and much cheaper receiver than SERIES.

The instrument comprises the tripod mounted receiver electronics and antenna, separate data recording unit and rubidium clock. It is more compact than either the MACROMETER V-1000 or the TI 4100. Released in early 1985, it has undergone only limited testing but the combination of its portability and ease of operation (see Chapter 3) makes it an attractive product for medium accuracy surveys. (FGCC tests were performed in October 1985.) The GPS Land Surveyor is specifically aimed at the lower accuracy end of the survey market. An externally obtained ephemeris is necessary to process the observations (for example, a S code navigation receiver such as the Trimble 4000A can be used to decipher the Navigation Message, from which the Broadcast Ephemerides can be extracted).

In the GPS Land Surveyor Model 2002, announced in mid-1985, the recorder has been replaced by a battery-powered field computer (which can also perform the data processing). Therefore, the instrument is more compact still. It is also cheaper than the 1991.

Other Receivers for Surveying

Aside from new MACROMETER and GPS Land Surveyor models already mentioned, a number of new geodetic GPS receivers, of the code-correlating variety, were unveiled at the First International Symposium on Precise Positioning with GPS (Goad 1985a), held at Rockville, Maryland (USA), during April, 1985. Among these are: the Litton LGSS from Aero Products Division, Litton Industries (Moorpark, California); the TR5S from Sercel (France); the Trimble 4000S from Trimble Navigation (Mountain View, California); and the WM 101 from Magnavox (Torrance, California) and Wild (Heerbrugg, Switzerland). Some of these instruments have already undergone some testing and most can be expected to be on the market in 1986. The main characteristics of each of these receivers are listed in Tables 2.1 and 2.2, along with those of the operational receivers described above. (A variety of S code navigation receivers have been built and are in the process of product development, or already available for purchase, but any discussion on these is outside the scope of this monograph.)

Table 2.1 GPS Receivers - Essential Characteristics.

Receiver	Observables	Frequency	Code?	Satellites Tracked
MACROMETER V-1000	Carrier phase	L1	None	6
MACROMETER II	Carrier phase	L1,L2	None	6
Texas Instruments TI 4100	Pseudo-range Carrier phase	L1,L2	S,P	4
GPS Land Surveyor Model 1991	P code range Transition phase	L1	None	All
GPS Land Surveyor Model 2002*	P code range Transition phase	L1	None	All
Litton LGSS	Pseudo-range Carrier phase	L1	S	8
Sercel TR5S	Pseudo-range Carrier phase	L1	S	5
Trimble 4000S	Pseudo-range Carrier phase	L1	S	4
Magnavox/Wild WM 101**	Pseudo-range Carrier phase	L1	S	9

* Future upgrades include carrier phase measurement and dual-frequency capability.

** Future upgrade to dual-frequency capability.

2.3.2 PERIPHERAL EQUIPMENT

The following is a basic list of peripheral equipment required for the field operation of GPS receivers presently in use:

The MACROMETER V-1000 requires a 350 watt petrol generator, an automotive battery, a time receiver, a time interval counter, cassette recorder, battery charger, 12 volt DC to 110 AC inverter, and telephone modem, if information must be transmitted back and forth between the office and field.

The Texas Instruments TI 4100 requires two large automotive batteries to power the receiver and cassette recorder (or field computer) used for data recording.

The GPS Land Surveyor requires a time receiver, time interval counter, and special battery chargers to charge internal batteries. A GPS navigation receiver to acquire the "Broadcast Ephemeris" and a portable computer is also necessary if computations are to be performed in the field. (A field computer is part of the GPS Land Surveyor Model 2002 configuration.)

2.3.3 COMPUTER REQUIREMENTS

Some type of computer is necessary for performing the tasks listed under "Survey Planning", in Section 3.2. These include prediction of rise and set times of the GPS satellites, preparation of "sky plots", and carrying out any simulations that may be required. Post-observation processing of some sort may also be necessary, for example, combining individually measured GPS baselines by means of a network adjustment, converting ellipsoidal heights to orthometric values, transforming coordinates from the GPS reference system into a local datum, etc. A powerful microcomputer is probably adequate for all of these duties.

The MACROMETER requires specially-formatted "A-files" to select the satellites to be tracked (see Section 3.3.1). These files can only be produced by the specially configured P-1000 computer which is sold by the manufacturer. The observations recorded by the receiver also can be decoded only by the P-1000. However, with the data translated by the P-1000, subsequent computations may be performed on the P-1000 or any other computer, provided suitable software is available.

Initially the Texas Instruments TI 4100 was not provided with satellite surveying software to process carrier phase observations. However, TI now offers baseline estimation software to run on a TI portable computer. NGS has developed multi-station software that is freely available to all users, but is intended to run on a minicomputer, although it could possibly be installed on a microcomputer. The Magnet 4100 program (Hatch & Larson 1985), written under contract by Magnavox for the G.S.I. company (to which they have exclusive rights), runs on a portable TI computer. Additional software and hardware is necessary for the transcription of the data cassettes, if these are used. (It is possible to record observations directly into computer mass storage.) Moreover, it is not necessary to subscribe to an ephemeris service unless the quality of the "Broadcast Ephemerides" is not high enough to satisfy the survey accuracy requirements.

Table 2.2 GPS Receivers - Essential Characteristics (continued).

Receiver	Power Required	Size	Cost* $US
MACROMETER V-1000	Operating 115 V AC 350 watts Standby 12 V DC 18 watts	Receiver 69x53x64 cm 45 kg Antenna 91x91x15 cm 16 kg + power generator	$110 000 each
MACROMETER II	Operating 24 V 180 watts	Receiver (2 units) 27 kg (combined) Antenna 91x91x15 cm + batteries	?
Texas Instruments TI 4100	Operating 22-32 V DC 110 watts Standby 10 watts	Receiver 46x38x20 cm 25 kg Antenna 28x18 cm 2 kg + batteries	$140 000 each
GPS Land Surveyor Model 1991	24 V DC 12 watts	Main Unit 66x36x23 cm 23 kg Clock/Battery 46x36x15 cm 14 kg Antenna 28x28x8 cm .5 kg	$125 000 each
GPS Land Surveyor Model 2002	internal battery	Main Unit/Antenna 23x42(dia.) cm 7 kg Clock/Battery 42x28x26 cm 18 kg	$60 000 each
Litton LGSS	113/230 V AC 20-30 V DC	Receiver 41x23x20 cm 5 kg	?
Sercel TR5S	24 V DC	Receiver 36x53x45 cm 27 kg	?
Trimble 4000S	115/230 V AC 20/32 V DC	Receiver 22x45x48 23 kg	$40 000 each
Magnavox/Wild WM 101	10.5-15 V DC 20 watts	Receiver 17x51x39 cm 14 kg	< $100 000 each

* These figures, representing single receiver costs, are only approximate. Prices have generally been negotiable. Two receivers are required for relative positioning. A computer and software are also needed, increasing cost further.

In the case of the GPS Land Surveyor, the necessary software is provided with the receiver and all processing is performed on an IBM-compatible computer. At present both portable and desktop Compaq computers are being used with the Model 1991. The Compaq has been programmed to store the Broadcast Ephemerides data received by a Trimble 4000A GPS receiver (if this configuration is employed). If desired, computations can then be performed on-site. Alternatively, data tapes can be returned to a central office for later processing using a suitable GPS ephemeris (the Broadcast Ephemerides or a post-processed ephemeris). The Model 2002 includes a compact field IBM-PC compatible computer.

2.3.4 ANTENNA CONSIDERATIONS

The main concern with antenna design is the effect of multi-path interference, that is, the tendency for some fraction of the signal to reach the receiver antenna via reflection off the ground or some other surface. The signature of multi-path effects depends on the antenna characteristics and height above the ground. A useful rule-of-thumb is that multi-path effects are about 10% of the signal wavelength, hence pseudo-ranging and P code transition phase measurements are affected more than carrier phase measurements. The antenna required to receive GPS signals need not be steerable (that is, directional), but its actual shape may be selected with the accuracy of a specific survey in mind. With carrier phase measurements, multi-path interference can be kept to the millimetre level using a simple dipole antenna and a 1 m square ground plane (Counselman & Gourevitch 1981), as employed by the MACROMETER. For centimetre level precision, a very compact, lightweight antenna is adequate.

3. SURVEYING WITH GPS

3.1 INTRODUCTION

There are presently four main techniques used by surveyors to determine the coordinates of points, namely:

(a) EDM traverse, triangulation or trilateration.
(b) Doppler (TRANSIT) satellite surveying.
(c) Inertial positioning.
(d) Analytical photogrammetry.

Traditional surveying with EDM and theodolite produces relative positioning of, on average, 1 part in 10^5 for distances ranging up to a few tens of kilometres. Although the EDM-theodolite technique has virtually reached its limit of accuracy, economy and efficiency, a significant advantage over the other technologies is the relatively low equipment costs. However, its most serious limitation is the requirement for station intervisibility. Inertial surveying methods are much faster than the traditional techniques and have the added advantage that stations need not be intervisible. However, a drawback, as far as the surveyor is concerned, is the large bulk and high cost of the equipment. TRANSIT surveying techniques can produce relative positioning accuracies of a few decimetres, which are essentially independent of baseline length. Therefore, closely spaced points have a lower proportional accuracy than points spaced far apart. The equipment is compact, easy to operate and inexpensive compared to inertial technology. However, it lacks the speed of inertial surveying and cannot match the accuracy of EDM for station separations less than a few tens of kilometres. Analytical photogrammetry requires that the points to be surveyed be photographed from the air (or with ground-based cameras) and that ground control points are available. Measurements are made on the photographic images and the coordinates are then determined after considerable computer processing. This technique has many applications in mapping and engineering surveys where a large number of points (often inaccessible) need to be coordinated.

GPS combines the high accuracy of conventional surveying with the convenience of satellite surveying. The likely impact of GPS on surveying and geodesy is discussed in Section 3.5. The main advantages of the GPS technique are:

(a) Millimetre instrumental accuracy plus a variable error of 1 to 2 ppm of the distance between points.
(b) Capable of three-dimensional surveying.
(c) Points do not have to be intervisible, thus obviating the need for trigonometrical beacons -- control station monumentation may even become redundant.
(d) High productivity resulting in low costs.

A significant requirement, and occasional disadvantage, of GPS surveying is that surveyed points must have a relatively unobstructed view of the sky above about 20 degrees from the horizon.

3.2 SURVEY PLANNING

Proper planning of a GPS survey is generally much more critical than for a conventional survey. This is particularly so with the current, limited constellation of satellites since the window of mutual visibility is only 4-6 hours per day. Furthermore, with the present (1985) high cost of GPS equipment ($US50 000-100 000 per unit) surveyors cannot afford to leave the equipment idle while they replan or reobserve a poorly executed survey. Experience has shown that careful planning and thorough reconnaissance is worth the effort.

3.2.1 PRE-SURVEY PLANNING

Prior to conducting a field reconnaissance a certain amount of office planning is essential. First, the points for which coordinates are desired should be plotted at a suitable scale. Next, the horizontal and vertical control points in the vicinity of the points being surveyed should also be plotted. The average length of line to be measured can then be scaled from the map along with the approximate coordinates of the points to be surveyed.

As an aid in reconnaissance, a satellite "sky plot", or visibility chart, should be prepared for each site, showing the tracks of the GPS satellites (their elevation and azimuth as a function of time) during the planned observation session -- see Figure 3.1.

The computation of the rise and set times of the GPS satellites is, of course, necessary for designing the observing schedule. This is particularly important for receivers that can only track, say, 4 satellites, as it is then possible to select beforehand the subset of satellites which should be tracked in order to give the best station-satellite geometry. It also ensures that all receivers being deployed on a survey are programmed to track the same set of satellites.

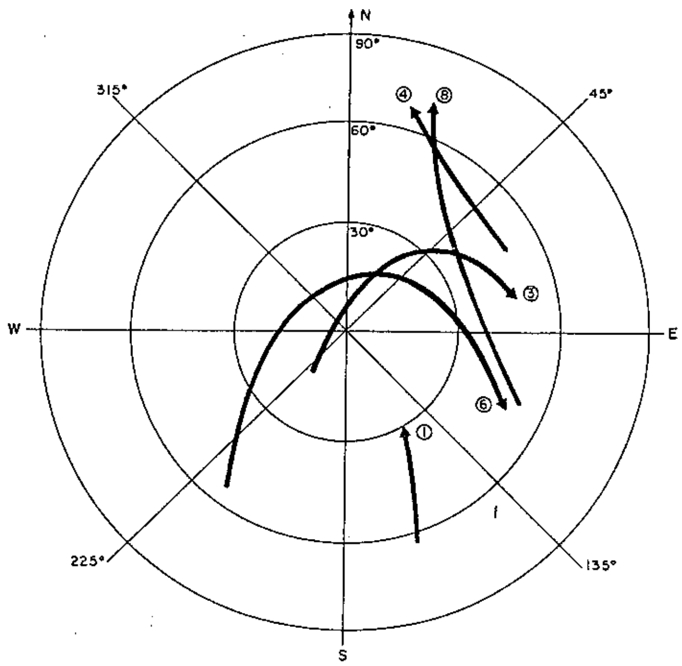
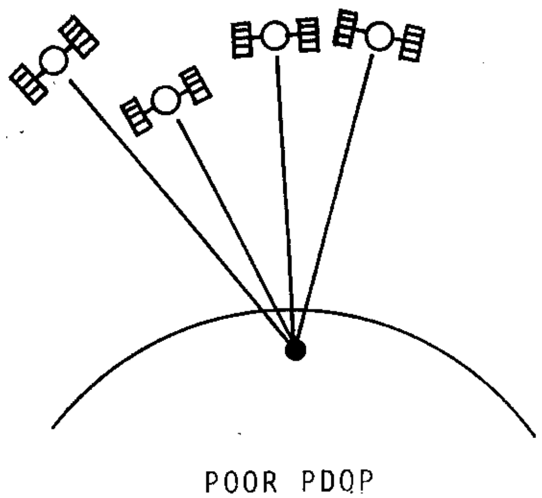


Figure 3.1 Typical Sky Plot of the GPS Satellites at a Site. (Adapted from Bock et al 1985)

Using these selected satellites and the approximate site coordinates, either a full computer simulation is performed, or the average PDOP (position dilution of precision) is calculated (Figure 3.2), to estimate the accuracy of the relative positions. Conversely, the simulation can indicate to the surveyor the length of time to be spent at each point to achieve a given accuracy. From this information, the travel time between points may be obtained. For example, typical MACROMETER observation times would be 2 hours at each of two points and one half hour travel time between points to achieve a few parts in 10^6 accuracy on 10 kilometre lines. The accuracy of the GPS technique is less sensitive than terrestrial techniques to network geometry. However, accuracy can be improved by introducing redundancy and enforcing closure of the polygons formed by chains of baselines. The network can be designed to take these considerations into account and the network precision also tested by computer simulation. This is particularly important when more than two GPS receivers are used simultaneously. If simulation programs are not available, past surveys could provide a good indication of achievable accuracies.



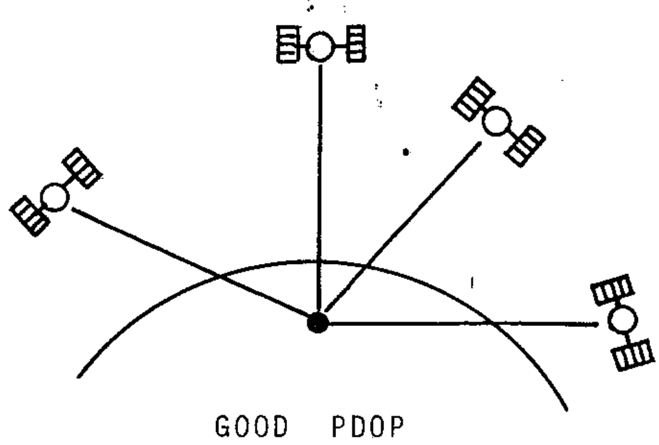


Figure 3.2 Position Dilution of Precision (PDOP).

The end product of pre-survey planning is a detailed observation schedule indicating which sites are to be occupied by which receivers, and in what order.

3.2.2 FIELD RECONNAISSANCE

Using the map and the estimated observation span, the field reconnaissance party can select the specific points and possibly erect the monuments. They should select points that are relatively free of obstructions and, if possible, easily accessible by vehicle. When the area is heavily wooded or there are man-made obstructions, as in urban areas, the reconnaissance party must pick points where the obstructions do not block the satellite signals during the observation session. These considerations are important when, for example, satellites may appear in only one quadrant of the sky, or when their elevations are too low at some time during the observing period. A "sky plot" of the paths of each satellite during the observation period is therefore helpful in determining where obstructions can be tolerated. In heavily wooded or urban areas the surveyor may have to resort to a tall tripod to elevate the GPS antenna to achieve the required visibility.

3.3 SURVEY EXECUTION

The various types of GPS survey equipment were discussed in Chapter 2. The survey procedure will depend to a certain extent upon the equipment used. There are two basic types of GPS receivers: code-correlating and codeless instruments. Presently three GPS instruments have been used for geodetic surveying: the MACROMETER V-1000 and the GPS Land Surveyor 1991, which are both codeless receivers, and the Texas Instruments TI 4100, which is a code-correlating receiver.

A GPS survey consists of three separate operations: pre-observation time setting, observation data recording, and post-observation time checking and data unloading.

3.3.1 PRE-OBSERVATION

The accuracy achievable in a survey depends, in part, on how well the receivers' clocks are synchronised to each other and with an accurate external standard. Codeless instruments require that Coordinated Universal Time (UTC) be independently determined to a few milliseconds to ensure 1 ppm or better accuracy (see Section 5.3.1). This is usually accomplished by synchronising the receiver's clock to UTC using an external time source such as an appropriate GPS time receiver (for example, the Trimble 4000A, which tracks and acquires time information from the GPS satellites) or an atomic clock. Other time sources include short wave radio time-tick transmissions, the domestic television network, communication satellites, etc.

The clock within each codeless receiver needs to be synchronised to the clocks of all other receivers to within about a microsecond, corresponding to about a millimetre position error. This can be done in the following manner. Let us assume that three GPS instruments A, B and C are to be used to perform the survey. The three units are physically collocated at a site (for example, at the base camp) and the clock of unit A is first synchronised to UTC to the nearest several milliseconds. Next the clocks in units B and C are individually synchronised with the clock in unit A. All units are then intercompared, using a time interval counter, to verify that the clocks are synchronised. This procedure is repeated at the start of each day.

With the MACROMETER V-1000, 110 volt (350 watts) of power must be supplied and the main unit turned on to perform this synchronisation. The "A-file" needed to operate the instrument is loaded from cassette at this time. The A-files must be prepared for each site and loaded into each MACROMETER prior to the occupation of each point. The MACROMETER A-file is a file that is prepared on the special P-1000 computer using approximate satellite ephemeris data and the approximate (nearest 10 kilometres) station coordinates. The file contains sixty discrete observation times and the corresponding expected satellite signal frequencies of each of the six satellites to be received. This file allows the MACROMETER's internal processor to sort out each satellite and assign it to the proper receiver channel (one to six). Only the six preselected satellites will be tracked during the observation session. There is little flexibility for changing the observing schedule in the field, or observing a different site.

Since the GPS Land Surveyor Model 1991 does not track individual satellites, but simply records all visible satellite signals, predictions of time, Doppler frequencies, etc. are not required. However, synchronisation of the receiver clock to UTC and to the other receivers is still necessary.

Code-correlating instruments such as the TI 4100 are implicitly tied to GPS time via the pseudo-ranges and, therefore, do not require synchronisation and need not be brought together at the start of each day. After instrument "warm-up", the approximate coordinates of the selected satellites for some epoch and the approximate UTC time are entered manually by the operator. Following this initialisation, the TI 4100 can search and acquire the satellite signals. The coordinates of only one satellite need to be entered, because upon receiving the signal from a GPS satellite, and reading the Navigation Message (Section 7.4.2), the approximate positions of all other GPS satellites are obtained and the instrument can automatically acquire the other selected satellite signals. The specific satellites being tracked can be altered (either manually by the operator or automatically by the receiver) during an observation session to ensure that the best geometry is maintained.

3.3.2 OBSERVATIONS

Following the pre-observation time setting, the codeless GPS receivers are transported to their first observation station. The antenna is placed on a tripod over the point and connected by cable to the receiver. For MACROMETER V-1000 observations, a 600 watt petrol generator is operated, the main unit is plugged into the generator and several function buttons are pushed. The unit then operates automatically. During the observation span certain values are displayed and satellite channel lights operate to indicate that proper observations are being obtained. The observation period for the MACROMETER is typically of the order of one or more hours for few ppm accuracy but can be less for lower accuracy work. Indeed, the accuracy is approximately inversely proportional to the observation span. The MACROMETER is then transported to the next point, a new A-file is loaded and the observation sequence repeated.

For the GPS Land Surveyor, the briefcase sized rubidium clock is connected to the recording unit and the start and stop times of the observation sequence are keyed into the recording unit. A blank (formatted) tape is loaded into the tape drive so that data can be recorded every 15 seconds. The rest of the observation procedure is automatic. Once the pre-observation time setting is carried out at the beginning of the day, the GPS Land Surveyor can be moved to any site without requiring instrument initialisation (such as loading A-files, etc.).

The setup procedure for code-correlating receivers is much the same as for the GPS Land Surveyor. The data is automatically recorded, at some preset sampling rate, onto cassette, cartridge or directly onto the mass storage medium of a field computer. The observation time can vary from less than half hour to several hours. A significant advantage of the TI 4100 receiver is its ability to maintain "lock" on the satellites even as it is being transported from one site to the next. This eliminates the need to re-enter approximate satellite coordinates at each site.

A site log containing a description of the site, weather data and any other information relevant to the observations, is prepared while observations are being made. Loss of measurements may occur because of obstructions to the satellite signals, satellites not transmitting, antenna malfunctions, poor cable connections or receiver malfunction. Well-trained personnel can deal with these nuisance problems, or, at the very least, note their occurrence.

3.3.3 POST-OBSERVATIONS

In the case of the MACROMETER, the satellite data are dumped from the internal bubble memory onto tape for safekeeping immediately following each individual observation session. For the other receivers the cassette is simply removed and labelled. However, before processing can begin, cassettes may need to be copied/decoded onto 9-track tape, or some other storage medium.

At the end of the day's observations all codeless receivers should be brought together so that the clocks can be checked. Checks are carried out by comparing the clocks with a time interval counter. In this manner the drift of each clock with respect to the unit at the other end of the line being measured can be computed. This clock drift value can be used when computing coordinates. This is not necessary for code-correlating receivers.

3.4 POST-SURVEY PROCEDURES

Basically, all GPS surveying instruments measure either the (ambiguous) carrier phase or P code transition phase of each satellite and count the number of integer cycles between each measurement. For example, if one hour of data were taken and the period divided into 60 epochs, the instrument would record the phase and integer cycles since the first measurement received from each satellite. Typically two or more receivers are deployed on a survey and measurements are made (simultaneously) to a number of GPS satellites. In most computation schemes some differencing is carried out and these differences form the "observable" used to obtain the difference in coordinates between the two points (see Section 5.1). In the case of code-correlating instruments the pseudo-ranges (either from P code or S code) are also measured and recorded. These additional data can be helpful in resolving the cycle ambiguity in the more precise carrier phase measurements and in synchronising the receiver clock to GPS time.

The procedures used for processing the observations will depend upon the receiver type, observables used (carrier phase, P code transition phase, pseudo-range, and any combination of these), and type of computer available. However, certain elements are common to all of these procedures: the data must reside on the processing computer (some data transfer/copy and pre-processing may therefore be required); an ephemeris for each satellite must be available for the times of observations; and a software package capable of estimating the site positions to the accuracy required for the survey must be available. A more technical description of how this processing is carried out is given in Section 5.2.3. Here we give only an outline of the basic elements of the solution process.

3.4.1 EPHEMERIDES

To process the observed data, the coordinates of each satellite for the exact time of each observation are required in some reference frame (see Chapter 4). The ephemeris information needs to be accurate enough to satisfy the accuracy specifications for the survey (see Section 1.4).

The problem encountered is storing all these ephemeris values, so in lieu of this, either Keplerian elements (see Section 7.4.2) for the mid-time of observation or X,Y,Z values at few minute intervals (Section 7.4.1) are stored in computer memory. These elements (six for each satellite) or coordinates can be expressed in the earth-centred / earth-fixed, or inertial systems. The first step in using the ephemeris data may therefore be to transform them to the coordinate system in which the station coordinates are expressed (this involves rotations to account for precession, nutation, pole position, and UT1-UTC -- see Chapter 4). The next step is to compute X,Y,Z values for each observation time. This can be accomplished with an orbit generating program or polynomial interpolator.

Since the codeless receivers cannot decipher the Navigation Message, and recover the Broadcast Ephemerides, ephemeris information must be obtained from an independent source. To minimise the delay between time of data collection and time of data reduction (and therefore allow computations to be done in the field), the Broadcast Ephemerides as recorded by a relatively inexpensive GPS navigation receiver (for example, the Trimble 4000A) could be used. Such a scheme is suggested for the GPS Land Surveyor. In the event that the quality of the "Broadcast Ephemeris" is degraded, or where accuracies better than a few parts in 10^6 are sought, post-processed ephemerides may be preferred.

3.4.2 MEASUREMENT PROCESSING

From the approximate values of the site coordinates, the ephemerides for the GPS satellites, and the epochs of the observations (from the input data file), the theoretical values of the observations at each epoch are computed. Subsequently, the observed values are compared to the computed values, and an improved set of site coordinates is obtained using the Least Squares procedures and the modelling equations given in Chapter 5. The time required for the data processing depends on the accuracy required, the software available, and the amount of data lost due to sky-obstructions and equipment malfunctions. If only moderate accuracy is desired and batch processing techniques are used, the computations might take less than a half-hour per point. The actual computer solution only takes a few seconds or minutes, but data organisation and inspection of the results can add significantly to the computation time. If (non-automatic) editing of the data is required, the data should be processed interactively or in multiple batch runs. This could take several hours.

During the first three years of GPS satellite surveying mostly baseline-by-baseline solutions were performed. When several baselines had been surveyed, the relative coordinates obtained from the individual baseline solutions, together with their respective variance-covariance matrices, were then combined into some type of network adjustment. Recently, surveyors have begun using the multi-station approach, in which all lines are processed at the same time. This approach is more rigorous, less time consuming, and perhaps more accurate, and also allows for the possibility of improving the accuracy of the satellite ephemerides.

Usually all data is taken back to the field office where, after the arrival of the post-processed GPS ephemerides or by using the recorded Broadcast Ephemerides, processing can be carried out on a portable computer. Code-correlating receivers automatically record the Broadcast Ephemerides (codeless receivers can use a relatively inexpensive S code navigation receiver for this purpose) and at the end of each day preliminary processing can verify the quality of the observations before the survey party leaves the area. The provision of post-processed ephemerides could take from a few hours to several weeks, depending, to some extent, on the accuracy of ephemerides. The processing of MACROMETER data must be, at least partially, performed on the specially configured P-1000 computer. The GPS ephemerides are usually provided by an **ephemeris service** (to which the MACROMETER user subscribes), in the form of specially formatted files acceptable to the P-1000 software. This is a slight inconvenience as independently obtained GPS ephemerides cannot be used in the P-1000 software. Software now being written to handle GPS data from different types of receivers will diminish the differences between receivers in respect to measurement processing and allow true "multi-instrument" analyses to be performed.

Several researchers (for example, Beutler et al 1985, Nakiboglu et al 1985, Masters & Stolz 1985) have developed GPS processing software that not only computes the receiver coordinates but also improves the accuracy of the satellite orbital positions. This processing of multi-station, multi-instrument data can significantly improve the survey results obtained from less than optimal ephemeris data. The method would permit surveyors to set up their own local tracking network, even using a degraded Broadcast Ephemerides, and obtain high accuracy.

3.4.3 ANALYSIS OF RESULTS

GPS surveys are usually more accurate than the networks to which they are connected. For network densification projects, it may be necessary to tie the GPS work to a minimum of three points at the corners of the area being surveyed. This connection to prior surveys allows the GPS work to be scaled and oriented to the existing control. The transformation from the GPS reference system (for example, WGS84 or some other earth-centred / earth-fixed system) to the local geodetic datum can be performed either by occupying previously surveyed points, or by using published values for the transformation parameters. In the latter case, the process of scaling and orienting of the GPS survey after transformation will partially absorb systematic differences between the GPS positions and the local control. When doing this, it is wise to make direct measurements to these existing points to verify their accuracy. Often large discrepancies are discovered in the published control coordinates, leading to distortions in the GPS survey results after adjustment.

A minimally constrained Least Squares adjustment of the GPS work alone should be performed if possible in order to determine its overall precision and consistency. After performing the minimally constrained adjustment and verifying the precision of the data, a constrained adjustment can be performed to fix the GPS work to the local control (after transforming the coordinates into the same reference system and verifying the quality of the local control).

For "corridor surveys" such as for transmission lines, a simple traverse analysis should suffice. The quality of the GPS survey is primarily reflected by the misclosures at the end of the traverse. Nevertheless, the remarks made above concerning verification of control coordinate data and the transformation parameters are still valid.

3.4.4 GPS ELEVATIONS

The most misunderstood aspect of GPS surveying pertains to the elevations or height information they produce. Chapter 6 gives more detailed information on this subject, but one rather simple approach that has been used quite effectively is to occupy a number of benchmarks surrounding the area being surveyed. In this way, these benchmarks will have two elevations: the published elevation and the GPS "height" (based on holding one benchmark fixed). The difference between the two elevations is the geoid height. The geoid can then be approximated by a surface of "best-fit" (usually a plane) passing through the geoid spot-heights (see Section 6.2.4).

Once the geometry of the geoid surface is established, the geoidal elevations of the remaining points in the area can be determined by interpolation. Experience has shown that this rather simple approach for correcting GPS heights, so that they are equivalent to the usual orthometric elevations, works quite well for areas up to 50 x 50 kilometres. Care should be exercised in mountainous terrain or where large gravity anomalies occur. In this case the procedures described in Chapter 6 should be used.

3.4.5 GPS AZIMUTHS

A major disadvantage of methods other than conventional line-of-sight surveying is that the azimuth at a point is not determined during the survey. The most expedient method of overcoming this disadvantage is to observe an astronomical azimuth at the point at the time GPS observations are being made. If this is impractical or the site occupancy time is too short to permit astronomical observations, a pair of intervisible GPS points can be observed and the second point used as an azimuth mark for the other point in subsequent surveys. This technique has been successfully used on a number of transmission line surveys to define azimuths to an accuracy of a few seconds of arc.

3.5 THE IMPACT OF GPS ON SURVEYING

Although the GPS satellite constellation is only about one third complete it has already been used for a variety of survey applications. The total number of points actually surveyed over the last few years numbers in the thousands. GPS is already competitive with the TRANSIT system and is likely to challenge the traditional surveying techniques in the near future. What then is the likely impact of the GPS technology on the whole range of surveying activities? The following discussion is taken mostly from Rizos et al (1984).

3.5.1 GEOPHYSICAL, ENGINEERING, LARGE SCALE CADASTRAL AND MAPPING SURVEYS

The overwhelming majority of civilian GPS surveys will probably be carried out in support of geophysical prospecting activities, engineering projects, large scale coordination, land parcel surveys, and for map control, to name a few. Indeed, GPS is ideally suited for surveys associated with the establishment of a coordinated cadastre. The relative accuracy requirements are likely to be of the order of 1 part in 10^4 to 1 part in 10^5.

The majority of these prospective users will have had no previous experience with satellite surveying. Therefore, the introduction of GPS into surveying practices will be mainly influenced by economic 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 for many tasks, the surveyor would have to satisfy himself that the GPS positioning technology is superior to the conventional EDM and theodolite procedures.

Although a definitive statement on the relative competitiveness of GPS and EDM-theodolite technologies is not yet possible, some criteria for evaluating competitiveness can be mentioned:

(a) Receiver cost.
(b) Ease of operation.
(c) Productivity.

Receiver cost

The average cost of a good quality integrated EDM and optical theodolite instrument is about $10 000. A "total station" with an automatic data acquisition and recording facility is approximately double the price of EDM-theodolite instruments. 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 between now and the 1990s. Prices of over one hundred thousand dollars have been quoted for first generation GPS receivers. However, instrument prices are falling and it is not unlikely that the cost of a GPS receiver will be comparable with that of a "total station" within a few years. (A pair of receivers would be needed for relative positioning, thus 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, the processing software should be resident in the receiver, or on a portable field computer, and require the minimum of operator intervention. Post-processing, if required, should be possible using microcomputers of the type that can be found in any survey office.

According to the rule-of-thumb given in Section 1.4, a relative positioning accuracy of 1 part in 10^4 implies that an orbit error of the order of 2 km can be tolerated, whilst an accuracy of 1 part in 10^5 is compatible with an orbit error of 200 m. These orbital accuracy requirements are of a very modest nature and easily satisfied by the Broadcast Ephemerides. They can also be satisfied by ephemerides generated using relatively unsophisticated orbit computation software -- allowing a surveyor to precompute ephemerides adequate for his positional needs. For the applications covered by these accuracies, access to the Broadcast Ephemerides is no longer an issue for concern.

The essential characteristics of a GPS receiver suitable for 1 part in 10^4 to 1 part in 10^5 surveying should therefore be low cost, compact size, minimal concern over the availability of satellite ephemerides, rapid field operation and data processing (preferably in the field). Code-correlating receivers have the advantage in that they can automatically access the Broadcast Ephemerides and require far less programming and initialisation prior to survey than do the codeless receivers. A number of such instruments are at various stages of development and expected to be on the market by 1986 (see Chapter 2).

Productivity

As surveyors, we are prone to compare GPS efficiency with that of EDM-theodolite procedures in regards to time spent in the field, acquiring data. The fallacies in such comparisons are illustrated by the following three types of surveys:

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, point positioning with one receiver or relative positioning with respect to a distant master 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 contruction activity or for surveying rural land boundaries. In these cases 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 observed.

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. In such applications inertial surveying technology may have a decided advantage. A vehicle equipped with inertial surveying instrumentation can rapidly "traverse" distances of 10 km or so and its requirement for frequent stops -- for "zero velocity updates" -- lends itself easily to the task of establishing closely spaced control networks. Nevertheless, the instrumentation is at present quite bulky and although positioning accuracies of 20 to 40 cm over distances of 5 to 100 km appear to be achievable (Schwarz 1981), it remains to be seen to what extent the inertial surveying technology is a competitor to GPS.

Clearly, when GPS is forced to operate in a pseudo-traversing mode, EDM-theodolite techniques maintain an edge in competitiveness. However, there probably is a critical station-separation above which GPS would be the more efficient technique. Initially, interstation distances of a few tens of kilometres (less in rugged terrain) are likely to be most efficiently bridged using GPS. As the surveyor becomes more familiar with GPS and receiver costs fall, GPS could be routinely used for traverses of 10 km or less in length. 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 value, as already mentioned.

3.5.2 FIRST ORDER GEODETIC SURVEYS

We can confidently expect GPS to replace TRANSIT for establishing, maintaining and densifying geodetic networks. Over distances of a few hundred kilometres, multi-station TRANSIT gives relative positioning accuracies of a few decimetres (Kouba 1983). Unlike TRANSIT, GPS can give accuracies of 1 ppm even over distances of a few tens of kilometres. Furthermore, GPS is much faster, with measurement times as short as half an hour 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 trig stations in unsurveyed regions can be increased over traditional techniques. The increased flexibility of GPS also permits control stations to be established in easily accessible places rather than being confined to hilltops as has hitherto been the case.

According to the rule-of-thumb already mentioned, the orbital information necessary to support 1 ppm surveying should be accurate to about 20 m. At present the Broadcast Ephemerides appears to be adequate for most surveying applications, although the accuracy is known to degrade occasionally without warning. Since immediate positioning results are not needed for the application considered here, post-processed ephemerides could be used. These could be obtained from an "ephemeris service" such as discussed, in the Australian context, by

Rizos et al (1985).

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 involved in high precision surveys with GPS (see Section 3.5.4). 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? The entire geodetic network may ultimately need to be strengthened and redefined using GPS.

3.5.3 CRUSTAL MOTION 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 300 km 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 generally associated with crustal movement surveys is a few parts in 10^7 or better (that is, a few centimetres in 100 km).

Until about a decade ago, the terrestrial techniques provided the only geodetic measurements on present day crustal movements. Recent measurements using Very Long Baseline Interferometry (VLBI) and Satellite Laser Ranging (SLR) have resulted in accuracies of several centimetres over distances of many thousands of kilometres. Unfortunately, the required equipment and personnel costs discourage 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 300 km in length cannot be justified economically.

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. 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 ultimately be applied to many of the geophysical problems presently studied using ground techniques over 5 to 30 km baselines.

The vertical accuracy, while less than that attainable in the horizontal coordinates, will have a major impact on geodynamic studies. The cost is much less than for precise levelling for distances greater than a few kilometres, and the accuracy will be higher over longer distances or for paths involving large vertical relief. 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 baseline measurements. Dual-frequency receivers capable of measuring and recording phase on the L1 and L2 frequencies are therefore the main candidates 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. One promising approach is to use Water Vapour Radiometers (WVR) to measure the water vapour content of the atmosphere and use this data to correct the GPS measurements (Resch et al 1982).

The main 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, ephemerides accurate to a few metres are required. Only carefully computed post-processed ephemerides would satisfy these accuracy requirements. Abbot et al (1985) have used a 3 station tracking network, equipped with dual-frequency receivers and high-stability atomic clocks, to obtain orbits accurate to a few parts in 10^7 within the region of the network. Stolz et al (1984) have shown that this accuracy or better could be obtained with a 3 to 4 station tracking network well distributed across the Australian region. Within a few years there may be a sufficient number of precise GPS tracking stations established by different countries and agencies, to determine accurate orbits over much of the globe. Studies indicate that very good results can be expected with as few as six worldwide tracking stations (Melbourne & Thornton 1984).

3.5.4 TOWARDS STANDARDS AND SPECIFICATIONS FOR GPS SURVEYS

What has been attempted above is the classification of potential GPS users into broad categories, ranging from those satisfied with 1 part in 10^4 accuracies to those engaged in ultra-high precision crustal motion surveys requiring accuracies of 1 part in 10^7 or better. However, this scheme is not sufficiently detailed and in many ways too arbitrary to form the basis for developing the standards and specifications to which future GPS surveys must adhere. In particuler, no attempt having been made to relate the categories to those of existing survey standards and practices. A first attempt at drafting a comprehensive set of standards and specifications for GPS surveying has been made by Hothem & Williams (1985a, 1985b) of the U.S. National Geodetic Survey (NGS).

One of the issues that needs to be tackled is that in many countries the first-order standard for geodetic surveys is an accuracy of 1 part in 10^5. However, GPS relative accuracies are typically better than this standard. Furthermore, conventional standards for (relative) vertical accuracies refer to orthometric elevations and do not include criteria for the ellipsoidal height differences yielded by GPS, or the geoid height differences required to relate these to orthometric elevations.

Since the present survey standards are not adequate for classifying high precision relative positioning surveys, Hothem & Williams (1985a) have proposed a two-tier system of classification:

- a general classification for general purpose surveys, incorporating the present classifications for geodetic control, engineering, mapping and cadastral surveys. The accuracy standards are based on minimum distance, azimuth, geoid height and ellipsoidal height accuracies. The results of general classification surveys are based on a constrained adjustment of GPS stations tied to the local network control.

- a special classification for special high precision geodetic surveys for, for example, crustal motion surveys, land subsidence monitoring and precise engineering surveys. Special classification surveys are adjusted independently of local network control and are therefore based only on the internal three-dimensional consistency of the GPS network.

Table 3.1, adapted from Hothem & Williams (1985b), summarises the proposed NGS accuracy standards for the general and special survey classifications. Once the classification scheme, with the associated accuracy standards, has been established, specifications can be drafted covering such aspects of GPS surveys as:

(a) **Network Design and Geometry.** This includes such factors as minimum and maximum station spacing, minimum number of connections to existing horizontal and vertical control, minimum spacing between azimuth and station marks, etc.

(b) **Instrumentation and Calibration.** Including advice on 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.

(c) **Field Procedures.** Specifying such things as the number of receivers to observe simultaneously, length of observation span, number of satellites to be observed simultaneously, cutoff elevation angle, measurement of meteorological parameters, etc.

(d) **Office Procedures.** Specifying the number of triangle and loop closures, model for constrained Least Squares adjustment, observation weights, error analysis tests, data editing techniques, etc.

The reader is referred to Hothem & Williams (1985b) for the preliminary NGS specifications that address the abovementioned factors. The NGS standards and specifications are likely to be adopted, in part or as a whole, by other national geodetic agencies.

Table 3.1 General and Special Classification Accuracy Standards for GPS Surveys (proposed NGS scheme).

Survey Categories	Classification		Minimum Accuracy			
	Order	Class	Distance	Azimuth (sec)	Geoid ht. diff. (cm)	Ellip. ht. diff. (cm)
SPECIAL						
Global Dynamics	1	I	$1:10^8$	NOT	APPLICABLE	
Regional Dynamics	1	I	$1:10^8$	"	"	
Primary Geodetic Network	2	II	$1:10^7$	"	"	
Local Dynamics	2	II	$1:10^7$	"	"	
Network Connections, Densification	3	I	$1:10^6$	"	"	
High Precision Engineering Surveys	3	II	1:500 000	"	"	
GENERAL						
Secondary Geodetic Network	1	I	1:300 000	1	2+d (1 ppm)*	1+d (1 ppm)
Cadastral Survey (Urban)	1	II	$1:10^5$	2	"	"
Large Scale Engineering and Mapping Projects	1	III	$1:10^5$	2	"	"
Cadastral Surveys, Mapping	2	I	1:50 000	4	3+d (1 ppm)	2+d (1 ppm)
" " "	2	II	1:20 000	6	"	"
Geophysical Exploration	3	I	$1:10^4$	10	4+d (1 ppm)	"
" " "	3	II	1:5000	10	"	"

* d (1 ppm) is one millionth of the GPS station separation in kilometres, expressed in centimetres, e.g. d = 10 km, then d (1 ppm) = 1 cm.

3.5.5 SUMMARY

No one can foresee all the effects of GPS, on the entire range of surveying tasks. However there is no doubt that GPS will replace the TRANSIT satellite surveying system for geodetic surveys. Furthermore, for a considerable portion of densification of primary control, GPS will probably displace conventional traversing and trilateration with EDM and theodolite.

For high precision crustal motion surveys GPS provides a much needed technique to bridge the gap between terrestrial methods and VLBI-SLR. Traditional EDM-theodolite procedures will probably be retained for distances less than 10 km or so, while the more established space techniques of VLBI and SLR are better suited for measurement of distances greater than a few hundred kilometres.

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 about 10 km in rugged terrain or less than a few tens of kilometres in flat terrain. For longer distances GPS is a potential competitor. However this is more likely to occur if the cost of GPS receivers falls to a level comparable with "total station" instrumentation.

4. TIME, EARTH ROTATION, REFERENCE SYSTEMS AND GEODETIC DATUMS

Two reference coordinate systems are required in satellite surveying, one rigidly attached to the earth (earth-fixed or terrestrial) and the other fixed with respect to space (space-fixed or celestial). In order to determine positions in the earth-fixed coordinate system from GPS observations we need to know the positions of the satellites in the terrestrial system. However, satellite ephemerides are usually expressed in space-fixed coordinates because the equations of motion of the satellites are formulated and solved in the celestial system since it more closely approximates an inertial frame of reference (see Chapter 7). Therefore, the transformation relationship between the earth-fixed and space-fixed coordinate systems needs to be known. A further problem arises from the fact that the satellite datums are global in nature, whereas traditional terrestrial geodetic networks are defined from observations taken over only small regions of the earth's surface. Thus the correct relationship between the global satellite and local geodetic network needs to be established. We consider these aspects below.

4.1 DEFINITIONS OF TIME

Three different systems of time are used in satellite surveying: dynamical time, atomic time, and 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. When we generate ephemerides for a GPS satellite, we implicitly use dynamical time. Atomic Time is time kept by atomic clocks. It is the basis of a uniform time scale on the earth. (A time scale is defined by the period, or its inverse the frequency, of the basic oscillation of the frequency-determining element, which is measured, and the origin of the time scale, which is defined and agreed upon by international convention.) Sidereal Time is measured by the earth's rotation about its axis. Although sidereal time was once used as a measure of time, through astronomical observations, it is much too irregular by today's standards and therefore should not be thought of as a measure of time at all. Rather, it is a measure of the angular position of a site on the earth with respect to a space-fixed reference frame (though in keeping with traditional practice, its units are seconds of time rather than seconds of arc). Within each of these broad categories, there are specific measures of time used in satellite surveying -- Figure 4.1 illustrates the relationship between the various time scales in common use. In the remainder of this section each of these measures is discussed, together with how they are obtained in practice and when they are used.

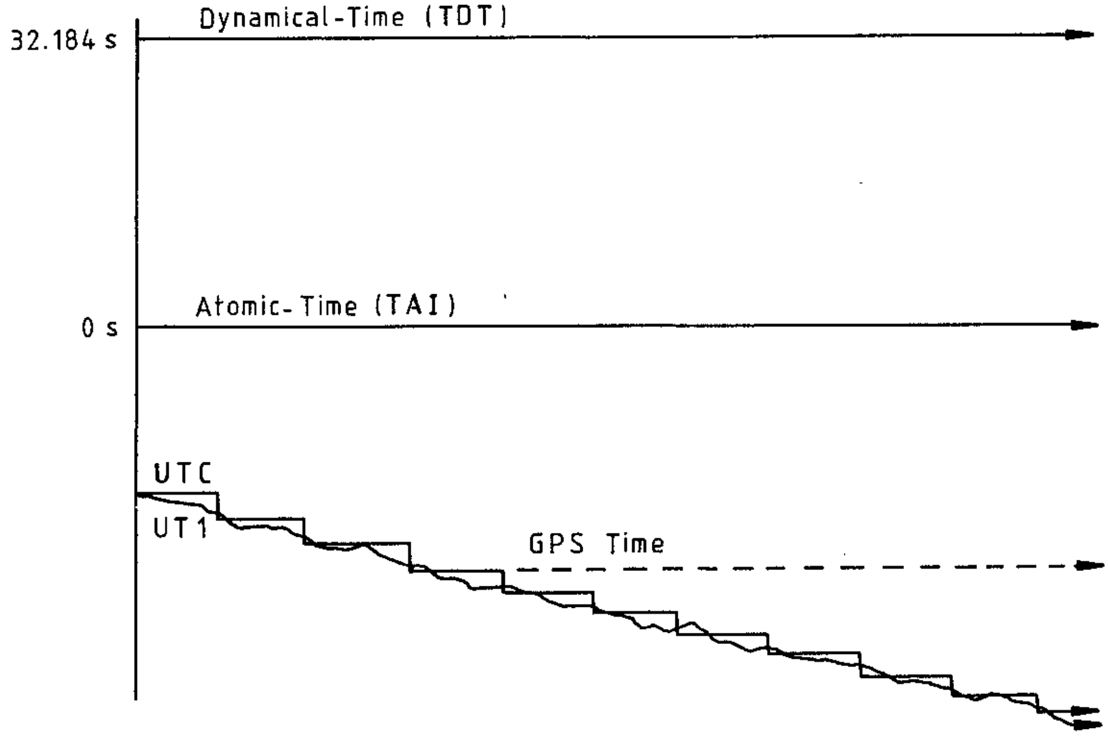


Figure 4.1 Time Standard Relationships.
(Adapted from Conley 1984)

4.1.1 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 (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 nearly the same perturbations.

For satellite orbit computations, we can 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).

The predecessor of TDB was known as Ephemeris Time (ET). In the terminology of General Relativity, TDB corresponds to Coordinate Time, and TDT to Proper Time.

4.1.2 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 de l'Heure (BIH) in Paris of data from atomic standards of many countries. It is a continuous time scale and serves as the practical definition of TDT, being related to it by:

$$TDT = TAI + 32.184 \text{ s} \quad (4.1)$$

The fundamental unit of TAI (and therefore TDT) is the SI second, defined as "the duration of 9 192 631 770 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 86 400 seconds and the Julian Century as 36 525 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 ("leap seconds") when necessary, normally at the end of June or December of each year. Each of the world's time centres keeps a local realisation of UTC, the epoch and rate of which relative to UTC(BIH) are monitored and corrected periodically. UTC(AUS) is maintained by the Division of National Mapping in cooperation with most of Australia's time and frequency laboratories. Its relationship to UTC(BIH) is known to about 5 microseconds. UTC may be determined by tuning in to radio time signals such as VNG, which have an uncertainty of about 1 millisecond due to changes in the propagation delay through reflection caused by the ionosphere.

The time signals broadcast by the GPS satellites are synchronised with atomic clocks at the GPS Master Control Station, soon to be located at the Consolidated Space Operations Center (CSOC), Colorado Springs, Colorado (circa May 1986). These clocks are in turn synchronised periodically with UTC. GPST was set to UTC at 0hr on 6 January 1980 and is not incremented by leap seconds. Therefore, there will be integer-second differences between the two. At the time of writing (November 1985) GPST = UTC + 4 seconds. There is a constant offset of 19 seconds between the GPST and TAI time scales, that is, at any instant:

$$GPST + 19s = TAI \quad (4.2)$$

4.1.3 SIDEREAL TIME

Sidereal time is a measure of the angle between a particular meridian of longitude and a point fixed in space (loosely speaking, the intersection of the earth's equator and the plane of its orbit -- the vernal equinox). The most common form of sidereal time is Universal Time (UT1) (not to be confused with UTC, which is a form of atomic time). The precise definition of sidereal time or UT1 is complicated because of the motion both of the equator and the earth's orbital plane with respect to inertial space, and because of the irregularity of the earth's rotational motion itself. We shall elaborate on the steps by which the rotation angle is computed after discussing reference systems in the following section.

4.2 CELESTIAL AND TERRESTRIAL REFERENCE SYSTEMS

The modelling of the GPS observable and the estimation of station positions depend upon the precise definitions of the coordinate systems in which the observations are made and subsequently analysed. It is the objective of this section to present the mathematical expressions which are used to relate the celestial (or inertial) system in which the satellite's motion is expressed to the terrestrial system in which the observing site's position is defined. We describe the earth's motion in space, and define the conventional celestial and conventional terrestrial reference systems used, before giving the formulae by which these systems are related in practice.

4.2.1 THE EARTH'S MOTION IN SPACE

The earth's pole of rotation is not fixed in space, but "precesses" and "nutates" due principally to the torques exerted by the gravitational fields of the moon and sun on the earth's equatorial bulge. Precession is the slow circular motion of the pole with respect to inertial space with a period of about 26 000 years. Nutation is a more rapid motion, superimposed on the precession, and comprised of oscillations ranging in period from 14 days to 18.6 years. Both are described by the motion of the instantaneous equator and equinox with respect to the fixed equator and equinox of a given epoch (date) -- see Figure 4.2. Precession changes the celestial longitude of a point on the earth's surface by about 50 arcsec per year. Nutation affects both celestial longitude and celestial latitude and has a maximum amplitude of about 20 arcsec.

When the effects of nutation are removed, the resulting fictitious equator and equinox are called the mean equator and equinox, and their positions at any instant are called the mean equator and equinox-of-date. The Conventional Celestial Reference System (CCRS) is defined by the equator and equinox at a particular date or epoch -- see Figure 4.3. By resolution of the International Association of Geodesy (IAG) and the International Astronomical Union (IAU), the CCRS used after 1 January 1984 is defined by the equator and equinox at 12h TDB on 1 January 2000 (Julian Date 2451545.0), designated J2000. (The Julian Day number is the number of the day in a consecutive count beginning so far back in time that every date in the historical era can be included. This uninterrupted series of days began at Greenwich mean noon on 1 January 4713 BC.) Applying nutation to the mean equator and equinox of a given date yields the true equator and equinox of that date.

Two additional transformations are necessary to relate the CCRS to an earth-fixed system. One is a rotation about the true pole-of-date, through the angle between the true equinox-of-date and the adopted point of zero longitude on the earth, nominally the Greenwich meridian. The other is a transformation between the true pole-of-date and the z-axis of the earth-fixed system, by accounting for polar motion. (Polar motion consists of a slow shift of the earth figure relative to its axis of rotation. Plots of this motion show a secular trend, a strong periodic component known as the "Chandler term", with a period of approximately 14 months, and a component with an annual period.) The

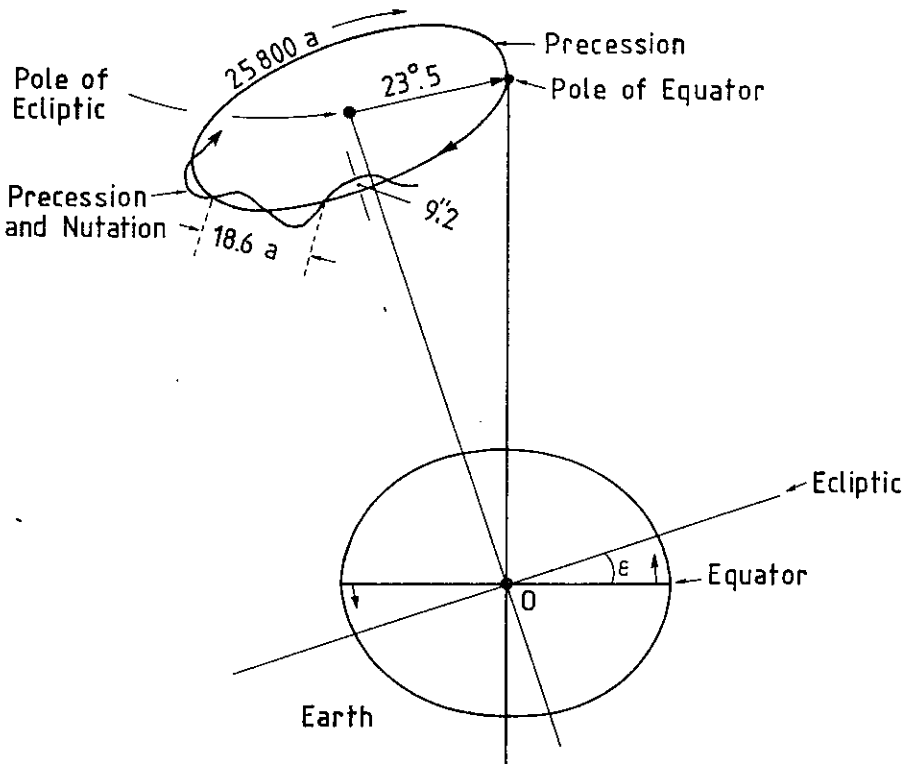


Figure 4.2 Precession and Nutation.
(Adapted from Torge 1980)

relationships used for these transformations (precession, nutation, rotation, and polar motion) effectively define the CCRS by describing its relationship to the terrestrial system. The practical realisation of the CCRS is a little more complicated, involving the satellite ephemerides, empirical measurements of the earth's rotation (UT1) and polar motion, and the definition of the terrestrial system itself. This is discussed in Section 4.2.3.

4.2.2 PRECESSION, NUTATION, POLAR MOTION AND SIDEREAL TIME

Precession

The transformation of coordinates from the mean equator and equinox-of-date at epoch t_i to the mean equator and equinox-of date of epoch t_j is accomplished via the precession matrix P, whose elements are functions of the three angles ζ_A, θ_A, and z_A, defined by the IAU(1976) system of constants (Kaplan 1981):

$$\zeta_A = (2306''.2181 + 1''.39656T - 0''.000139T^2)t$$
$$+ (0''.30188 - 0''.000344T)t^2 + 0''.017998t^3$$

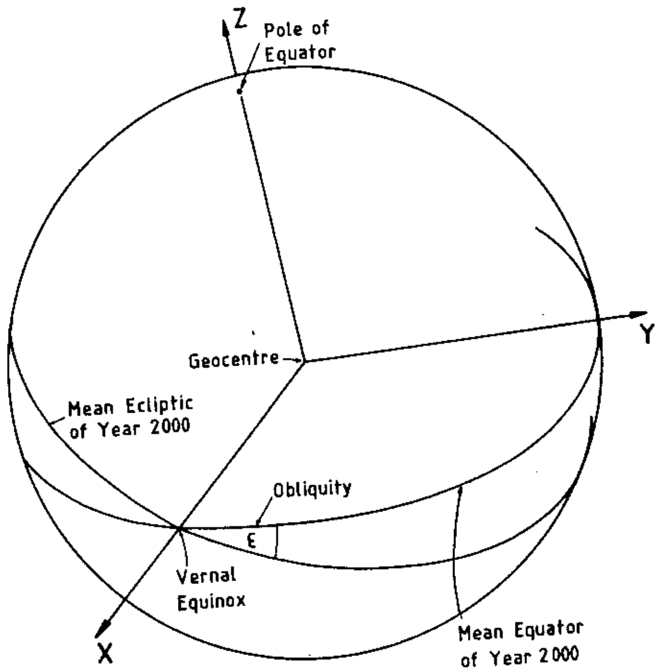


Figure 4.3 The Conventional Celestial Reference System (CCRS).

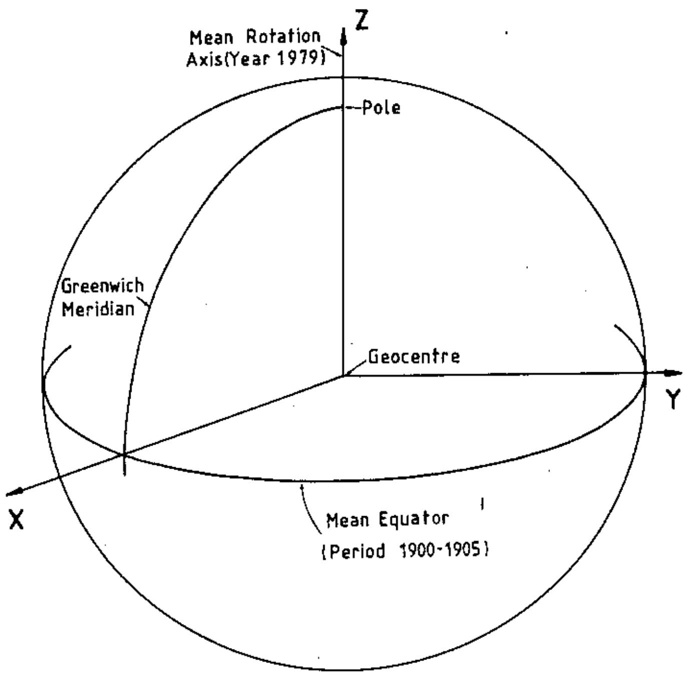


Figure 4.4 The Conventional Terrestrial Reference System (CTRS).

$$z_A = (2306''.2181 + 1''.39656T - 0''.000139T^2)t + (1''.09468 + 0''.000066T)t^2 + 0''.018203t^3$$

$$\theta_A = (2004''.3109 - 0''.85330T - 0''.000217T^2)t - (0''.42665 + 0''.000217T)t^2 - 0''.041833t^3 \quad (4.3)$$

where T is the interval measured in Julian centuries of TDB, between J2000 and epoch t_i, and t is the interval measured in Julian centuries of TDB between epochs t_i and t_j. The transformation from mean-of-date coordinates at one epoch, ρ, to mean-of-date coordinates at another epoch, ρ', is given by:

$$\rho' = P\,\rho \quad (4.4)$$

where

$$P = R_3(-z_A)\,R_2(\theta_A)\,R_3(-\zeta_A) \quad (4.5)$$

and R_i is the matrix effecting a right-handed rotation about the i axis (see, for example, Mueller 1977).

Nutation

The transformation of coordinates from the mean equator and equinox-of-date to the true equator and equinox-of-date (at the same epoch) is accomplished via the nutation matrix **N**, whose elements are functions of the <u>mean obliquity of date</u> ε, the <u>nutation in longitude</u> $\Delta\psi$ and the <u>nutation in obliquity</u> $\Delta\varepsilon$. The mean obliquity is:

$$\varepsilon = (84381''.448 - 46''.8150T - 0''.00059T^2 + 0''.001813T^3) + (-46''.8150 - 0''.00117T + 0''.005439T^2)t + (-0''.00059 + 0''.005439T)t^2 + 0''.001813t^3 \quad (4.6)$$

Expressions for $\Delta\psi$ and $\Delta\varepsilon$, using the IAU(1976) system of constants and the 1980 IAU Theory of Nutation, are given in Kaplan (1981) -- see Table 4.1 for an extract of the nutation constants. The transformation from mean-of-date coordinates, ρ', to true-of-date coordinates, ρ'', is given by:

$$\rho'' = N\,\rho' \quad (4.7)$$

where

$$N = R_1(-\varepsilon-\Delta\varepsilon)\,R_3(-\Delta\psi)\,R_1(\varepsilon) \quad (4.8)$$

Earth Rotation

The angle of the earth's rotation with respect to the true equinox of date is called <u>Greenwich Apparent Sidereal Time</u> (GAST) and is usually assigned the symbol θ. The computation of GAST involves a mixture of conventional expressions, to account for nutation and the mean rate of

Table 4.1 The 1980 IAU Theory of Nutation.

Series for nutation in longitude $\Delta\Psi$ and obliquity $\Delta\epsilon$, referred to the mean equator and equinox of date, with T measured in Julian centuries from epoch J2000.0.

	ARGUMENT l l' F D Ω	PERIOD (DAYS)	LONGITUDE (.0001")		OBLIQUITY (.0001")	
1	0 0 0 0 1	6798.4	−171996	−174.2T	92025	8.9T
2	0 0 0 0 2	3399.2	2062	.2T	−895	.5T
3	-2 0 2 0 1	1305.5	46	0.0T	−24	0.0T
4	2 0 -2 0 0	1095.2	11	0.0T	0	0.0T
5	-2 0 2 0 2	1615.7	−3	0.0T	1	0.0T
6	1 -1 0 -1 0	3232.9	−3	0.0T	0	0.0T
7	0 -2 2 -2 1	6786.3	−2	0.0T	1	0.0T
8	2 0 -2 0 1	943.2	1	0.0T	0	0.0T
9	0 0 2 -2 2	182.6	−13187	−1.6T	5736	−3.1T
10	0 1 0 0 0	365.3	1426	−3.4T	54	−.1T
11	0 1 2 -2 2	121.7	− 517	1.2T	224	−.6T
12	0 -1 2 -2 2	365.2	217	−.5T	−95	.3T
13	0 0 2 -2 1	177.8	129	.1T	−70	0.0T
14	2 0 0 -2 0	205.9	48	0.0T	1	0.0T
15	0 0 2 -2 0	173.3	−22	0.0T	0	0.0T
16	0 2 0 0 0	182.6	17	−.1T	0	0.0T
17	0 1 0 0 1	386.0	−15	0.0T	9	0.0T
18	0 2 2 -2 2	91.3	−16	0.1T	7	0.0T
19	0 -1 0 0 1	346.6	−12	0.0T	6	0.0T
20	-2 0 0 2 1	199.8	−6	0.0T	3	0.0T
⋮		⋮	⋮	⋮	⋮	⋮
96	-1 -1 0 2 1	35.0	1	0.0T	0	0.0T
97	0 0 -2 0 1	13.6	−1	0.0T	0	0.0T
98	0 0 2 -1 2	25.4	−1	0.0T	0	0.0T
99	0 1 0 2 0	14.2	−1	0.0T	0	0.0T
100	1 0 -2 -2 0	9.5	−1	0.0T	0	0.0T
101	0 -1 2 0 1	14.2	−1	0.0T	0	0.0T
102	1 1 0 -2 1	34.7	−1	0.0T	0	0.0T
103	1 0 -2 2 0	32.8	−1	0.0T	0	0.0T
104	2 0 0 2 0	7.1	1	0.0T	0	0.0T
105	0 0 2 4 2	4.8	−1	0.0T	0	0.0T
106	0 1 0 1 0	27.3	1	0.0T	0	0.0T

the earth's rotation (including the effect of precession) and an empirical term to account for the small irregular variations. The mean rotation is defined by Greenwich Mean Sidereal Time (GMST) at midnight on the day of interest, $\bar{\theta}_o$, and its rate of change, $d\bar{\theta}/dt$. Conventional formulae for these values are given in the IAU(1976) system of constants (Kaplan 1981):

$$\bar{\theta}_o = 6^h 41^m 50^s.5481 + 8640184^s.812866T + 0^s.093104T^3 - 6^s.2 \times 10^{-6}T^3$$

$$\frac{d\bar{\theta}}{dt} = 1.002737909350795 + 5.9006 \times 10^{-11}T - 5.9 \times 10^{-5}T^2 \qquad (4.9)$$

GMST is given by:

$$\theta_o = \bar{\theta}_o + (d\bar{\theta}/dt)UT1 \qquad (4.10)$$

where UT1 = UTC + (UT1-UTC) incorporates the irregular variations of earth rotation which are known only from actual measurement, using conventional astrometric techniques (for example, photographic zenith tubes) or, more accurately, with radio interferometric observations of quasars or laser ranging to artificial satellites and the moon. UT1-UTC is computed, tabulated, and distributed by national and international time services such as the BIH or the U.S. Naval Observatory -- see Table 4.2. The last step in computing GAST is to correct GMST for the nutation of the equinox:

$$\theta = \theta_o + \Delta\psi\cos\varepsilon \tag{4.11}$$

Polar Motion

The final transformation from the celestial to the terrestrial system accounts for polar motion. The BIH and other organisations use data from conventional astronomic and space techniques to compute, as a function of time, the angles x and y. These angles define the position of the pole-of-date with respect to the mean pole of 1903, which is called the Conventional International Origin (CIO) -- see Table 4.2.

We can now write the complete transformation between the CCRS and what we shall call the Conventional Terrestrial Reference System (CTRS) -- see Figure 4.4:

$$r = R_2(-x)\ R_1(-y)\ R_3(\theta)\ N\ P\ \rho \tag{4.12}$$

where r gives the (earth-fixed) CTRS coordinates and ρ are the (space-fixed) CCRS coordinates. The computational procedure for transforming from the CCRS to the CTRS is described in Section 5.1.6.

It must be emphasised that the Conventional Terrestrial Reference System as defined here is an earth-fixed coordinate system oriented to the BIH zero meridian and the CIO pole. The CTRS label should not be used for satellite reference systems which are not explicitly related to the BIH/CIO system. For example, the reference system to which TRANSIT positions are tied is rotated in longitude by 0."80 to the east of the CTRS.

4.2.3. PRACTICAL REALISATION OF GPS REFERENCE SYSTEMS

In Section 4.2.1 the concepts of the conventional terrestrial and conventional celestial reference systems for GPS surveying were described. In practice the situation is more complicated.

Conventional Terrestrial Reference System

The CTRS is defined by a primary (global) network of tracking stations. A survey can be related to the CTRS by occupying secondary sites whose positions are known with respect to the primary network. There will be uncertainties associated with the relative coordinates of sites in both the primary and secondary networks, and with the orientation of these networks with respect to the Greenwich meridian. With some care however, a CTRS may be defined with uncertainties at the level of 10 cm or less in most areas of the world. Such a system is possible because of the large number of observations made over the past 5-10 years by

Table 4.2 BIH Evaluation of the Coordinates of the Pole and Universal Time.

EARTH'S ROTATION

1 - COORDINATES OF THE POLE AND UNIVERSAL TIME (BIH evaluation)
Units : arc second for x,y , second for UT1.

Date (0h UTC)			smoothed values				raw values		
1985		MJD	x	y	UT1R-UTC	UT1R-TAI	x	y	UT1-UTC
Final Circular D values									
JUN	3	46219	-0.1221	0.4364	-0.41607	-22.41607	-0.1168	0.4363	-0.41769
JUN	8	46224	-0.1089	0.4468	-0.42360	-22.42360	-0.1059	0.4468	-0.42346
JUN	13	46229	-0.0954	0.4565	-0.43033	-22.43033	-0.0971	0.4549	-0.43090
JUN	18	46234	-0.0816	0.4651	-0.43645	-22.43645	-0.0863	0.4649	-0.43900
JUN	23	46239	-0.0673	0.4727	-0.44198	-22.44198	-0.0709	0.4729	-0.44022
JUN	28	46244	-0.0526	0.4791	-0.44703	-22.44703	-0.0547	0.4780	-0.44857
JUL	3	46249	-0.0375	0.4845	0.54839(1)	-22.45161	-0.0393	0.4861	0.54819
The following values are preliminary (to be updated in next issue of Circular D)									
JUL	8	46254	-0.0220	0.4889	0.54414	-22.45586	-0.0262	0.4860	0.54346
JUL	13	46259	-0.0059	0.4922	0.53987	-22.46013	-0.0047	0.4918	0.53935
JUL	18	46264	0.0105	0.4944	0.53560	-22.46440	0.0120	0.4942	0.53966
JUL	23	46269	0.0284	0.4956	0.53170	-22.46830	0.0276	0.4958	0.53687
JUL	28	46274	0.0478	0.4958	0.52790	-22.47210	0.0444	0.4969	0.53266

Note. In UT1R, the effects of zonal tides with periods shorter than 35 days are removed; UT1R-UT1 (smaller than 0.0025s in absolute value) should be added after quadratic interpolation of UT1R. Section 3 of this Circular gives the daily interpolation of x,y and UT1.

very accurate space techniques; in particular, very long baseline interferometry (VLBI) observations of extragalactic radio sources and laser ranging observations to the LAGEOS satellite and the moon.

From satellite laser ranging (SLR), the coordinates of the National Mapping Laser Ranging Station at Orroral Valley (ACT) and the NASA Laser Ranging Station at Yarragadee (WA) are known with respect to each other and to laser ranging stations in North America and Europe with an uncertainty of about 5 cm (Stolz & Masters 1983). From VLBI observations, the coordinates of the NASA Deep Space Network Tracking Station at Tidbinbilla (ACT) and the Parkes radio astronomy telescope (NSW) are also known with respect to each other and similar radio antennas in North America and Europe with an uncertainty of about 5 cm (Harvey et al 1983). Moreover, observations have been carried out to tie together the SLR and VLBI reference systems, and each with other systems such as that used in TRANSIT satellite observations. Thus by locating GPS receivers at the SLR and VLBI tracking stations, a network of sites within Australia could be tied into the CTS.

Conventional Celestial Reference System

The realisation of the CCRS for GPS is more difficult and prone to greater uncertainties. Some of the space geodetic techniques refer to a nearly inertial system: lunar laser ranging because the moon's motion is tied to the solar system barycentre through the gravitational attraction of the sun; and VLBI because the observed radio stars are so distant from the earth that they have negligible proper motion. However, the non-gravitational forces (for example, solar radiation pressure, unbalanced attitude control thrusts, and drag effects) perturb the orbits these satellites so that they provide a poor realisation of an inertial system. In GPS surveying we maintain access to the CCRS only by frequently regenerating the satellite orbits using the best possible models for the forces and by tracking them from stations whose coordinates are known more precisely from other space techniques.

UT1 and Polar Motion

Another potentially weak link in tying GPS measurements to the CCRS and CTRS is the transformation expressions used to relate the two systems. If conventional expressions are used to relate the one system to the other, as is the case for precession and nutation, and these are used consistently, no significant error or uncertainty is introduced into the reduction of GPS observations. However, two parts of the transformation are empirical: UT1 and pole position. If incorrect values of these quantities are used to determine orbits of the GPS satellites (that is, for modelling the relationship between the earth-fixed tracking station coordinates and the space-fixed satellite coordinates), the celestial system defined by these orbits will be in error by the same amount. This error is also present when transforming accurate GPS ephemerides into an earth-fixed reference system using these UT1 and pole position values (Equation 4.12). (The BIH UT1 and pole position values would be "incorrect" if the earth-fixed reference system is not the CTRS as defined above, as, for example, in the case of TRANSIT satellite datum(s) -- see discussion in Section 4.3.)

Fortunately, "rapid service" values of UT1 and pole position with uncertainties less than 0.02 arcsec are available from the BIH and other organisations within 10 days of the time when the observations were made, and predicted values with uncertainties less than 0.05 arcsec are available by the time a survey is being performed. A given angular error in earth rotation will cause an equal angular error in the orientation of the surveyed baseline. An error of 0.05 arcsec produces a baseline error of less than 3 parts in 10^7, so the accuracy of UT1 and pole values is not a problem for most applications. The surveyor should take some care, however, that he doesn't inadvertently use predicted values instead of "observed" values, especially when they are more than a few weeks old.

4.3 RELATING THE CONVENTIONAL TERRESTRIAL SYSTEM TO GEODETIC DATUMS

Local Datums

Conventional surveying is confined to the earth's surface. Classical geodetic computations are performed on a mathematical figure which is chosen to fit specific regions. This procedure requires an ellipsoid which gives the "best fit" (geometrically) to the geoid for the region. A geodetic datum is usually defined by the semi-major axis (a) and flattening (f) of the selected ellipsoid, the geodetic coordinates of the origin or fundamental station and the geoid separation at the fundamental station. A geodetic datum therefore has no specific, predefined relationship with the geocentre or the earth's rotation pole; put another way, its definition is quite arbitrary while its selection is subject only to convenience. The steps involved in defining a geodetic datum using conventional ground-based surveys and astronomically determined azimuths are outlined in, for example, Torge (1980) and Vanicek & Krakiwsky (1982).

Local geodetic datums have little relevance outside the local region for which they are defined. Over the years, scores of such datums have been established, for regions as small as an island to as large as a continent. For example, the Australian Geodetic Datum 1984 (AGD84) is defined by an ellipsoid with semi-major axis 6378160 m and flattening 1/298.25, and the coordinates of the fundamental station -- Johnston origin -- (Allman & Veenstra 1984):

latitude	=	25° 56′ 54."5515 S
longitude	=	133° 12′ 30."0771 E
height	=	571.2 m above ellipsoid
	=	566.3 m above geoid.

Satellite Datums

Satellite datums, on the other hand, are global in nature and defined by:

(a) physical (or dynamic) models such as the adopted gravity field model of the earth, models for the other satellite perturbing forces (see Chapter 7), and fundamental constants such as GM (Gravitational constant times the earth's mass), rotation rate of the earth, velocity of light, etc.

(b) 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.

In essence, the satellite datum is defined by the above mentioned models and maintained by the satellite ephemerides (the coordinates of the orbiting "trig stations") expressed in an earth-fixed reference system. In the following discussion, the expressions "satellite datum" and "earth-fixed reference system" are interchangeable.

Surveyors are most familiar with the satellite datums of the TRANSIT system. Some confusion has arisen from the (incorrect) labelling of the TRANSIT satellite datum as the WGS72 (World Geodetic System 1972) system. In fact, there are a number of satellite datums associated with TRANSIT, reflecting the different combinations of the gravity field models (and associated geodetic constants), earth rotation models and tracking station coordinates that have been used. Each datum may differ from the CTRS in orientation, in the location of the origin (although nominally at the geocentre) and in scale. The TRANSIT Broadcast Ephemeris uses the WGS72 gravity field model and the NWL 10D station coordinate set. The satellite datum is known as the NWL 10D system. The Precise Ephemeris (post-processed ephemeris) datum has changed with improvements in the gravity field model and the tracking station coordinates. Most recently, the NSWC 10E-1 gravity field model and the NSWC 9Z-2 station coordinate set has been employed to generate the Precise Ephemeris. This constitutes the NSWC 9Z-2 satellite datum. Even if identical tracking station coordinates and the same dynamic and earth rotation models were used to derive the Broadcast and Precise Ephemeris, the datums would still not be the same because of the different number and distribution of tracking stations (each with coordinate errors) that are used for the production of each ephemeris. That is, the datums are defined to be the same, but their practical realisation leads to them being distinct from each other.

During the initial deployment and testing phase of GPS, the Broadcast Ephemerides and post-processed ephemerides computed by the U.S. Department of Defense (see Section 7.4) have been given nominally in the WGS72 system. Beginning in October 1985, both sets of ephemerides are computed in the new World Geodetic System 1984 (WGS84). In order to maintain compatability with user navigation equipment now in the field, the Broadcast Ephemerides may continue to be given in the WGS72 system for some time to come. Documentation of WGS84 will be available from the U.S. Defense Mapping Agency in about March 1986. The new system incorporates an improved gravity field model, an origin much closer to the geocentre, and an improved set of station coordinates. WGS84 will define a global reference system the accuracy and consistency of which will be more than adequate for most surveying applications. However, the precise transformation relationship between WGS84 and a particular local geodetic datum will need to be established.

4.3.1 TRANSFORMATION MODELS

The transformation from one coordinate system to another can be represented by scale, rotation and translation terms. There are a number of ways in which this can be achieved. The following is a discussion of transformation models useful in satellite surveying. The reader is referred to Harvey (1985) for details.

The most general transformation suitable for surveying is the affine transformation. However, the conformal and orthogonal transformation models are preferred in satellite surveying. An affine transformation transforms straight lines to straight lines and parallel lines remain parallel. However, the size, shape, position, and orientation of lines in a network are changed. 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 factor. There are 12 unknowns, so at least four sites common to both systems are needed to determine the transformation parameters. A conformal transformation is an affine transformation for which the scale factor is the same in all directions. Shape is preserved, so angles are not changed, but the lengths of lines and the position of points in the transformed system may be changed. An orthogonal transformation is a conformal transformation in which the scale factor is unity. The angles and distances within the network are preserved and only positions of points change after transformation.

The general conformal transformation model relating the coordinates X (X,Y,Z) of a point in one system to the coordinates x (x,y,z) in the other system is defined by:

$$X = s R x + T \qquad (4.13)$$

where s is the scale factor, R is an orthogonal rotation matrix and T (Tx,Ty,Tz) are the translation components. There are seven parameters -- three rotation angles, three translation components and a scale factor. The translation terms are the coordinates of the origin of net x in net X.

For small rotations (angles less than about 3") R can be approximated by:

$$R = \begin{bmatrix} 1 & \kappa & -\theta \\ -\kappa & 1 & \omega \\ \theta & -\omega & 1 \end{bmatrix} \quad (4.14)$$

where ω, θ and κ are the anticlockwise rotation angles about the x,y, and z axes respectively, as viewed from the positive end of the axis. Considerably larger angles can be tolerated if the network is not extensive.

The minimum number of observations required to solve for the seven parameters of a conformal transformation is seven. Three points common to each net will yield a redundancy of two. However, it is desirable to have at least four common sites to determine the seven parameters in a least squares adjustment. Initially all seven parameters should be estimated. Omitting some parameters is justified if the estimates of some parameters are statistically insignificant (see Harvey 1985) or if the redundancy is small (for example with only three common stations). The estimated parameter should not be applied to transform points outside the area for which they were defined.

In the presence of observational and computational errors, the accuracy of the estimated transformation parameters may vary considerably, depending on the spatial distribution of the points used. For example, the points should not be colinear because the components of rotations about axes parallel to the line of points cannot be determined. Ideally the points should be geographically well distributed.

Bursa-Wolf Model

Equation 4.13 defines the Bursa-Wolf model for a conformal 7-parameter transformation (Bursa 1966, Wolf 1963). The parameters of the model are highly correlated, since the region of common points usually covers only a small portion of the earth's surface. In other words, over a limited area it is difficult to distinguish between displacements due to the translation terms (Tx,Ty,Tz) and those due to rotations (ω,θ,κ). This is not so great a problem for continental transformations.

Other Transformation Models

The Molodensky-Badekas model (Badekas 1969) removes the high correlations that may exist between the model parameters by relating them to the centroid of the network or some other convenient point within the network. This model gives the same answers for the baseline lengths and angles of the survey network, and for the scale and rotation parameters, as the Bursa-Wolf model. However, the translation parameters are different and have higher *a posteriori* precisions. Harvey (1985) develops an ellipsoidal transformation model in which the same 7 parameters are determined. However ellipsoidal coordinates and their corresponding variance-covariance matrices are employed. When correctly applied, all three models produce identical results for the adjusted coordinates of the network, the variance-covariance matrices, scale and rotation terms, and their variance-covariance matrices.

4.3.2 PRACTICAL ASPECTS

Neither the satellite reference system or the terrestrial datum is perfect. Both contain systematic errors which affect the transformation model, thereby producing distortions. Conformal transformations guard against undue distortions over small regions. However, when applied to a large network a conformal transformation model may significantly distort scale and orientation. Therefore, the entire network should be tested for significant local distortions in scale and orientation, as a conformal transformation will tend to smooth these out. Distortions within a continental network such as the AGD84 are not uniform and it may be more appropriate to solve for the transformation parameters which best fit sub-regions of the total net.

Transforming between a local geodetic datum and a satellite datum will be a routine operation for those using GPS. Essentially, the surveyor has two options: he can

(a) use published values for the transformation parameters, or
(b) determine his own transformation parameters.

In GPS data processing, one or more stations are held fixed, while the positions of the other stations relative to the fixed stations are determined in the GPS reference system. As long as the same transformation model (and parameters) is used to transform the fixed station coordinates (local datum) into the GPS datum (earth-fixed reference system in which the GPS ephemerides are expressed) as is used to subsequently transform the new station coordinates out of the GPS system, no difficulties will arise. However, care must be taken to ensure that the transformation parameters are "close" to the correct values so that no distortions are introduced in the data analysis. By ensuring that the transformation parameters are reasonably good, small errors in the fixed station coordinates, expressed in the GPS datum, will not unduly affect the precision of the coordinates of the other surveyed stations (note that an error of 20 m in the coordinates of fixed stations relative to the GPS datum -- the GPS ephemerides system -- is equivalent to an orbit "error" of this amount; consequently 1 ppm errors will be introduced into the whole network as a result of this).

There are two possible approaches for determining the transformation parameters directly:

1. using one fixed station whose coordinates in the GPS datum have been determined with GPS in the point-positioning mode, or

2. using *a priori* transformation parameters (for example, from previous TRANSIT surveys) to transform the fixed station.

The former approach is better. Unless the *a priori* transformation model is good, the computed transformation parameters will not necessarily reflect the true relationship between the GPS satellite datum and the local datum. Unless the surveyor is convinced of the quality of the local geodetic datum, the transformation parameters (obtained directly or from another source) should not be used for transforming points outside the area defined by the common stations used in the determination of the transformation parameters.

5. GPS DATA PROCESSING

In this Chapter, we describe the post-processing of the observations for position information. Precisely how are positions or position differences obtained from the carrier phase measurements recorded in the receivers? What are the dominant sources of error in these measurements and how may their effect be reduced? These matters are addressed here.

5.1 GPS PHASE OBSERVABLES

Below we derive the theoretical model for the carrier phase -- the most frequently used GPS observable for surveying applications. The phase measurements at a number of sites can be differenced to reduce the effect of common-mode oscillator, atmospheric, and orbit errors. The theoretical formulations for these observations are derived and the parameters which they eliminate are described. Techniques which do not depend on differencing are described in Section 5.2.6. The models for the observables are functions of the time delay -- the time taken by the signal to travel from the satellite to the receiver. Accordingly, a step-by-step description of the propagation time delay computation is given.

5.1.1 ONE-WAY PHASE

The output from a single phase-tracking channel of a GPS receiver is the difference between the phase of the received carrier signal of a satellite and the phase of the local oscillator (LO) within the receiver. It is often called the carrier beat phase. The received carrier phase differs from the transmitted phase due to the relative motion of the satellite and ground station (the Doppler effect), fluctuations in pathlength due to the atmospheric refraction and multipath, and phase noise introduced by the receiver. We can write the phase $\phi_{ij}(t_j)$, at station j, for satellite i, as follows:

$$\phi_{ij}(t_j) = \phi_{rij}(t_j) - \phi_{LOj}(t_j) + n_{ij} + \phi_{noise} \qquad (5.1)$$

where the units of all the ϕ's are cycles, and

- t_j = time of reception of the satellite signal at station j
- ϕ_{rij} = carrier phase received at station j from satellite i
- ϕ_{LOj} = phase of the receiver oscillator
- ϕ_{noise} = random measurement noise
- n_{ij} = an integer, representing the inherent n-cycle ambiguity in the observed phase (see Section 5.2.4).

Strictly speaking, t_j in Equation 5.1 is measured in dynamical time (TDT) -- see Section 4.1. However, in GPS data processing, we can ignore the difference between dynamical time and a uniform terrestrial atomic time scale. Hence, we may think of t_j as TAI, UTC, or GPST. Nevertheless, the time kept by a GPS receiver may exhibit significant variations from a uniform time scale.

In the case of codeless GPS receivers which reconstruct the carrier phase by squaring the signal or related proprietary techniques, the frequency of the observed signal is <u>double</u> the nominal carrier frequency. Consequently, all quantities in Equation 5.1 and in the following equations should be multiplied by two. If the observed quantity is the P code transition phase, then the following development is still generally valid, but the observed frequency is 10.23 MHz and not the L1 or L2 carrier frequencies.

The signal received at time t_j is related to the signal transmitted by the i-th satellite at time t_i by:

$$t_i = t_j - \tau_{ij}(t_j) \tag{5.2}$$

where τ_{ij}, the <u>propagation time delay</u>, includes both the geometric delay and the delay introduced by the troposphere and ionosphere (Section 5.1.6). The phase of the received signal ϕ_{rij}, is related to the phase of the transmitted signal ϕ_{tij}, by:

$$\phi_{rij}(t_j) = \phi_{tij}(t_j - \tau_{ij}) \tag{5.3}$$

The transmitter frequency, though nominally constant, in fact varies due to instabilities in the satellite's oscillator. For accurate point positioning (but usually not for relative positioning -- see Section 5.1.2 and 5.1.4), these variations must be modelled. The phase of the satellite transmitter can be represented as a function of time by:

$$\phi_{ti}(t+\Delta t) = \phi_{ti}(t) + \dot{\phi}_{ti}(t)\Delta t + \tfrac{1}{2}\ddot{\phi}_{ti}(t)\Delta t^2 + \ldots \tag{5.4}$$

where $\dot{\phi}$ and $\ddot{\phi}$ are respectively the time rates of change and acceleration of the phase, that is, the frequency and frequency drift. Using Equations 5.3 and 5.4, we expand Equation 5.1 thus:

$$\begin{aligned}\phi_{ij}(t_j) &= \phi_{ti}(t_j) - \dot{\phi}_{ti}(t_j)\tau_{ij}(t_j) + \tfrac{1}{2}\ddot{\phi}_{ti}(t_j)\tau_{ij}^2(t_j) - \ldots \\ &\quad - \phi_{LOj}(t_j) + n_{ij} + \phi_{noise}\end{aligned} \tag{5.5}$$

We now model the transmitter frequency as follows:

$$f_i(t) = f_o + a_i + b_i(t - t^o) \tag{5.6}$$

where t^o is a reference epoch chosen, for example, to be near the centre of the span of observations. The nominal transmitter frequency f_o (= 1575.42 MHz for the L1 signal), is the same for all GPS satellites.

In general, for a given satellite, the frequency offset a_i and drift coefficient b_i will change for observations spanning a few hours. For Cesium oscillators the fractional frequency accuracy a_i/f_o, is usually better than 10^{-12} over 6 hours and b_i/f_o is less than $10^{-15}s^{-1}$ over this period. For Rubidium oscillators a_i/f_o is $\cong 10^{-11}$ and $b_i/f_o \cong 10^{-14}\ s^{-1}$. Using Equation 5.6 we can evaluate the transmitter-phase terms in Equation 5.5, that is:

$$\phi_{ti}(t_j) = \int_{t^o}^{t_j} [f_o + a_i + b_i(t_j-t^o)]\, dt$$

$$= (f_o+a_i)(t_j-t^o) + ½b_i(t_j-t^o)^2 + \phi_{ti}(t^o) \qquad (5.7)$$

$$\dot{\phi}_{ti}(t_j) = d\phi/dt|_{t_j} = f_i(t_j) = f_o + a_i + b_i(t_j-t^o) \qquad (5.8)$$

$$\ddot{\phi}_{ti}(t_j) = d^2\phi/dt^2|_{t_j} = df_i/dt|_{t_j} = b_i \qquad (5.9)$$

Station clock instabilities, which affect both the epoch of the observations and the phase of the local oscillator, can be modelled in a similar manner.

If the <u>true</u> epoch of reception is t_j (TDT or UTC), but the receiver clock <u>reads</u> t'_j, then we can write:

$$t'_j - t_j = \Delta t = q_j + r_j(t_j-t^o) + ½s_j(t_j-t^o)^2 \qquad (5.10)$$

We now use Equation 5.10 to evaluate $\phi_{LO}(t)$ in Equation 5.5. Since the frequency f_{LO} of the LO is constant with respect to its own time base t', we write:

$$\phi_{LO}(t) = \phi_{LO}(t^o) + f_{LO}q + \int_{t^o}^{t} f_{LO}dt' \qquad (5.11)$$

Since $dt'/dt = 1 + r_j + s_j(t-t^o)$ (Equation 5.10):

$$\phi_{LO}j(t) = \phi_{LO}j(t^o) + f_{LO}q + f_{LO} \int_{t^o}^{t} [1 + r_j + s(t-t^o)]\, dt$$

$$= \phi_{LOj}(t^o) + f_{LO}q_j + f_{LO}(t-t^o) + f_{LO}r_j(t-t^o) + ½f_{LO}s_j(t-t^o)^2 \qquad (5.12)$$

Combining Equations 5.7 to 5.9 and Equation 5.12 gives the desired expression for the <u>carrier beat phase observable</u>:

$$\begin{aligned}\phi_{ij}(t_j) = &- [f_o + a_i + b_i(t_j-t^o)]\, \tau_{ij}(t_j) + ½b_i\tau_{ij}^2(t_j)\\ &+ \phi_{ti}(t^o) + f_o(t_j-t^o) + a_i(t_j-t^o) + ½b_i(t_j-t^o)^2\\ &- \phi_{LOj}(t^o) - f_{LOj}(t_j-t^o) - f_{LOj}q_j - f_{LOj}r_j(t_j-t^o)\\ &- ½f_{LOj}s_j(t_j-t^o)^2\\ &+ n_{ij} + \phi_{noise} \qquad (5.13)\end{aligned}$$

In Equation 5.13, only the terms on the first line depend on the propagation delay τ_{ij}, and hence on the geometry of the observations. The first term in line 1 is simply the product of transmitter frequency, defined by Equation 5.6, and the propagation delay. The second term in line 1, which is proportional to τ_{ij}^2, is small and can be neglected for GPS observations. (Note: We sometimes call $\phi_{ij}(t_j)$ the one way phase observable because it involves the signal transmitted by a satellite and received by a particular receiver on the ground. This terminology serves to distinguish $\phi_{ij}(t_j)$ from the various differenced observables discussed below.)

The effect of instabilities in the satellite and receiver oscillators are accounted for by the terms appearing in lines 2, 3, and 4. The combination of the first term on line 2 and the first term on line 3, represents the phase difference between transmitter and receiver oscillators at time t^o. This is an unknown constant which is determined at the data processing stage (Section 5.2). The integer ambiguity n_{ij} is indistinguishable from this constant. Hence a combined term comprising the initial offset of satellite and receiver clock and the ambiguity term is obtained. The expression "resolving the ambiguity" refers to procedures whereby the n_{ij} term is explicitly determined separately from the initial clock offset terms.

The second terms on lines 2 and 3 will cancel if we chose a nominal LO frequency f_{LOj}, equal to the nominal transmitter frequency f_o. Some receivers (eg. the TI 4100) can be programmed with f_{LO} offset from f_o so that the observed phase is biased by $(f_o-f_{LO})(t-t^o)$. The remaining terms in Equation 5.13 represent polynomials accounting for the different time behaviour of the transmitter and receiver clocks. Depending on whether observations to several satellites from one ground station or observations to the same satellite from several ground stations are available, we could choose to determine either the transmitter-frequency coefficients a_i and b_i, or the receiver-clock coefficients rj and sj.

We now write Equation 5.13 as:

$$\phi_{ij}(t_j) = - f_t(t_j)\ \tau_{ij}(t_j) + C_{ij}(t_j) + \phi_{bias} + \phi_{noise} \qquad (5.14)$$

where

$$f_t(t_j) = f_o + a_i + b_i(t_j-t^o) \qquad (5.15)$$

The time-dependent clock terms are of the form:

$$C_{ij}(t_j) = f_o\ [\ -q_j + \alpha(t-t^o) + \tfrac{1}{2}\beta(t-t^o)^2] \qquad (5.16)$$

where

$$\alpha = a_i/f_o - r_j \qquad\qquad \beta = b_i/f_o - s_j$$

and

$$\phi_{bias} = \phi_{ti}(t^o) - \phi_{LOj}(t^o) + n_{ij}$$

It is convenient to express Equation 5.16 in terms of the UTC time at which the signal is received on the ground. This time is obtained from the receiver clock time and the model given in Equation 5.10. We shall substitute this model into the expression for the observable when we form the partial derivatives of the observables with respect to the model parameters (Section 5.2.2). However, we first derive expressions for the differenced phase observables and examine which terms in Equation 5.13 cancel or nearly cancel.

5.1.2 BETWEEN-STATIONS DIFFERENCES

The between-stations or singly-differenced phase observable $\Delta\phi_i$ for satellite i, is defined as the difference between the received phases at sites 2 and 1, that is:

$$\Delta\phi_i = \phi_{i2}(t_2) - \phi_{i1}(t_1) \tag{5.17}$$

where the phase is sampled when the site clocks read the same time (that is, $t_1' = t_2'$). Nevertheless, the observation epochs (in UTC) will not be identical since the clocks are not perfectly synchronised and run at different rates.

From Equation 5.13 we obtain, after neglecting the term proportional to τ_{ij}^2 and assuming $f_o = f_{LO1} = f_{LO2}$:

$$\begin{aligned}\Delta\phi_i = & - f_o[\tau_{i2} - \tau_{i1}] \\ & - [a_i + b_i(t_1-t^o)](\tau_{i2}-\tau_{i1}) - b_i(t_2-t_1)\tau_{i2} \\ & + a_i(t_2-t_1) + \tfrac{1}{2}b_i(t_2-t_1)[2(t_1-t^o) + (t_2-t_1)] \\ & - f_o(r_2-r_1)(t_1-t^o) - f_o r_2(t_2-t_1) - \tfrac{1}{2}f_o(s_2-s_1)(t_1-t^o)^2 \\ & \qquad - \tfrac{1}{2}f_o s_2(t_2-t_1)[2(t_1-t^o) + (t_2-t_1)] \\ & - [\phi_{LO2}(t^o)-\phi_{LO1}(t^o)] - f_o(q_2-q_1) + (n_{i2}-n_{i1}) + \phi_{noise}\end{aligned} \tag{5.18}$$

The time argument appearing in the delays τ_{ij} is omitted for convenience. Observe that the dominant transmitter-clock terms in Equation 5.13 (the product of a_i or b_i and powers of t_j-t^o) have cancelled in the between stations difference. In order to evaluate the magnitude of each of the terms in Equation 5.18 we assume $a/f_o \cong 10^{-11}$, $b/f_o \cong 10^{-14}$ s^{-1} for a Rubidium, and $a/f_o \cong 10^{-12}$, $b/f_o \cong 10^{-15}$ s^{-1} for a Cesium oscillator. If Cesium oscillators are available at the ground stations as well, then $r \cong 10^{-12}$, $s \cong 10^{-14}$ s^{-1}, and $t_2-t_1 \cong 10^{-5}$s. On the other hand, if crystal oscillators are employed, and the epoch cannot be synchronised with a broadcast time signal, then $r \cong 10^{-8}$, $s \cong 10^{-12}s^{-1}$, and $t_2-t_1 \cong 1$ s. Typically, $t-t^o \cong 10^4$ and $\tau_{ij} \cong 10^{-1}$s, and for transcontinental baselines (≅3000 km), $\tau_{i2}-\tau_{i1} \cong 10^{-2}$s.

After substituting the above values into Equation 5.18, we find:

a) that of the geometric terms (lines 1-2), only the first is significant;

b) that provided the ground station clocks can be synchronised to better than about 0.05 second, all of the terms relating to transmitter oscillator behaviour (line 3) produce changes which are less than 10^{-2} cycles; and,

c) that only the first and third the terms on line 4 are important when Cesium oscillators are employed at the ground stations. For crystal oscillators these terms can be quite large, and our simple polynomial model is probably inadequate.

The primary application of the singly-differenced observable is for orbit determination, where the ground stations are equipped with atomic frequency standards and are separated by thousands of kilometres. When Cesium oscillators are available Equation 5.18 reduces to:

$$\begin{aligned}\Delta\phi_i = &- f_o[\tau_{i2} - \tau_{i1}] \\ &- [a_i + b_i(t_1-t^o)](\tau_{i2}-\tau_{i1}) \\ &- f_o(r_2-r_1)(t_1-t^o) - \tfrac{1}{2}f_o(s_2-s_1)(t_1-t^o)^2 \\ &- [\phi_{L02}(t^o)-\phi_{L01}(t^o)] - f_o(q_2-q_1) + (n_{i2}-n_{i1}) \\ &+ \phi_{noise}\end{aligned} \tag{5.19}$$

As before, line 1 represents the geometric effect. Line 2 models the drift in the transmitter frequency which over transcontinental baselines and for a Cesium or Rubidium oscillator is less than 1 cycle in 6 hours. Line 3 accounts for the effect of a frequency offset between the local oscillators, which can be quite large, often dominating singly-differenced phase residuals. Line 4 describes the initial station oscillator phase difference (including the part due to clock offset), and the integer part of the total phase difference. These three terms are indistinguishable from one another and are therefore lumped together and estimated as a <u>bias</u>.

5.1.3 BETWEEN-SATELLITES DIFFERENCES

The between-satellites differenced phase observable $\Delta\phi_j$ for site j, is defined as the difference between the phases of the received signals from satellite 2 and satellite 1, that is:

$$\Delta\phi_j = \phi_{2j}(t_j) - \phi_{1j}(t_j) \tag{5.20}$$

From Equation 5.13 and neglecting the term proportional to τ_{ij}^2:

$$\begin{aligned}\Delta\phi_j = &- f_o[\tau_{2j} - \tau_{1j}] \\ &- [a_2\tau_{2j}-a_1\tau_{1j}] - [b_2\tau_{2j}-b_1\tau_{1j}](t_j-t^o) \\ &+ (a_2-a_1)(t_j-t^o) + \tfrac{1}{2}(b_2-b_1)(t_j-t^o)^2 \\ &+ [\phi_{t2}(t^o)-\phi_{t1}(t^o)] + (n_{2j}-n_{1j}) \\ &+ \phi_{noise}\end{aligned} \tag{5.21}$$

where line 1 represents the geometric effect, lines 2 and 3 contain the effect of frequency differences between the two satellites' oscillators, and line 4 represents both the initial transmitter oscillator phase difference and the integer part of the total phase difference. The transmitter frequency effects appearing in lines 2, which are scaled by the propagation delay, can be ignored in the presence of the much larger effects on line 3, which are scaled by powers of $(t-t^o)$. As before, the initial phase and integer differences are lumped together and estimated as a single bias. Note that the receiver oscillator errors are eliminated in this observable.

5.1.4 DOUBLE-DIFFERENCES

The doubly-differenced phase observable $\Delta^2\phi$ is obtained, by differencing the between-stations differenced observable (Equation 5.19) between the satellites:

$$\begin{aligned}\Delta^2\phi &= \Delta\phi_2 - \Delta\phi_1 \\ &= - f_o[\tau_{22} - \tau_{21} - \tau_{12} + \tau_{11}] \\ &\quad - [a_2 + b_2(t_2-t^o)](\tau_{22}-\tau_{21}) \\ &\quad + [a_1+b_1(t_1-t^o](\tau_{12}-\tau_{11}) \\ &\quad + (n_{22} - n_{21} - n_{12} + n_{11}) + \phi_{noise} \end{aligned} \qquad (5.22)$$

The geometric effect represented by the product of the nominal transmitter frequency and the propagation delay differences appears on line 1. With Cesium oscillators on board the satellites, the effect of transmitter frequency drift (lines 2 and 3), can usually be ignored for baselines less than 100 km long. However, for longer baselines, which might be used in orbit determination, and for observations involving satellites with less accurate oscillators, the terms on lines 2 and 3 can be significant. The terms appearing on lines 2 and 3 should be included in the model for the doubly-differenced observable if the transmitter oscillators are intentionally "dithered" after 1989 by the U.S. Defense Department as a means of limiting real-time accuracy. Note that the unknown initial phases of all satellite and receiver oscillators are eliminated. Hence, except for model error and noise, the phase "bias" is an integer. Under certain conditions, discussed in Section 5.2.4, this integer value can be determined and subsequently held fixed when estimating baseline coordinates. This procedure is possible only with the doubly-differenced observable.

5.1.5 BETWEEN-EPOCHS DIFFERENCES

Phase measurements can also be differenced between different observing epochs to form between-epochs differences. However, some of the information content is thereby removed from the data resulting in less precise estimates for relative positions than would be obtained from the other differenced observables. Nonetheless, the between-epoch differenced observable is not without merit. In particular, because a single anomalous phase measurement affects only two consecutive epoch differences the between-epochs differences are very useful for cycle slip editing (Section 5.2.5).

With the between-epochs observable, (also called the triple difference), the quality of relative position determination is only slightly degraded when compared to the solution using one-way phase observations where the integer-cycle ambiguities are not resolved but a "bias" parameter is estimated (Goad & Remondi 1984). On the other hand, resolving the ambiguities produces baseline coordinates which are more accurate than those obtained from between-epoch observations by a factor of two (for observing spans of 4 or more hours) to five (for observing spans of 2 hours or less) (Bock et al 1985, Hatch & Larsen 1985).

5.1.6 COMPUTING THE TIME DELAY

The phase observables are functions of the transmitter and oscillator frequencies and the propagation delay τ, which contains the geometric information essential for GPS surveying. The basic problem is to calculate the time taken by a signal which is emitted from the GPS satellite to travel to the receiver on the ground. The models required are the positions of the GPS satellites and the receivers and the atmospheric retardation of the signal.

The calculations should, strictly speaking, be performed in a solar system barycentric (ie. nearly inertial) reference frame and within the framework of General Relativity (see Section 4.1). However, a geocentric inertial frame and a purely Newtonian theory have been found adequate in practice. If the satellite ephemerides have been generated in a particular space-fixed reference frame, for example the Conventional Celestial Reference Frame (CCRS) defined in Section 4.2.3, then the earth-fixed receiver coordinates, in the rotating Conventional Terrestrial Reference System (CTRS) must be transformed to the CCRS at each observation epoch before computing the propagation delay. An equivalent approach is to transform the satellite ephemerides from the CCRS to the CTRS. This approach is more efficient if the number of tabular values in the ephemerides is less than the number of observation epochs to be processed. In the description that follows, we shall assume that the latter procedure is to be used.

Coordinates of the Receiver

Given the earth-fixed coordinates (for example, latitude, longitude and height) of the receiver in a conventional geodetic datum, the first step is to employ standard transformations to obtain geocentric Cartesian coordinates (r=x,y,z) in the CTRS (see Section 4.3). If necessary, we correct the receiver coordinates, for the effects of solid-body tides according to formulae given in, for example, Melbourne et al (1983). The solid-body or earth tide is the heaving of the earth's surface due to the direct gravitational attraction of the sun and moon, and although the tidal displacements of the sites vary by approximately 10-50 cm, they would nearly cancel in relative positioning over distances of a few hundred kilometres or less. The effects of ocean loading (the deformation of the earth's surface due to the waxing and waning of ocean tides) is, in general, an order of magnitude less than that of the solid-body tide and can be neglected for all but high precision applications.

Coordinates of the Satellite at signal transmission

If the satellite coordinates s'(t) are given in the space-fixed system (CCRS) the first step is to transform them into the earth-fixed system (CTRS). From Equation 4.12

$$s = R_2(-x_p)\, R_1(-y_p)\, R_3(GAST)\, N\, P\, s' \qquad (5.23)$$

where x_p and y_p are the coordinates of the instantaneous rotation pole with respect to the z-axis of the CTRS, and R_1 and R_2 are matrices effecting right-handed rotations about the x and y axes respectively. Values for x_p and y_p are obtained from, for example, the BIH (Table 4.2). R_3(GAST) is the rotation about the instantaneous rotation axis through the Greenwich Apparent Sidereal Time, N is the nutation matrix, and P is the precession matrix, all defined in Section 4.2.

The GAST changes at a rate equal to the earth's rotation and must be evaluated at each epoch of the ephemeris. The precession and nutation matrices change more slowly (≅0".14/day), and the error incurred by evaluating them only once over a 4 hour observing session is only 0.1ppm.

With the GPS "Broadcast Ephemerides" the user does not need to transform the satellite coordinates as they are already provided in an earth-fixed system though this is not the CTRS. However, the ephemerides are tabulated according to GPST (Section 4.1); so before interpolating the satellite coordinates at observation time t_j(UTC) the corresponding t_j(GPST) must be obtained.

In order to compute the position of the satellite at transmission it is necessary to know the time t_i when transmission occurs, and therefore it is necessary to know the space position of the transmitter at this time. This circular problem is solved iteratively commencing with the approximation:

$$t_i(\mathrm{TAI}) = t_j(\mathrm{TAI}) - \tau_o \qquad (5.24)$$

where τ_o is an approximate value of time delay. With this approximation for t_i we scan the satellite ephemeris to interpolate $s(t_i)$ the coordinates of the satellite in the CTRS. With these coordinates and the coordinates of the receiver, we compute a second approximation to the propagation delay:

$$\tau_{ij} = |s(t_i) - r|/c \qquad (5.25)$$

where the vertical bars denote the magnitude of the vector, and c is the speed of light. The iteration is repeated if the agreement between the first and second approximations is greater than some accuracy criterion set for the calculations. Roughly three orders of magnitude are gained for each iteration. The effect of the atmosphere is of the order of 10^{-7} of the one-way time delay and can be neglected during this iteration cycle.

The first step in scanning the satellite ephemeris is to convert the time of reception, t_j', as recorded by the receiver clock, to dynamical time t_j, using our best model for the station clock behaviour (see Section 5.1.1). Any conventional offsets or theoretical variations between the available atomic time scale (for example UTC) and the time scale of the satellite ephemeris are also taken into account (see Section 4.1). We prefer to compute the satellite ephemeris in TAI and to time-tag GPS observations with UTC. In this case we compute t_j(TAI) from t_j(UTC):

$$t_j(\mathrm{TAI}) = t_j(\mathrm{UTC}) + (\mathrm{TAI-UTC}) \qquad (5.26)$$

where TAI-UTC is obtained from an international time service such as the BIH (Table 4.2). The coordinates of the satellite at time t_j(TAI) are now obtained by interpolation.

Atmospheric Delay

The GPS signals are refracted by the troposphere and the ionosphere. This means that the signal is received on the ground a little later than if the intervening space were a vacuum: that is, the signal is delayed. The troposphere is the lower part of the earth's atmosphere in which the temperature decreases with increasing height. It extends to a height of about 10 km. The tropospheric delay is calculated and applied once the above iteration has converged. This correction procedure is described in Section 5.3.2. The ionosphere is a region of ionised air and free electrons in the earth's upper atmosphere extending from a height of about 50 km to 1000 km. In relative positioning the delay caused by the ionosphere is either ignored or determined from dual-frequency measurements (see Section 5.4). In the latter case the correction is subsequently added to the theoretical time delay or used to correct the phase observable.

5.2 ANALYSIS OF GPS OBSERVATIONS

Given a model for the observable and the partial derivatives of the observable with respect to station coordinates and other parameters, we can use least squares or other estimation techniques to determine improved values of these coordinates and parameters. In this section we review least squares with emphasis on those aspects which have particular application to the analysis of GPS observations; we derive partial derivatives of the phase observable with respect to the parameters of interest; and finally we discuss computational procedures, ambiguity resolution and data editing.

5.2.1 LEAST SQUARES ESTIMATION

The first step in least squares analysis is to formulate the observation equation which expresses the observations ℓ at each observation epoch in terms of the parameters $\mathbf{x}$:

$$c_t = \ell_t + v_t \qquad (5.27)$$

where c_t is the modelled or theoretical observable at time t, ℓ_t is the observation at time t and v_t is the observation error. This equation is linearised using Taylor's theorem to give:

$$c_a + \mathbf{A}dx = \ell + v \qquad (5.28)$$

where $\mathbf{A}$ (often referred to as the design matrix), is the matrix of partial derivatives of the observations with respect to the parameters, c_a is the vector of theoretical observables calculated using *a priori* values x_a of the parameters $\mathbf{x}$, v is the vector of post-fit residuals and dx is the vector of small corrections to the *a priori* values of the parameters. The partial derivatives of the carrier phase observable with respect to the various parameters are given in Section 5.2.2.

The least squares estimates $\hat{x}$ of the parameters are obtained by minimising the weighted sum of the squared residuals (WSSR) $v^T W v$, where $W = P^{-1}$ and P is the variance-covariance matrix for the observations. The solution to the least squares problem is given by:

$$\hat{x} = x_a + dx$$

where

$$dx = N^{-1} A^T W (\ell - c_a) \tag{5.29}$$

and the normal equation matrix is

$$N = A^T W A$$

The normal equation matrix is inverted using any standard method such as Gaussian elimination or Grout-Cholesky. However, a solution can only be obtained if both P and N are non-singular. Usually P will be diagonal and therefore non-singular. On the other hand, the observations from a given survey are not necessarily sensitive to all the parameters in the theoretical model. In this case N is singular. For example, GPS "between-stations" difference observations (Section 5.1.2) contain little information on the reference system origin. Accordingly only relative positions can be estimated. Additional information must be introduced to obtain absolute positions. One way of achieving this is to constrain (hold fixed) the a priori coordinates of one station. This will produce a minimally constrained least squares solution for the geodetic positions. Free network adjustments are minimally constrained and are implemented in conventional geodetic networks by constraining or fixing the mean position and mean spatial orientation of the a priori coordinates for points in the network (Vanicek & Krakiwsky 1982). Free network adjustments are special cases of Bayesian Least Squares where we minimise (Bossler 1972):

$$v^T W v + dx^T W_{xx} dx \tag{5.30}$$

$W_{xx} = P_{xx}^{-1}$, and P_{xx} is the variance covariance matrix for the a priori values of the parameters. The least squares solution is given by:

$$dx = (A^T W A + W_{xx})^{-1} [A^T W(\ell - c) + W_{xx} x_a] \tag{5.31}$$

Equation 5.31 can be thought of as the combination of two independent least squares adjustments of two different sets of observations. In this way, W_{xx} and x_a are obtained from one adjustment and A, W and ℓ from the other. Often, x_a is set to zero and Equations 5.30 and 5.31 are implemented simply by adding the weight matrix W_{xx} to the normal equation matrix. The procedure is commonly adopted when orbits and geodetic positions are estimated simultaneously. Usually W_{xx} is diagonal, with the diagonal terms equal to the reciprocal values of the a priori variances of the parameters.

Partitioned Solutions

Partitioned estimation schemes reduce the number of parameters in computer memory at any one time. They are especially useful for the block diagonal normal matrices which normally occur in GPS surveying. Partitioning is also employed to eliminate nuisance parameters such as the clock terms, to determine the ambiguity biases and to build up bigger networks from small networks or networks from single baseline measurements.

GPS observations are made simultaneously to four or more satellites at regular time intervals or epochs. For example, the Macrometer instrument is programmed to collect data at 60 epochs over an observing span of a few hours. The parameters of the problem can be conveniently split into those that are sensitive only to the observations at a specific epoch and those that are sensitive to the observations at each and every epoch (that is, they are common to all observations). In subsequent discussions we refer to these parameters as "<u>epoch parameters</u>" and "<u>common parameters</u>", respectively. Examples of the common parameters are the station positions, the satellite coordinates and measurement biases. Oscillator errors are examples of epoch parameters. The full set of observation equations for the one-way phase observable has the following form:

$$\begin{bmatrix} A_{c1} & A_{e1} & & & & & & 0 \\ A_{c2} & & A_{e2} & & & & & \\ A_{c3} & & & A_{e3} & & & & \\ A_{c4} & & & & A_{e4} & & & \\ \cdot & & & & & \cdot & & \\ \cdot & & & & & & \cdot & \\ \cdot & & 0 & & & & & \cdot \\ A_{cn} & & & & & & & A_{en} \end{bmatrix} \begin{bmatrix} x_c \\ x_{e1} \\ x_{e2} \\ x_{e3} \\ x_{e4} \\ \cdot \\ \cdot \\ \cdot \\ x_{en} \end{bmatrix} = \begin{bmatrix} \ell_{e1} \\ \ell_{e2} \\ \ell_{e3} \\ \ell_{e4} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \ell_{en} \end{bmatrix} \tag{5.32}$$

where at epoch i the matrix A_{ci} contains the partials for the common parameters x_c, and A_{ei} contains the partials for the epoch parameters x_{ei}, and ℓ_{ei} are the observations. Dropping the epoch time subscript for convenience and ignoring the zeros, the observation equations at a particular epoch become:

$$[\, A_c \mid A_e \,] \begin{bmatrix} x_c \\ \hline x_e \end{bmatrix} = \ell_e \tag{5.33}$$

A suitable algorithm permits the contributions to the "common" part of the normal matrix to be evaluated and the "epoch" part to be eliminated for the observations at each epoch. Only the partials and observations pertaining to the one observation epoch are therefore required in computer memory at any stage of the computations. The partitioned normal equation matrix at that epoch can be written as:

$$\begin{bmatrix} N_{cc} & N_{ce} \\ N_{ec} & N_{ee} \end{bmatrix} \begin{bmatrix} x_c \\ x_e \end{bmatrix} = \begin{bmatrix} b_c \\ b_e \end{bmatrix} \tag{5.34}$$

where $b_c = A_c^T W_e(\ell_e - c_e)$ and

$$b_e = A_e^T W_e(\ell_e - c_e)$$

The epoch terms are eliminated using:

$$N_{cc}' = N_{cc} - N_{ce}\, N_{ee}^{-1}\, N_{ec}$$

and (5.35)

$$b_c' = b_c - N_{ce} N_{ee}^{-1} b_e$$

The transformed normal equations are:

$$\begin{bmatrix} N_{cc}' & 0 \\ N_{ec} & N_{ee} \end{bmatrix} \begin{bmatrix} x_c \\ x_e \end{bmatrix} = \begin{bmatrix} b_c' \\ b_e \end{bmatrix} \tag{5.36}$$

The part of the transformed normal equation matrix pertaining to the common terms (top left-hand partition) is independent of the epoch parameters. Hence, the epoch equations (lower right-hand partition) can be disregarded, unless for some reason they are required, in which case the appropriate matrices are written on a file for later back-substitution. This procedure is repeated for each epoch after which the solution for the common parameters alone is obtained. The solution is identical to that which would be obtained by solving Equation 5.32 directly. However, as the zero off-diagonal terms occurring in the observation and normal equations are not processed, the partitioning technique is much more efficient than the direct solution of Equation 5.32.

Orthogonalisation

Orthogonalisation algorithms have been devised to solve the least squares problem without forming normal equations (Lawson & Hanson 1974). Householder transformations and Gram-Schmidt algorithms are employed to orthogonalise the observations. These methods operate directly on the partitioned observation equations producing "derived" observations:

$$A_c' = QA_c$$

$$A_e' = QA_e = 0$$

$$\ell' = Q\ell \tag{5.37}$$

where the matrix Q is orthogonal. The "new" observation equations can be written as follows:

$$[\, A_c' \mid 0 \,] \begin{bmatrix} x_c \\ x_e \end{bmatrix} = \ell' \tag{5.38}$$

These algorithms can be used to eliminate the nuisance parameters, like clock terms, from the GPS phase observable, thus providing an alternative to simple differencing (see Section 5.2.6). If the observations at each epoch have equal variances and are uncorrelated, the orthogonalised observations will also be uncorrelated. Orthogonalisation therefore provides one method for dealing with the correlation problems that occur with simply differenced observations (Section 5.2.6).

Statistical Tests

Statistical tests are useful for assessing the confidence levels of the estimated parameters. The variance-covariance matrix of the estimated parameters is obtained from the law of propagation of variances. Usually $C_{xx} = \sigma_o^2 G$ where σ_o^2 is the variance factor and G is the cofactor matrix given by the inverse of the normal equations:

$$G = N^{-1} \tag{5.39}$$

or, if a priori constraints were applied to the parameters:

$$G = (N + W_{xx})^{-1} \tag{5.40}$$

The variance factor is estimated from:

$$s^2 = (v^T W v + dx^T W_{xx} dx)/r \tag{5.41}$$

where r represents the degrees of freedom of the solution. This variance factor is statistically tested using either Chi-Squared tests or F-tests (Mikhail 1976, Patterson 1984, Harvey 1985). The assumptions that have been made about either the models or the observations are incorrect if these tests fail at some confidence level.

After adjustment, the observations are tested for outliers. For correlated residuals, it is inappropriate to use a simple rejection criterion based on 2.5 or 3 times the standard deviation of the residuals. Nevertheless, because of its simplicity, this criteria is often applied. Pope (1976) and Vanicek & Krakiwsky (1982) describe more elaborate methods for outlier detection. Multivariate tests are then applied to determine whether the estimated parameters or groups of parameters are significant (Vanicek & Krakiwsky 1982, Harvey 1985).

In least squares analysis, certain model parameters are often held fixed, for example some ground station coordinates. The effect of errors in these fixed parameters on the estimated parameters is determined by partitioning the observation equations as follows:

$$A_1 x + A_2 y = \ell - c$$

This set of equations is solved to yield:

$$E_{xx} = - N^{-1} A_1^T W A_2 E_{yy} \tag{5.42}$$

where E_{xx} and E_{yy} are the uncertainties in the estimated and fixed parameters respectively. A_1 is the matrix of partial derivatives for the estimated parameters and the matrix A_2 comprises the partial derivatives for the fixed parameters. The variance-covariance matrix C_{xx} for the estimated parameters can also be expressed in terms of the separate contributions from the observations and the fixed parameters:

$$C_{xx} = N^{-1} + N^{-1} A_1^T W A_2 C_{yy} A_2^T W A_1 N^{-1} \tag{5.43}$$

C_{yy} is the variance-covariance matrix of the fixed parameters. Equations 5.42 and 5.43 are useful for studying the effect of modelling errors on GPS position determinations. In effect, this is accomplished by assuming that modelling errors are errors in the fixed parameters.

5.2.2 CALCULATION OF THE PARTIAL DERIVATIVES

We require the partial derivatives of the phase observables with respect to the geometric parameters, the clock terms and the biases. These represent the amount by which the observed phase or differenced phase would change for a given change in the parameters, taken one at a time. The expressions appearing below are derived by partially differentiating the expressions for the observables given in Sections 5.1.1 to 5.1.6.

Geometric Parameters

The partial derivative of phase with respect to the site coordinates and satellite positions is obtained from Equation 5.14, that is:

$$\frac{\partial \phi_{ij}}{\partial x} = -f_t \frac{\partial \tau_{ij}}{\partial x} \tag{5.44}$$

where f_t is the transmitter frequency defined by Equation 5.15 and x is the parameter of interest. The partials for the between-stations, between-satellites, and double differences are similarly obtained:

$$\frac{\partial \Delta\phi_i}{\partial x} = -f_t \left(\frac{\partial \tau_{i2}}{\partial x} - \frac{\partial \tau_{i1}}{\partial x} \right) \tag{5.45}$$

$$\frac{\partial \Delta\phi_j}{\partial x} = -f_t \left(\frac{\partial \tau_{2j}}{\partial x} - \frac{\partial \tau_{1j}}{\partial x} \right) \tag{5.46}$$

$$\frac{\partial \Delta^2\phi}{\partial x} = -f_t \left(\frac{\partial \tau_{22}}{\partial x} - \frac{\partial \tau_{21}}{\partial x} - \frac{\partial \tau_{21}}{\partial x} + \frac{\partial \tau_{11}}{\partial x} \right) \tag{5.47}$$

The partials of τ are obtained from Equation 5.24, that is:

$$\tau_{ij} = | R_{ij} | /c = R_{ij} /c \tag{5.48}$$

where R_{ij} is the range vector, defined by:

$$R_{ij} = s_i(t_i) - r_j \tag{5.49}$$

The sought partials are:

$$\frac{\partial \tau_{ij}}{\partial x} = \frac{1}{c} \frac{\partial R_{ij}}{\partial x} = \frac{1}{c} \frac{R_{ij}}{R_{ij}} \cdot \frac{\partial R_{ij}}{\partial x} \tag{5.50}$$

The partials of the site coordinates are obtained from Equation 5.49:

$$\frac{\partial R_{ij}}{\partial x} = - \frac{\partial r_j}{\partial x} \tag{5.51}$$

where

$$\frac{\partial r_j}{\partial \phi} = \begin{vmatrix} -\nu \cos\lambda \sin\phi \\ -\nu \sin\lambda \sin\phi \\ \nu \cos\phi \end{vmatrix} \tag{5.52}$$

$$\frac{\partial r_j}{\partial \lambda} = \begin{vmatrix} -\nu \sin\lambda \cos\phi \\ \nu \cos\lambda \cos\phi \\ 0 \end{vmatrix} \tag{5.53}$$

$$\frac{\partial r_j}{\partial \nu} = \frac{r_j}{\nu} \tag{5.54}$$

and ϕ, λ and ν are respectively geocentric latitude, east longitude and geocentric radius of site j. The partials for other coordinate types (for example, geodetic latitude, longitude and height) are obtained in a similar manner. The partials for the satellite positions are:

$$\frac{\partial R_{ij}}{\partial x} = \frac{\partial s_i(t_i)}{\partial x} \tag{5.55}$$

where the partials of the satellite positions $\partial s_i(t_i)/\partial x$, at the time of observation with respect to a particular orbital parameter x, must be obtained from the model used to generate the satellite ephemeris (see Chapter 7). If the ephemerides are given in the space-fixed system (CCRS), then the partials and the coordinates must be transformed using Equation 5.23.

Clock Terms

In the undifferenced and singly-differenced observable, the errors in satellite and receiver clock parameters enter mainly through the terms that do not depend on the propagation delay τ (see Equations 5.13, 5.18 & 5.21). However, most of these terms cancel in the doubly-differenced observable (Equation 5.22), and we need only be concerned with the effect of a timing error on τ. The partial derivative of the undifferenced phase observable with respect to the clock parameter is, from Equation 5.14:

$$\frac{\partial \phi_{ij}}{\partial x} = \frac{\partial C_{ij}(t_j)}{\partial x} \tag{5.56}$$

The effect of transmitter clock errors is omitted since these are scaled by τ, which is negligible compared to $(t_j - t^o)$. Using Equation 5.16, we compute the partial $\partial C_{ij}(t_j) / \partial x$, for the transmitter and receiver clock errors with respect to each of the parameters in our model.
Hence:

$$\partial C_{ij}(t_j) / \partial a = (t_j - t^o) \tag{5.57}$$

$$\partial C_{ij}(t_j) / \partial b = ½(t_j - t^o)^2 \tag{5.58}$$

$$\partial C_{ij}(t_j) / \partial q = -f_o \tag{5.59}$$

$$\partial C_{ij}(t_j) / \partial r = -f_o(t_j - t^o) \tag{5.60}$$

$$\partial C_{ij}(t_j) / \partial s = -½ f_o(t_j - t^o)^2 \tag{5.61}$$

The significant terms for the partial derivatives of the between-stations differences with respect to clock parameters are obtained from Equation 5.19. The partials with respect to the parameters describing the behaviour of the transmitter frequency are small when compared to the same partials obtained from the undifferenced and between-satellites differences. Since the latter observables are superior for determining transmitter frequency parameters, we omit these partials here. However, the partials with respect to receiver clock parameters cannot be neglected. They are:

$$\partial\Delta\phi_i/\partial r_1 = -\partial\Delta\phi_i/\partial r_2 = f_o(t_1-t^o) \tag{5.62}$$

$$\partial\Delta\phi_i/\partial s_1 = -\partial\Delta\phi_i/\partial s_2 = \tfrac{1}{2}f_o(t_1-t^o)^2 \tag{5.63}$$

For between-satellites differences, the receiver clock parameters vanish. The partials with respect to the transmitter frequency parameters follow from Equation 5.21:

$$\partial\Delta\phi_j/\partial a_1 = -\partial\Delta\phi_j/\partial a_2 = -(t_1-t^o) \tag{5.64}$$

$$\partial\Delta\phi_j/\partial b_1 = -\partial\Delta\phi_j/\partial b_2 = -\tfrac{1}{2}(t_1-t^o)^2 \tag{5.65}$$

where only the largest terms are considered. The transmitter-frequency offsets and drift-rates, a_1, a_2, b_1 and b_2 are the only clock parameters that appear explicitly in the doubly-differenced phase observable (Equation 5.22). The respective partials are much smaller than those obtained from the between-satellites differences. Hence, it is far better to estimate a_1, a_2, b_1 and b_2 using between satellite differences.

No explicit terms for the receiver clock parameters remain in the double-difference observable. However, the double-differences are still sensitive to q_2-q_1, the differences in clock epochs described by Equation 5.10. The sensitivity arises from the significant change in propagation delay with changes in reception time. Thus:

$$\frac{\partial\Delta^2\phi}{\partial q_1} = f_o\left(\frac{\partial\tau_{21}}{\partial t} - \frac{\partial\tau_{11}}{\partial t}\right) \tag{5.66}$$

$$\frac{\partial\Delta^2\phi}{\partial q_2} = -f_o\left(\frac{\partial\tau_{22}}{\partial t} - \frac{\partial\tau_{12}}{\partial t}\right) \tag{5.67}$$

The delay rates, $\partial\tau/\partial t = d\tau/dt$, are computed by differentiating Equation 5.48:

$$\frac{d\tau_{ij}}{dt} = \frac{1}{c}\frac{dR_{ij}}{dt} = \frac{1}{c}\frac{R_{ij}}{R_{ij}}\cdot\frac{dR_{ij}}{dt} = \frac{1}{c}\frac{R_{ij}}{R_{ij}}\cdot\left(\frac{ds_i}{dt} - \frac{dr}{dt}\right) \tag{5.68}$$

Biases

The partial derivative of the phase observables with respect to a bias parameter is unity.

5.2.3 PROCESSING SOFTWARE

The observations collected in the field may consist of only carrier phase (at the L1 frequency or both L1 and L2) in the case of a codeless receiver or both carrier phase and pseudo-range (at L1, or with P-code access at L1 and L2) for the code correlating receivers. Each observation is time tagged and for the code-correlating receivers the Broadcast Ephemeris information is recorded. Eventually all the files are transported, or transferred via some communication link, to a central processing facility.

Data management is important once the field data is stored in the central computer system. In particular, the data file structure should be designed to permit ready access and identification for processing. This requirement also holds for all ancillary information: a priori station coordinates, earth rotation and polar motion data, ephemerides, meteorological readings, etc.

For maximum efficiency, the data from all stations occupied during the same session is sorted into sequential order. Simultaneity of the time-tags for the data from the different sites must be ensured to allow the oscillator drifts to be eliminated using differencing or some equivalent technique. For the code-correlating receivers, the pseudo-ranges can be used to determine the receiver clock offset from GPS time. Codeless receivers are manually synchronised to UTC but may have a constant clock offset or large drift which is determined by regular synchronisation of the receiver clocks (see Chapter 3). Preliminary estimates of the ionospheric delays are obtained from dual-frequency pseudo-ranges.

Ephemerides covering the entire field campaign are required to process the observations. The "Broadcast Ephemerides" contained in the Navigation Message are adequate for most applications. Other options for obtaining ephemeris information are discussed in Section 7.4.

The phase data recorded in the receiver can be processed as one-way phases or as differenced phases. The choice depends on the task in mind (see Table 5.1). It is not the intention to present details here. Merely a general outline of the computation procedure is provided. After deciding on the type of differencing to be applied, the processing steps are as follows:

Read ephemeris files, site files and a priori information.

Start loop over observing sessions (if multi-session).

Start loop over measurement epochs.

Start loop over sites and satellites at each epoch.

- Read in phase observation and time-tag.
- Read a priori station position, ephemeris information and transformations between various reference systems.
- Calculate theoretically modelled phase or phase differenced observable using time tag and relevant a priori information.
- Correct observations for ionospheric and tropospheric refraction (if necessary).

- Compute difference between the observed and modelled phase (residual).

- Evaluate partial derivatives of theoretical observable with respect to parameters.

End station-satellite loop.

- Calculate differences of residuals and partials. Increment normal equation matrix.

End measurement epoch loop.

End observation session loop (if multi-session).

- Increment normal equation matrix to account for effect of *a priori* information on parameters.

- Solve normal equations.

- Update parameters and calculate statistical information.

- Calculate and plot residuals using updated parameters.

- Edit observations.

- Search for cycle slips and repeat solution if required.

- Resolve ambiguities (if possible) and repeat solution with ambiguity biases held fixed.

Output

- Site coordinates and variance-covariance matrix.

Optional

- Transform station coordinates to local datum using known set of transformation parameters or parameters determined from points with known positions in both satellite and local datums (Section 4.3).

- For large networks combine separate solutions in a final adjustment (Section 5.2.5).

- Determine parameters required to transform ellipsoidal heights to orthometric heights or use gravimetry to calculate relative geoid heights (Chapter 6).

Table 5.1 Parameters estimatable from phase observables.

Observable	Clock Errors Eliminated ? Receiver	Clock Errors Eliminated ? Satellite	Estimatable Parameters
Carrier beat phase	No	No	Position, orbit, bias (combined cycle ambiguity and initial clock offset), receiver and satellite clock terms
Between-stations difference [1]	No	Yes	Relative position, orbits, differenced bias (combined differenced cycle ambiguity and initial differenced receiver clock offset), Difference between receiver clock terms
Between-satellites difference	Yes	No	Position, relative orbits, differenced bias (combined differenced cycle ambiguity and initial difference of satellite clock offsets)
Double difference	Yes	Yes	Relative position, relative orbits, differenced bias (differenced cycle ambiguity)
Between-epochs difference [2]	Yes	Yes	Relative position

[1] Also referred to as single difference or interferometric phase

[2] Also referred to as triple difference or doppler

5.2.4 AMBIGUITY RESOLUTION

Phase observations are ambiguous by virtue of the unknown integer number of cycles between the receiver and the satellite. Once the satellite signals have been acquired by the receiver the whole number of cycles are tracked and counted. Therefore, the initial (unknown) integer number of cycles are the same over a particular observing session and can be represented by a single bias term. A new bias term is introduced if the receiver loses lock on the satellite. This cycle slip problem is dealt with in Section 5.2.5.

The (undifferenced) carrier beat phase and single-difference phase observables always contain unknown constant terms arising from the satellite and station oscillators (Section 5.1). As the unknown integer number of cycles will be indistinguishable from the initial clock terms it makes no sense to attempt to resolve the integer-cycle ambiguity. The two terms are combined and solved for as a single parameter. This term is not usually an integer. The observed phases can also be differenced between epochs avoiding this problem. However, since the clock and bias terms cancel in doubly-differenced phases the integer-cycle ambiguity can be resolved provided the observations are precise enough. With double differences the ambiguity biases are estimated along with the geodetic parameters. Departures from integer values can be attributed to measurement noise and errors in the theoretical model for the observable. We know of no simple method for forcing them to be integers. Ideally, values for the bias parameters obtained from a preliminary solution will be close to integers and uncertain by less than 1 cycle. These estimates are rounded to the nearest integer and the least squares solution is repeated with the biases held fixed at these values. Statistical tests are then applied to decide whether the "best" integer biases were selected. For baselines less than 20-30 km in length the precision of the estimated biases is usually high enough to allow the integer values to be resolved (Bock et al 1985). If the uncertainties in the integer values exceed one cycle then sub-optimal estimation schemes such as proposed by Langley et al (1984) must be used. After a preliminary solution the integer values which fall within the confidence interval of each initially estimated bias are scanned. The integer bias which minimises the WSSR (Section 5.2.1) is then selected and the procedure is repeated for all ambiguity biases. Langley et al (1984) describe an efficient algorithm for determining the WSSR without reinverting the normal equation matrix. The technique is based on the fact that all the required information is available after the biases and geodetic parameters have been estimated. The observation equations are partitioned as follows:

$$A_1 dx_1 + A_2 x_2 = f + v \qquad (5.69)$$

where x_1 are the geodetic parameters, x_2 are the ambiguity biases and $f = (\ell^1 - c)$ are the pre-fit residuals. After the first iteration in which both x_1 and x_2 are estimated, new observation equations are formed:

$$A_1 dx_1' = f' - v' \qquad (5.70)$$

where $f' = f - A_2 x_2^*$ and x_2^* are the rounded integer biases. As the x_2^* are held fixed, we have:

$$dx_1' = (A_1^T W A_1)^{-1} A_1^T W f' \qquad (5.71)$$

The WSSR is given by:

$$WSSR = v'^T W v' \quad (5.72)$$

$$= e_1 + e_2^T x_2^* + (x_2^*)^T E_3 x_2^* \quad (5.73)$$

where

$$e_1 = f^T W f - u_1^T N_{11}^{-1} u_1$$

$$e_2^T = -2(u_2 - u_1^T N_{11}^{-1} N_{12})$$

$$E_3 = N_{22} - N_{12}^T N_{11}^{-1} N_{12}$$

and

$$N_{ij} = A_i^T W A_j$$

$$u_i = A_i^T W f$$

The quantities e_1, e_2 and E_3 are obtained in the first iteration. Hence, it is a simple matter to calculate the WSSR for any combination of biases x_2^* without actually reinverting the normal equation matrix.

5.2.5 CYCLE SLIP EDITING

Cycle slips occur when for some reason, such as an obstruction of the line of sight, tracking stops for a moment and is subsequently resumed. The fractional phase measurement after reacquisition of the signal is the same as if tracking had been maintained. However, the integer number of cycles is different. With dual-frequency measurements, cycle slips can occur at either frequency, which further complicates the problem. Oscillator errors usually dominate in one-way phase measurements which means that only obvious cycle slips are detectable. On the other hand, double-differences are relatively free from oscillator errors. Therefore they are suitable for isolating cycle slips. Techniques for dealing with the problem abound. The principles of some of these methods are outlined below.

One approach is to calculate residuals holding fixed the best available station positions. Suitable station positions are obtained by first processing the between-epochs differenced phases, which are largely free from the effects of cycle slips. Cycle slips are then visually identified as discontinuities in the double-difference residuals. The times at which the cycle slips occurred are noted and the observations corrected, and reprocessed. Usually interactive graphics are employed to plot the residuals. However, the method can be time consuming and tedious.

Beutler et al (1984) commence editing where gaps have occurred in the data. A polynomial function is fitted to the data around the gaps. If no cycle slips have occurred the zero-order terms of the polynomials should be the same over a whole session of observations. Hence, any differences in the zero order terms can be used to identify the cycle slips. The gaps in the data are again located by means of interactive graphics. Though an improvement on the above method, this process can also be quite time consuming, particularly if a lot of cycle slips are present. Also the method is not sensitive to small cycle slips.

Remondi (1985a) describes an automated cycle slip editing procedure. Good preliminary station positions are determined from the doubly differenced phases which have also been differenced between epochs (triple differences). Discontinuities in the double-differenced phases are then isolated by searching the between-epochs differenced residuals for outliers. The outliers are subsequently assumed to be the cycle slips and are rounded to the nearest integer value. The receiver channel in which the slip occurred is identified and the cycle slip removed from all subsequent observations on that channel. If the base-station base- satellite concept proposed by Goad (1985b) is used, the cycle slips pertaining to the base-site or base-satellite are transferred to the subsequent data pertaining to all other sites and satellites. However, this incorrect transfer of the cycle slip to the other receiver channels cancels out during double-differencing or when the station-satellite clock offsets at each epoch are estimated. Plots of the double-difference residuals before and after editing can be helpful to ensure that the procedure has worked correctly. Once the cycle slips have been edited from the data, the station positions are recomputed using the edited data.

5.2.6 ALTERNATIVES TO DIFFERENCING

As each one-way phase contributes to more than one differenced observable, the differenced observables are correlated. For multi-station networks these correlations are difficult to keep track of. From this point of view, algorithms which operate on the undifferenced phase observable and for which the oscillator errors are eliminated have distinct advantages. Goad (1985b), Hatch & Larsen (1985) and Wu (1984), have developed such algorithms. In these methods, the clock errors are not modelled as a polynomial function (Equation 5.16). Instead, they are regarded as nuisance parameters, and eliminated at each epoch. In effect, the clock offsets or oscillator errors are determined from the measurements made at a single epoch. This approach is sometimes referred to in the literature as "single difference" (Goad & Remondi 1984) or "undifferenced" (Goad 1985b) processing to indicate the manner in which the observable enters the normal equations. However, the reader should note that measurement differencing achieves the same effect without explicitly determining these parameters at each epoch.

Two distinct methods of achieving this implicit differencing have been devised. In the first method the partitioning scheme described in Section 5.2.1 is applied to the normal equations to eliminate the oscillator terms at every epoch (Goad 1985b, Hatch & Larsen 1985). In this way, the epoch parameters are eliminated as the observations for a particular epoch are processed. Hence, only the equations for the epoch parameters being processed are stored in computer memory. Typically, the epoch terms are the oscillator errors. Literally hundreds of epoch parameters can be solved for using a minimum of computer memory. Moreover, the unnecessary additions and multiplications of zero terms is avoided. The equations for the common parameters are kept in computer memory for the entire process. They include terms for the station positions, satellite orbital parameters and biases.

On the other hand, Bock et al (1985) and Wu (1984) transform the data acquired at a particular epoch to derived observations containing no clock terms by means of orthogonalisation algorithms. The principles of the orthogonalisation methods were discussed in Section 5.2.1. Partitioning, orthogonalisation and differencing give the same answers provided the variance-covariances of the observations are correctly propagated in the solution procedure. If only the minimum complement of receivers and satellites is available to estimate the clock terms (two receivers and two satellites), orthogonalisation and simple differencing are in fact identical algorithms.

The pros and cons of either method are difficult to quantify. Differencing is simple to implement on a computer and results in fewer parameters to be estimated. However, the correlations between differenced observations are difficult to keep track of in a multi-station computation. The correlation problem does not arise with orthogonalisation. However, unmodelled systematic errors, like cycle slips in the one-way phases, are difficult to detect in the orthogonalised observations. However, these errors can often be isolated by examining differences between the non-orthogonalised observations. Partitioning also overcomes the correlation problems of differencing but the solution requires considerable computer file space if the epoch parameters are needed. One advantage of partitioning over the other techniques is that unequal precisions for the observations from different receivers can be readily modelled.

5.2.7 NETWORK ADJUSTMENT

Historically, most GPS survey data has been processed on a baseline by baseline basis. Networks are formed by combining individual baselines in a subsequent network adjustment. The change from single baseline solutions to network solutions gives increased redundancy and closure. Therefore, positions obtained from a rigorous network adjustment in which all the data comprising individual baselines are adjusted simultaneously should be superior to single baseline calculations. Increased redundancy also allows field mistakes, like incorrectly placed antennae, to be isolated.

Multi-station data can be processed either:

(a) as single baselines, and then combined in a rigorous adjustment with the appropriate variance-covariance matrices (Bock et al 1985), or

(b) in a one-step multi-station adjustment of all sites observed simultaneously -- in the same "session" (Beutler et al 1984, Nakiboglu et al 1985, Masters & Stolz 1985), or

(c) by a combination of (a) and (b) in which several smaller networks are combined in a rigorous adjustment (the "multi-session" option).

Multi-session procedures are an extension of the multi-station approach and partitioning schemes can be used to make these procedures extremely efficient (Section 5.2.1).

The software required to process single baselines is simple compared to that required in the other options. However, the correlations between the individual baselines are ignored and this may cause problems for high accuracy surveys. The single-step multi-station method lends itself to orbit improvement along with the position determination (Section 7.3). However, the pros and cons of simultaneous orbit improvement and position determination have not been fully investigated. Masters & Stolz (1985) found that orbit adjustment seems to improve the geodetic solutions only when poor quality ephemerides are available or when long baselines are included. This is substantiated by the results of Beutler et al (1985). Nonetheless further study is required in this area.

5.3 ERROR ANALYSIS

Any physical process which affects the observations and is not properly accounted for in the theoretical model of the observables produces errors in the parameter estimation procedure. The main error sources in GPS surveying are:

(1) instrumental effects, including multi-path and instabilities in the transmitter and receiver clocks;

(2) propagation medium effects; and

(3) deficiencies in the dynamical models used to determine the relative motions of the GPS satellites and the observing sites.

Satellite clock, receiver clock and orbital errors usually dominate in undifferenced and singly-differenced observations. In doubly-differenced observations, particularly carrier phase measurements, instrumental effects are much reduced and one is concerned primarily with orbit and propagation medium errors.

5.3.1 INSTRUMENTAL EFFECTS

In the category of instrumental errors, we include those resulting from both the transmitting and receiving systems. The most important errors for (undifferenced) carrier beat phase and singly-differenced phase observations are the fluctuations in the oscillators. The crystal oscillators normally used in GPS receivers have a fractional stability of no better than 1 part in 10^{10}, which implies variations in the carrier beat phase of 100 cycles in an hour. With atomic oscillators the stability is better -- typically 1 part in 10^{11} for a Rubidium oscillator and 1 part in 10^{12} for a Cesium oscillator over an interval of a few minutes to several hours. However, even a stability of 10^{12} would produce a change of 1 metre in the signal pathlength in one hour. This is adequate for relative position determination where stations are separated by 1000 km or more, for low accuracy orbit determination (3-10ppm), and for point positioning, but not for most surveying applications. Hence, for the most important application of singly-differenced observations -- orbit determination -- even more stable atomic oscillators are desired: hydrogen masers, which have a fractional stability of a few parts in 10^{14} or better.

In the doubly-differenced observable (Section 5.1.4) commonly used for surveying, the effects of frequency-standard fluctuations are considerably reduced. Nevertheless, there are three types of clock error that are important:

1) an epoch offset from UTC,
2) an epoch offset between the two receivers, and
3) a rate difference between the two receivers.

An epoch offset from UTC is common to both receivers, will result in the ephemeris being interpolated for the wrong time. The error introduced into the baseline measurement is given by the product of the timing error and the satellite angular velocity (about 1.5×10^{-4} rad s^{-1}). Hence, we need to synchronise the receiver clocks to UTC within about 7 milliseconds to keep the error below 1 ppm.

Synchronisation of the two receiver clocks should produce more precise results because the error is no longer the same for both receivers. The error introduced into the observable is the product of the timing error and the radial satellite velocity (1 km s^{-1}). The receiver clock offset should be kept below 3 microseconds to reduce this error below 1 cm. For long baselines or low-accuracy surveys, offsets of 10 microseconds can be tolerated. On the other hand, clock synchronisation is critical for high-accuracy, short-baseline work.

For a rate difference of 1 part in 10^9 (1.5 Hz at L1), the epoch difference will grow to 4 microseconds after 1 hour. Thus, the difference in drift rate between the two receiver oscillators is usually not a problem, unless the oscillators are badly out of adjustment or have not been warmed up properly. In any case, rate differences can be estimated from the between-station differenced phase observable. Hence, it is good practice to examine these single-differences as a part of the double-difference analysis.

If the receiver clocks are properly synchronized, the largest instrumental error remaining in doubly-differenced phase observations is likely to be signal multipath. Multipath is caused by reflection of the signal from the ground or other objects before reaching the antenna. The effect on the measured phase depends on the nature of the surrounding terrain, the height of the antenna above the ground and the antenna design. The magnitude of the effect, which is generally no more than a few centimetres (Young et al 1985), can be reduced to a few millimetres with some antenna designs (Counselman & Gourevitch 1981). For an antenna located a few metres above the ground, the period of multi-path variation is of the order of 10 minutes, so that some averaging occurs over longer observation spans.

5.3.2 PROPAGATION MEDIUM

The effect of the troposphere on the propagation delays can be modelled using measurements of the surface temperature, pressure and relative humidity (or dewpoint). For observations 20° or more above the horizon, the delay is given with sufficient accuracy by (Saastamoinen 1973):

$$\tau_{atm} = 7.595\times10^{-12} \sec z \,[P + (\frac{1255}{T} + 0.05)e - \tan^2 z\,] \qquad (5.74)$$

where

τ_{atm} is the atmospheric delay in seconds
z is the zenith angle of the satellite
P is the pressure in millibars
T is the temperature in °K (=°C + 273.16) and
e is the partial pressure of water vapour in millibars.

The partial pressure of water vapour may be computed from the relative humidity, RH (as a fraction of 1.), by:

$$e = 6.108\ RH \exp \left[\frac{(17.15T - 4684)}{T - 38.45} \right]$$

and the pressure P, at height above sea level, h (in kilometres) is given in terms of the surface pressure P_s, and temperature T_s, by:

$$P = P_s \left[\frac{T_s - 4.5h}{T_s} \right]^{7.58}$$

The part of the model accounting for the dry component of the troposphere is accurate to better than 1 cm. The effect of water vapour, which can be tens of centimetres at microwave frequencies is more difficult to model mainly because the surface measurements of relative humidity do not accurately reflect the distribution of water vapour along the signal path. Large clouds of water vapour in particular can introduce significant fluctuations in signal pathlength. (A useful rule of thumb is that the effect of water vapour on pathlength is about 3 cm when the cloud colour is just discernibly grey).

For surveys of points separated by less than a few tens of kilometres the tropospheric delay is nearly the same at both ends of a baseline and will cancel in the between-stations difference or double difference observable. Hence it is important to use the same model for computing the tropospheric delay at each end of the baseline during the data analysis. The best approach is to model the delay using a single measurement of pressure, temperature and humidity at both stations. If the stations are at different heights above sea level, the pressure used for each station should be corrected for this difference. Temporal changes in pressure can usually be measured accurately and these can be incorporated in the model (although they are usually too small to be important in observations over a few hours). On the other hand, temporal changes in water vapour along the pathlength are not so well defined by surface measurements of humidity; use of these measurements will serve mainly to introduce noise into the model. A preferred method in the presence of large and changing cloud formations is to observe for long enough to average out the fluctuations. Nevertheless, under a wide range of conditions, including rain, the error in baseline coordinates introduced by the troposphere will be less than 1 ppm.

The effect of the ionosphere on signal pathlength can be kept below 1 cm using dual-frequency observations (see Section 5.4). With single-frequency observations made in a normal, daytime, mid-latitude environment, the effect on baselines will be roughly 1-2 ppm for a two hour observation span. For data spans of 15 minutes or so, the effect may be several times worse. Hence, to achieve survey accuracies of about 5 ppm, a dual-frequency instrument may not be mandatory. Moreover, at night, the ionospheric delay is a factor of 5 or more less than for daytime observations. Accordingly, 1 ppm accuracy can usually be achieved by scheduling nighttime observations of two hours duration.

5.3.3 EPHEMERIS ERRORS

The nature of the GPS ephemerides is discussed in Section 7.4. An error in the satellite ephemeris produces the same error in a surveyed baseline scaled by the ratio of the baseline length to the satellite altitude (20 200 km). Of course, this is only approximately true and the precise effects of orbital errors will depend on the geometry of the observations. If more than four satellites are observed, then the effect of orbital error should be considerably reduced.

The surveyor should bear in mind that the broadcast ephemeris for a particular satellite may be degraded from time to time. This may be the case for several days after an orbit correction manoeuvre has taken place or during the annual, two-month long "eclipse" period when a satellite enters the earth's shadow (Fliegel et al 1985). During this eclipse period, the satellite often experiences non-gravitational accelerations caused by irregular thermal emissions from the satellite and unbalanced attitude-control thrusts. When this occurs, it is not unusual for the satellite ephemeris to be in error by more than 100 m, corresponding to a baseline measurement error of 5 ppm. These orbital manoeuvres are introduced without warning. Hence some sort of orbit monitoring service would be helpful to determine when this occurs. We recommend that the surveyor build some redundancy into his observations (ie. observe more than the minimum number of satellites) to reduce the effect of this problem.

5.4 DUAL-FREQUENCY OBSERVATIONS

With the improved orbital data that is expected to become available in the near future (better than 1 ppm), the ionosphere will become the dominant source of error in single-frequency observations. Measurements can be made at both L-band frequencies to remove the ionospheric effects on GPS phase and pseudo-range observations. Removing these effects may result in high accuracy (1-5 ppm) from much shorter observing times (see Ladd et al 1985).

5.4.1 ELIMINATING IONOSPHERIC EFFECTS

The ionospheric contribution to the variation of the pathlength of a radio signal is proportional to the highly-variable ion concentration of the ionosphere and inversely proportional to the square of the signal frequency. The effect on phase delay, defined as phase divided by frequency, is equal to, but opposite in sign, to the effect on group delay, the derivative of phase with respect to frequency. The group delay is associated with pseudo-range measurements. For phase observations the ionospheric delay is:

$$\tau_{ion} = \phi_{ion}/f = -\frac{1.35\text{x}10^{-7}}{f^2} N_e \qquad (5.75)$$

where τ_{ion} is in seconds, ϕ is in cycles, f is the signal frequency in Hz, and N_e is the integrated electron content over the signal path in electrons m^{-2}. N_e is a function of the electron density and the elevation angle of the satellite. The electron density is a function of latitude, time of day, season of the year, and the current epoch within the approximate 11-year sunspot cycle. Variations in N_e lead to pathlength changes from less than a metre to tens of metres. In differential positioning the effects at each end of the baseline mostly cancel, since both sites observe the satellite through virtually the same ionosphere. This relationship degrades nearly linearly with site separation, so that the effect on baseline coordinates is roughly a constant fraction of the baseline length.

The "ionosphere-free" dual-frequency observable is readily obtained from Equation 5.75. We first express the geometric propagation delay (τ_g) in terms of the observed phases ϕ_1 and ϕ_2 for L1 and L2, respectively:

$$\tau_g = \phi_1/f_1 + 1.35\times10^{-7}N_e/f_1^2$$
$$\tau_g = \phi_2/f_2 + 1.35\times10^{-7}N_e/f_2^2 \qquad (5.76)$$

which gives:

$$\phi_{ion(L2)} = (f_1/f_2)\ \phi_{ion(L1)} \qquad (5.77)$$

The total observed "phase pathlength" (in cycles) observed at each frequency is the sum of the geometric phase pathlength, and the ionospheric correction:

$$\phi_1 = f_1\ \tau_g + \phi_{ion(L1)} \qquad (5.78)$$

$$\phi_2 = f_2\ \tau_g + \phi_{ion(L2)} \qquad (5.79)$$

Substituting Equation 5.77 in Equation 5.79 and subtracting the result from the product of Equation 5.78 and the factor R ($=f_1/f_2$) eliminates the ionospheric correction. Multiplying the result by f_2 gives:

$$f_1\ \phi_1 - f_2\ \phi_2 = f_1^2\ \tau_g - f_2^2\ \tau_g \qquad (5.80)$$

The geometric delay τ_g in Equation 5.80 can be expressed:

$$\tau_g = \frac{f_1\ \phi_1 - f_2\ \phi_2}{f_1^2 - f_2^2}$$
$$= \frac{1}{f_1}\ \frac{\phi_1 - R\ \phi_2}{1 - R^2} \qquad (5.81)$$

Alternatively, by adding and subtracting ϕ_1 from the right-hand-side of Equation 5.81 and combining terms, we obtain a "combined", ionosphere-free, phase observable, ϕ_c, that is:

$$\phi_c = f_1\ \tau_g = \phi_1 - \frac{R\ (\phi_2 - R\ \phi_1)}{1 - R^2} \qquad (5.82)$$

Substituting R = (1227.6/1575.42) = 0.779, we get:

$$\phi_c = \phi_1 - 1.984(\phi_2 - 0.779\ \phi_1) \qquad (5.83)$$

This <u>ionosphere-free phase observable</u> can be used for geodetic positioning in exactly the same manner as single-frequency observables except that the resolution of ambiguity biases is more complicated (see Section 5.4.2). For short baselines, the ionosphere-free phase observable is noisier than the corresponding single frequency measurement. This results from the fact that the likely dominant error source will be multi-path which is magnified when the two are combined (since it affects the L1 and L2 phase differently). Accordingly, it may be preferable to use single frequency measurements for short baseline work.

Other strategies have been devised to deal with the ionospheric effect. Least squares filtering techniques are employed to take advantage of the high correlation introduced by the ionosphere between observing epochs. Dual-frequency pseudo-ranges, if they are available, can also be used to determine preliminary values for the ionospheric effect. These values can in turn be used as a priori estimates for the ionospheric effect on carrier phase observations.

5.4.2 AMBIGUITY RESOLUTION

With dual-frequency measurements, ambiguity resolution is much more complex than is the case for single-frequency observations (discussed in Section 5.2.4). Bender & Larden (1985) have developed the theory of dual-frequency ambiguity resolution and propose strategies for getting around the problem. If the ambiguity bias terms are included, Equation 5.83 should be written as:

$$\begin{aligned}\phi_c &= \phi_1 + n_1 - 1.984[\phi_2 + n_2 - 0.779(\phi_1 + n_1)] + \phi_e \\ &= \phi_1 - 1.984(\phi_2 - 0.779\phi_1) + (2.546n_1 - 1.984n_2) + \phi_e\end{aligned} \tag{5.84}$$

where n_1, ϕ_1 are respectively the integer and fractional phase values at the L1 frequency and n_2, ϕ_2 are respectively the integer and fractional phase values at the L2 frequency. ϕ_e represents the errors in our models arising from sources other than the ionosphere (for example: troposphere, satellite orbits, and clocks).

Equation 5.84 applies to doubly-differenced phase as well as for the undifferenced carrier beat phase observations. The third term in Equation 5.84, which arises from errors in our estimate of the unknown number of integer cycles at the L1 and L2 frequencies can assume only discrete non-integer values, in contrast to single frequency measurements where it can only be an integer. This is also true for double-differences if the values of n are the combined integer biases for the double-difference observable. In some cases the interval between the consecutive values of the bias terms will be less than one cycle, making it difficult to determine the correct values for n_1 and n_2. For example, with an error of 1 cycle at L1 (Δn_1=1) and 1 cycle at L2 (Δn_2=1) (or for Δn_1=-1, Δn_2=-1), the change in ϕ_c is 0.56 cycle. An error of (3,4) for (n_1,n_2) leads to a change in ϕ_c of 0.30 cycles; and an error of (7,9) for (n_1,n_2) leads to a change in ϕ_c of only 0.03 cycles!

Dual-frequency ambiguity resolution is usually performed for baseline lengths for which other sources of error, mainly satellite orbits and the troposphere, contribute less than a half-cycle. The reader will recall that for a code-correlating receiver, such as the TI 4100, one cycle for L1 equals 19 cm whereas for some codeless receivers, such as the Macrometer II, one cycle equals 9.5 cm (see Chapter 2). A 2 m orbital error produces a 1 cm baseline error for a 100 km baseline. The main contribution to the tropospheric error arises from water vapour and this might be 1-5 cm. Under normal daytime conditions, the ionospheric error in single-frequency observations probably does not exceed 2 ppm, corresponding to 20 cm for this baseline. Using observations made only at L1, the error in n_1 (Equation 5.84) should be less than 7 cycles and probably less than 3 cycles. With code-correlating receivers, pseudo-range measurements at L1 and L2 could be used to restrict the range of values of n_1 and n_2 to within acceptable limits. If pseudo-range measurements are employed or the the critical combinations (3,4) and (7,9) of n_1 and n_2 can be ruled out, the error in the ambiguity reduces to $\Delta n_1 = \Delta n_2 = \pm 1$, which makes the minimum possible difference between two choices of ambiguity equal to 0.56 cycles. Hence in order to resolve the ambiguity, the total contribution from troposphere, ionosphere, and orbits in the combined observable, ϕ_c, should be less than about 10 cm with code-correlating receivers or 5 cm with codeless receivers.

6. HEIGHT DETERMINATION USING GPS

6.1 INTRODUCTION

A satellite positioning system such as GPS provides ground coordinates of a receiver in a Cartesian coordinate system with the origin at the earth's geocentre. More precisely, the origin, orientation and scale of this Cartesian coordinate system are defined by the "fixed" coordinates of an ensemble of tracking stations that collect the data used to compute the GPS satellite orbits (Section 4.2.3).

The Cartesian coordinates may then be transformed into geodetic latitude, longitude and ellipsoidal height, with respect to some reference ellipsoid, or spheroid, by well known relations (Torge 1980). The reference ellipsoid may have its origin at the geocentre or it may be an arbitrarily located ellipsoid, such as one implicit in the definition of a regional geodetic datum (see Section 4.3). The Australian National Spheroid is an example of a non-geocentric reference ellipsoid.

Surveyors and engineers are in most cases interested in the orthometric height as measured above some datum reference surface, identified as the geoid. The geoid is one of a whole family of equipotential, or level, surfaces of the earth's gravity field. Most geodetic measurements, by virtue of their reliance on the local horizon, are influenced by the earth's gravity field. Level surfaces, as the name implies, are surfaces of constant gravitational potential. (Level surfaces are also known as geops.) The gravity vector, or the direction of the vertical, at any point is perpendicular to the geop passing through that point.

Orthometric height has greater "physical" meaning than the essentially "geometrical" ellipsoidal height. Orthometric height has traditionally been determined by the technique of levelling in which increments in height are obtained from the intersection of the line-of-sight of a level instrument, tangential to the geop passing through the level axis, on two graduated staves, as illustrated in Figure 6.1. Orthometric height information is needed for precise engineering operations such as dams, pipelines, tunnels, etc., which deal with fluids and their flow. There is no analogous physical interpretation for ellipsoidal heights.

It was recognised last century that the mean surface of the oceans was a good approximation to a level surface of the earth's gravity field and this surface was selected as the datum surface for elevations. The geoid has become a very useful concept for the practical determination of orthometric heights and statements such as "heights above mean sea level" and "height above geoid" are considered equivalent in the context of most surveying applications.

The difference between the ellipsoidal height h, the quantity that is measured with GPS, and the orthometric height H, the quantity that is required, is the geoid height or geoid undulation N, given simply by:

$$H = h - N \tag{6.1}$$

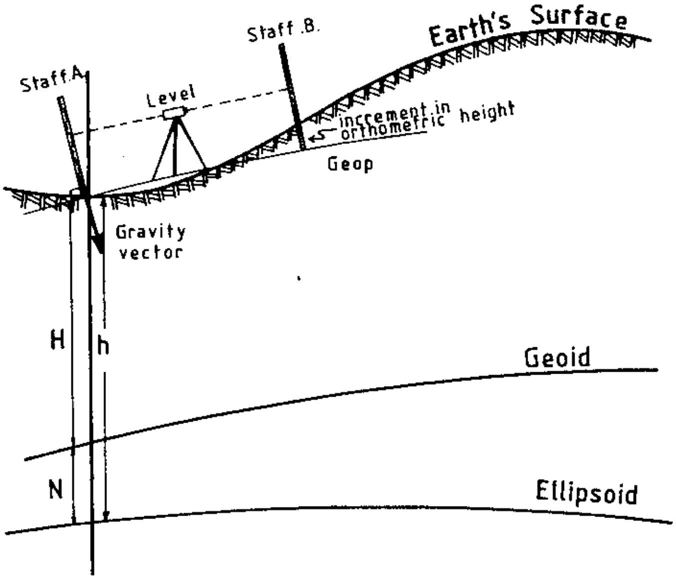


Figure 6.1 Geoid, Ellipsoid, Earth's Surface and Levelling.

An important task of physical geodesy has therefore been to "determine" the geoid or, more precisely, the height of the geoid above a reference ellipsoid. As a result of density variations within the earth, the geop surfaces, including the geoid, have irregular shapes. The magnitude of geoid undulations with respect to a "best-fitting" reference ellipsoid is of the order of ±100 m. Figures 6.2a and 6.2b are geoid contour maps showing different amounts of detail. In geologically disturbed areas it is not uncommon to have geoid slopes of the order of ten metres in a hundred kilometres or so.

For GPS survey applications, a geoid model should provide geoid heights with an accuracy commensurate with that of the ellipsoidal height, if the accuracy of the derived orthometric heights is not to be reduced. GPS point-positioning, using 30 minutes to 4 hours of observations on 4 to 5 satellites, can provide accuracies at the 5 to 10 metre level for each coordinate (Bock et al 1984). Using dual-frequency carrier phase measurement and higher precision orbits, point-positioning at the few metre level may eventually be possible on a routine basis. Consequently, the geoid height needs to be known to this accuracy or better. Differential positioning will be the most widely used procedure for GPS surveying and therefore only the relative geoid height between two stations will be required. The determination of height when GPS is deployed in the translocation mode is discussed in Section 6.4.

6.2 GEOID COMPUTATION TECHNIQUES

Geoid models can take the form of point values, profiles or a continuous surface, and exhibit varying degrees of detail. For example, the

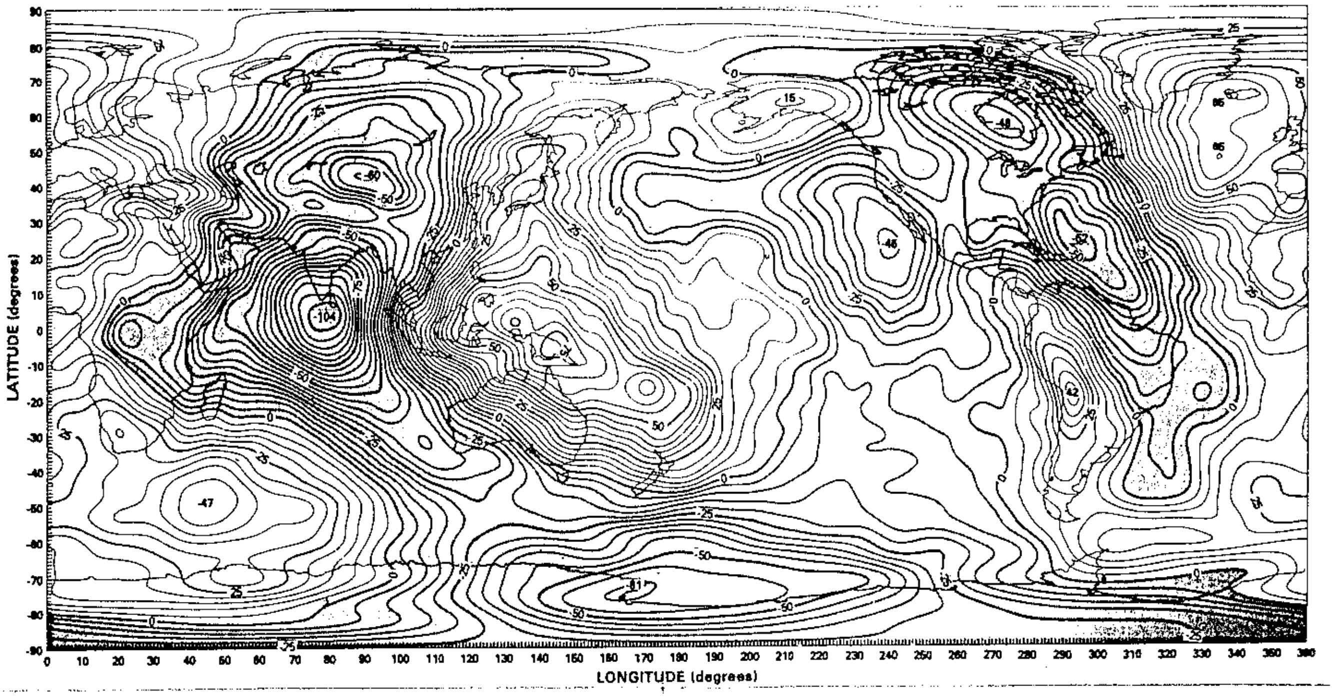


Figure 6.2a Geoid Model Inferred from Spherical Harmonic Coefficients Truncated at Degree and Order n'=36.

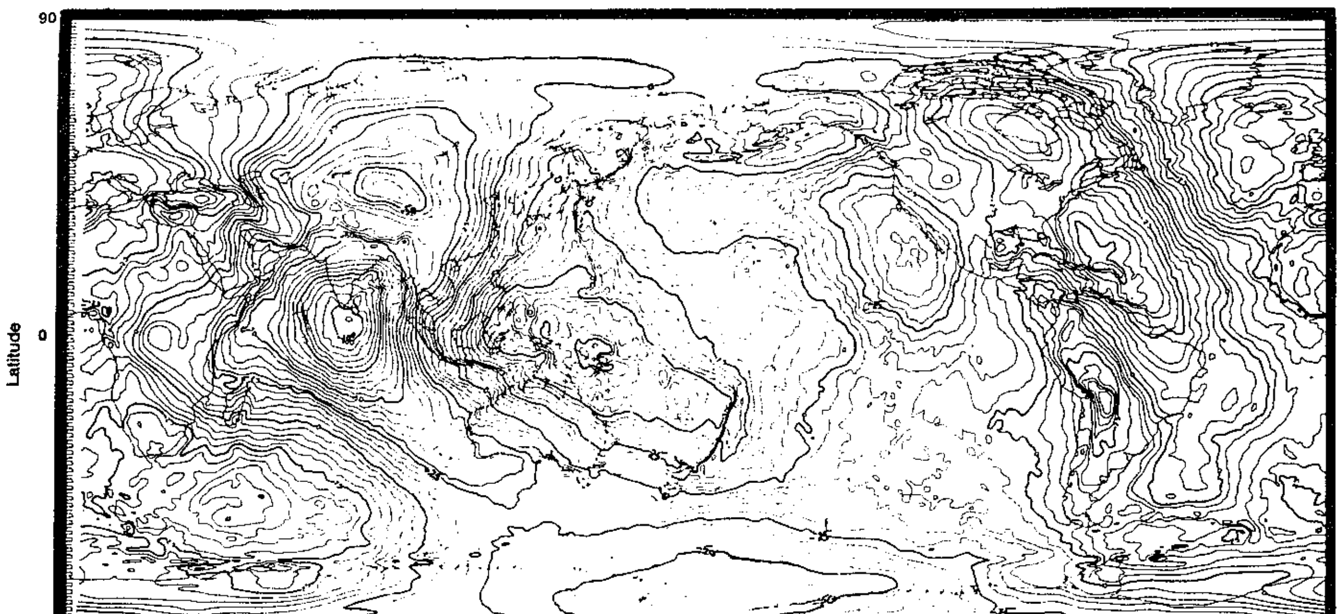


Figure 6.2b Geoid Model Inferred from Spherical Harmonic Coefficients Truncated to Degree and Order n'=180.

spectrum of geoid height information contained in a model may be essentially of short, medium or long wavelength: short wavelength information is sensitive to geoid features of extent < 100 km, medium wavelengths contain information on trends in the geoid over distances between 100 and 1000 km and long wavelengths represent geoid features of extent > 1000 km. The models may be global in nature or representative of only a local area.

All techniques of geoid determination on land utilise one or more of the following types of data:

(a) gravity measured at the surface of the earth.
(b) tracking data to near-earth artificial satellites.
(c) three-dimensional satellite derived position.
(d) astronomical observations.
(e) geodetic levelling.

In some of these techniques, for example, gravity measurements, the geoid heights are obtained indirectly. In the other cases the geoid shape is established directly from linear and angular measurement or three-dimensional positions, or it is inferred from a knowledge of the earth's geopotential, and combinations of some or all of the above.

Geoid models may therefore be classified according to the technique employed for their computation: astrogeodetic geoids, geoid models represented by geopotential models ("satellite geoids"), gravimetric geoids and geoids determined from three-dimensional position fixes and levelling data. It is not the intention here to give anything more than a brief review of the common geoid computation techniques. For further details the reader is referred to Rizos (1980) and the references contained therein.

6.2.1 ASTROGEODETIC GEOIDS

The normal to the reference ellipsoid passing through a point A intersects the local vertical (the normal to the geop through A) at some angle ε, called the deflection of the vertical. The angle ε is measured in the plane, at some arbitrary azimuth α, containing both normals, such that:

$$\varepsilon = \xi \cos\alpha + \eta \sin\alpha \tag{6.2}$$

where ξ, η are the components of the deflection of the vertical in the direction of the meridian and prime vertical respectively.

The increment dN in geoid height over a distance ds gives the slope of the geoid with respect to the ellipsoid:

$$dN = -\varepsilon'.ds \tag{6.3}$$

where ε' is the deflection of the vertical referred to the geoid, and is related to the ground level deflection ε by formulae given in Heiskanen & Moritz (1967).

The ground level deflections are computed by comparing the geodetic and astronomic coordinates at the same point. The relationship between geodetic latitude ϕ, geodetic longitude λ, on the one hand, and astronomic latitude Φ, astronomic longitude Λ, on the other is:

$$\xi = \Phi - \phi$$
$$\eta = (\Lambda - \lambda)\cos\phi \qquad (6.4)$$

The geodetic coordinates are established by GPS,TRANSIT or terrestrial measurements. The astronomic coordinates are determined directly from observations to stars. Although the astronomic coordinates of a point on the earth are unique this is not necessarily the case with the geodetic coordinates. Geodetic coordinates may refer to a geocentric ellipsoid or they may refer to some arbitrarily defined reference ellipsoid as in the case of the Australian Geodetic Datum (Allman & Veenstra 1984).

If the deflections of the vertical along a profile AB are known the shape of the geoid can be obtained by integrating Equation 6.3, that is,

$$N_B = N_A - \int_B^A \varepsilon'.ds \qquad (6.5)$$

Selecting suitable profiles and adopting a value for the geoid height at the origin of profiles allows a geoid map to be constructed. The method of determining a geoid model using Equation 6.5 is known as <u>astrogeodetic levelling</u>. A detailed description of the method is given by Heiskanen & Moritz (1967), while the practical aspects, accuracies obtainable, etc. are discussed in Bomford (1971).

Because of its high cost and limited accuracy, the technique of astro-geodetic geoid determination is of limited relevance in the context of converting GPS heights to orthometric values.

6.2.2 <u>GEOIDS AND GRAVITY FIELD MODELS</u>

The geoid height at a point q with spherical coordinates (ϕ_q,λ_q,R_q), can be approximated by the truncated spherical harmonic series:

$$N_q = \frac{Gm_e}{R_q\gamma_q}\sum_{n=2}^{n'}\left(\frac{a_e}{R_q}\right)^n\sum_{m=0}^{n} P_{nm}(\sin\phi_q)(C_{nm}\cos m\lambda_q + S_{nm}\sin m\lambda_q) \qquad (6.6)$$

where a_e is the mean radius of the earth, Gm_e is the product of the gravitational constant and mass of the earth, γ_q is normal gravity at the point, $P_{nm}(\sin\phi)$ is the Legendre function of degree n and order m, and C_{nm},S_{nm} are dimensionless spherical harmonic coefficients.

The irregular mass distribution of the earth -- manifested by a non-symmetric gravity field -- is by far the dominant cause of the perturbations of near-earth satellite orbits from an ellipse (see Section 7.2). Therefore the coefficients C_{nm},S_{nm} can be determined from the analysis of the orbital perturbations of artificial satellites. However, an inspection of Equation 6.6 reveals that the effect of the higher degree harmonics is rapidly diminished as the term $(a_e/R)^n \ll 1$ for large n and for satellite altitudes $R>a_e$. Thus only harmonics up to some limiting degree n' (about 36) can be adequately determined by satellite perturbation analysis. The magnitude of the coefficients decreases with increasing degree according to the approximate relation:

$$C_{nm},S_{nm} \cong 10^{-5}/n^2 \qquad (6.7)$$

The finite set of coefficients C_{nm}, S_{nm} together with the constants a_e and Gm_e constitute the **geopotential** or **gravity field model** of the earth. Recent gravity field models include GEM L2 (Lerch et al 1983) and GRIM3 L1 (Reigber et al 1985).

Although harmonic coefficients of high degree ($n > n'$) cannot be reliably estimated from satellite observations, they still contribute significantly to the geoid height signal at the surface of the earth. Therefore any evaluation of N based on Equation 6.6 with satellite-derived harmonic coefficients will result in an over-smoothed estimate. For example, the most recent gravity field models produced from the analysis of satellite perturbations and surface gravity data are truncated to about degree and order 36 (approximately 1360 coefficients), equivalent to a half wavelength resolution of about 500 km. Figure 6.2a illustrates the global geoid inferred from gravity field model to degree and order 36.

Very High Degree Geopotential Models

In recent years gravity field models of very high degree and order ($n'=180$) have been developed (Lerch et al 1981, Rapp 1981), based on the analysis of surface gravity data and altimeter-derived marine geoid heights. Such models consist of a very large number of coefficients ($\cong$32760) and special computer algorithms are required to evaluate N using Equation 6.6. They offer very high geoid model resolution, equivalent to 1°x1° block values, or half wavelengths of the order of 100 km. The accuracies over these wavelengths are estimated to be of the order of ±1-2 m. Figure 6.2b illustrates the geoid inferred from a geopotential model complete to degree and order 180. Note the marked increase in geoidal detail over that found in Figure 6.2a.

A gravity field model based on a spherical harmonic representation has a number of advantages as a source of geoid information:

(1) It is global in nature and therefore valid for any geographic location.

(2) It is a more convenient geoid model than a map, as it consists merely of a set of coefficients which can be used to evaluate geoid height as and where it is required.

(3) Such a geoid model is geocentric in nature and therefore N can be easily related to the GPS datum.

6.2.3 GRAVIMETRIC GEOIDS

In 1849 G.G. Stokes presented a practical method for the computation of geoid heights using **surface gravity measurements**. An outline of the method can be found in most geodesy textbooks (for example, Heiskanen & Moritz 1967) and in its most elementary form can be represented by the equation:

$$N_p = \frac{R}{4\pi\gamma} \iint f(\psi)\ \Delta g . d\sigma \qquad (6.8)$$

where N_p is the geoid height at the point P, ψ is the angle between geocentric radii to P and the element of surface area $d\sigma$, R and γ are the mean earth radius and mean gravity respectively, Δg is the gravity anomaly associated with surface element $d\sigma$ and $f(\psi)$ is Stokes' function (IBID 1967, p.94). The gravity anomaly is observed gravity at the geoid minus the normal gravity at the equivalent point on the ellipsoid. The integration in Equation 6.8 is performed over the whole globe, requiring values of the gravity anomaly to be known everywhere on the geoid.

Most areas, outside the developed countries and their continental shelf areas, are devoid of gravity data. However the contribution to Equation 6.8 from gravity anomalies of areas remote from P decreases rather rapidly with increasing ψ and is relatively small for $\psi > 30°$. For computational ease the evaluation of Stokes' integral is nowadays only carried out for some spherical cap, with angular radius ψ_0 (ranging from 1.5° to over 30°), centred on the computation point. Contributions to the geoid height from areas beyond ψ_0 are most easily obtained by using spherical harmonic models of the geopotential (Rizos 1980). Therefore errors in gravimetrically-derived geoid height will arise from the following sources: errors in the gravity anomaly values, the precise scheme implemented for Equation 6.8. and errors in the gravity field model for areas beyond ψ_0.

6.2.4 GEOMETRICALLY DETERMINED GEOIDS

If both the ellipsoidal heights (from GPS or TRANSIT observations) and orthometric heights (from precise levelling) are known at a number of points within an area, then a geoid contour map can be constructed by fitting a surface through these points (using Equation 6.1). This model may then be used to interpolate the geoid height at intermediate locations where the orthometric height is not known, and for which it is required (Figure 6.3).

This procedure is analogous to the astrogeodetic technique described above. In both techniques "absolute" coordinates (the astronomical coordinates in the one case and the ellipsoidal height in the other) are compared with geodetic quantities (ϕ,λ,h) to infer information on the geoid shape. This information is then used to develop a geoid model for a region. Aside from the observational errors, the major limitation to accuracy in both techniques is the assumption that the geoid slope is smooth between adjacent stations.

Geoid from TRANSIT Data

At present most of the ellipsoidal height information has been obtained from TRANSIT satellite observations. Many TRANSIT stations have been established in Australia and elsewhere in the world for strengthening existing first order control networks. However, the station spacing has been such that the preparation of a meaningful geoid model from TRANSIT-derived ellipsoidal height has not been possible. Nevertheless, these sparsely distributed point-geoid height estimates do provide a valuable dataset when used in combination with an independently derived geoid model. In particular the TRANSIT geoid heights can be used to determine the regional biases present in global geoid models inferred from spherical harmonic coefficients (Section 6.2.2).

For example, Allman (1982) has developed a geoid model for SE Asia and the Pacific region combining the GEM 10C geopotential model (Lerch et al 1981) with over 600 TRANSIT-derived geoid heights. Although this model is still essentially a smoothed representation of the geoid (no additional geoid detail over and above that already contained in the geopotential model has been added), it has been translated and tilted to

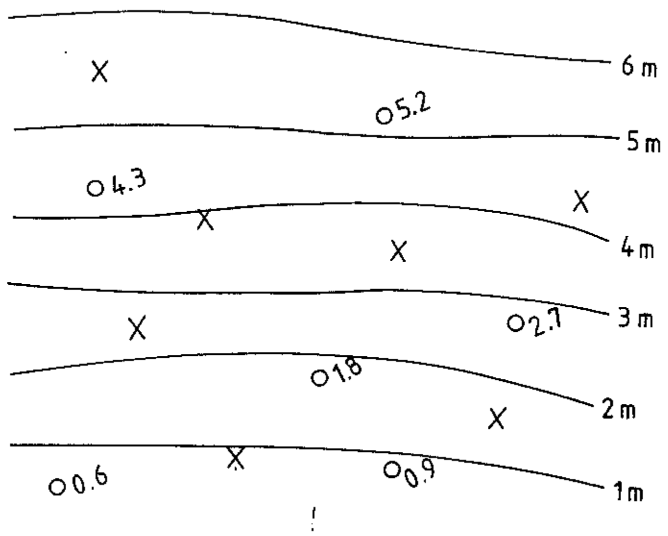


Figure 6.3 A Geoid Contour Map from Point-Geoid Height Values.

° Points of known orthometric and geoid height (Equation 6.1)
x Points of GPS survey where geoid heights are required

better conform to the regional geoid trends as represented by the TRANSIT data.

Geoid from GPS

The same surface fitting technique can be used with GPS-derived geoid heights. An important difference is that GPS stations are likely to be closer together, of the order of a few tens of kilometres, and the local geoid can be estimated directly if levelling data are available in the area. This, together with its relative simplicity makes the method a very attractive and practical one for correcting GPS heights.

Instead of constructing a geoid contour map directly from the benchmark information (as in Figure 6.3), it is also possible to use the two sets of elevations (that is, levelled and GPS-derived) to define two distinct planes. The published, levelled elevations are referred to the geoid and the GPS elevations are referred to an ellipsoidal surface. If both elevations are made equal at one benchmark then, in general, the other benchmarks will have two elevation values. A simple mathematical model:

$$N = AX + BY + C \tag{6.9}$$

can be used to determine the east-west and north-south tilt of the two (actually curved) surfaces and their separation. These values are designated in Equation 6.9 by the coefficients A,B,C respectively, and X,Y are, for example, UTM coordinates. In such a simple surface fitting technique only three well-distributed benchmarks with both levelled and GPS elevations are needed. Once the coefficients A,B,C have been calculated, they can be used in Equation 6.9 to estimate the geoid height of any other point in the area. Naturally, if more than three benchmarks have orthometric and ellipsoidal height values, the three unknowns A,B,C can be solved by the Least Squares method. Alternatively, a more complex (curved) surface can be used in place of the plane model in Equation 6.9.

Estimating geoid heights using simple surface fitting techniques is already the most popular technique for GPS vertical surveying when a local, high quality, high resolution geoid model is not available (see, for example, Collins & Leick 1985). However, at present there are no guidelines regarding the benchmark spacing necessary to satisfy particular accuracy standards. However, this method appears to be adequate for areas up to 50 x 50 km where the geoid is smooth.

6.3 REMARKS ON "ABSOLUTE" GEOID DETERMINATION

Which surface is the geoid? Can it be recognised by some physical feature such as mean sea level? Rizos (1980) discusses these issues in some detail. However, it is not the definition of the geoid in an abstract or mathematical sense that is the problem; rather, the difficulty lies in identifying the geopotential surface to which the orthometric heights are referred. The geoid-corrected ellipsoidal heights can be directly compared with orthometric heights determined by traditional levelling methods, once this has been done.

Astrogeodetic geoids suffer because one has to adopt some value of geoid height for a point within the astrogeodetic levelling network. In the case of gravimetric solutions of the geoid, the situation regarding the datum or "zero degree" effect is quite complex. Stokes' integral is insensitive to effects of "zero degree" and therefore errors in estimating the potential of the geoid and in the definition of the height and gravity datum contribute to the uncertainty in the absolute geoid height. For geoids inferred from geopotential models the problem is that the geoid does not correspond to any physical entity at the earth's surface. On the other hand, such problems do not exist with geoids estimated geometrically. Because the orthometric heights at a number of stations are themselves used to infer the shape of the geoid, this geoid model is implicitly related to the zero height surface in the immediate vicinity of the GPS survey, and will absorb most of the systematic errors present in the height control.

Datum uncertainties and other long wavelength errors disappear, to a large extent, in relative geoid determinations and therefore weaknesses in the techniques of absolute geoid determinations are not a serious problem.

Australian Height Datum

Mean sea level (MSL) deviates from a geopotential surface by ±1-2 m due to inhomogeneities in sea water density and ocean circulation. Where the entire height system is based on one datum benchmark, be it a tide gauge or a trig control station, the situation is relatively simple. However, the Australian Height Datum (AHD) is based on zero elevation assigned to MSL at 30 tide gauges distributed around the Australian coastline (Roelse et al 1971). The zero orthometric height surface deviates by perhaps ±1-2 m from a geop of "best-fit" to these 30 MSL estimates. For Australia, orthometric heights determined using GPS-derived ellipsoidal height and an independently calculated geoid model will therefore not be related to the AHD to better than a metre or two. Moreover, this error will be a function of position. In such a case, a geoid model defined locally by a well spaced network of GPS stations of known AHD height will be the best approach for forcing height data obtained from GPS surveys to be compatible with the local height system. All other geoid models will suffer from metre level datum uncertainties.

6.4 RELATIVE HEIGHTS FROM GPS

The problem in relative positioning with GPS then to derive the orthometric height differences (ΔH) between points 1 and 2 from the ellipsoidal height differences (Δh). These differences can be obtained if geoid undulation differences (ΔN) are given, that is:

$$\begin{aligned}(H_2 - H_1) &= (h_2 - h_1) - (N_2 - N_1) \\ \Delta H &= \Delta h - \Delta N\end{aligned} \tag{6.10}$$

GPS is capable of determining ellipsoidal height differences over short to medium length baselines (<100 km) to accuracies of about 1-2 parts per million of the baseline length (that is, 5-10 cm over a 50 km baseline). The precision of the geoid height change between the two endpoints of a baseline should match that of the GPS survey so that the precision of the orthometric height differences computed from Equation 6.10 is not seriously degraded. However, although the accuracy requirements are more stringent in the case of relative geoid height determinations, the changes in geoid height can also be computed to a higher accuracy. This is because the dominant error sources in geoid determination techniques produce nearly equal geoid height errors at both endpoints and therefore tend to cancel when (N_2-N_1) is formed. In addition, as mentioned already, the datum or zero degree problem is eliminated. What are the strengths and weaknesses of the various techniques for relative geoid height determination?

6.4.1 RELATIVE GEOID HEIGHTS BY THE ASTROGEODETIC TECHNIQUE

The major shortcomings of astrogeodetic geoids are:

(1) Astrogeodetic traverses were observed in the process of establishing the basic geodetic control network. Additional astronomical observations are unlikely to be made.

(2) The limited accuracy of astronomical observations and atmospheric refraction corrections. For example, an uncertainty of 1" in astronomic latitude or longitude results in an uncertainty in the slope of the geoid of 0.1 m in 20 km (5 ppm).

(3) The accuracy of geoid determination diminishes as a function of increasing astrogeodetic station spacing. In Australia, changes of 20" or more can occur over distances of 50-100 km, with more rapid changes in mountainous and geologically disturbed areas.

(4) Because astrogeodetic stations are normally sited on hilltops, the geoid grades inferred from the deflections of the vertical are not necessarily representative of the area as a whole.

6.4.2 RELATIVE GEOID HEIGHTS FROM GEOPOTENTIAL MODELS

Estimating ΔN by differencing the geoid heights obtained from geopotential models (Equation 6.6) suffers from the problem that only the general trend of the geoid surface over large areas can be determined. Such models are global representations of the long wavelength features of the geoid and even very high degree models can only resolve geoid features with extents of about 100 km (Figure 6.2b). However, this method of determining relative geoid height is extremely simple to implement (the software can be built into GPS receivers). Where the geoid shape over an area is smooth and the accuracy requirements for the geoid height differences are not stringent as, for example, in most engineering and geophysical surveys, geopotential models will be useful for converting GPS heights to orthometric heights.

6.4.3 RELATIVE GEOID HEIGHTS BY THE GRAVIMETRIC TECHNIQUE

Geoid heights can be computed from terrestrial gravity data. The height of the geoid is evaluated at each end of a GPS baseline and the geoid height difference is computed. Provided the quality and coverage of the surface gravity data satisfies certain conditions and special computational techniques are used, geoid height differences with decimetre accuracies or better are possible. Kearsley (1984) presents a detailed discussion of data requirements and computational procedures necessary for high precision relative geoid height determinations. Kearsley (1984) showed that the utility of the gravimetric technique can be significantly improved if an integration cap of only 1° to 2° in extent is used and the remote zone contribution is represented by a high degree geopotential model (n'=180). For further convenience, the remote zone contribution could be precomputed on a grid and these values stored or mapped to allow subsequent interpolation. The interpolated value could then be added to the inner cap geoid height contribution computed from local gravity data (Equation 6.8) to complete the computation procedure. Further, the geoid heights could, in fact, be computed across the entire continent on, say, a 0.5°x 0.5° grid, tabulated and the data made available to all surveyors. Their task would be reduced to simply interpolating the geoid heights from the gridded data. Geoid height differences of subdecimetre accuracy have been obtained using this technique for GPS surveys in the U.S.A. (Engelis et al 1984) and in Germany (Engelis et al 1985).

6.4.4 RELATIVE GEOID HEIGHTS FROM GPS

Geoid height differences can be most directly determined for GPS satellite surveys using the GPS technique itself. This geometric method requires that the stations in a survey network with unknown orthometric heights be connected by GPS to benchmarks of known height. There is no confusion concerning the reference ellipsoid to which the geoid heights refer nor any concerning the relationship between the geoid and the height datum. However, as with the astrogeodetic technique, the required density of GPS stations collocated with vertical benchmarks will be a function of (a) the accuracy required for the orthometric heights to be determined, and (b) the variability of the geoid in the area. Therefore when high accuracy is sought and/or the local geoid is highly variable, the network of GPS benchmarks should be more closely spaced in order to define the geoid-ellipsoid separation.

6.4.5 SUMMARY

Astrogeodetic models of the geoid will play only a minor role in converting GPS heights to orthometric heights. They may be useful in areas where the geoid shape is smooth and the GPS survey is close to several astrogeodetic stations.

Geopotential models are most useful for long baselines (>50 km), when used in regions where the geoid is smooth, or where local fluctuations in geoid height are too small to affect the accuracy of the survey. They would therefore be most suitable for 2nd and 3rd order surveys over large geographic extents and in areas where vertical benchmarks are sparse or nonexistent. However, they are important in defining the remote zone contribution for gravimetric geoid determinations.

The gravimetric and geometric techniques of geoid height determination are the more suitable for supporting high accuracy vertical surveying using GPS. The two techniques have been compared in a number of tests and agreements at the few centimetre level have been obtained (Engelis et al 1984, Engelis et al 1985).

Which of the two techniques is preferable and what are the respective advantages and disadvantages? If gravity data are available, the gravimetric technique has great versatility because it requires no additional fieldwork. However, it places stringent requirements on gravity coverage, is computationally extensive and requires advanced geodetic knowledge. It is most likely to be the approach adopted in government departments or within those organisations with the necessary expertise. On the other hand, the computations could be performed once, by some geodetic agency, and the data, in the form of a fine geographic grid, made available to surveyors on request.

Provided the geoid is smooth, the geometric technique is easy to implement, but it may require additional sites, apart from those originally planned, to be surveyed with GPS. Where these additional sites are available, in the density and accuracy required, there is no problem. In areas where vertical benchmarks are scarce and the geoid surface is not smooth this method breaks down.

7. SATELLITE ORBITAL MOTION

The prediction of the motion of the celestial bodies was one of the first problems attacked in astronomy. Earliest predictions were based upon empirical rules. However, the introduction of Kepler's laws of planetary motion greatly refined these theories. Newton's laws of motion and gravitation have provided a basis for theoretical calculations, which, except for very small relativistic effects, produce an accurate representation for the orbit. Sophisticated computer programs have now been developed to solve the GPS orbit determination problem. In this chapter, we present the fundamentals of orbital theory for two bodies in relative motion. Perturbations introduced by disturbing effects are described and the degree of sophistication required for modelling the perturbing forces acting on the GPS satellites is evaluated. Procedures for orbit generation (also known as orbit prediction) are briefly reviewed and the process of orbit computation (also known as orbit improvement or orbit correction) is outlined. The computation of GPS ephemerides from the information contained in the Navigation Message (the "Broadcast Ephemerides") is described. Finally, schemes for generating post-processed ephemerides are presented.

7.1 MOTION OF A NEAR-EARTH ORBITING SATELLITE

After a satellite has separated from the 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. In this section we first regard the motion in the gravitational field of a spherically symmetric earth. This produces an elliptical orbit which is fixed in space. Due to the effects of other gravitational and non-gravitational forces which perturb the orbit, the actual orbit of the satellite departs from this ideal. These disturbing forces produce variations with time in the parameters describing the orbit. This leads to the equations of perturbed satellite motion which need to be solved to provide the satellite ephemeris.

7.1.1 UNPERTURBED SATELLITE MOTION

The problem of determining the motion of a homogeneous mass m_1 relative to the centre of mass of another homogeneous mass m is called the two-body or central-force-motion problem, and is described by the second order differential equations:

$$\ddot{r} = - \frac{G(m + m_1)}{|r|^3} r \qquad (7.1)$$

where

r = vector from centre of mass of central body to centre of mass of second body.

G = gravitational constant.
m = mass of central body.
m_1 = mass of second body.

For artificial satellites, m_1 is negligible relative to m and consequently:

$$\ddot{r} = - \frac{Gm}{|r|^3} r \qquad (7.2)$$

The beauty of the two-body problem is that it can be solved analytically. It is one of the few such problems in celestial mechanics which has a known complete solution when the motion of the body is expressed in elliptic elements. But perhaps the most useful property of the two-body problem is that it represents, with reasonable accuracy, the observed motion of near-earth orbiting spacecraft. (The contribution of additional accelerations, though small, must nevertheless be taken into account for accurate predictions of the motion of GPS satellites -- Section 7.2.)

It can easily be shown that the solution of Equation 7.2 is the equation of an ellipse (see, for example, Kaula 1966):

$$r = \frac{a(1-e^2)}{1+e \cos f} \qquad (7.3)$$

where r is the distance of the orbiting body from the central body; a is the semi-major axis; e is the eccentricity, $=(a-b)/2a$, where b is the semi-minor axis; and f is the true anomaly, the geocentric angle between the directions to the orbiting body and to the point of closest approach, or perigee -- see Figure 7.1. While a and e completely define the shape of the Kepler ellipse, its orientation in space must be specified by three angles defined with respect to a space-fixed reference coordinate system (for example the CCRS defined in Section 4.2).

The spatial orientation of the orbital ellipse is shown in Figure 7.2. The angle Ω is the right ascension of the ascending node of the orbit on the equatorial plane, measured eastward from the vernal equinox; i is the inclination of the orbit to the equatorial plane; and ω is the argument of perigee, measured in the plane of the orbit from the ascending node. With the perigee defined in space, we have only to locate the orbiting body's position in the ellipse relative to the perigee.

The true anomaly f is related to the eccentric anomaly E (Figure 7.1) by:

$$\tan(f/2) = [(1+e)/(1-e)]^{1/2} \tan(E/2) \qquad (7.4)$$

The true anomaly, or eccentric anomaly, specifies geometrically the position of the body within its orbital ellipse, but we need also to specify its position in time. This is accomplished via the time of perigee passage t_p.

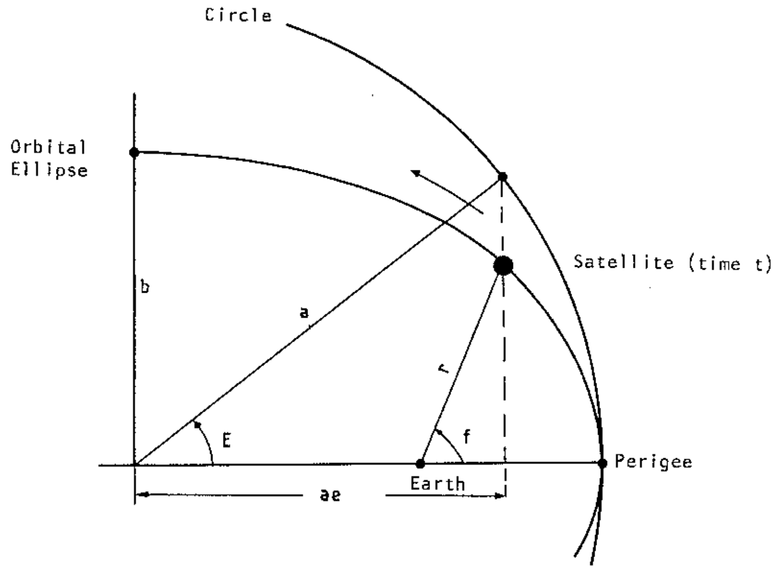


Figure 7.1 The Orbital Ellipse

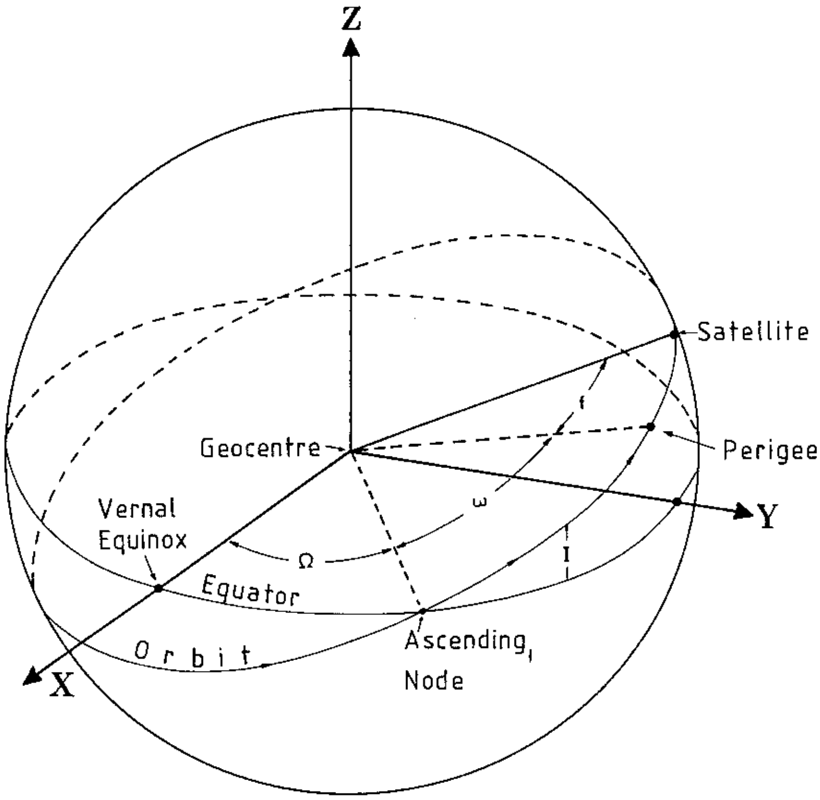


Figure 7.2 The Orbit in Space

Kepler's Third Law relates the mean motion (or mean angular velocity) to the semi-major axis and the gravitational constant μ (= Gm in Equation 7.2):

$$\mu = a^3 n^2 \qquad (7.5)$$

where n is the mean motion (= $2\pi/T$, and T is the orbital period of the satellite). An approximate value of Gm for the earth is 398 600 km^3/s^2. If we now introduce the mean anomaly M:

$$M = n\,(t - t_p) \qquad (7.6)$$

the relationship between the time of perigee passage and the true anomaly (through the intermediate quantities of E and M) is given by Kepler's Equation:

$$E - e \sin E = M \qquad (7.7)$$

Kepler's Equation must be solved iteratively. Note that M increases linearly with time, while the time rate of change of E and f is, in general, non-linear. Since the orbits of GPS satellites are nearly circular (e $\cong$ 0), Equations 7.5 and 7.6 can be used to calculate the approximate orbital motion of the satellites (that is, M $\cong$ E). The set of elliptic elements a, e, i, Ω, ω and M (or f or E) are called the Keplerian Elements, and their geometry is illustrated in Figure 7.2. For unperturbed or two-body motion, a, e, i, Ω and ω are constants and only the anomaly is considered time-dependent.

The equations of motion are of 2nd order and three in number. Since we have six arbitrary constants of integration, it is not surprising that six elliptic parameters are required to completely define two-body motion. These six constants may also be taken as x, y, z, $\dot{x}$, $\dot{y}$, and $\dot{z}$, but no closed expressions exist for the solution of the two-body problem in rectangular coordinates. Relations exist between the elliptic and rectangular coordinates so that given either set of constants, they may be transformed to the other (see Schanzle 1980).

7.1.2 PERTURBED SATELLITE MOTION

The actual orbit of a satellite departs from the Keplerian orbit due to the effects of various "disturbing" accelerations of gravitational and non-gravitational origin (Figure 7.3), including:

1. The non-sphericity of the central body (harmonics of the earth), $\ddot{r}_g$.
2. The attraction of additional bodies, $\ddot{r}_{ls}$, for example, the moon and sun (Planetary accelerations are negligible for near-earth orbiting spacecraft such as GPS satellites).
3. Atmospheric drag, $\ddot{r}_d$.
4. Direct and reflected solar radiation pressure, $\ddot{r}_{sp}$ and $\ddot{r}_a$.
5. Earth and ocean tides, $\ddot{r}_e$ and $\ddot{r}_o$.

These perturbing accelerations produce variations in the orbital elements with time (orbital perturbations). The perturbed orbit can be

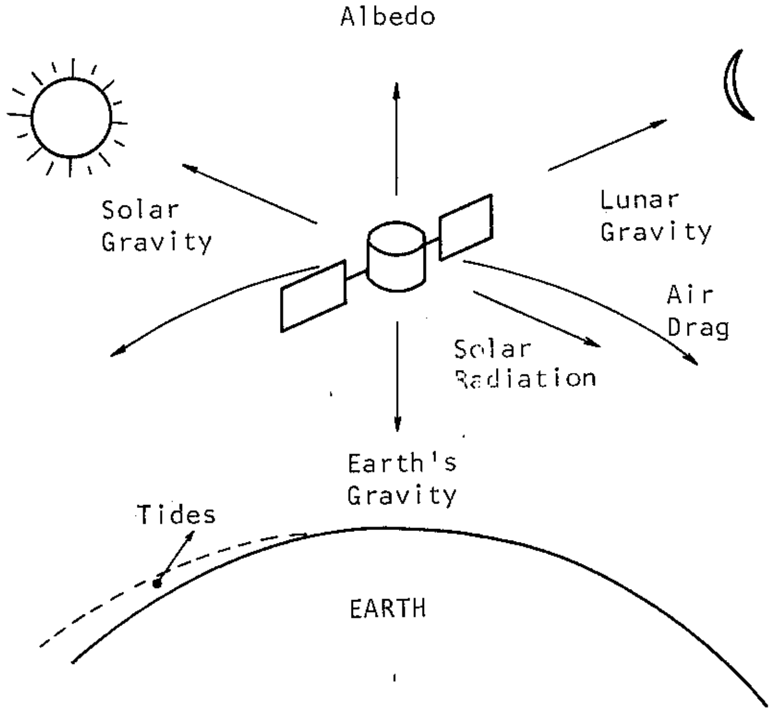


Figure 7.3 Perturbing Forces acting on the GPS Satellites.

viewed as the envelope of Keplerian ellipses (or <u>osculating ellipses</u>) which are defined at any instant by the current or instantaneous orbital elements a, e, i, Ω, ω and M.

The total perturbing acceleration on a near-earth satellite is:

$$\ddot{r}_t = \ddot{r}_g + \ddot{r}_{ls} + \ddot{r}_d + \ddot{r}_{sp} + \ddot{r}_a + \ddot{r}_e + \ddot{r}_o \qquad (7.8)$$

and the equations of motion (Equation 7.2) of the satellite about the earth become:

$$\ddot{r} = -\frac{Gm_e r}{|r|^3} + \ddot{r}_t \qquad (7.9)$$

where the first term represents the central or two-body acceleration discussed earlier and the remaining term is the total perturbing acceleration.

7.2 MOTION OF GPS SATELLITES

Models for the perturbing forces operating on the satellite are required in order to solve the equations of motion. This force modelling capability is a prerequisite for (a) orbit generation (Section 7.3.3), and (b) orbit improvement (Section 7.3.4). What degree of sophistication is needed for modelling the forces acting on GPS satellites in order to achieve 2-20 m orbit accuracy and thus satisfy positioning accuracies of 0.1-1 ppm?

The effects of perturbing forces on GPS orbits were studied by Rizos & Stolz (1985). GPS satellites are at such high altitudes that atmospheric drag is negligible. Accordingly only the effects of the earth's non-sphericity, the direct and indirect attraction of the sun and moon, and the direct and indirect effect of solar radiation pressure were evaluated. To understand the behaviour of GPS satellites over short arcs (a few hours) and longer arcs (1 to 2 days) in response to various perturbing forces, these authors generated a number of test orbits first with and then without the perturbing acceleration under study. The satellite position differences were obtained and plotted, and are summarised in Table 7.1.

Table 7.1 Effect of Perturbing Forces on GPS Satellites.

Source	Acceleration (m/sec^2)	Perturbation (m)	
		3-hour arc	2-day arc
Earth's non-sphericity:			
(a) C_{20}	5×10^{-5}	≅ 2 km	≅ 14 km
(b) other harmonics	3×10^{-7}	5 - 80	100 - 1500
Point-mass effects of sun and moon	5×10^{-6}	5 - 150	1000 - 3000
Earth's tidal potential:			
(a) earth tides	1×10^{-9}	-	0.5 - 1.0
(b) ocean tides	1×10^{-9}	-	0.0 - 2.0
Solar radiation pressure	1×10^{-7}	5 - 10	100 - 800
Albedo pressure	1×10^{-9}	-	1.0 - 1.5

7.2.1 EARTH'S GRAVITATIONAL POTENTIAL

In satellite surveying, the non-central part of the geopotential W, which varies as a function of the latitude, longitude and geocentric distance, is usually represented by a spherical harmonic expansion similar to Equation 6.6, that is:

$$W = Gm_e \sum_{n=2}^{n'} \frac{a_e^n}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi)(C_{nm}\cos m\lambda + S_{nm}\sin m\lambda) \qquad (7.10)$$

where a_e is the equatorial radius of the earth, r the geocentric radius to the satellite (with latitude, longitude ϕ,λ), Gm_e is the product of the gravitational constant and the mass of the earth, $P_{nm}(\sin\phi)$ is the Legendre function of degree n and order m, and C_{nm}, S_{nm} are spherical harmonic coefficients of the earth's gravity field (Section 6.2.2), known to some maximum degree and order n'.

In rectangular coordinates the perturbing acceleration is:

$$\ddot{r}_g = \ddot{x}_g i + \ddot{y}_g j + \ddot{z}_g k \qquad (7.11)$$

where

$$\ddot{x}_g = \frac{\partial W}{\partial x} = \frac{\partial W}{\partial r} \cdot \frac{\partial r}{\partial x} + \frac{\partial W}{\partial \phi} \cdot \frac{\partial \phi}{\partial x} + \frac{\partial W}{\partial \lambda} \cdot \frac{\partial \lambda}{\partial x} \qquad (7.12)$$

and the i, j, and k are unit vectors respectively along the x,y, and z axes of the CTS. Similar expressions hold for $\ddot{y}_g$ and $\ddot{z}_g$.

The coefficient C_{20}, representing the earth's flattening, is approximately three orders of magnitude larger than any other harmonic coefficient. The main effect of C_{20} is to produce periodic and secular variations in the arguments of perigee and node, causing, for example, the slow precession of the line of nodes of a satellite's orbit around the equator.

The acceleration due to the central force term in Equation 7.9 is about 0.5 m/s^2, while C_{20} produces a perturbing acceleration of about $5 \times 10^{-5} m/s^2$. The remaining terms in the geopotential model contribute approximately $3 \times 10^{-7} m/s^2$ to the total perturbing acceleration vector.

Zonal geopotential coefficients (those of the form C_{n0}) produce primarily long period (periods > 1 day) and secular satellite perturbations, while the even degree zonals produce mainly secular perturbations. The non-zonal terms (those with $m \neq 0$, representing the geopotential variation with longitude) give rise primarily to perturbations with periods of less than a day.

The case of <u>resonance</u>, when successive groundtracks (the path of the satellite's orbital plane traced out on the earth) of the satellite are exactly separated by an interval equal to the wavelength of the geopotential harmonic, deserves special mention. After a number of satellite revolutions, the satellite groundtrack repeats itself and the satellite's motion is perturbed in an identical manner, thus magnifying the earlier perturbations. In the case of GPS orbits, the harmonic coefficients of even order lead to resonances of very long period and significant amplitude, but their effects on arcs of up to a few days in length is much reduced.

A number of gravity field models complete to degree and order 20 and beyond, have been developed for use in orbital dynamics, (Lerch et al 1983, Reigber et al 1985). Because GPS satellites are in high stable orbits of about 20 000 km altitude, they are much less affected by short wavelength features in the geopotential than satellites at lower altitude. Thus only a subset of the model coefficients need be used for computing GPS orbits. Rizos & Stolz (1985) found that orbits generated with a geopotential model complete to degree and order 8 -- referred to as an (8,8) model -- comprising 81 coefficients, do not deviate from those generated with a much more complete model by more than a centimetre after two days. Therefore, an (8,8) model can be used as a "reference" model for GPS satellite arcs of a few revolutions.

In Figure 7.4 the effects on two GPS orbits of truncating the geopotential model to degree and order (4,4) as opposed to degree and order (8,8) are shown. Also shown are the reduced orbital errors that result when the (4,4) model is augmented with zonal and even order coefficients to degree 8. Note the large improvement obtained by using the augmented (4,4) model, and only a slight change in the case of orbit B. Clearly, a (4,4) model is adequate if 20 m accuracy orbits over two days are required. However, the (4,4) model is generally not accurate enough for 2 m orbits, except for arcs of only a few hours duration.

Modelling the earth's gravity field for high precision GPS orbits is not a great problem because orbit generating algorithms based on the numerical integration principle (Section 7.3.2) can easily accommodate geopotential models of degree and order 10 and above. Rizos & Stolz (1985) did not study the effect of errors in the model coefficients, but these are not expected to be significant for low degree and order terms.

7.2.2 DIRECT EFFECT OF SUN AND MOON

The perturbing acceleration due to the masses of the sun and moon is:

$$\ddot{r}_{ls} = Gm_s \left[\frac{(r_s - r)}{|r_s - r|^3} - \frac{r_s}{|r_s|^3} \right] + Gm_l \left[\frac{(r_l - r)}{|r_l - r|^3} - \frac{r_l}{|r_l|^3} \right] \qquad (7.13)$$

where G is the gravitational constant, m_s and m_l are the masses of the sun and moon respectively, r_s and r_l are the geocentric position vectors to the sun and moon respectively, and r is the geocentric position vector to the satellite.

The perturbations of the satellite position due to luni-solar attraction, although mainly of long period, are nevertheless significant. The magnitude of the acceleration acting on the GPS satellites is of the order of $5 \times 10^{-6} m/s^2$. Position errors of the order of 1 to 3 km can result if these effects are ignored. After only 3 hours, errors of 50-150 m in each of the radial, crosstrack and alongtrack components of the orbit are produced.

The geocentric position vectors of the sun and moon, r_s and r_l, can be obtained either by evaluating analytical expressions that describe the motion of these bodies (see, for example, Explanatory Supplement, 1974) or, more accurately, by interpolating from a numerically derived ephemeris such as produced by the Jet Propulsion Laboratory, Pasadena, California (JPL, M. Standish, personal communication). Note that these position vectors refer to a particular celestial reference system.

7.2.3 EARTH AND OCEAN TIDE EFFECTS

Earth and ocean tides change the earth's gravitational potential, in turn producing additional accelerations on the GPS satellites. In its simplest form, the perturbing acceleration due to the solid earth tides caused by either the sun or moon, is given by:

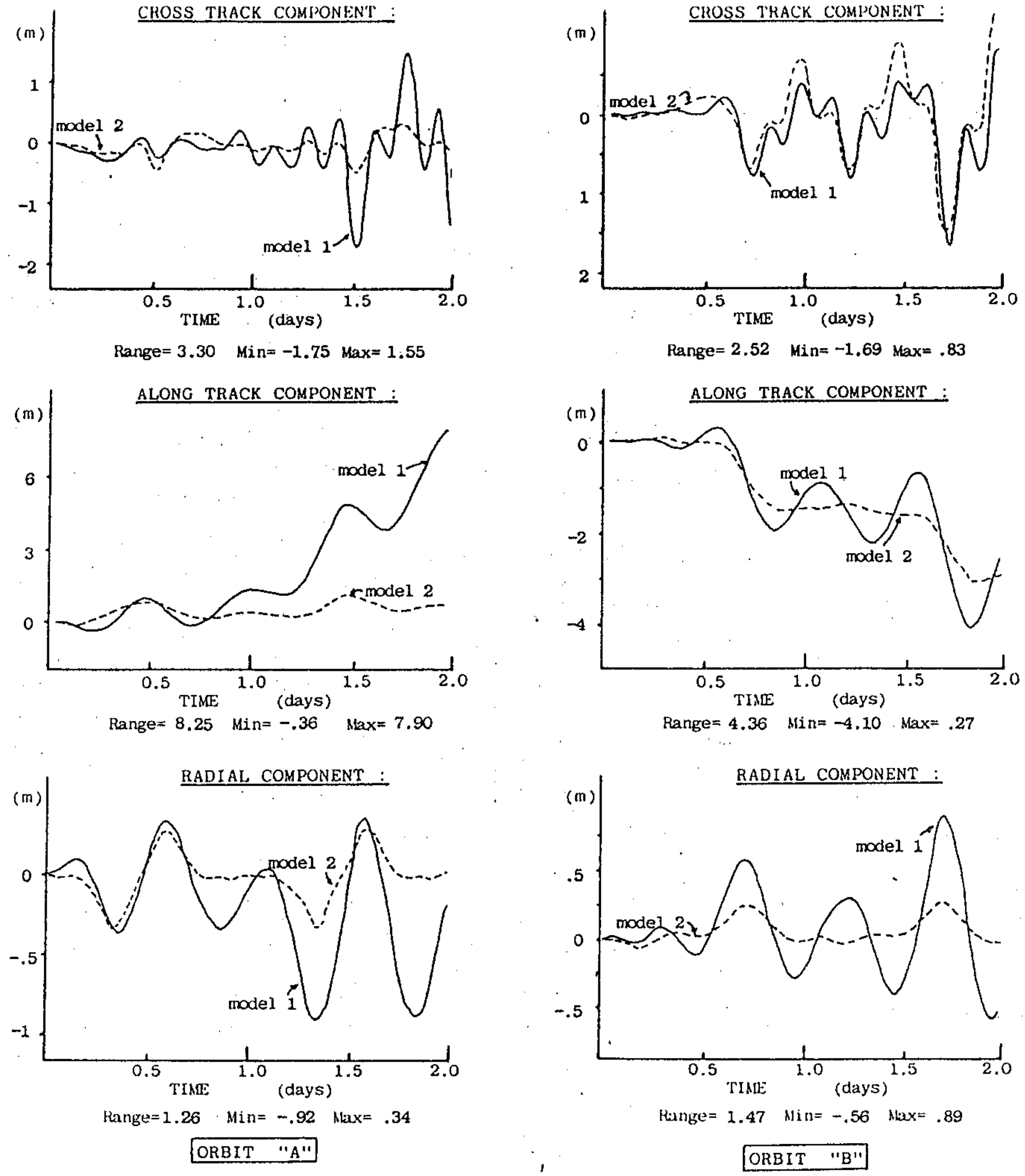


Figure 7.4 Two examples of discrepancies in the position of a GPS satellite from a reference orbit on using: (a) a (4,4) geopotential model, or (b) a (4,4) model with even order coefficients to degree 8. The reference orbit is generated using a geopotential model complete to degree and order 8. Orbital elements: a=26559012.0 m, e=.00884, i=63.44°, Ω=51.49° (orbit A), Ω=141.49° (orbit B).

$$\ddot{r}_e = \frac{k_2 \, Gm_d}{2\,|r_d|^3} \, \frac{a_e^5}{|r|^4} \left[(3-15\cos^2\theta) \frac{r}{|r|} + 6\cos\theta \, \frac{r_d}{|r_d|} \right] \qquad (7.14)$$

where k_2 is the Love number of degree 2, m_d is the mass of the disturbing body, r_d is the geocentric position vector to the disturbing body, a_e is the mean earth radius, θ is the angle between the geocentric position vectors r and r_d, and the remaining quantities are defined in Equation 7.13. The magnitude of the acceleration is of the order of 10^{-9}m/s². Neglecting the earth tide effect would produce orbital errors of 0.5-1 m after two days. This is well below the error tolerances considered here.

The ocean tide effect is more difficult to model. The potential at a point A in space, due to a mass load dm at a point p arising from a tide height of h, is:

$$U_A = \frac{Gdm_p}{a_e} \sum_{n=0}^{n'} (1 + k'_n) \, P_{no}(\cos\psi) \qquad (7.15)$$

where k'_n are load deformation coefficients of degree n, ψ is the geocentric angle between A and p, $P_{no}(\cos\psi)$ are the associated Legendre functions of degree n and

$$dm_p = \rho_o \, h(p,t).d\sigma \qquad (7.16)$$

where ρ_o is the average ocean density, t is time and $d\sigma$ is the element of surface area. The total ocean potential at A is the global integral of Equation 7.15. The perturbing acceleration is found by applying Equation 7.15 to Equations 7.11 and 7.12, as was done in the case of the geopotential. A model of the ocean tide height h(p,t) as a function of time and place is required. Schwiderski (1978) gives one such model in which the tidal constituents are expressed in terms of spherical harmonics.

Perturbing acceleration due to the ocean tide is of the order of 10^{-9}m/s². Moreover, the contribution of the ocean tides to the orbital perturbations of GPS satellites varies from around the submetre level to over 2 m, after two days. Therefore, it is not necessary to account for the ocean tide effect for arclengths considered here.

7.2.4 DIRECT AND REFLECTED RADIATION PRESSURE

The non-gravitational forces arising from the sun's direct radiation pressure and that portion of it that is reflected back from the surface of the earth are the most difficult to model for the GPS satellites. A simplified model for the perturbing acceleration due to direct solar radiation pressure on a spherical satellite is given by:

$$\ddot{r}_{sp} = \nu \, P_s \, C_r \, \frac{A}{m} \, r_s^2 \, \frac{(r - r_s)}{|r - r_s|^3} \qquad (7.17)$$

where ν is an eclipse factor, such that $\nu=0$ when the satellite is in earth's shadow, $\nu=1$ when it is in sunlight, and $0<\nu<1$ if it is in the penumbra region. P_s is the solar constant (=4.65 x 10^{-5} dyne/cm²), C_r

is a factor depending on the reflective properties of the satellite and A/m is the area-to-mass ratio of the satellite. All other quantities are defined in Equation 7.13.

For GPS satellites the magnitude of $\ddot{r}_{sp}$ is of the order of 10^{-7}m/s^2. Depending on the orientation of the orbit in relation to the sun, the perturbations of the satellite's position can range from less than 100 m to over 800 m after two days. Arcs of a few hours duration can be perturbed by 5-6 m. Therefore it is essential that an accurate solar radiation model be used where high precision orbits are required, even for relatively short arcs.

The principal uncertainties in the model defined by Equation 7.17 arise from the following:

(a) the solar "constant" is not constant.
(b) defining a model for the earth's shadow and penumbra.
(c) the use of a single C_r factor for the GPS satellite.
(d) determining the effective cross-sectional area A.

The solar constant has been observed to vary by about 7% over a year and this problem can be overcome by utilising data on observed solar activity.

A simple cylindrical model can account for the effect of the earth's shadow but the sharp discontinuities in acceleration experienced by the satellite on entering and leaving the shadow zone may lead to instability in the numerical integration process (see Section 7.3.2). A "regularising factor" applied to ν is therefore useful in order to smooth the abruptness of the discontinuity as the satellite moves through the penumbra region.

The GPS satellite is an irregularly shaped object constructed of materials with different reflective properties. In addition, as it orbits the earth, the surface area of the satellite that is illuminated by the sun changes. It is therefore inappropriate to assign a single value for C_r and A to GPS satellites. A solar radiation model which takes into account the separate contributions to the perturbations from the engine assembly, the solar panels, the antenna array and the main body unit is required (Fliegel et al 1985). The main value of such an elaborate model is to predict the GPS orbital motion over a week or more. For shorter time periods a simple "flat-plate" model, with C_r estimated from observations, seems to be adequate.

A portion of the solar radiation received by the earth is reflected back from it. The ratio of reflected radiation to the incoming solar flux is called the <u>albedo</u>. The main problem here is that the distribution of land areas, open ocean, and cloud cover, each of which are characterised by a different albedo value, make the albedo radiation pressure difficult to model. Nevertheless, the effect of albedo pressure on the GPS satellite orbit is only about 1-2% of that of the solar radiation pressure.

Rizos & Stolz (1985) tested a latitude dependent model for the albedo based on a second degree zonal harmonic representation with a seasonal factor (Schutz et al 1982). The magnitude of the perturbing acceleration is approximately 10^{-9}m/s^2, which is of the same order of magnitude as earth and ocean tides. Moreover, it appears that the albedo perturbing acceleration is not sensitive to the orientation of the orbit in space. The orbital perturbations were found to be of the order of 1-1.5 m after two days and can therefore be ignored for most applications.

7.3 COMPUTATION OF SATELLITE ORBITS

Direct analytical solution of the equations of perturbed motion (Equation 7.9) is not possible. Historically, solutions to this problem have been obtained using two principal approaches:

(a) the perturbation model is truncated so that an analytical solution is possible, or
(b) the entire perturbation model (Equation 7.8) is included and the equations of motion are solved by numerical integration techniques.

Analytical solutions are usually expressed in elliptic elements (a, e, i, Ω, ω and M), while inertial rectangular coordinates and velocities are preferred with numerical integration.

7.3.1 ANALYTICAL SOLUTION OF EQUATIONS OF MOTION

Numerous analytical theories for orbital perturbations of satellites have been developed. All are approximate and all deal with the perturbations due to zonal and non-zonal geopotential harmonics, the luni-solar attraction, solar radiation pressure, etc., separately. The analytical expressions are based on complex theories of classical mechanics, truncated infinite series, and involve extremely large amounts of algebra.

In analytical solutions the three second order differential equations (Equation 7.9) are replaced with six first order equations expressed in terms of the elliptic elements. Moreover the orbit is represented in two parts:

(a) a reference orbit, for example, provided by the two-body motion (Section 7.1.1), and
(b) small perturbations superimposed upon the reference orbit.

The perturbations of the orbital elements can be further divided into <u>secular</u> and <u>periodic</u> components. <u>Long period</u> perturbations have periods greater than an orbit revolution -- 12 hours for GPS satellites -- while <u>short period</u> perturbations have periods which are less than an orbital period. Table 7.2, taken from Schanzle (1980), describes the perturbations for each of the elliptic elements.

The range of available analytical theories encompasses those in which only the perturbations due to the second degree zonal harmonic of the geoptential are considered, to those in which the effects of many higher order geopotential coefficients, the attraction of the sun and moon, and of solar radiation pressure, are included.

Table 7.2 Perturbations of the Elliptic Elements.

Parameter	Perturbations		
	Secular	Long Period	Short Period
a	No	No	Yes
e	No	Yes	Yes
i	No	Yes	Yes
Ω	Yes	Yes	Yes
ω	Yes	Yes	Yes
M	Yes	Yes	Yes

The main advantage of the analytical solution is that the method provides qualitative insights into the satellite motion: that is, it allows the secular, long and short period terms to be identified and studied. The main disadvantages are:

1. Since they are obtained by truncating infinite series, all analytical solutions are approximate.

2. The computational efficiency of analytical solutions is relatively low, since many trigonometric functions are involved.

3. The perturbations due to non-gravitational forces such as solar radiation pressure are not continuous functions, and are therefore very difficult to represent analytically.

7.3.2 NUMERICAL SOLUTION OF EQUATIONS OF MOTION

With all the disadvantages of analytical solutions of the equations of motion, it is not surprising that a numerical approach is preferred when high precision orbits are required. In this approach the perturbing accelerations which act on the satellite are modelled as accurately as possible. Then, given the position and velocity of the satellite at some starting epoch t_o, the equations of motion are numerically integrated to obtain position and velocity at some later epoch t. For most numerical integration schemes, we replace the three second order deifferential equations 7.9 by six first order differential equations, three for position and three for velocity:

$$\frac{dr}{dt} = \dot{r} \qquad (7.18a)$$

$$\frac{d\dot{r}}{dt} = -\frac{Gm_e r}{|r|^3} + \ddot{r}_t \qquad (7.18b)$$

Numerical integration software which will handle a large number of equations of the form of 7.18 is readily availbale on most computers. The user need only specify the initial values of the dependent variables (left-hand sides of Equations 7.18) and provide a subroutine to evaluate the right-hand sides. For the orbit problem the initial values are the inertial position (r=x,y,z) and velocity ($\dot{r}$=$\dot{x}$,$\dot{y}$,$\dot{z}$) vectors; and the right-hand sides are the velocities (for Equation 7.18a) and the accelerations (for Equation 7.18b) given in Equation 7.9.

Numerical integration schemes are divided into "single-step" and "multi-step" procedures. The Runga-Kutta approach is a single-step procedure, while predictor-corrector type numerical integrators are multi-step procedures. The reader is referred to Cappellari et al (1976) for details on numerical integration algorithms.

The primary advantages of numerical integration are that it can be easily applied to any differential equation and that it is very efficient on modern computers. There are no approximations to the equations of motion as is the case with the analytical approach, but problems can arise with round-off and truncation errors. The main disadvantage is that many intermediate values of the satellite's position must be computed (together with the complex perturbing forces at each step) in order to get up to the particular time of interest. That is, this is a <u>multi- step</u> procedure, as opposed to the analytical solution which is performed only once. (Even if one needs the satellite coordinates at many epochs, the number of steps required for the numerical integration is usually larger, by a factor of 10 or so, than the number of tabular values needed for interpolation.)

7.3.3 <u>ORBIT GENERATION</u>

The terms "orbit generation" or "orbit prediction" refer to the computational process by which a satellite's position and velocity (the <u>state vector</u>) is computed at specified intervals of time. In order to meet varying precision and efficiency requirements, different computational techniques can be used, ranging from an analytical solution to the high precision numerical integration methods. The term "almanac" is commonly used to distinguish the very approximate orbit needed, for example, to support GPS pre-survey planning (Section 3.2.1) from the high accuracy orbit (or "ephemeris") required to process the GPS observations (Section 5.2.3). This distinction between <u>almanac</u> and <u>ephemeris</u> also applies to the Navigation Message transmitted by each GPS satellite. In addition to containing the (predicted) ephemeris for the transmitting satellite (Section 7.4.2), the Navigation Message also contains an almanac for every other satellite in the GPS constellation, to aid the receiver in acquiring the remaining satellite signals.

<u>GPS Almanac Generation</u>

The almanac is used to determine the "rise" and "set" times of the satellite, and to produce "sky plots" (Figure 3.1) showing the approximate satellite paths across the sky. Accordingly, a high efficiency, comparatively low accuracy, analytical technique is sufficient to generate the almanac.

Analytical orbit theories usually only take into account the effect of the earth's geopotential on the satellite motion. Two well known techniques are based on the theories of Brouwer and Kaula (Brouwer & Clemence 1961, Kaula 1966). In the Brouwer theory, the short and long period, and secular orbital perturbations (Table 7.2) are modelled by approximating the earth's mass distribution by a uniformly stratified sphere (equivalent to a point mass concentrated at the geocentre) augmented by low degree zonal harmonics of the geopotential (coefficients C_{nm}, n=1,2,3...,m=0). Kaula's development, which is more general, takes into account the first-order perturbations of all the harmonic coefficients C_{nm}, S_{nm}.

For generating an almanac, it is sufficient to use an approximation to the Brouwer theory, in which only the secular variations of Ω, ω, M due to the second degree zonal harmonic are estimated. The Kepler ellipse maintains its shape, but the orbital plane slowly precesses about the equator and the perigee point slowly precesses around the orbital ellipse. The satellite position at any instant t_i can be estimated from the (Keplerian) starting elements at time t_o, according to:

$$
\begin{aligned}
a(t_i) &= a(t_o) \\
e(t_i) &= e(t_o) \\
i(t_i) &= i(t_o) \\
\Omega(t_i) &= \Omega(t_o) + \dot{\Omega}\Delta t \\
\omega(t_i) &= \omega(t_o) + \dot{\omega}\Delta t \\
M(t_i) &= M(t_o) + \dot{M}\Delta t + n\Delta t
\end{aligned}
\qquad (7.19)
$$

where $\Delta t = t_i - t_o$, n is the mean motion (Equation 7.5) and:

$$
\begin{aligned}
\dot{\Omega} &= -3\,\gamma\,\cos i \\
\dot{\omega} &= 1.5\,\gamma\,(5\cos^2 i - 1) \\
\dot{M} &= 1.5\,\gamma\,(1-e^2)^{1/2}\,(3\cos^2 i - 1)
\end{aligned}
\qquad (7.20)
$$

where

$$\gamma = -\frac{C_{20} a_e^2}{2a^2(1-e^2)^2}\,n \qquad (7.21)$$

and a_e is the earth's equatorial radius.

The present GPS satellites are in orbits with a ranging from 26 557 to 26 562 km, e ranging from .002 to .011 and i ranging from 62.5° to 64°. (The operational GPS satellites will be inclined at 55° to the equator.) Typical values of the secular rates are therefore:

$$\gamma \cong 0.0226 \text{ degs/day}$$

$$n \cong 721.9 \text{ to } 722.2 \text{ degs/day}$$

$$\dot{\Omega} \cong -.030 \text{ degs/day}$$

$$\dot{M} \cong -.013 \text{ degs/day}$$

Because the inclinations are close to the "critical inclination" ($\cong$ 63.4°) for which $5\cos^2 i - 1$ in Equation 7.20 vanishes, the perigee point is almost stationary ($\dot{\omega} \cong 0$).

The GPS almanacs generated from Equations 7.19, 7.20 & 7.21 can be in error by tens of kilometres after a few days, and after a month the error could be of the order of hundreds of kilometres. Keeping in mind that the GPS satellites travel at approximately 3 km/s, the orbit error after one month's forward prediction using the above relations is approximately equivalent to an error of 1 minute in the rise/set time of the satellites. Therefore, starting elements that are even a few months old can still be used to generate GPS almanacs of adequate accuracy for supporting pre-survey planning.

GPS Ephemeris Generation

The best analytical theories for the GPS satellites produce orbits with an accuracy of 20-60 m (1-3 ppm) (see, for example, Abbot et al 1983). For higher accuracy, numerical integration techniques must be used. The equations of motion are usually integrated in an inertial (non-accelerating) reference frame using Dynamical Time as the independent variable (see Section 4.1.1). This inertial reference frame is commonly defined by the true equator and equinox at some (arbitrary) epoch in time. The starting elements must be transformed into this reference system before orbit generation can begin.

The numerical integration technique most often used in ephemeris generation is a multi-step predictor/corrector algorithm of the Adams-Cowell type (Cappellari et al 1976). The number of previous satellite state vectors used in the prediction/correction procedure is typically between 7 and 12, and the step-size used for GPS orbits is about 150 seconds. Usually, the orbital elements (the satellite state vectors) are only required at intervals of about 15-20 minutes. Therefore, a number of intermediate satellite positions and velocities have to be computed, between each specified pair of ephemeris values, and the state vector interpolated from these intermediate values.

The numerical solution of the equations of motion (Equation 7.9) requires the perturbing accelerations acting on the satellite to be accurately modelled. These models are not error free. Moreover, the integrator algorithm introduces round-off errors in the computer. Hence, the orbital accuracy will degrade with arclength.

Section 7.2 describes the degree of sophistication necessary to maintain orbits of 20 m accuracy for arclengths ranging from a few hours to several days. In summary:

1. An (8,8) geopotential model is more than adequate for 2-day orbital arcs while models to degree and order (4,4) can be used for arcs of a few hours duration.

2. The gravitational effects of the sun and moon should be included. Even for short arcs, perturbations in satellite position of over 100 m can be expected.

3. Perturbations in satellite position arising from solar radiation pressure can range from less than 100 m to over 800 m after two days. Tailored solar radiation pressure models are necessary for arcs more than a few days in length.

Nevertheless, the estimates of the initial position and velocity of the satellite must also satisfy this accuracy. If longer arclengths are used or the initial starting elements are poorly known, an "orbit improvement" must be carried out using observations taken from ground stations whose coordinates are known.

7.3.4 ORBIT COMPUTATION

The terms "orbit computation", "orbit improvement" and "orbit correction" all refer to the procedure by which the "best" orbit is estimated from satellite observations obtained at sites of known position or whose coordinates are to be estimated simultaneously with the orbital parameters. If the site coordinates are known a priori and observations are taken over several orbital periods, the estimated orbital parameters will provide a good representation of the orbit outside the span of the observations. If, on the other hand, the orbital parameters are estimated in a "short-arc" mode, simultaneously with site coordinates, the orbit generated from those parameters will represent the satellite's motion well over the span of the observations but will extrapolate poorly outside that span.

The procedures for estimating orbital parameters are essentially the same as those described in Section 5.2 for site coordinates and clock parameters. One needs only to add the partial derivatives of orbital parameters to the normal equations. For GPS satellites these parameters usually include the position and velocity at some starting epoch (for a numerically integrated model of the orbit) or a set of mean elliptic elements (for an analytic model), and one or two parameters accounting for non-gravitational forces on the satellite. The partial derivative of the observations with respect to a satellite parameter are given by Equations 5.44 to 5.50. The partials of the range vector R with respect to parameter x are:

$$\frac{\partial R}{\partial x} = \frac{\partial s}{\partial x} \tag{7.22}$$

where s is the satellite state vector (position and velocity vectors) expressed in the earth-fixed system. If the ephemeris (including its partial derivatives) has been generated in a space-fixed system, then the coordinates (r) and partials ($\partial r/\partial x$) must first be transformed to the earth-fixed system using Equation 4.12.

The partial derivatives $\partial r/\partial x$, (as a function of time in the space fixed system) are obtained by integrating the partial derivatives of the equations of motion (Equations 7.18a & 7.18b):

$$\frac{\partial}{\partial x}\left(\frac{dr}{dt}\right) = \frac{d}{dt}\left(\frac{\partial r}{\partial x}\right) = \frac{\partial \dot{r}}{\partial x} \tag{7.23a}$$

$$\frac{\partial}{\partial x}\left(\frac{d\dot{r}}{dt}\right) = \frac{d}{dt}\left(\frac{\partial \dot{r}}{\partial x}\right) = \frac{Gm_e}{|r|^3}\left[\frac{3r}{|r|^2}\left(r \cdot \frac{\partial r}{\partial x}\right) - \frac{\partial r}{\partial x}\right] + \frac{\partial \ddot{r}_t}{\partial x} \tag{7.23b}$$

For x_{oj}, j=1...3, the initial positions and x_{oj}, j=3...6, the initial velocities, the initial conditions of Equation 7.23 are given by:

$$\frac{\partial r_i}{\partial x_{oj}} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

and

$$\frac{\partial \dot{r}_i}{\partial x_{oj}} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \qquad (7.24)$$

where i=1-3 and j= 1-6.

For all other parameters (for example, solar radiation pressure), the initial conditions are zero.

Equations 7.23a and 7.23b are called the variational equations. If orbits are generated by numerical integration, these equations (six for each parameter x) can be integrated simultaneously with the equations of motion. Thus the total number of equations to be integrated for each satellite is 6(n+1), where n is the number of orbital parameters to be estimated. Since many of the quantities on the right-hand sides of the equations of motion and the variational equations are common to all of the equations and therefore need be evaluated only once per step, the computation time required does not grow linearly with the number of equations. A further saving in computing time results from the fact that for most partials the last term on the right-hand side of Equation 7.23b is small compared to the first term. Therefore the former need not be included except for the parameters which appear explicitly in the expression for the perturbing acceleration.

When analytical techniques are used, the state vector is replaced by the Keplerian elements (a,e,i,Ω,ω,M), and the partial derivatives are obtained analytically (for example, see Nakiboglu et al 1985).

7.4 GPS EPHEMERIDES

A knowledge of the GPS satellite orbits for the time span of the survey is essential to analyse the data collected and to determine the positions of the receivers. The data analysis procedure was described in Section 5.2.3. Just as the coordinate accuracy resulting from a conventional survey is directly related to the quality of the local control coordinates, the accuracy of the final receiver coordinates (or coordinate differences) is a function of the accuracy of the GPS ephemerides. The approximate relationship between orbit accuracy and relative positional accuracy was given in Section 1.4.

How are the GPS ephemerides calculated? In principle, all are produced using the orbit computation procedures described in Section 7.3.4. However, we should distinguish between post-processed ephemerides, which define the satellite positions for the period during which the tracking is carried out, and extrapolated ephemerides, such as those contained in the Navigation Message. The latter are predictions of the satellite positions at some future date and are based on (post-processed) reference ephemerides from the immediate past.

7.4.1 POST-PROCESSED EPHEMERIDES

NSWC Ephemerides

The first post-processed ephemerides have been made available to the surveying community by NGS. The GPS orbit determination is at present carried out at the Naval Surface Weapons Center (NSWC), Dahlgren, Virginia, for DMA (Swift 1985) using 7-days of pseudo-range observations collected from up to eight globally distributed tracking stations. These ephemerides are sometimes referred to as the "Precise Ephemerides", using the analogy of the TRANSIT system Precise Ephemeris. A number of different orbit computation schemes have been tried and there is still considerable experimentation with different processing software, different tracking schemes, and even different organisational structures within both NSWC and DMA. A definitive statement about who will be responsible in the future for the computation of the post-processed ephemerides is not possible at this stage.

At present, all requests for these ephemerides by United States non-military users must be directed to NGS. NGS will provide the ephemerides without restriction to domestic users. In Australia the request must be routed through the Director of Survey - Army. The ephemeris data is available in several formats to satisfy a variety of user needs (Remondi 1985b, or write to National Geodetic Information Center, NGS, NOAA, Rockville, Md. for further information).

Up to now the positions and velocities of the satellites have been given in the WGS72 earth-centred, earth-fixed reference frame and the accuracy of the NSWC ephemerides has usually been 10-20 m, adequate to support surveying at the level of 0.5-1 ppm.

Other Sources of Ephemerides

A number of U.S. agencies, academic institutions and private firms are developing computer software for the computation of GPS orbits in order to support their own requirements, particularly if:

(1) the NGS-distributed ephemerides were not accurate enough for the GPS surveying tasks required,
(2) the NGS-distributed ephemerides were not timely enough, or
(3) the U.S. government placed unacceptable restrictions on the distribution of these ephemerides.

In this event, foreign countries may proceed to establish their own orbit computation capability, or in association with their neighbours, to support GPS surveying in their area of the world (Rizos et al 1985). However, to establish the capability for independently determining (or even merely verifying) GPS orbits involves considerable investment in hardware (the tracking stations and central computer facility) and software. Nevertheless, even if restrictions were introduced, most surveying needs (see Section 3.5.1) could still be satisfied by using the Broadcast Ephemerides. Only if the quality of these ephemerides were degraded (in line with the degradation of the S code) would there be an urgent need to independently process orbits.

High precision GPS surveys (with relative accuracies better than 1 ppm) require ephemerides accurate at the few metre level. Such ephemerides have been already computed, for example using data from GPS receivers installed by the U.S. Air Force Geophysics Laboratory at VLBI observatories equipped with hydrogen maser frequency standards (Abbot et al 1985). In addition, to ensure high accuracies for the Crustal Dynamics Project and to provide an indirect tracking capability (via the GPS satellites) for its own space missions, NASA has proposed the establishment of a six-station global GPS tracking network. It may also be possible to establish temporary tracking stations in an area to support local crustal movement surveys and would be dismantled on completion of the surveying campaign.

7.4.2 THE BROADCAST EPHEMERIDES

The term "Broadcast Ephemerides" refers collectively to the predicted GPS satellite positions that are contained within the Navigation Message which is transmitted by each GPS satellite. The steps used in computing the Broadcast Ephemerides have been outlined by Russell & Schaibly (1980):

Step 1. A reference orbit is generated for each GPS satellite for a period of several weeks, from past tracking data; that is the orbits are extrapolated into the future.

Step 2. The predicted orbital elements are used as starting values in a Kalman filter estimation program to provide current estimates of the satellite positions and velocities. In other words, at the end of each day, the tracking data acquired by the GPS Monitor Stations are used to update that day's portion of the reference orbit. At this stage, the updated satellite state vectors for the end of the day are calculated.

Step 3. From these updated state vectors, the orbits are again extrapolated for up to two weeks into the future.

Step 4. The orbital information that is transmitted in the Navigation Message is based on curve fits to these long extrapolated ephemerides. The curves are fitted to four to six hours of ephemeris data.

Step 5. These short arcs are then transformed from Cartesian inertial coordinates into pseudo-Keplerian elements and the data relevant to a particular satellite is uploaded to the satellite and stored. Each hour, new elements are transmitted, replacing the previous elements in the Navigation Message.

Based on the relative positional accuracies achieved using the "Broadcast Ephemerides", a conservative guestimate of the accuracy of the broadcast orbital data is 40-100 m. More optimistic accuracy estimates of less than 20 m are quoted in the official literature. As far as orbit accuracy is concerned, a number of comments are appropriate:

(a) The upload is usually performed on a daily basis. Therefore, the portions of ephemeris data relevant to the second through fourteenth day are not normally transmitted (except when upload is not possible).

(b) Each day's tracking data are used, in conjunction with the reference orbits, to predict the following day's satellite ephemerides. Therefore, it is likely that the ephemeris accuracy will degrade as a function of time since upload (start of day) and possibly as a function of the age of the reference orbits as well (since start of the week).

The Navigation Message

The Navigation Message is a 50 bit per second data bit stream modulated on the GPS signals (both frequencies). The data message is contained within a data frame that is 1500 bits long, made up of five subframes, each 300 bits long. Each subframe contains GPS system time, the S to

P code handover information and a number of check flags. The information in the Navigation Message is organised in the following way (Van Dierendonck et al 1980):

(a) The first subframe (also known as Data Block 1) contains the satellite clock correction parameters and ionospheric propagation delay model parameters, as computed by the GPS Master Control Station.

(b) The second and third subframes (Data Block 2) contain the satellite ephemeris parameters.

(c) The fourth subframe contains alphanumeric information for the users. This is also known as the Message Block.

(d) The fifth subframe (Data Block 3) contains the almanac information, clock correction parameters and health status of all the GPS satellites (one satellite per subframe). The almanac is required by the user to acquire the other satellites.

The entire frame repeats itself every 30 seconds, except that Data Block 3 rotates through 25 subframes of data (each satellite's almanac, clock correction parameters, etc. resides on one subframe). Every hour (nominally) the information in Data Block 1 and 2 are refreshed with ephemeris data that applies to the new period. Data Block 3 and the Message Block are changed at upload time (nominally each day).

Functional Representation of the 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. The perturbations comprise periodic (periods of a half hour or less to a few months) and secular components (Section 7.2). 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 7.3.

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 present, Sunday midnight.) The Keplerian representation has physical meaning (Section 7.1.1) but additional parameters are required to model the perturbations about the Keplerian orbit. The following parameters are introduced:

- Δn, represents the secular change in mean anomaly (or argument of perigee).

- $\dot{\Omega}$, describes the secular drift of the ascending node in the equatorial plane (mainly due to the second degree zonal harmonic C_{20} -- Equation 7.19).

- $C_{uc}, C_{us}, C_{rc}, C_{rs}, C_{ic}, C_{is}$, 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.

Table 7.3 Ephemeris Representation Parameters.
(Van Dierendonck et al 1980)

M_o	Mean anomaly at reference time
Δn	Mean motion difference from computed value
e	Eccentricity
$\sqrt{A}$	Square root of the semi-major axis
Ω_o	Right ascension at reference time
i_o	Inclination angle at reference time
ω	Argument of perigee
$\dot{\Omega}$	Rate of right ascension
C_{uc}	Amplitude of the cosine harmonic correction term to the argument of latitude
C_{us}	Amplitude of the sine harmonic correction term to the argument of latitude
C_{rc}	Amplitude of the cosine harmonic correction term to the orbit radius
C_{rs}	Amplitude to the sine harmonic correction term to the orbit radius
C_{ic}	Amplitude of the cosine harmonic correction term to the angle of inclination
C_{is}	Amplitude of the sine harmonic correction term to the angle of inclination
t_{oe}	Ephemeris reference time
AODE	Age of Data (Ephemeris)

Note that these parameters are obtained from a curve fit to the predicted satellite ephemeris over an interval of 4-6 hours. They are not true Keplerian elements as they only describe the ephemeris over the period 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.)

The equations given in Table 7.4 allow the user to derive the earth-centred / earth-fixed Cartesian coordinates from the broadcast orbital information indicated in Table 7.3. The earth-fixed reference system is based on the WGS72 (and later on the WGS84). The algorithm in Table 7.4 is implemented in every code-correlating GPS receiver, both of the navigational and geodetic variety.

Table 7.4 Broadcast Ephemeris Computational Procedure. (Van Dierendonck et al 1980)

A	$= (\sqrt{A})^2$	Semi-major axis	
n_o	$= \sqrt{\frac{\mu}{A^3}}$	Computed mean motion	
t_k	$= t - t_{oe}$*	Time from epoch	
n	$= n_o + \Delta n$	Corrected mean motion	
M_k	$= M_o + nt_k$	Mean anomaly	
M_k	$= E_k - e \sin E_k$	Kepler's equation for eccentric anomaly	
$\cos \upsilon_k$	$= (\cos E_k - e)/(1 - e \cos E_k)$	True anomaly	
$\sin \upsilon_k$	$= \sqrt{1 - e^2} \sin E_k/(1 - e \cos E_k)$		
ϕ_k	$= \upsilon_k + \omega$	Argument of latitude	
δu_k	$= C_{us} \sin 2\phi_k + C_{uc} \cos 2\phi_k$	Argument of latitude correction	2nd harmonic perturbations
δr_k	$= C_{rc} \cos 2\phi_k + C_{rs} \sin 2\phi k$	Radius correction	
δi_k	$= C_w \cos 2\phi_k + C_{is} \sin 2\phi_k$	Correction to inclination	
u_k	$= \phi_k + \delta u_k$	Corrected argument of latitude	
r_k	$= A(1 - e \cos E_k) + \delta r_k$	Corrected radius	
i_k	$= i_o + \delta i_k$	Corrected inclination	
x'_k	$= r_k \cos u_k$	Positions in orbital plane	
y'_k	$= r_k \sin u$		
Ω_k	$= \Omega_o + (\dot{\Omega} - \dot{\Omega}_e)t_k - \dot{\Omega}_e t_{oe}$	Corrected longitude of ascending node	
x_k	$= x'_k \cos \Omega_k - y'_k \cos i_k \sin \Omega_k$	Earth fixed coordinates	
y_k	$= x'_k \sin \Omega_k + y'_k \cos i_k \cos \Omega_k$		
z_k	$= y'_k \sin i_k$		

μ = WGS 84 value of the earth's universal gravitational parameter

Ω_e = WGS 84 value of the earth's rotation rate

* t is GPS system time at time of transmission, i.e., GPS time of reception corrected for transit time (range/speed of light). Furthermore, t_k must be the actual total time difference between the time t and the epoch time t_{oc}, and must account for beginning or end of week crossovers. That is, if t_k is greater than 302,400, subtract 604,800 from t_k. If t_k is less than 302,400 add 604,800 to t_k.

REFERENCES

ABBOT, R.I., CEFULA, P.J. & TSE, S.F., 1983, "Satellite Orbital Theory for a Small Computer", Astrodynamics, 54, p.543-572.

ABBOT, R.I., BOCK, Y., COUNSELMAN III, C.C., KING, R.W., GOUREVITCH, S.A. & ROSEN, B.J., 1985, "Interferometric Determination of GPS Satellite Orbits", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.63-72.

ALLMAN, J.S., 1982, "A Geoid for South-East Asia and the Pacific", Aust.J.Geod.Photo.Surv., 36, p.59-63.

ALLMAN, J.S. & VEENSTRA, C., 1984, "Geodetic Model of Australia 1982", Tech.rept.no.33, Division of National Mapping, Canberra, 58pp.

BADEKAS, J., 1969, "Investigations Related to the Establishment of a World Geodetic System", Rept.no.124, Dept. of Geodetic Science, The Ohio State University, Columbus.

BENDER, P. & LARDEN, D.R., 1985, "GPS Carrier Phase Ambiguity Resolution over Long Baselines", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.357-362.

BEUTLER, G., DAVIDSON, D.A., LANGLEY, R.B., SANTERRE, R., VANICEK, P. & WELLS, D.E., 1984, "Some Theoretical and Practical Aspects of Geodetic Positioning Using Carrier Phase Difference Observations of GPS Satellites", Tech.rept.no.109, University of New Brunswick, Fredericton, N.B., July, 79pp.

BEUTLER, G., GURTNER, W., BAUERSIMA, I. & LANGLEY, R., 1985, "Modelling and Estimating the Orbits of GPS Satellite", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.99-112.

BOCK, Y., ABBOT, R.I., COUNSELMAN III, C.C., GOUREVITCH, S.A., KING, R.W. & PARADIS, A.R., 1984, "Geodetic Accuracy of the Macrometer V-1000", Bull.Geodes., 58, p.211-221.

BOCK, Y., ABBOT, R.I., COUNSELMAN III, C.C., GOUREVITCH, S.A. & KING R.W., 1985, "Establishment of Three-dimensional Geodetic Control by Interferometry with the Global Positioning System", J.Geophys.Res., 90, p.7689-7703.

BOMFORD, G., 1971, "Geodesy", (3rd ed.), Oxford Univ. Press, 731pp.

BOSSLER, J.D., 1972, "Bayesian Inference in Geodesy", Ph.D. thesis, Dept. of Geodetic Science, The Ohio State University, 79pp.

BOSSLER, J.D., 1984, "The Impact of VLBI and GPS on Geodesy", EOS, Trans.Am.Geophys.U., 64, p.569-570.

BROUWER, D. & CLEMENCE, G.M., 1961, "Methods of Celestial Mechanics", Academic Press, New York, 598pp.

BUENNAGEL, L.A., MacDORAN, P.F., NEILLAN, R.E., SPITZMESSER, D.J. & YOUNG, L.E., 1984, "Satellite Emission Range Inferred Earth Surveying (SERIES) Project: Final Report on Research and Development Phase, 1979 to 1983", NASA JPL pub. 84-16, March 1984, 46pp.

BURSA, M., 1966, "Fundamentals of the Theory of Geometric Satellite Geodesy", Travaux de l'institut Geophysique de l'academie Teheco-slovaque des Sciences, No.241.

CANNON, E., LACHAPELLE, G. & GOAD, C.C., 1985, "Recovery of a 1700 km Baseline Using Dual Frequency Carrier Phase Measurements", in proc. First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.593-602.

CAPPELLARI, J.O., VELEZ, C.E. & FUCHS, A.J., 1976, "Mathematical Theory of the Goddard Trajectory Determination System". Doc.X-582-76-77, Goddard Space Flight Center, Greenbelt, Md.

COLLINS, J., 1984, "GPS Satellite Surveying", ACSM Bulletin, 88, p.23-26.

COLLINS, J. & LEICK, A., 1985, "Analysis of Macrometer Networks with Emphasis on the Montgomery County (PA) Survey", in proc. First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.677-694.

CONLEY, R., 1984, "Reference Frames, Earth Models and GPS", GPS Joint Program Office, Directorate of Systems Engineering, YEE 84-040, 27th April 1984, 137pp.

COUNSELMAN III, C.C. & GOUREVITCH, S.A., 1981, "Miniature Interferometer Terminals for Earth Surveying: Ambiguity and Multipath with GPS", IEEE Trans. Geosci. Remote Sens., GE 19, p.244-252.

CUPPELS, G. & COLLINS, J., 1985, "The Montgomery County Pennsylvania Global Positioning Geodetic Control Project", in proc. First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.705-711.

DELIKARAOGLOU, D., BECK, N., McARTHUR, D. & LOCHHEAD, K., 1985, "On the Establishment of 3-D Geodetic Control by Interferometry with the TI-4100 GPS Receiver", in proc. First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.645-656.

ENGELIS, T., RAPP, R.H. & TSCHERNING, C.C., 1984, "The Precise Computation of Geoid Undulation Differences with Comparison to Results Obtained from the Global Positioning System", Geophysical Research Letters, 11(9), p.821-824.

ENGELIS, T., RAPP, R.H. & BOCK, Y., 1985, "Measuring Orthometric Height Differences with GPS and Gravity Data", Manuscripta Geodaetica, 10, p.187-194.

Explanatory Supplement to the Astronomical Ephemeris, H.M. Stationery Office, London, 1974, 533pp.

FLIEGEL, H.F., FEESE, W.A., LAYTON, W.C. & RHODUS, N.W., 1985, "The GPS Radiation Force Model", in proc. First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.113-120.

GOAD, C.C. (convenor), 1985a, "Proceedings of the First International Symposium on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985", U.S. Department of Commerce, NOAA, Rockville, Md., May 1985, 946pp.

GOAD, C.C., 1985b, "Precise Relative Position Determination Using Global Positioning System Carrier Phase Measurements in a Non-Difference Mode", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.347-356.

GOAD, C.C. & REMONDI, B.W., 1984, "Initial Relative Positioning Results Using the Global Positioning System", Bull.Geodes., 58, p.193-210.

GOAD, C.C., SIMS, M.L. & YOUNG, L.E., 1985, "A Comparison of Four Precise Global Positioning System Geodetic Receivers", IEEE Trans. Geosci. Remote Sens., GE 23(4), p.458-466.

HARVEY, B.R., 1985, "The Combination of VLBI and Ground Data for Geodesy and Geophysics", Unisurv S-27, University of New South Wales, Kensington, N.S.W., 244pp.

HARVEY, B.R., STOLZ, A., JAUNCEY, D.L., NIELL, A.E., MORABITO, D.D. & PRESTON, R.A., 1983, "Results of the Australian Geodetic VLBI Experiment", Aust.J.Geod.Photo.Surv., 38, p.39-51.

HATCH, R. & LARSON, K., 1985, "Magnet-4100 GPS Survey Program Processing Techniques and Test Results", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.285-298.

HEISKANEN, W.A. & MORITZ, H., 1967, "Physical Geodesy", W.H.Freeman & Co., San Francisco, 364pp.

HEISTER, H., SCHOEDLBAUER, A. & WELSCH, W., 1985, "Macrometer Measurements 1984 in the Inn Valley", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.567-578.

HENSON, D., COLLIER, A. & SCHNEIDER, K., 1985, "Geodetic Applications of the Texas Instruments TI 4100 GPS Navigator", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.191-200.

HOTHEM, L.D. & FRONCZEK, C.J., 1983, "Report on Test Demonstration of Macrometer Model V-1000 Interferometric Surveyor", Federal Geodetic Control Commission, Rept.FGCC-IS 83-2, NGS/NOAA, Rockville, Md., 36pp.

HOTHEM, L.D. & WILLIAMS, G.E., 1985a, "Factors To Be Considered in Development of Specifications for Geodetic Surveys Using Related Positioning GPS Techniques", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.633-644.

HOTHEM, L.D. & WILLIAMS, G.E., 1985b, "Proposed Revisions to Geodetic Survey Standards and Preliminary Specifications for Geodetic Surveys Using Relative Positioning GPS Techniques", unpublished doc., NGS, May 1985, 16pp.

HOTHEM, L.D., GOAD, C.C. & REMONDI, B.W., 1984, "GPS Satellite Surveying - Practical Aspects", The Canadian Surveyor, Autumn, p.177-192.

IUGG, 1980, "The Geodesist's Handbook", Bull.Geodes., 54, 3, 463pp.

JANICZEK, P.M. (ed.), 1980, "Global Positioning System", papers published in Navigation, reprinted by the Institute of Navigation, Washington, D.C.

KAPLAN, G.H., 1981, "The IAU Resolutions on Astronomical Constants, Time Scales, and the Fundamental Reference Frame", U.S. Naval Observatory, Circ. 163, Washington, D.C.

KAULA, W.M., 1966, "Introduction to Satellite Geodesy", Blaisdell, Waltham, Mass., 124pp.

KEARSLEY, A.H.W., 1984, "Precision Limitations and Data Requirements for the Determination of Relative Geoid Heights from Gravimetry", Univ. of Uppsala, Inst. of Geophysics, Dept. of Geodesy Rept. No.26, 110pp.

KLEUSBERG, A., LANGLEY, R.B., SANTERRE, R., VANICEK, P. & WELLS, D.E., 1985, "Comparison of Survey Results from Different GPS Receiver Types", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.579-592.

KOUBA, J., 1983, "A Review of Geodetic and Geodynamic Satellite Doppler Positioning", Rev. Geophys. & Space Phys., 21(1), p.27-40.

KRAKIWSKY, E.J., 1981, "A Synthesis of Recent Advances in the Method of Least Squares", Publ. No.1003, Dept. of Surv. Eng., Univ. of Calgary, Calgary.

LADD, J.W., COUNSELMAN III, C.C. & GOUREVITCH, S.A., 1985, "The Macrometer (TM) II Dual-Band Interferometric Surveyor", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.175-180.

LANGLEY, R., BEUTLER, G., DELIKARAOGLOU, D., NICKERSON, B., SANTERRE, R., VANICEK, P. & WELLS, D., 1984, "Studies in the Application of the Global Postioning System to Differential Positioning", Tech.rept.no.108, University of New Brunswick, Fredericton, N.B.

LAWSON, C.L. & HANSON R.J., 1974, "Solving Least Squares Problems", Prentice-Hall, Englewood Cliffs, N.J., 340pp.

LEEMAN, R.W., MACAULAY, H.D. & FLETCHER, L.B., 1985, "GPS - A Users Critique", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.613-624.

LERCH, F.J., PUTNEY, B.H., WAGNER, C.A., & KLOSKO, S.M., 1981, "Goddard Earth Model Development for Oceanographic Applications (GEM 10B and 10C)", Marine Geodesy, 5(2), p.145-187.

LERCH, F.J., KLOSKO, S.M. & PATEL, G.B., 1983, "A Refined Gravity Field Model from LAGEOS". NASA Tech.Mem. 84986, Goddard Space Flight Center, Greenbelt, Md.

MacDORAN, P.F., 1983, "SERIES-GPS, Codeless Pseudo-Ranging Positioning Technology", CSTG Bulletin 5, p.46-55.

MacDORAN, P.F., SPITZMESSER, D.J. & BUENNAGEL, L.A., 1982, "SERIES: Satellite Emission Range Inferred Earth Surveying", in proc. Third Int. Geodetic Symp. on Satellite Doppler Pos., Las Cruces, New Mexico, Feb.8-12, p.1143-1164.

McARTHUR, D., BECK, N., LOCKHEAD, K. & DELIKARAOGLOU, D., 1985, "Precise Relative Positioning with the Macrometer V-1000 Interferometric Surveyor: Experiences at the Geodetic Survey of Canada", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.533-542.

MARTIN, T.V., OH, I.H., EDDY, W.F. & KOGURT, J.A., 1976, "Geodyn Systems Description", NAS-5-22849, Task 009, Goddard Space Flight Center, Greenbelt, Md.

MASTERS, E.G. & STOLZ, A., 1985, "Multi-station GPS with Orbit Improvement", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.395-402.

MELBOURNE, W.G. & THORNTON, C.L., 1984, "Proposed NASA Roles and Missions Using the GPS", Jet Propulsion Lab., Pasadena, California (unpublished).

MELBOURNE, W., ANDERLE, R., FEISSEL, M., KING, R., McCARTHY, D., SMITH, D., TAPLEY, B. & VICENTE, R., 1983, "Project MERIT Standards", U.S. Naval Observatory, Circ. 167, Washington, D.C.

MIKHAIL, E.M., 1976, "Observations and Least Squares", Dun-Donnelly, New York. 497pp.

MOREAU, R., BEUTLER, G., LECLERC, J.G., LABREQUE, B., LANGLEY, R.B. & SANTERRE, R., 1985, "The Quebec 1984 Macrometer Test: Comparison of Phase-Difference Processing and Network Solution Methods", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.557-566.

MUELLER, I.I., 1977, "Spherical and Practical Astronomy", Frederick Ungar Publishing Co., New York, 615pp.

NAKIBOGLU, S.M., KRAKIWSKY, E.J., SCHWARZ, K.P., BUFFETT, B., & WANLESS, B., 1985, "A Multi-Station, Multi-Pass Approach to Global Positioning System Orbital Improvement and Precise Positioning", Final Contract Rept 85-003, Geodetic Survey of Canada, Ottawa, April, 77pp.

PATTERSON, R., 1984, "The Effect of an Estimated Variance Factor on the Size of Error Ellipses", Aust.J.Geod.Photo.Surv., 41, p.37-60.

POPE, A.J., 1976, "The Statistics of Residuals and the Detection of Outliers", NOAA Tech. Rept. NOS65 NGS1.

RAPP, R.H., 1981, "The Earth's Gravity Field to Degree and Order 180 Using SEASAT Altimeter Data, Terrestrial Data and Other Data", Rept.no.322, Dept. of Geodetic Science, The Ohio State University, Columbus.

REIGBER, C., MUELLER, H., BOSCH, W., BALMINO, G. & MOYNOT, B., 1985, "GRIM Gravity Model Improvement Using Lageos (GRIM3-L1)", J.Geophys.Res., 90, p.9285-9300.

REMONDI, B., 1985a, "Modeling the GPS Carrier Phase for Geodetic Applications", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.325-336.

REMONDI, B., 1985b, "Distribution of Global Positioning System Ephemerides by the National Geodetic Survey", proc. First conf. on Civilian Applications of GPS - Inst. of Navigation, Greater Philadelphia Section, September 12, 1985.

RESCH, G.M., CHAVEZ, M.C. & YAMANE, N.I., 1982, "Description and Overview of an Instrument Designed to Measure Line-of-Sight Delay Due to Water Vapour", Jet Propulsion Lab. TDA Progress Rept., 42-72, Oct.-Dec.

RIZOS, C., 1980, "The Role of the Gravity Field in Sea Surface Topography Studies", Unisurv S-17, School of Surveying, University of New South Wales, Kensington, N.S.W., 299pp.

RIZOS, C. & STOLZ, A., 1985, "Force Modelling for GPS Satellite Orbits", in proc First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.87-98.

RIZOS, C., STOLZ, A. & MASTERS, E.G., 1984, "Surveying and Geodesy in Australia with GPS", Aust.Surveyor, 32, p.202-225.

RIZOS, C., MASTERS, E.G. & STOLZ, A., 1985, "GPS Ephemeris Considerations for the Australian Surveyor", Aust.Surveyor, 32, p.574-583.

ROELSE, A., GRANGER, H.W. & GRAHAM, J.W., 1971, "The Adjustment of the Australian Levelling Survey 1970-1971", Tech.rept.no.12, Division of National Mapping, Canberra.

RULAND, R. & LEICK, A., 1985, "Application of GPS in High Precision Engineering Survey Work", in proc First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.483-494.

RUSSELL, S.S. & SCHAIBLY, J.H., 1980, "Control Segment and User Performance", special issue of Navigation on GPS, p.74-80.

SAASTAMOINEN, J., 1973, "Contributions to the Theory of Atmospheric Refraction, Bull.Geodes., 107, p.13-34.

SCHANZLE, A.F., 1980, "Lectures on Orbital Mechanics and Trajectory Analysis", unpublished lecture notes, E.G.&G. Washington Analytical Services Center, Riverdale, Md.

SCHUTZ, B.E., TAPLEY, B.D. & EANES, R.J., 1982, In "Sources of LAGEOS Model Errors". Pres. at Third Crustal Dynamics Working Group Meeting, Goddard Space Flight Center, Greenbelt, Md., Oct 1982.

SCHWARZ, K.P. (ed.), 1981, Proceedings Second Int. Symp. on Inertial Technology for Surveying and Geodesy, Banff, June 1-5, 664pp.

SCHWIDERSKI, E.W., 1978, "Global Ocean Tides, Part 1, A Detailed Hydrodynamical Model", Rept. TR-3866, U.S. Naval Surface Weapons Center, Dahlgren, Va.

SPILKER, Jr., J.J., 1978, "GPS Signal Structure and Performance Characteristics", Navigation, 25(2), p.121-146.

STOLZ, A. & MASTERS, E.G., 1983, "Satellite Laser Range Measurements of the 3200 Km Orroral-Yarragadee Baseline", Aust.Surveyor, 31, p.557-562.

STOLZ, A., MASTERS, E.G. & RIZOS, C., 1984, "Determination of GPS Satellite Orbits for Geodesy in Australia", Aust.J.Geod.Photo.Surv., 40, p.41-52.

SWIFT, E.R., 1985, "NSWC's GPS Orbit/Clock Determination System", in proc First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.51-62.

TORGE, W., 1980. "Geodesy", Walter de Gruyter, Berlin, 254pp.

VAN DIERENDONCK, A.J., RUSSELL, S.S., KOPITZKE, E.R. & BIRNBAUM, M., 1980, "The GPS Navigation Message", special issue of Navigation on GPS, p.55-73.

VANICEK, P. & KRAKIWSKY, E.J., 1982, "Geodesy: The Concepts", North Holland, Amsterdam, 691pp.

WELLS, D., 1985, "Recommended GPS Terminology", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.903-923.

WOLF, H., 1963, "Geometric Connection and Re-orientation of Three-Dimensional Triangulation Nets", Bull.Geodes., 68, p.165-169.

WU, J.T., 1984, "Elimination of Clock Error in a GPS Based Tracking System", AIAA/AAS Conference, August 20-22.

YOUNG, L.E., NEILAN, R.E. & BLETZACKER, F.R., 1985, "GPS Satellite Multipath: An Experimental Investigation", in proc. of the First Int. Symp. on Precise Positioning with the Global Positioning System -- Positioning with GPS-1985, U.S. Dept. of Commerce, NOAA, Rockville, Md., May 1985, p.423-432.