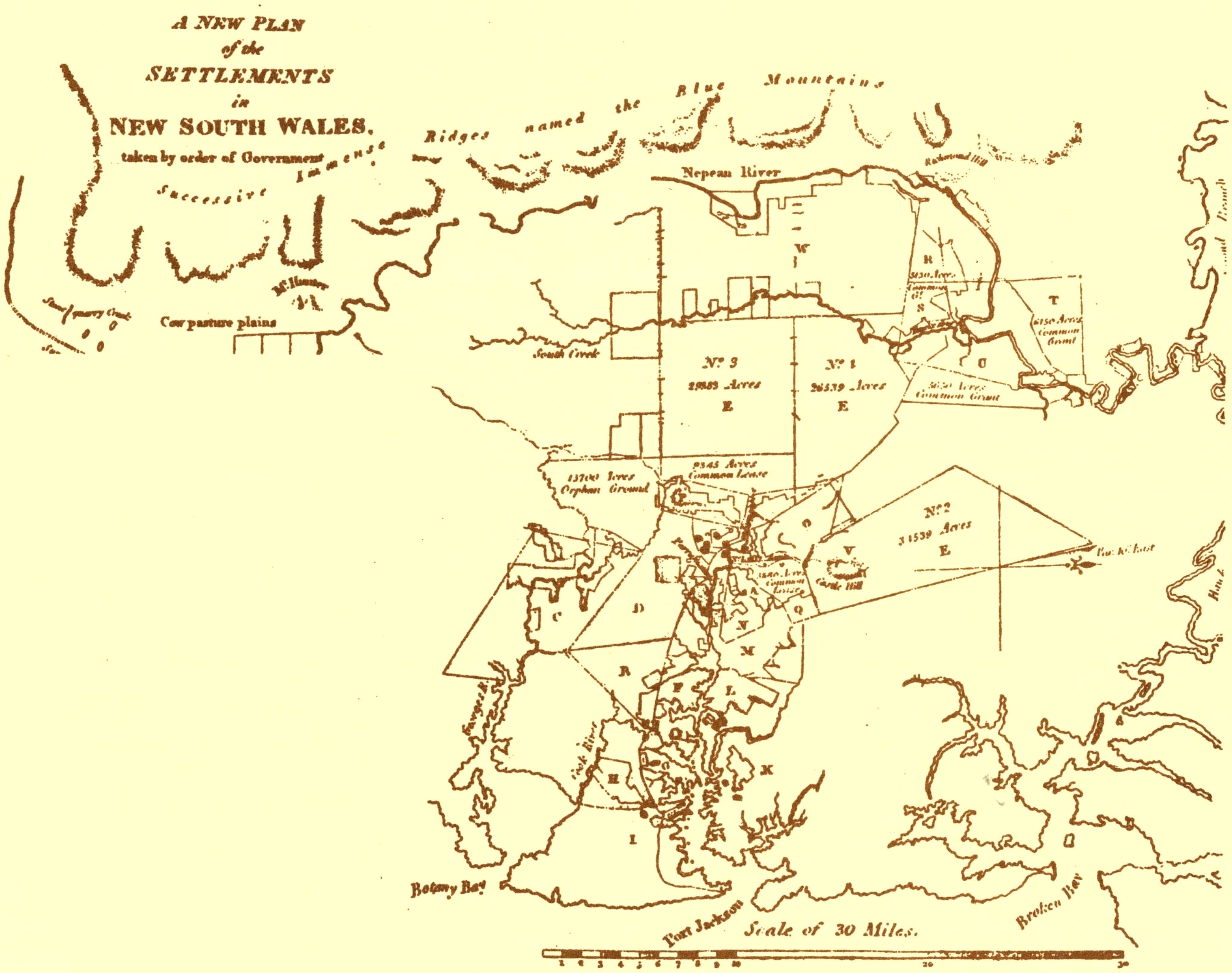
CONTRIBUTIONS TO GPS STUDIES

by

D.B. Grant, C.C. Mazur, B. Merminod, C. Rizos, A. Stolz

Editor: C. Rizos



UNISURV S-38, 1990

Reports from

SCHOOL OF SURVEYING

THE UNIVERSITY OF NEW SOUTH WALES KENSINGTON, NSW AUSTRALIA 2033

UNISURV REPORT S-38, 1990

CONTRIBUTIONS TO GPS STUDIES

C. Rizos (Ed.)

Received: August, 1990

SCHOOL OF SURVEYING
UNIVERSITY OF NEW SOUTH WALES
P.O. BOX 1
KENSINGTON N.S.W. 2033
AUSTRALIA

National Library of Australia

Card No. and ISBN 0 85839 056 6

TABLE OF CONTENTS

DEALING WITH GPS BIASES: SOME THEORETICAL AND SOFTWARE CONSIDERATIONS

D.B. Grant, C. Rizos, A. Stolz

TIME AND THE GLOBAL POSITIONING SYSTEM

C. Rizos, D.B. Grant

RESOLUTION OF THE CYCLE AMBIGUITIES

B. Merminod

A STUDY OF THE VARIABILITY AND PRECISION OF GPS BASELINE MEASUREMENTS

C.C. Mazur

DEALING WITH GPS BIASES: SOME THEORETICAL AND SOFTWARE CONSIDERATIONS

D.B. Grant[1]
C. Rizos
A. Stolz

School of Surveying
University of New South Wales
P.O.Box 1
Kensington, N.S.W., 2203, Australia

ABSTRACT

In this paper the basic observation model is derived and different processing strategies that can be implemented for the determination of station coordinates will be briefly described. In particular, the rank deficiencies present in the system of GPS normal equations are discussed in the context of conventional geodetic network design and analysis practices. Methods of eliminating rank deficiencies, particularly those arising from the various bias parameters, as used in various GPS processing software, will be described.

[1]Now at the Dept. of Survey and Land Information, Wellington, New Zealand.

1. INTRODUCTION

Up until the mid 1980's the standard surveying techniques for determining the coordinates of points were based on the theodolite and electronic distance measurement (EDM) technologies. These measurement technologies were the basis of the methods of triangulation and trilateration used for the establishment of geodetic control networks. Although in the 1970's a lot of survey control was also established using the TRANSIT Doppler satellite system, the overall impact of satellite positioning technology on everyday survey operations was relatively minor. On the other hand, the use of satellites for specialised geodetic applications has a comparatively long history going back to the launch of the first artificial satellite in 1957. The analysis of tracking data to artificial earth satellites, initially with optical systems and then subsequently with microwave and laser tracking systems, has made a significant contribution to the mapping of the earth's gravity field. Nowadays, the techniques of satellite geodesy can determine of the coordinates of selected points on the earth's surface to such a high accuracy that the relative movement of continents can be measured.

For over 20 years, traditional terrestrial geodetic practices for the establishment of survey control, and the techniques of satellite geodesy, supporting earth gravity studies, and more recently for crustal movement surveys, have operated side-by-side with a minimum of interaction. However, with the advent of the Global Positioning System (GPS) this neat separation of the roles of terrestrial and satellite geodesy is no longer possible. In particular, GPS is the first satellite positioning technology to seriously challenge traditional terrestrial techniques for the establishment of geodetic survey control.

Although GPS is obviously a satellite technology, there are a number of special features of the GPS observable (and hence in the data processing strategies employed) that set it apart from both terrestrial geodesy and conventional satellite technologies such as Satellite Laser Ranging (SLR). In particular:

- GPS is a multi-satellite system, in which each satellite can be considered an orbiting "trig station" with known coordinates.
- The GPS observable is a one-way biased range.
- To support survey operations, two or more ground receivers are usually employed.

In contrast, conventional satellite techniques are based on one tracking station - one satellite configurations, in which positioning is carried out in an "absolute" sense, and the modelling of satellite motion is fundamentally physical in nature and is critical to the method's success. The analysis techniques of satellite geodesy have their origins in Celestial Mechanics. Furthermore, the "raw" SLR measurement, for example, is very accurate and not significantly "biased".

GPS is essentially a technique for relative positioning sharing many similarities with the method of trilateration, but in which the known trig stations have been replaced by stations that are in orbit. For most applications, GPS is a geometric technique that owes more to conventional notions of geodetic network observation and analysis than it does to the "traditional" satellite techniques such as SLR. However, the presence of a number of biases in the measurement demand that special solution strategies be used.

2. GPS OBSERVATION EQUATIONS

The GPS satellites transmit on two L-band frequencies: L1 at 1575.42 MHz and L2 at 1227.60 MHz. The observation equation for the carrier beat phase is developed below, and is valid for measurements made on both the L1 and L2 frequencies.

2.1 Carrier Beat Phase

2.1.1 Clock Phase Error

In developing the observation equations for carrier beat phase it is useful to suppose initially that every clock, or oscillator, can be compared directly with a "perfect" oscillator having a known and constant frequency f_0 in our chosen reference time scale (for example GPS Time). The phase of the "perfect" oscillator at time T is represented by $\Phi(T)$. The behaviour of phase with time in general obeys the relation:

$$\Phi(T) = \Phi(T_0) + \int_{T_0}^{T} f(T)\, dT \tag{2-1}$$

where $f(T)$ is the time dependent frequency of the oscillator and T_0 is the time at some arbitrary initial reference epoch. Because the frequency of the "perfect" oscillator is constant at f_0 we can write:

$$\Phi(T) = \Phi(T_0) + f_0\,(T - T_0) \tag{2-2}$$

The oscillators in the satellites and receivers are used to generate timed signals such as the P code, C/A code and observation time-tags. It is common therefore, to think of them as clocks and to consider the errors caused by the variation of frequency as clock errors. This use of the term "clock error" can cause difficulties when we wish to distinguish between the carrier phase error caused by variations in the receiver oscillator frequency, and errors in the observation time-tag which are also affected by these frequency variations. It is preferable to use the terms **clock phase error** (or clock range error in the case of pseudo-range) for the former and **time-tag error** for the latter (RIZOS & GRANT, 1990). Although the carrier phase and observation time-tags are generated by the same oscillator they are generally treated as if they were independent in GPS processing. For instance, the time-tags are often corrected according to the clock error estimates obtained from a pseudo-range point position. These corrections are not generally accompanied by an equivalent correction to the observed carrier phase.

If we use $\phi(T)$ to represent the phase of our real (imperfect) satellite or receiver oscillator at time T we can define the clock phase error $\varepsilon(T)$ in terms of the phase error $(\phi(T) - \Phi(T))$:

$$\varepsilon(T) = \frac{1}{f_0}\,(\phi(T) - \Phi(T)) \tag{2-3}$$

or

$$\phi(T) = \Phi(T) + f_0\, \varepsilon(T) \tag{2-4}$$

Of course, in using equations (2-3) or (2-4) we must account for both the integral and fractional parts of the phase. This point is important when we consider how phase is actually measured.

2.1.2 Signal Transit Time

We now consider the transmission of a signal from a satellite i to a receiver j. The **time of reception** of a signal at receiver j is T_j. The **time of transmission** of that signal from satellite i is T_j^i. The **transit time** of the signal from i to j is $\tau_j^i(T_j)$ and is defined by:

$$\tau_j^i(T_j) = T_j - T_j^i \tag{2-5}$$

The beat phase formed as an observation in the receiver is the difference between the phase of the received signal and the phase of the local receiver oscillator. The phase of the received signal at the time of reception T_j is equal to the phase of the transmitted signal at the time of transmission T_j^i. Therefore:

$$\begin{aligned}\phi_{bj}^i(T_j) &= \phi_{rj}^i(T_j) - \phi_{loj}(T_j) \\ &= \phi^{ti}(T_j^i) - \phi_{loj}(T_j)\end{aligned} \tag{2-6}$$

where :

- $\phi_{bj}^i(T_j)$ = the carrier beat phase for receiver j, satellite i, at reception time T_j
- $\phi_{rj}(T_j)$ = the received signal phase from satellite i at receiver j at time T_j
- $\phi_{loj}(T_j)$ = the local oscillator phase of receiver j at time T_j
- $\phi^{ti}(T_j^i)$ = the transmitted signal phase from satellite i at transmission time T_j^i

From equation (2-4) we can write an expression relating phase and clock phase error of the local oscillator:

$$\phi_{loj}(T_j) = \Phi(T_j) + f_0\, \varepsilon_j(T_j) \tag{2-7}$$

where $\varepsilon_j(T_j)$ is the clock phase error of receiver j. Using equations (2-2), (2-4) and (2-5), an expression relating transmitted phase and satellite clock phase error, at reception time can be formed:

$$\begin{aligned}\phi^{ti}(T_j^i) &= \Phi(T_j^i) + f_0\varepsilon^i(T_j^i) \\ &= \Phi(T_j) + f_0\varepsilon^i(T_j^i) - f_0\tau_j^i(T_j)\end{aligned} \tag{2-8}$$

where $\varepsilon^i(T)$ is the clock phase error attributable to the oscillator of satellite i. Combining equations (2-6), (2-7) and (2-8) leads to an expression for **carrier beat phase**:

$$\phi_{bj}^i(T_j) = -f_0[\tau_j^i(T_j) - \varepsilon^i(T_j^i) + \varepsilon_j(T_j)] \tag{2-9}$$

The transit time component in equation (2-9) is made up of two parts. The main part is derived from the true **geometric range** ρ_j^i between the satellite at the time of transmission and the receiver at the time of reception. This can be determined from the position vectors of the satellite and the receiver, as defined in the same reference system. Given the speed of the signals (c, the speed of light) we can calculate the time taken for the signal to travel this distance (τ_j^i). The value we use for the speed of light c is the speed in a vacuum and hence this is the primary means of defining scale in GPS. It is the satellite and receiver position information contained in the geometric range that allows us to use GPS for positioning. The second part of the transit time term accounts for the extra time taken for the signal to travel through the earth's atmosphere. This is caused by a change in velocity but can also be modelled as a time delay, a phase delay or as an increase in range. We can incorporate this as a phase correction term ϕ_{atmos}. (This may be further broken down into terms for different parts of the atmospheric delay; ϕ_{ion} for the ionosphere, ϕ_{trop} for the troposphere.) Hence:

$$f_0\, \tau_j^i(T_j) = (\, f_0 / c\,)\, \rho_j^i(T_j) + \phi_{atmos} \tag{2-10}$$

Combining equations (2-9) and (2-10) we obtain a model for carrier beat phase:

$$\phi_{bj}^i(T_j) = -(\, f_0 / c\,)\, \rho_j^i(T_j) + f_0\, [\varepsilon^i(T_j^i) - \varepsilon_j(T_j)] - \phi_{atmos} \tag{2-11}$$

2.1.3 Measured Carrier Beat Phase

The <u>measured</u> carrier beat phase differs from the above <u>modelled</u> beat phase in a number of important respects. Firstly, the carrier beat phase measurements will have random noise (and signal interference such as multipath) associated with the measurement process. This is contained in an additional term ϕ_{noise}. Secondly, the definition of clock phase error in equation (2-4) depends on the integral and fractional part of the carrier phase. When the carrier beat phase is measured in the receiver the measurement is ambiguous with regards to the number of integer cycles. This integer is set in a counter to some arbitrary value when the satellite signal is first acquired (for example zero). There is therefore an unknown integer number of cycles n_j^i difference between the measured carrier beat phase and the model beat phase in equation (2-11). This will be unique to a particular satellite-receiver pair. Furthermore, it will be a constant for as long as the receiver continues to track and count the integer number of cycles from the time the satellite signal is first acquired. At any epoch other than the initial measurement epoch, the instrument measures the fractional phase $\phi_{fj}^i(T_j)$ and, in addition, takes a reading $C_R(T_j)$ on the counter. This combined fractional and integer phase observation is therefore referred to as **integrated**

carrier beat phase, and n_j^i is the **cycle ambiguity**. Thirdly, if the receiver at any time fails to track the signal correctly and the ambiguity changes between observation epochs then a **cycle slip** $S(T_j)$ has occurred.

The model of the measured carrier beat phase therefore includes these three additional terms:

$$\begin{aligned} \phi_{bj}{}^i(T_j) &= \phi_{fj}{}^i(T_j) + C_R(T_j) + S(T_j) \\ &= -\,(f_0 / c)\,\rho_j{}^i(T_j) + f_0\,[\varepsilon^i(T_j{}^i) - \varepsilon_j(T_j)] \\ &\quad + n_j{}^i - \phi_{atmos} + \phi_{noise} \end{aligned} \qquad (2\text{-}12)$$

This is the basic observation equation for integrated carrier beat phase. We do not explicitly indicate the dependence of ϕ_{atmos} and ϕ_{noise} on the receiver, satellite or epoch, but it should be noted that these terms will vary slightly for every observation. The change in observed carrier beat phase with time is therefore equal to the change in any of the following quantities:

- satellite - receiver geometric range,
- satellite - receiver clock phase error difference,
- net number of cycle slips,
- atmospheric delay or,
- measurement noise.

2.1.4 Mathematical Model for Integrated Carrier Beat Phase

Equation (2-12) is an example of a <u>physical</u> model of the integrated carrier beat phase observation. In order to develop an appropriate parameter model for carrier phase data processing, a <u>mathematical</u> model needs to be set up first. The mathematical model incorporates only those terms of the physical observation model that will be parametrised for the adjustment, but without explicitly defining the functional models for the terms. Equation (2-12) can therefore be written in the form:

$$\begin{aligned} \phi_{bj}{}^i(T_j) &= -\,(f_0 / c)\,\rho_j{}^i(T_j) + f_0\,[\varepsilon^i(T_j{}^i) - \varepsilon_j(T_j)] \\ &\quad + n_j{}^i - \phi_{atmos} \end{aligned} \qquad (2\text{-}13)$$

Note that the fractional and integer part of the observation are now combined into a single term $\phi_{bj}{}^i(T_j)$, and the noise and cycle slip terms have been dropped.

The mathematical model for pseudo-range, or "code phase" as it is sometimes known, is also equation (2-13), but without the cycle ambiguity term.

2.2 The Nature of GPS Observation Model Biases

The basic GPS observables, carrier beat phase as well as pseudo-ranges, are essentially **biased ranges**. The ideal quantity required for GPS data processing is the true range ρ_j^i . The true range contains all the geometric information necessary to determine the receiver coordinates and/or the satellite position. One of the challenges in GPS processing is therefore to develop data analysis strategies that best handle the measurement biases. The biases arise from a number of sources:

- clock errors in both the receiver clocks (ε_j) and satellite clocks (ε^i),
- the atmospheric refraction delay (ϕ_{atmos}), and
- cycle ambiguities (n_j^i), in the case of carrier phase observations.

Upon adopting equation (2-13) as the basic measurement model, and proceeding with the normal linearisation process required for a least squares adjustment, a further class of biases is introduced into the residual quantity: measured range (phase or code) minus modelled (biased) range. This "O-C" quantity now also includes errors in the *a priori* information contained in the modelled range, primarily the errors in the assumed known receiver and satellite coordinates (used to derive the "calculated" value for ρ_j^i). These additional biases can be:

(a) explicitly included in equation (2-13) in the form of corrections to the *a priori* information (receiver and/or satellite coordinates) in the event of these parameters being adjusted, or
(b) neglected (for example when no receiver or satellite coordinate adjustment is performed), in which case these biases are now unmodelled errors.

A further challenge in GPS processing is therefore to develop strategies that either reliably estimate the additional model biases introduced into a least squares adjustment, or at the very least, minimise the effects of neglected systematic and random (error) biases. The basic observable in GPS data processing can therefore be considered to contain the following information:

(1) geodetic parameters of interest, such as the receiver or satellite coordinates being estimated,

(2) explicit biases in the measurements that can be parametrised in equation (2-13), and

(3) residual errors that may contaminate the "O-C" quantity and which are not parametrised in the observation model, arising from measurement noise, cycle slips, or faulty *a priori* knowledge of the biased range parameters.

In particular we wish to direct our attention to the parameters that are explicitly modelled: the geodetic parameters, and the remaining "nuisance" parameters (cycle ambiguity, clock errors, atmosphere). Some of these biases are specific to one receiver site, others are specific to a particular satellite, while others are related to specific receiver-satellite pairs. Within this group of nuisance parameters, the bias subset comprising the cycle ambiguity and clock phase error terms will be referred to collectively in this paper as the **"clock biases"**.

2.3 Accounting for Biases

Dealing with the "nuisance" parameters in GPS solutions based on the mathematical model (2-13) is not an inconsiderable problem. In fact much of the development effort in GPS software has gone into investigating ways of best reducing the computational burden imposed by having to determine the biases present in the GPS range/phase measurements. Several approaches can be taken:

(1) They can be estimated as explicit (nuisance) parameters.

(2) Those biases linearly correlated across different datasets can be eliminated by differencing.

(3) The biases can be directly measured, for example using observations on a second frequency in the case of the ionospheric delay.

(4) The biases can be modelled (possibly using external information), for example, as is sometimes attempted for the tropospheric delay.

(5) They can be ignored.

2.3.1 Dealing with Clock Biases and the Fundamental Differencing Theorem

The only options, however, that are applicable to the case of the "clock biases" are (1) and (2). The questions that then immediately spring to mind are:

- Do the nuisance parameter approach (option (1)) and the data differencing approach (option (2)) both lead to the same result?
- Which is the "best" approach?

The **fundamental differencing theorem** is useful in this regard:

> **"Linear biases can be accounted for either by reducing the number of observations so that the biases cancel, or by adding an equal number of unknowns to model the biases. Both approaches give identical results."** (LINDLOHR & WELLS, 1985)

Thus the effect of the clock phase errors (ε_j and ε^i) and cycle ambiguity (n_j^i) may be eliminated by differencing the basic equation (2-13) between satellites and between receiver sites, observed simultaneously, and perhaps between observation epochs as well; or by estimating them explicitly, and the result would be equivalent.

The observation modelled by equation (2-13) is sometimes called "one-way phase" or "undifferenced phase" to distinguish it from the more commonly used single-, double- or triple-differenced phase observations described in section 2.4. The term "undifferenced phase" is, strictly speaking, a misnomer because we can see in equation (2-6) that it is formed by differencing the incoming and local (receiver) oscillator phases. Nevertheless we can refer to the two approaches to dealing with the clock biases as being the "undifferenced" or "nuisance parameter" approach on the one hand, and the "differenced" approach on the other.

2.3.2 "Undifferenced" & "Differenced" Approaches to Handling Clock Biases

However, one comment on the equivalence between the "undifferenced" and the "differenced" approach is worth making here. The double-differencing approach (difference between satellites to eliminate ε_j , and between receivers to eliminate ε^i) does introduce mathematical correlations in the resultant double-difference observables. Only if this correlation is reflected in the variance-covariance (VCV) matrix of the double-difference observables will equivalence between this method and the undifferenced approach be assured. The same holds true for triple-differences. As a consequence, to ensure this equivalence it is necessary to explicitly construct the VCV matrix at each observation epoch (of course, in practice, advantage is taken of the fact that the weight matrix would only change with a change in the satellite/receiver combinations used to construct the double- or triple-differences).

In the undifferenced approach, there is more flexibility in the modelling options. For example, the clock phase errors in equation (2-13) are a function of time. If the satellite clock, or receiver clock, or both, had a stable, predictable behaviour (as when external atomic frequency standards such as cesium or rubidium oscillators are used), a parameter model based on a clock polynomial could be used, and the coefficients estimated, together with the cycle ambiguity, as session parameters. In general however, the clock phase errors are assumed to possess the characteristics of "white noise" and are explicitly modelled by independent bias terms on an epoch-by-epoch basis. The total number of clock bias parameters is (R+S)T+RS , where R is the number of receivers, S the number of satellites and T is the number of epochs in a session of data. There is therefore a potentially large number of these bias parameters to be estimated. In the case of 4 receivers, 4 satellites and 60 (one minute) measurement epochs, the total number of clock bias parameters would be 496!

Further discussion on the relative merits of the differenced and undifferenced approaches for GPS processing and on the clock phase error modelling options is given in section 2.5.

2.4 <u>Eliminating GPS Biases by Observation Differencing</u>

A common technique in GPS phase adjustment is to form new observables from differences of the carrier beat phase observations in which the unwanted clock biases have been eliminated.

2.4.1 Between - Satellite Differences

Receiver clock phase errors may be eliminated by forming the difference between simultaneous observations from one receiver to two satellites. The operator ∇ indicates a **between-satellite difference**. The differenced observable can be written as:

$$\nabla\phi_j^{12}(T_j) = \phi_{bj}^{1}(T_j) - \phi_{bj}^{2}(T_j) \tag{2-14}$$

where we assume that the two observations from receiver j to satellites 1 and 2 have been made at the same time T_j . We can write the differenced observation using equation (2-13):

$$\begin{aligned}\nabla\phi_j^{12}(T_j) = & -(f_0/c)\,(\rho_j^{1}(T_j) - \rho_j^{2}(T_j))\\ & + f_0\,[\,\varepsilon^{1}(T_j^{1}) - \varepsilon^{2}(T_j^{2})\,]\\ & + n_j^{1} - n_j^{2} - \nabla\phi_{atmos}\end{aligned} \tag{2-15}$$

Note that the receiver clock phase error has been eliminated and the clock bias terms that remain are the between-satellite clock phase error and the between-satellite cycle ambiguity. In as far as setting up the design matrix for a least squares adjustment based on the between-satellite difference, we now have a choice in the cycle ambiguity modelling. We can persist with individual ambiguity modelling for n_j^1 and n_j^2 , or we can adopt a new definition for the ambiguity parameter consisting of the between-satellite cycle ambiguity k_j^{12} (= n_j^1 - n_j^2), see section 3.4.3.

2.4.2 Between - Station Differences

The satellite clock phase errors may be (almost) eliminated by forming the difference between observation from two receivers to one satellite at the same time. The operator Δ indicates a **between-station difference** also sometimes referred to, ambiguously, as a single-difference. Although the nominal observation or reception times of the two measurements may be the same, the actual time of measurement may differ from one receiver to the next because of receiver time-tag errors. We can construct the differenced observation from:

$$\Delta\phi_{12}^{i}(t) = \phi_{b1}^{i}(T_1) - \phi_{b2}^{i}(T_2) \tag{2-16}$$

where t is the nominal receiver clock time, and T_i is the true time of reception of the signal at receiver i . Using equation (2-13):

$$\begin{aligned}\Delta\phi_{12}^{i}(t) = & -(f_0/c)\,(\rho_1^{i}(T_1) - \rho_2^{i}(T_2))\\ & + f_0\,[\,\varepsilon^{i}(T_1^{i}) - \varepsilon^{i}(T_2^{i})\,]\\ & - f_0\,[\,\varepsilon_1(T_1) - \varepsilon_2(T_2)\,]\end{aligned}$$

$$+ n_1^i - n_2^i - \Delta\phi_{atmos} \tag{2-17}$$

The satellite clock phase errors may not completely cancel because they refer to different transmission times. The transit times to the satellite are different because of the different satellite - receiver ranges, and this difference may be up to 1 millisecond for a 300 km baseline. However, as the satellites use stable atomic oscillators, it is usual to assume that the satellite clock phase errors are identical and thus cancel when the between-station difference is formed (RIZOS & GRANT, 1990):

$$\begin{aligned} \Delta\phi_{12}^i(t) = \; & -(f_0/c)\,(\rho_1^i(T_1) - \rho_2^i(T_2)) \\ & - f_0\,[\varepsilon_1(T_1) - \varepsilon_2(T_2)] \\ & + n_1^i - n_2^i - \Delta\phi_{atmos} \end{aligned} \tag{2-18}$$

Note that the clock bias terms that remain are the between-station clock phase errors and the between-station cycle ambiguity. As in the case of the between-satellite difference, we can persist with individual ambiguity modelling for n_1^i and n_2^i , or we can adopt a new definition for the ambiguity parameter consisting of the between-station cycle ambiguity k_{12}^i $(= n_1^i - n_2^i)$.

2.4.3 Double - Difference Observable

This is the term usually applied to the observable which has been formed by differencing between satellites and stations. This double-difference phase may be created by forming the between-satellite difference phase and then differencing between stations or by first forming the between-station difference and then differencing between satellites:

$$\begin{aligned} \phi_{DD}(t) &= \nabla\Delta\phi_{12}^{12}(t) \\ &= \nabla\phi_1^{12}(T_1) - \nabla\phi_2^{12}(T_2) \\ &= \Delta\phi_{12}^1(t) - \Delta\phi_{12}^2(t) \end{aligned} \tag{2-19}$$

If we again accept the assumption that satellite clock phase errors cancel in a between-station difference, then we can use equation (2-18) to derive the double-difference observation equation:

$$\begin{aligned} \phi_{DD}(t) = \; & -(f_0/c)\,(\rho_1^1(T_1) - \rho_1^2(T_1) - \rho_2^1(T_2) + \rho_2^2(T_2)) \\ & + n_1^1 - n_1^2 - n_2^1 + n_2^2 - \Delta\nabla\phi_{atmos} \end{aligned} \tag{2-20}$$

The only clock biases remaining in this equation are the integer cycle ambiguities. We can explicitly model the 4 individual cycle ambiguities for n_1^1, n_1^2, n_2^1 and n_2^2 , or we can adopt a new definition for the ambiguity parameter consisting of the double-differenced cycle ambiguity k_{12}^{12} (= $n_1^1 - n_1^2 - n_2^1 + n_2^2$), see section 3.4.3.

2.4.4 Triple - Difference Observable

Bias parameters which are constant with time may be eliminated by forming the **between-epoch difference**. The operator δ is used to represent this difference and is applied to double-difference observables to obtain the **triple-difference** observables. If the nominal reception times, in receiver clock time, are t_a and t_b , then applying equation (2-20):

$$\begin{aligned}
\phi_{TD}(t_{ab}) &= \delta\Delta\nabla\phi_{12}{}^{12}(t_{ab}) \\
&= \phi_{DD}(t_a) - \phi_{DD}(t_b) \\
&= -(f_0/c)\,(\,\rho_1{}^1(T_{a1}) - \rho_1{}^2(T_{a1}) - \rho_2{}^1(T_{a2}) + \rho_2{}^2(T_{a2})\,) \\
&\quad + (f_0/c)\,(\,\rho_1{}^1(T_{b1}) - \rho_1{}^2(T_{b1}) - \rho_2{}^1(T_{b2}) + \rho_2{}^2(T_{b2})\,) \\
&\quad - \delta\Delta\nabla\phi_{atmos} \qquad (2\text{-}21)
\end{aligned}$$

Note that all the clock bias terms, including the integer ambiguities, have been eliminated. If our assumption that the integer ambiguity is constant in time is false then an extra term will be required for the cycle slip. In fact, not only have the (unknown) clock biases been removed, but also the effect of other biases arising from atmospheric delay, satellite ephemeris error and receiver coordinate error, are substantially reduced by triple differencing, making this observation especially useful for cycle slip detection.

2.5 GPS Observation Modelling: the Differenced and Undifferenced Approach

All forms of the phase observation model, be it the one-way phase equation (2-13), or the various differenced observation equation models described in section 2.4, are valid for GPS phase adjustment. In fact, any combination of between-satellite, between-station and between-epoch differencing operation can be applied to construct a GPS observable. The only requirement is that the resulting mathematical observation model must account for all the clock biases either by differencing them all out (for example, the triple-difference), or explicitly modelling them all, or a combination of the two (such as in the case of the single- and double-differences).

2.5.1 Clock Bias Modelling

It should be pointed out however that the explicit clock bias modelling approach (in the undifferenced or single-difference observation models) introduces an additional problem, namely: what parameter model should be adopted for the clock bias(es)? The parameter model may be a constant for the whole observation session, or a new (and independent) constant for each observation epoch, or something between these two extremes, such as a truncated power series in time. The integer cycle ambiguity is modelled as a constant bias term n_j^i (unless a cycle slip occurs in which case a new ambiguity term should be introduced), whereas the clock phase errors (ε_j and ε^i) are often modelled as discrete biases on an epoch-by-epoch basis. This is the approach normally adopted with the use of the one-way phase observable and although the number of clock phase error parameters to be estimated is large, these may be eliminated from the GPS solution at each epoch (GOAD, 1985) using partitioning techniques such as the "Helmert-blocking" procedure (CROSS, 1983). Note that epoch-by-epoch modelling of the clock phase errors is also possible for partially differenced observables. For example, in a between-satellite-difference (equation (2-15)) in which the receiver clock phase error has been differenced out, the remaining satellite clock phase errors can be modelled in this manner.

The resulting **normal equation matrix** from any of the above GPS observables, can be partitioned in the following way:

$$\mathbf{NEM} = \begin{bmatrix} \mathbf{N_{cc}} & \mathbf{N_{ce}} \\ \mathbf{N_{ec}} & \mathbf{N_{ee}} \end{bmatrix} \tag{2-22}$$

where the $\mathbf{N_{ee}}$, $\mathbf{N_{ec}}$ and $\mathbf{N_{ce}}$ submatrices contain the contributions of the design submatrix $\mathbf{A_e}$ of the clock bias (or "epoch") parameters, and $\mathbf{N_{cc}}$ the contribution of only the non-bias (or "common") parameters in the design submatrix $\mathbf{A_c}$. Reducing the normal equation matrix using the techniques described in, for example, CROSS (1983):

$$\mathbf{NEM} = \begin{bmatrix} \mathbf{N^*_{cc}} & \mathbf{0} \\ \mathbf{0} & \mathbf{N^*_{ee}} \end{bmatrix} \tag{2-23}$$

the resultant $\mathbf{N^*_{cc}}$ submatrix leads to the non-bias parameter solution (receiver coordinates and/or satellite state parameters).

2.5.2 Conditions for Equivalence between Results of GPS Phase Adjustment

Although the undifferenced and the various differenced observation equations are all theoretically equivalent, in general they will not lead to identical non-bias results unless:

(1) the clock bias parameter models are equivalent, and
(2) the mathematical correlations introduced by differencing are taken into account in the VCV matrix of the observables.

As the non-bias solution is $\mathbf{x_c} = \mathbf{N^{*}_{cc}}^{-1} \cdot \mathbf{b^*_{cc}}$, these two conditions are in fact those necessary to ensure that the product defining the solution for $\mathbf{x_c}$ is invariant with respect to the choice of design submatrices $\mathbf{A_c}$ and $\mathbf{A_e}$, and the VCV of the observations.

In relation to (1) above, by explicitly estimating the clock phase errors on an epoch-by-epoch basis, there is no assumption made as to the structure and stochastic behaviour of these errors over the entire period of the observation span. In other words the clock errors are treated as "white noise", and hence are not serially correlated. In contrast, the use of a truncated power series in time implies that the clock errors are predictable and their structure known a priori. In general, however, GPS satellite and receiver clock errors cannot be modelled satisfactorily with truncated polynomials, in which case the $\mathbf{b}^*_{\mathbf{cc}}$ vector will be different depending upon the clock phase error model used. Hence the coordinate results obtained will also vary with the clock phase error model used <u>even if the same differenced observable is used</u>. As a corollary, exact equivalence between approaches which use explicit clock phase error modelling (one-way and single-differences) and those that have eliminated those same clock errors through differencing (double- and triple-differences) is only possible if the clock phase error is modelled as an independent clock error parameter at each epoch.

In section 2.3 the "undifferenced" technique of GPS phase adjustment was loosely defined as that approach which makes use of the one-way phase observable (no matter what the explicit clock phase error model may be), to contrast it with the "differenced" approach which is characterised by the use of reduced observables in which some or all of the clock biases have been removed from the observation modelling. The "undifferenced" approach in which the clock phase errors are independently modelled on an epoch-by-epoch basis is also sometimes referred to as the **implicit differencing** technique (KING et al, 1987), in order to distinguish it from the "undifferenced" approach based on a less general clock phase error model such as a polynomial.

Condition (2) was mentioned in section 2.3. In short, given a differencing operator $\mathbf{D}$ that creates a vector of linearly independent double-difference observables ϕ_{DD} (equation (2-20)) from a vector of undifferenced phases ϕ_{UD} (equation (2-13)):

$$\phi_{DD} = \mathbf{D}\, \phi_{UD} \qquad (2\text{-}24)$$

and a VCV matrix of the undifferenced phases $\mathbf{Q}_{UD}$, the VCV matrix of the double differences $\mathbf{Q}_{DD}$ is, by the law of propagation of variances:

$$\mathbf{Q}_{DD} = \mathbf{D}\, \mathbf{Q}_{UD}\, \mathbf{D}^T \qquad (2\text{-}25)$$

If $\mathbf{Q}_{DD}$ is used in an adjustment of the double-difference observations ϕ_{DD} , the result will be the same as if $\mathbf{Q}_{UD}$ had been used in an adjustment of the observations ϕ_{UD} with the clock phase error parameters being estimated explicitly at each epoch. This is also true for any other type of differenced observable, as long as condition (1) is also satisfied.

2.5.3 The Roles and Merits of Various Phase Observables in GPS Adjustments

In summary, the undifferenced approach to GPS phase adjustment is identical to using single-, double- or triple-differences only if the mathematical correlations introduced during the differencing operation are included in the VCV matrix and any explicitly modelled clock biases are assumed to be offsets (on an epoch-by-epoch basis in the case of the clock phase errors, and on a session-by-session basis for the cycle ambiguity parameters).

Often, however, these conditions for exact equivalence are not met. Nevertheless the various differenced observables do have their role to play. The between-satellite and between-station differences, for example, are useful for cycle slip editing, while triple-differences are often used to obtain preliminary site coordinate solutions because of their relative insensitivity to cycle slips in the phase data. However, it is the double-differenced observable that is most commonly used in precise GPS phase adjustment. Generally the mathematical correlations are neglected, and hence the results are not strictly equivalent to using the undifferenced approach, but they are still suitable for most survey applications seeking relative accuracies of a few parts per million of the baseline lengths (BEUTLER et al, 1987).

The use of the undifferenced observable approach for GPS phase adjustment is nevertheless preferable for a number of reasons, including (WELLS et al, 1987):

(i) It is closer to the physical GPS measurement than any of the differenced observables.

(ii) The basic geodetic parameters remain the receiver and satellite state vectors, rather than the more complex receiver baseline or inter-satellite vectors.

(iii) There is no complicated "bookkeeping", as in the case of differenced observables.

(iv) There is no need to consciously incorporate correlations arising from the differencing operation, as an adjustment of undifferenced phase data is always rigorous.

(v) The undifferenced model allows one the flexibility of implementing alternate methods of modelling clock phase errors (for example, clock polynomials, or a Kalman filter), whereas if the observations are differenced to eliminate the clock errors then alternative methods of clock error modelling are not possible.

For these, and other reasons, the undifferenced model has been adopted in the University of New South Wales Satellite Measurement Analysis Software System (RIZOS & STOLZ, 1988). However, the various differenced GPS observables, together with the Fundamental Differencing Theorem, do provide us with a convenient way of analysing the rank defects arising from the explicit modelling of the clock biases (ε_j , ε^i and n_j^i).

3. GPS PHASE DATA PROCESSING

The conditions for equivalence in GPS phase adjustments discussed in section 2.5 do provide the software engineer with a choice as to the basic phase observable to be used in his GPS adjustment program. Furthermore, the use of a differenced phase observable permits the degree of "rigour" in the adjustment to be controlled. For example, for lower accuracy applications the triple-difference observable without the mathematical correlations taken into account (which would have made it equivalent to the "rigorous" double-difference) can be used. In fact it is ironic that the true utility of the triple-difference observable (its relative insensitivity to cycle slip errors) is revealed only when it is used in such a "non-rigorous" manner.

However, in any attempt to program the GPS observation equations, for either the differenced or undifferenced approach (with or without the appropriate correlations included) the software engineer is confronted with a major problem. The resulting normal equation matrix (NEM) is singular. A singular normal matrix indicates that the set of parameters to be estimated contains some information which has not been supplied, either through the observables or through the model. Such a matrix has a number of zero eigenvalues, and this number is equal to the **rank defect**.

If we partition the design matrix, obtained from any of the GPS phase observation models, into two matrices $\mathbf{A}_c$ and $\mathbf{A}_e$, the former representing the geodetic parameters of interest (non-bias) part and the latter the "nuisance" (clock bias) part of the model, the resulting partitioned NEM and reduced NEM are given in equations (2-22) and (2-23) respectively. Rank deficiencies exist in both the geodetic parameter and the clock bias submatrices, $\mathbf{N}^*_{cc}$ and $\mathbf{N}^*_{ee}$.

The rank defect in the $\mathbf{N}^*_{cc}$ submatrix is often found in geodetic adjustments, and is related to the **datum defect** of range-type observations. The rank defect in the submatrix associated with the clock biases has been noted in REMONDI (1985) and discussed in some detail in LINDLOHR & WELLS (1985). In GPS phase adjustment software both types of rank defects must be accounted for in an appropriate manner. In this paper, however, we will focus our attention mainly on the clock bias rank defects.

First we should emphasise a point made in section 2.4 about cycle ambiguity modelling in differenced observable modelling. We can either:

(a) model explicitly all the ambiguity parameters n_j^i and hence introduce 2 (in the single-differences) or 4 (in the double-differences) coefficient terms in each differenced observation equation, or

(b) model the ambiguity terms as a single "lumped" ambiguity parameter: k_j^{12} $(= n_j^1 - n_j^2)$ for between-satellite differences; k_{12}^i $(= n_1^i - n_2^i)$ for between-station differences; or k_{12}^{12} $(= n_1^1 - n_1^2 - n_2^1 + n_2^2)$ for double-differences.

and the results for the non-bias parameters are the same.

In order to compare and contrast the differenced and undifferenced observable modelling approaches, the following discussion assumes that individual ambiguity (n_j^i) and clock phase error (ε_j , ε^i) terms are explicitly modelled. It should nevertheless be borne in mind that for differenced observable modelling, option (b) above is often resorted to in practice, as for example in the TRIMVEC™ and PoPS™ programs (see section 4). Section 3.4.3 gives further details of this lumped ambiguity parameter approach.

The most straightforward approach to overcoming rank defect problems is to reduce the size of the NEM by fixing some of the parameters (and hence explicitly removing them from the solve-for parameter set) to *a priori* values in the case of the geodetic parameters of interest, or to arbitrary values such as zero for the nuisance or bias parameters. The number of parameters fixed should be equal to the number of rank defects that exist.

Before analysing the nature of clock bias rank deficiencies, and discussing possible remedies, we should attempt to resolve the following questions:

- How many rank deficiencies are there for the various differenced and undifferenced observables?
- What is the physical reason for these deficiencies?
- What is the effect of holding some biases fixed?
- How may we interpret the choice of fixed biases?

In order to influence the solution in a manner consistent with the above concerns, care must therefore be taken in the application of a remedy to the rank defect problem. In the parlance of conventional terrestrial geodetic network design and analysis, we are therefore concerned with:

(a) **Zero order design**, which deals with the estimability of the unknown parameters from the different possible observables. In this respect it is a problem of **datum definition**, and where the datum has not been completely defined there is a **datum defect.**

(b) **First order design**, which deals with the connectability of the parameters with one another through the medium of the observations. That is, how the datum definition information is transferred to all like parameters. It is essentially a problem of **configuration design**, and a defect in the design will lead to a **configuration defect.**

3.1 The Nature of the Datum Defects

A three-dimensional datum, or reference system, is defined geometrically by the origin (three components), orientation in space (three components) and scale (one component). As GPS measurements are (biased) ranges, or linear combinations of these, they can provide information on datum scale only. Hence the missing origin and orientation components are the datum defect. This is also the case with terrestrial geodetic networks determined from distance measurements alone. Therefore to overcome the six rank deficiencies in the normal equation matrix an equal number of constraints on the site coordinates must be introduced. These constraints may be in the form of fixed values of some site coordinates, or suitable weights on some or all coordinate component values.

However the situation with a GPS phase adjustment is more complex, and the datum defect problem is in many respects a four-dimensional one, rather than a conventional three-dimensional problem. In addition to requiring a definition of spatial (coordinate) origin, orientation and scale, we also need to define in some way the origin of the phase measurements and the implied time scale. The clock phase errors were introduced into the GPS observation equations by considering a perfect oscillator or clock with which the receiver and satellite oscillators could be compared. If we attempt to estimate the clock phase errors relative to this perfect oscillator we create two problems. Firstly, this perfect oscillator does not exist. Secondly, if it did exist we would have no means of comparing it directly with all our other oscillators. GPS observations, being differences between received phase signals and local receiver oscillator phase, are sensitive only to differences between satellite and receiver clock phase error (see equation (2-13)). Thus, in introducing the parameters of absolute clock phase error, we have created the equivalent of a datum defect. Therefore as in the case of the **coordinate datum defect**, some constraints on the satellite / receiver clock behaviour have to be introduced into a GPS phase adjustment to overcome this **phase datum defect**. One way to overcome this datum defect is to hold the clock phase errors of one **reference** oscillator fixed. The phase datum is then defined by holding the reference clock phase errors fixed (at an arbitrary value of zero) for each of the epochs in an observation session.

There is no phase datum defect in the case of double-differenced or triple-differenced observables because all clock errors were explicitly eliminated in the differencing process.

3.2 The Nature of the Configuration Defects

If the datum defects (coordinate and phase) are completely overcome, any remaining rank deficiencies in the normal equation matrix are due to a configuration defect. In fact an analysis of the normal equation matrix can give not only an indication of rank defects but, in addition, an insight into the configuration "strength". Network configuration design is based on the study of the solution covariance matrices implied from various simulated observation scenarios, in order to determine the "best" configuration for a network. However, a network with configuration defects (as opposed to merely a "weak" configuration) does require special attention as it implies that there is not enough information in the observations to connect all the parameters. Ideally, such a defect should not exist, and is in general avoidable. In the case of terrestrial geodetic networks, a configuration defect is due to poor planning or an oversight in the observation schedule that results in insufficient connecting observations for the determination of the coordinates of all the points in the network. Additional observations may then be made in order to both overcome the defect(s) and to give further "strength" to the solution. This can also occur in GPS phase adjustment. However, in the case of the GPS clock biases, a configuration defect will always occur in adjustments involving the one-way phase observable. They are absent from adjustments based on triple-differences (where there are no clock bias parameters present), or adjustments based on the double-differenced observable in which individual cycle ambiguity modelling is replaced by "lumped" ambiguity terms (see section 3.4.3).

The clock phase errors were introduced into the GPS observation equations by considering a perfect oscillator or clock with which the receiver and satellite oscillators could be compared. In attempting to estimate the clock phase errors relative to this perfect oscillator we create two problems. The first of these was the definition of the phase datum. The second problem is that although the phase datum defect is overcome by the definition of absolute phase in the reference oscillator, there are only ambiguous phase observations to determine the unambiguous clock phase errors (and cycle ambiguities) of all the other oscillators. There, therefore, remains a rank defect of 1 for each of the other R+S−1 oscillators. This is equivalent to a configuration defect where the **internal** phase "geometry" is not defined by the set of observations. We can now choose R+S−1 reference clock biases to fix in order to eliminate the remaining rank defects. These are chosen so that they provide a phase link between the reference oscillator and every other oscillator.

3.3 Analysing Rank Defects in the Clock Bias Parameters

With regards to the clock biases we can summarise the rank defect problem as follows: we have a **datum defect** caused by the attempt to estimate absolute clock phase error from observations of the difference between received satellite phase and receiver oscillator phase. We have a further **configuration defect** caused by the attempt to estimate unambiguous phase errors from ambiguous phase observations.

It was noted that in the undifferenced phase observable using the implicit double-difference method (where the clock phase errors were assumed independent at each epoch) there are T phase datum defects, corresponding to the definition of a reference clock at each of the T observation epochs in a session. There are no datum defects in the double-difference observable because all clocks were explicitly removed in the differencing process. It was then deduced that a further R+S−1 clock biases had to be constrained to overcome the phase configuration defects. Thus, the use of the implicit double-difference approach results in a total of T+R+S−1 rank deficiencies in the resulting normal equation matrix, while the explicit double-difference approach results in R+S−1 rank defects.

However, in order to develop the most appropriate strategies for overcoming rank defects in the clock bias parameters such an intuitive approach to rank defect analysis is of limited value. What is needed is a more mathematical technique for analysing rank defects in clock bias parameters. The Fundamental Differencing Theorem (section 2.3.1) provides us with just such a tool. Using the Fundamental Differencing Theorem we can equate the reduction in the number of linearly independent observations caused by differencing with the number of estimable parameters that have been eliminated. The method of using the Fundamental Differencing Theorem to determine the rank defects is from LINDLOHR & WELLS (1985). (It should be kept in mind that the Fundamental Differencing Theorem is only applicable for clock phase error models based on an independent estimate of the phase error at each epoch, and not for clock error polynomial models or filter solutions for the clock errors.)

A good starting point is the triple-difference observable, as the number of clock biases in this case is zero. Given R receivers, S satellites and T epochs as above, we have RST undifferenced phase observations. From these observations, (R–1)(S–1)(T–1) **linearly independent** triple-difference observables can be formed. (More than this number of different triple-differences can be formed but the extra ones are just linear combinations of this set.) The reduction in the number of observations from RST to (R–1)(S–1)(T–1) is RS+ST+RT–R–S–T+1. According to the Fundamental Differencing Theorem, this reduction in the number of observations has resulted in the same reduction in the number of estimable clock bias parameters. As there are zero clock biases in the triple-difference model, the number of clock bias parameters C_E that may be estimated in the undifferenced model in this example is therefore:

$$C_E = RST - (R-1)(S-1)(T-1)$$

$$= RS + ST + RT - R - S - T + 1 \qquad (3\text{-}1)$$

The clock biases are completely described by this number of parameters just as they are completely eliminated by reducing the number of observations by this amount. The total number of clock biases C_T in the undifferenced observation model is:

$$C_T = RS + ST + RT \qquad (3\text{-}2)$$

where RS is the number of cycle ambiguities n_j^i

ST is the number of satellite clock phase error parameters ε^i

RT is the number of receiver clock phase error parameters ε_j

The rank defect therefore is:

$$C_D = C_T - C_E \qquad (3\text{-}3)$$

$$= R + S + T - 1$$

This is equal to the rank defect that we intuitively determined earlier.

A similar process can be applied to the adjustment model based on the double-difference observable (section 2.4.3). The number of linearly independent double difference observations is (R–1)(S–1)T and the reduction in the number of observations when (R–1)(S–1)(T–1) triple-differences are formed is (R–1)(S-1). This is the number of estimable ambiguities:

$$C_E = (R-1)(S-1)T - (R-1)(S-1)(T-1)$$

$$= RS - R - S + 1 \qquad (3\text{-}4)$$

The number of cycle ambiguities in the double-difference observation model is:

$$C_T = RS \qquad (3\text{-}5)$$

The rank defect, from equation (3-3), therefore is:

$$C_D = R + S - 1 \qquad (3\text{-}6)$$

Let us consider the case of 2 receivers and 4 satellites, and assume a phase adjustment based on double-differenced observables. The number of estimable ambiguities is 3. The choice is, therefore, which ambiguities to estimate and which to hold fixed as **reference ambiguities**. For example, we could take:

Fixed: n_1^1 , n_1^2 , n_1^3 , n_1^4 , n_2^1

Estimated: n_2^2 , n_2^3 , n_2^4

In Table 1 the clock bias rank defect is given for all differenced observables for which a discrete epoch-by-epoch modelling of clock phase errors (if present) is used. Note that the clock bias rank defect can be computed easily from the relation:

$$C_D = C_T - (O_M - O_{TD}) \qquad (3\text{-}7)$$

where:

C_T is the total number of clock biases in the observation model.

O_M is the number of linearly independent observables that can be formed from the RST carrier beat phase observations.

O_{TD} is the number of linearly independent triple-difference observables. This is a constant = $(R-1)(S-1)(T-1)$

Table 1 CLOCK BIAS RANK DEFECTS

MODEL	Total No. of Clock Biases (1)	No. of Clock Biases Required (2)	Clock Bias Rank Defect (3)
Undifferenced			
(1)	RS+ST+RT	RST - (R-1)(S-1)(T-1) = RS+ST+RT-R-S-T+1	R+S+T-1
Single Difference			
(2) Between-stations Δ	RS+RT	(R-1)ST - (R-1)(S-1)(T-1) = RS+RT-R-S-T+1	R+S+T-1
(3) Between-satellites ∇	RS+ST	R(S-1)T - (R-1)(S-1)(T-1) = RS+ST-R-S-T+1	R+S+T-1
Double Difference			
(4) $\nabla\Delta$	RS	(R-1)(S-1)T - (R-1)(S-1)(T-1) = RS-R-S+1	R+S-1
Triple Difference			
(5) $\delta\Delta\nabla$	0	(R-1)(S-1)(T-1) - (R-1)(S-1)(T-1) = 0	0

[R receivers, S satellites, T epochs]

3.4 Strategies for Eliminating Rank Defects due to Clock Biases

There are a number of approaches to overcoming rank defects:

(1) The selection of some reference clock bias parameters not to be estimated.
(2) Estimating only the linear combinations of clock bias parameters resulting from the differencing process.
(3) The introduction of external constraints on the parameters via the VCV matrix, or the use of additional observations such as pseudo-ranges.
(4) The use of the pseudo-inverse.

The most straightforward approaches are option (1) and (2). In these closely related techniques the rank deficiencies are overcome by holding an equivalent number of reference clock biases fixed (that is, not estimating them) or forming differenced bias parameters in which an equivalent number of rank defects are eliminated, and estimating only the "lumped" differenced bias parameters. These are also the approaches adopted in most GPS phase adjustment software (section 4). Options (3) and (4) are not considered in this paper. The use of the pseudo-inverse approach is one of the ambiguity resolution options implemented in software by RIZOS et al (1990)

If we were interested only in the estimates of the non-bias parameters, such as the receiver coordinates and/or satellite orbit parameters, the choice of reference clock bias parameters is completely arbitrary. Thus, if the equivalence relations given in section 2.5.2 are satisfied, any phase adjustment based on the undifferenced or differenced approach, and for which the minimum clock bias parameters are held fixed as reference biases, will give the same results for the non-bias parameters. The choice of reference clock biases only affects the estimates of the other clock biases even though the reference clock bias parameters are fixed to values which are substantially different from the "true" values. This only becomes important if cycle ambiguity resolution is to be attempted (see section 3.4.5).

3.4.1 Fixing Reference Clock Bias Parameters

The following discussion is only relevant for GPS observable models in which the clock biases are explicitly modelled as individual clock phase error or cycle ambiguity parameters. In section 3.4.3 the "lumped" (or differenced) clock bias modelling approach is described.

As mentioned earlier, for adjustments using undifferenced and single-differenced observables, one option for overcoming the phase datum defect is to hold T clock phase error parameters fixed for the chosen reference clock. The remaining R+S–1 configuration defects in these observable models (and in the between satellite / station double-difference observable) can be overcome, for example, by holding R+S–1 clock bias parameters fixed for one epoch. There are other possibilities, though many of them cannot neatly separate the parameters fixed to define the phase datum from those fixed to overcome the phase configuration defect.

In each of the three categories of clock bias parameters (receiver clock, satellite clock, cycle ambiguity) we have identified below different classes of reference clock bias parameters to be held fixed, that can be associated with any of the four phase observable models given in Table 1 (a triple-difference observable model is not considered), and the sum of which must equal the rank defect for that model:

Receiver clocks

0	No parameters held fixed. All RT clock parameters to be estimated.
R	All receiver clocks fixed for one epoch. Not necessarily the same epoch for each clock. R(T–1) parameters to be estimated.
T	One clock held fixed in every epoch. Not necessarily the same clock for every epoch. (R–1)T parameters to be estimated.
R+T–1	Combination of the above two cases. One clock fixed for every epoch and every receiver clock fixed for one epoch. (R–1)(T–1) parameters to be estimated.
none	There are no receiver clock parameters. They have been removed by between-satellite differencing.

Satellite clocks

0	No parameters held fixed. All ST clock parameters to be estimated.
S	All satellite clocks fixed for one epoch. Not necessarily the same epoch for each clock. S(T–1) parameters to be estimated.
T	One clock held fixed in every epoch. Not necessarily the same clock for every epoch. (S–1)T parameters to be estimated.
S+T–1	Combination of the above two cases. One clock fixed for every epoch and every satellite clock fixed for one epoch. (S–1)(T–1) parameters to be estimated.
none	There are no satellite clock parameters. They have been removed by between-station differencing.

Integer ambiguities

0	No parameters held fixed. All RS ambiguity parameters to be estimated.
R	One ambiguity fixed for every receiver. Not necessarily all to same satellite. R(S–1) parameters to be estimated.
S	One ambiguity fixed for every satellite. Not necessarily all to the same receiver. (R–1)S parameters to be estimated.
R+S–1	Combination of the above two cases. One ambiguity fixed for every satellite and every receiver. (R–1)(S–1) parameters to be estimated.

There are therefore potentially a large number of different ways of selecting reference clock bias parameters to eliminate the clock bias rank defect. Table 2 lists various ways of selecting reference clock bias parameters, based on the options of fixing the various clock bias parameters outlined above.

Table 2 CHOICES OF REFERENCE CLOCK BIASES

MODEL	Rank Defect	Reference Clock Biases: Stn. Clock	Sat. Clock	Ambiguity
Undifferenced				
(1)	R+S+T−1	T	0	R+S−1
(2)	R+S+T−1	0	T	R+S−1
(3)	R+S+T−1	R+T−1	0	S
(4)	R+S+T−1	0	S+T−1	R
(5)	R+S+T−1	R+T−1	S	0
(6)	R+S+T−1	R	S+T−1	0
Between-Stations				
(7)	R+S+T−1	T	none	R+S−1
(8)	R+S+T−1	R+T−1	none	S
Between-Satellites				
(9)	R+S+T−1	none	T	R+S−1
(10)	R+S+T−1	none	S+T−1	R
Between Receivers/Satellite				
(11)	R+S−1	none	none	R+S−1

[R receivers, S satellites, T epochs]

Table 2 summarises 11 symmetric methods of selecting the reference clock biases. Some of these have been included for completeness and may not be used in practice. We have 6 options of selecting reference clock biases for the undifferenced approach, 4 options for single-differences, and one option for double-difference processing. Is there an equivalence of results for the clock biases in these 11 reference clock bias parameter options? This is discussed below.

3.4.2 The Effect on Estimates of Clock Bias Parameters of Holding Some Fixed as Reference Clock Biases

If we use undifferenced observations we have 6 choices of reference clock bias parameter model. The distinction between receiver and satellite clock is an arbitrary one in terms of the definition of phase datum. In each of these 6 models the phase of one reference clock is fixed in every epoch and the remainder of the rank defect is eliminated by fixed clocks (models (1) and (2)), fixed ambiguities (models (5) and (6)) or a combination of both (models (3) and (4)). To compare these reference clock bias parameter models, we need to consider the effect that fixed parameters have on the other estimated clock bias parameters.

Firstly we need to make the distinction between the unknown true value of a parameter and its least squares estimate. If we assume for the moment that the observations do not have any systematic errors and are normally distributed then we can write (WELSCH, 1980):

$$E(\hat{x}) = x \tag{3-8}$$

where x is the true value of a parameter, $\hat{x}$ is the least squares estimate and E( ) is the expectation operator. This equation is only true for estimable parameters. GRAFAREND & SCHAFFRIN (1976) have proven that parameters are estimable if and only if they are invariant under any transformation which leaves the observations invariant. It is easily demonstrated that the clock biases are not estimable parameters. For example, an equal change of all oscillators in any epoch will alter the clock phase errors significantly but not the observed phases. If a change of one cycle is introduced into any oscillator for all epochs then by the definition of the ambiguity, a compensating change of one cycle will ensure that the observation remains unaltered, hence equation (3-8) is not true for clock biases.

However, the transformations described above do not produce any net change in the <u>total</u> clock bias $B_j^i(t_k)$ defined as:

$$B_j^i(t_k) = f_0\varepsilon^i(t_k) - f_0\varepsilon_j(t_k) + n_j^i \tag{3-9}$$

where, for simplicity, we have given a common time-tag (t_k) to the time dependent terms. From equations (3-9) and (2-13) we can see that any transformation which leaves the observed phases invariant must leave the total clock bias invariant. Therefore we can write:

$$E(\hat{B}_j^i(t_k)) = B_j^i(t_k)$$

$$= f_0\varepsilon^i(t_k) - f_0\varepsilon_j(t_k) + n_j^i \tag{3-10}$$

and thus

$$f_0\,E(\hat{\varepsilon}^i(t_k)) - f_0\,E(\hat{\varepsilon}_j(t_k)) + E(\hat{n}_j^i) = f_0\varepsilon^i(t_k) - f_0\varepsilon_j(t_k) + n_j^i \tag{3-11}$$

since the expectation of a sum of terms is the sum of the expectations of the individual terms (LIEBELT, 1967). When we "fix" a parameter, by selecting it as a reference, we set its expected value to the *a priori* value which, for the clock biases, is usually zero and thus by using equation (3-11) we can determine the effects on the other estimable clock bias parameters.

As an example, let us consider model (1), based on undifferenced phase observations with one (reference) receiver clock phase error parameter fixed every epoch and an ambiguity parameter fixed for every receiver and every satellite. We will specify these reference clock biases as follows:

- Ambiguities for receiver r fixed for all satellites.
- Ambiguities involving satellite s fixed for all receivers.
- Receiver clock t fixed for all epochs.

This can be expressed in terms of expectation values as:

$E(\hat{n}_r^i) = 0$ for all satellites i

$E(\hat{n}_j^s) = 0$ for all receivers j

$E(\hat{\varepsilon}_t(t_k)) = 0$ for all epochs k

Clock error of satellite s
Using equation (3-11) we have for receiver t and satellite s:

$$E(\hat{\varepsilon}^s(t_k)) - E(\hat{\varepsilon}_t(t_k)) + E(\hat{n}_t^s) / f_0 = \varepsilon^s(t_k) - \varepsilon_t(t_k) + n_t^s / f_0 \qquad (3\text{-}12)$$

and eliminating the reference clock bias parameters:

$$E(\hat{\varepsilon}^s(t_k)) = \varepsilon^s(t_k) - \varepsilon_t(t_k) + n_t^s / f_0 \qquad (3\text{-}13)$$

Clock errors of all receivers j
For satellite s and any other receiver j:

$$E(\hat{\varepsilon}^s(t_k)) - E(\hat{\varepsilon}_j(t_k)) + E(\hat{n}_j^s) / f_0 = \varepsilon^s(t_k) - \varepsilon_j(t_k) + n_j^s / f_0 \qquad (3\text{-}14)$$

Eliminating the reference clock bias parameter and using equation (3-13):

$$E(\hat{\varepsilon}_j(t_k)) = \varepsilon_j(t_k) - \varepsilon_t(t_k) + n_t^s / f_0 + n_j^s / f_0 \qquad (3\text{-}15)$$

Clock errors of all satellites i
For receiver r and any other satellite i:

$$E(\hat{\varepsilon}^i(t_k)) - E(\hat{\varepsilon}_r(t_k)) + E(\hat{n}_r^i) / f_0 = \varepsilon^i(t_k) - \varepsilon_r(t_k) + n_r^i / f_0 \qquad (3\text{-}16)$$

Substituting equation (3-15) into (3-16) for the case where j = r:

$$E(\hat{\varepsilon}^i(t_k)) = \varepsilon^i(t_k) - \varepsilon_t(t_k) + n_r^i / f_0 + n_t^s / f_0 - n_r^s / f_0 \qquad (3\text{-}17)$$

Ambiguities of receivers j and satellites i
Finally for any other satellite i and any other receiver j:

$$E(\hat{\varepsilon}^i(t_k)) - E(\hat{\varepsilon}_j(t_k)) + E(\hat{n}_j^i) / f_0 = \varepsilon^i(t_k) - \varepsilon_j(t_k) + n_j^i / f_0 \qquad (3\text{-}18)$$

and using equations (3-15) and (3-17):

$$E(\hat{n}_j^i) = n_j^i - n_j^s - n_r^i + n_r^s \qquad (3\text{-}19)$$

This result is revealing in two respects. Firstly, equation (3-19) shows that, for this example, the estimated ambiguities are in reality doubly-differenced ambiguities even though undifferenced observations were used. This combination of integer ambiguities is still an integer, but the ambiguities n_j^s , n_r^i and n_r^s are reference ambiguities and are therefore not estimated (they are fixed to *a priori* values, usually zero). Therefore reference clock bias parameter model (1) (and (2), which is similar) gives estimated cycle ambiguities with integral expectation values.

The second point of interest is in the method used. The fixed reference ambiguity between satellite s and the reference clock t allow us to derive an expression for the clock error of satellite s (equation (3-13)). Using the reference ambiguities from this satellite we derived expressions for all other receiver clocks (equation (3-15)) and thence all other satellite clocks (equation (3-17)). Hence the phase datum can be transferred from the fixed reference clock to all other clocks through the fixed reference ambiguities, thus overcoming the configuration defect that existed.

This is the so-called **base station - base satellite** approach to selecting reference clock bias parameters. In a complex observing session with satellites rising and setting and the possibility of receivers observing at slightly different times, such a simple scheme as that given in the above example may not be possible. By analysing the method used in the above derivations we see that any group of reference ambiguities will satisfy the rank defect or phase configuration defect provided that they provide a path from the reference clock to every other clock.

It is also worth noting that the estimated ambiguities do not depend on the reference clock. Only the clock phase error terms depend on this, and these are estimated for each epoch. Therefore it is not necessary to have the same reference clock at every epoch. If receiver t in the above example stops observing, one of the other clocks (including one of the satellite clocks) may be chosen as the new reference without affecting the estimated ambiguities or other non-bias parameters.

Examples of the paths provided by reference ambiguities for the transfer of the phase datum are shown in Figure 1. In these figures there are 5 satellites and 6 receivers so the number of fixed reference ambiguities required is 10 (= R+S−1). Figure 1a shows the reference ambiguities for an example the same as that given above. The other figures illustrate more complicated schemes and in these cases the expected values of the estimated parameters will have more complex forms than those given above. For example, it can be shown that the estimated ambiguity between receiver 1 and satellite 5 in Figure 1b will be a linear combination of 10 integer ambiguities. The expected value of the ambiguity will still be an integer and will still be unaffected by a change of reference clock between epochs.

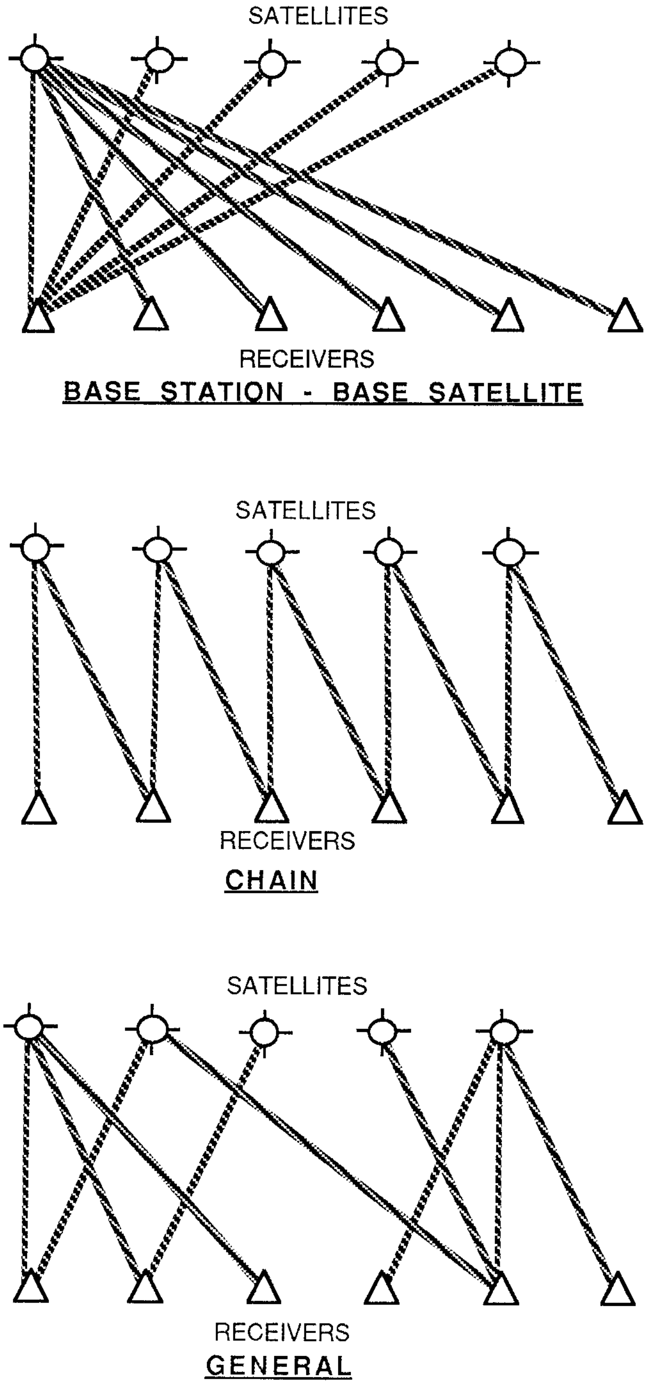


Figure 1 SOME CHOICES OF REFERENCE AMBIGUITIES

By using the same method as in the above example, we can also test the remaining 9 reference clock bias parameter models in Table 2. It can be demonstrated that for the 5 models (1), (2), (7), (9) and (11) that the estimated cycle ambiguities all have integral expectation values. All these models rely on identifying R+S−1 reference ambiguities which are held fixed in a phase adjustment. The other models, (3), (4), (5), (6), (8) and (10) do not give integral expectation values for the estimated ambiguities.

3.4.3 Alternative Ambiguity Modelling for Differenced Observations

An alternative and equivalent procedure for ambiguity modelling that can be used for single- or double-differenced observations is to define a new single ambiguity parameter that is a linear combination of the individual station-satellite ambiguity terms n_j^i. In the case of between-satellite differences the parameter k_j^{12} (= $n_j^1 - n_j^2$) could be used in the mathematical model (equation (2-15)); for between-station differences the parameter k_{12}^i (= $n_1^i - n_2^i$) could be used (equation (2-17)); and for double-differenced observations, either of the parameters k_j^{12} or k_{12}^i, or a double-differenced ambiguity k_{12}^{12} (= $n_1^1 - n_1^2 - n_2^1 + n_2^2$). The new ambiguity parameters k_j^{12}, k_{12}^i, k_{12}^{12} are referred to in this paper as "**lumped**" ambiguity parameters.

An important advantage in using such a differenced or lumped ambiguity modelling approach is that if only linearly independent differenced parameters are formed the rank defect in the clock bias part of the normal matrix $\mathbf{N_{ee}}$ is reduced. Following the procedure used to compute the rank deficiencies in Table 1, the number of clock biases required (column (2)) is unchanged, but the total number of clock biases present in the observations (column (1)) are:

- Between-station differences for k_{12}^i modelling:
  $(R-1)S + RT = RS + RT - S$
- Between-satellite differences for k_j^{12} modelling:
  $R(S-1) + ST = RS + ST - R$
- Double-differences for k_{12}^{12} modelling:
  $(R-1)(S-1) = RS - R - S + 1$

Now, taking column (2) in Table 1 minus the new column (1) (see equation (3-7), we get the reduced phase rank defect for differenced observations using lumped ambiguity parameters:

- Between-station differences: rank defect = $R+T-1$
- Between-satellite differences: rank defect = $S+T-1$
- Double differences: rank defect = 0

Note that in the case of linearly independent double-differences, there is <u>no rank defect in the clock bias parameters</u>.

Let us consider the example of 2 receivers and 4 satellites, and assume a phase adjustment based on double-differenced observables. In the case of the individual (undifferenced) ambiguity modelling approach, the number of estimable ambiguities is 3 (Table 1: column (1) minus column (3)), see discussion following equation (3-3). However, if double-differenced cycle ambiguities are modelled instead, all 3 lumped ambiguity parameters can be estimated:

$$k_{12}^{12} = n_1^1 - n_1^2 - n_2^1 + n_2^2$$
$$k_{12}^{13} = n_1^1 - n_1^3 - n_2^1 + n_2^3$$
$$k_{12}^{14} = n_1^1 - n_1^4 - n_2^1 + n_2^4$$

An alternative to using the double-differenced ambiguity parameter in an adjustment based on double-differenced observables is to use either the (partially) differenced k_{12}^i or k_j^{12} parameter. The rank defect is R–1 in the case of the k_{12}^i ambiguity model (between-station difference), and S–1 in the case of the k_j^{12} ambiguity model (between-satellite difference). The k_{12}^i ambiguity modelling option is the one used in the PoPS™ software (section 4.2).

3.4.4 Implications for GPS Software Design

A GPS phase adjustment in which all the clock bias parameters are estimated (except for the reference parameters that may have been held fixed at arbitrary values in order to overcome the rank defect) is often referred to as a "**bias-free**" or "**ambiguity-free**" solution. The cycle ambiguity parameters are supposed to be integers, but there is no known way to constrain an estimated parameter in a least squares adjustment to take on an integer value (KING et al, 1987). The combination of measurement noise and systematic errors will ensure that the estimates are not integers. The effect of measurement noise becomes less as the number of measurements increases but the effect of systematic errors is not so easily reduced.

However, under certain circumstances and with the appropriate reference clock bias parameter model, it is possible to obtain estimates of the cycle ambiguities (or linear combinations thereof) that are very near integer values. The circumstances under which this occurs are not well understood but it appears to be valid for lines under 30km (IBID et al, 1987). The estimates of the cycle ambiguities in models (1), (2), (7), (9) and (11) (Table 2) should be integers, unless systematic errors have affected them, and the ambiguities can be "**resolved**", the nearest integer value adopted as the most likely true value of the cycle ambiguity, and a new solution repeated for the other clock biases and geodetic parameters. This process is known as "**integer bias fixing**", but the term "**ambiguity resolution**" is preferred here. The second solution in which all or some of the ambiguities are held fixed to their integer values, is known as a "**bias-fixed**" or "**ambiguity-fixed**" solution, and should not be confused with the process described in section 3.4.1 of "fixing" the reference ambiguities in order to overcome clock bias rank deficiencies. Ambiguity resolution is discussed further in section 3.4.5.

The software engineer needs to therefore consider such issues as:

(a) **Should an ambiguity-fixed solution be attempted?** If the answer is "yes", either the undifferenced or differenced (though not the triple-difference) observation models can be used. Furthermore, certain reference clock bias parameter models have advantages. For example the estimates of the cycle ambiguities in models (1), (2), (7), (9) and (11) have integral expectation values, and if systematic errors are small it may be possible to resolve them to integers, and an ambiguity-fixed solution attempted.

(b) **What if satellites rise and set, and receivers start and stop observing at different times?** The examples used above were all rather simple ones where the same reference clock was used for all epochs, where the phases of clocks were all fixed in a reference epoch and where ambiguities were fixed to a single receiver and/or a single satellite, or some suitable combination. In GPS adjustment software, when the choice of parameters is made it may not be known which clock (if any) is present at all epochs or in which epoch (if any) all clocks are present. However any group of reference clock bias parameters in Table 2 will satisfy the rank defect, as long as they provide a path from the reference oscillator to every other oscillator. For example, for those reference clock bias models based on a reference clock, the selection of the reference clock is immaterial. It need not be the same oscillator for all epochs. Nor need it necessarily be the most stable clock.

(c) **What if a clock polynomial model or a filter solution is used for the clock phase errors?** For these options (not explicitly shown in Tables 1 & 2), relevant only for the undifferenced and single-differenced observation models, the reference clock should be the same for all epochs. Otherwise a change of reference clock will introduce an unmodelled step in the between-epochs change of clock phase errors of the other oscillators. All other considerations remain the same.

(d) **Will the choice of reference clock affect the estimated cycle ambiguity?** Such a query is relevant only for reference clock bias parameter models (1), (2), (3), (4), (7), (8), (9) and (10). The ambiguities do not depend on the selection of the reference clock. Therefore it is not necessary to have the same reference clock in every epoch. It should also be noted that the reference oscillator need not be that of the base station or base satellite used to define the reference cycle ambiguities.

(e) **Will the choice of reference clock biases affect the solution for non-bias parameters?** The method used to eliminate the total rank defect is arbitrary provided we are completely indifferent to the estimation of the remaining clock biases. If, however, we hope to find estimates for the ambiguities that are close to integers so we can then fix them to integers in a second ambiguity-fixed adjustment, our choice is no longer arbitrary. An ambiguity-fixed solution leads to a different non-bias parameter solution. There are two issues here (section 3.4.5). Firstly, certain reference clock bias parameter models make the task of ambiguity resolution easier (see (a) above). The second consideration is the choice of which clock biases (for a certain reference clock bias model) are to be held fixed. It may be that a certain set of reference clock biases (receiver/satellite clocks and ambiguities) are a better selection for ambiguity resolution than are others.

(f) **What about the processing of dual-frequency observations?** In the case of dual-frequency observations, the ionosphere-free combination of L1 and L2 observations is usually formed. The number of rank defects remains the same, and the same remedies can be applied. The problem of ambiguity resolution is, however, more complicated. None of the reference clock bias parameter models leads directly to integral expectation values for the estimated cycle ambiguities and special procedures have to be followed (see, for example, SCHAFFRIN & BOCK, 1988).

(g) **What if we use the double-differenced approach and model the cycle ambiguity as lumped terms?** What is the difference? The result for the geodetic parameters of interest is the same as if the individual undifferenced cycle ambiguities were estimated. One advantage is that there are no rank defects and hence there is no need to explicitly fix certain reference ambiguities. As in the case of the individual undifferenced ambiguity modelling approach, care needs to be taken with complex observing sessions. Ambiguity resolution can still be carried out.

(h) **What is the benefit of pseudo-range data?** Pseudo-range data, together with *a priori* receiver and satellite positions, contain information about the clock differences between receiver and satellite oscillators. It was the lack of such information that necessitated the introduction of the integer ambiguity parameters in the GPS phase observation models. Although the implications of this synergistic use of two types of observations have not been considered here, it may be that some choices of reference ambiguities are more appropriate than others in the presence of these additional observations.

In almost all respects, the 5 reference clock bias models (1), (2), (7), (9) and (11) of Table 2 are superior to any other alternative model.

Another advantage of these 5 models is that the same choice of reference clock bias parameters can be used for undifferenced, single-differenced and double-differenced observables (see Table 2). In particular, once an algorithm has been developed to select the reference ambiguities it may be used for all models. Furthermore, an algorithm for choosing reference ambiguity parameters is relatively simple even for complex observing sessions.

Considering the observations in the first epoch, the ambiguities from one station (the base station) to all satellites and from one satellite (the base satellite) to all receivers are chosen as the reference ambiguities. All other ambiguities are included amongst the parameters to be estimated. In subsequent epochs:

- if observations become available to a new satellite then new ambiguities are introduced except for one to an existing station (say the base station) which is not estimated,

- if a new receiver starts observing then new ambiguities are introduced except for one to an existing satellite (say the base satellite), which is a reference ambiguity and thus not estimated,

In this way a link is provided from the new clock to one of the existing clocks. This existing clock has already been linked to the reference clock and thus the phase datum is defined for all observables.

3.4.5 Cycle Ambiguity Resolution

According to the Fundamental Differencing Theorem the double-difference solution with free (unresolved) ambiguities is the same as a triple-difference solution (assuming the correlations are taken into account). If the cycle ambiguities in the double-difference solution are then fixed at integers and the adjustment repeated, this ambiguity-fixed solution will no longer be equivalent to the triple-difference solution. Such an ambiguity-fixed solution will be more precise (and more accurate if the correct integer ambiguity values are selected) than an ambiguity-free solution. There are of course considerably less free nuisance parameters to be estimated, but more importantly the solution is based on <u>unambiguous</u> range observations, rather than the (ambiguous) phase observations used in an ambiguity-free solution, and hence represents the most dramatic impact of external information in GPS phase adjustment.

As was noted in section 3.4.4, if we were only interested in eliminating the rank defect and were not interested in the estimates of the remaining clock phase errors and cycle ambiguities, then the selection of reference clock bias model (and the reference clock biases themselves), or whether to estimate differenced (lumped) ambiguities instead of individual ambiguities, is arbitrary. There are therefore three factors influencing the success, or otherwise, of resolving the cycle ambiguities to integer values:

(a) **The choice of reference clock bias parameter model.** It was shown in section 3.4.2 that for models (1), (2), (7), (9) and (11) of Table 2 the expectation values for the cycle ambiguities n_j^i were integers. <u>Therefore these models are preferred</u>. Ambiguity resolution is also possible with the other models but is more complex. A different set of parameters is required. For example, the ambiguities from a "base station" and a "base satellite" are fixed while the remaining ambiguity parameters are replaced by parameters which are double-differences of ambiguities in the case of (5) and (6) or single-differences of ambiguities in the other cases and thus have integral expectation values. These can then be held fixed to integer values and an ambiguity-fixed solution attempted.

(b) **The choice of reference ambiguities.** The selection of base station may be important. In particular, the ambiguities associated with stations close to the base receiver are more accurately determined than those associated with stations further away. <u>The base station should be chosen near the centre of the network to minimise the distances from it to other receivers</u>. In such a scheme, it may be possible to resolve all or most of the cycle ambiguities using a "boot-strapping" procedure. First the ambiguities from stations near the base receiver are resolved, and an ambiguity-fixed solution attempted in which the remaining (unresolved) cycle ambiguities are again estimated. A few more ambiguities from stations more distant from the original base receiver may then be resolved, and the solution again repeated. The procedure is then iterated until no further ambiguities can be resolved. This also means that a greater accuracy is obtained on those baselines that are near or include the base receiver.

(c) **The inclusion of pseudo-range data.** This data may accelerate the process of ambiguity resolution, for station separations up to 100km or more, and for observation sessions as short as 0.5 hours. The inclusion of pseudo-range data may also influence the choice of reference ambiguities ((b) above), as some choices of reference ambiguity may be more appropriate than others. More study needs to be done on this question.

4. ACCOUNTING FOR CLOCK BIASES IN GPS SOFTWARE

In this paper we have discussed two basic aspects of GPS phase data adjustment:

(1) the choice of GPS observation equation model.
(2) the remedy for rank defects in the clock bias parameters.

With respect to the former, the two choices are between the undifferenced observable modelling approach, based on the explicit modelling of all clock biases in the GPS one-way phase observation model, and the differenced observable modelling approach, in which some or all of the clock biases are eliminated through some form of observation differencing. A rigorous GPS solution will always be obtained using the one-way phase observation model. However the number of parameters to be estimated is potentially very large and special adjustment techniques need to be used to make the task manageable. Solution procedures based on differenced observable models, on the other hand, require a considerable amount of "bookkeeping" to control the differencing process, but do have the advantage of requiring less nuisance parameters to be estimated.

As to the issue of rank defects in the clock bias parameter models, the differenced observable modelling approach also has some advantages. Through the elimination of some of the clock bias parameters the size of the rank defect in the normal equation matrix is reduced. In fact, by considering the clock biases in the observable models as being lumped together into one parameter, the rank defect can also be reduced (for single-differences) or eliminated altogether (for double-differences). The process of double-differencing not only eliminates biases, it may also eliminate rank defects!

The choice between the undifferenced observable modelling approach and the differenced observable modelling approach is therefore not clearcut. If flexibility in observation modelling and exact mathematical rigour are not essential, the double-differenced observable model is a suitable candidate for phase adjustment, as both the satellite and receiver clock errors have been eliminated but not the cycle ambiguity. It is therefore the favoured approach for operational GPS software, particularly if it is designed to be instrument-specific. On the other hand, there is a trend for GPS software developed by government agencies or academic institutions for geodetic parameter estimation of the highest accuracy to adopt the undifferenced approach. We discuss below some GPS phase adjustment software; two commercial packages, one program developed by a government agency, and two programs developed at the School of Surveying, University of N.S.W.

4.1 TRIMVEC™

The TRIMVEC™ program was developed to process GPS observations made using the TRIMBLE 4000S series of receivers (TRIMBLE NAVIGATION, 1986). It is an example of a commercial GPS adjustment program written to conform to rather narrow specifications: it is instrument-specific, runs on a micro-computer and can process data from only two receivers (a baseline) at any one time. This is reflected in the phase adjustment model used. It makes use of the double- and triple-differenced observables formed between the 2 receivers and the up to 4 (for the 4000S) or 5 (for the 4000SX) satellites tracked by the receiver. (A multi-station version of the TRIMVEC™ program is now available to process data from more than 2 receivers simultaneously.)

The pseudo-range data is first used to synchronise the observation time-tags to the GPS Time scale (RIZOS & GRANT, 1990). A triple-difference solution (no correlations modelled) is then performed in order to derive accurate *a priori* receiver coordinates, and to correct the data for cycle slips. An adjustment of phase data using the double-differenced observable model is then carried out. This final adjustment is an ambiguity-free solution, and if the ambiguities can be resolved to integer values, the adjustment is repeated to obtain an ambiguity-fixed solution.

The means by which the phase configuration defect is handled by a single-baseline, double-differencing program such as TRIMVEC™ can be discussed at two levels: a conceptual level, and with respect to what is actually implemented within the software. Firstly we can tackle the problem of estimating the cycle ambiguities using the approach outlined in section 3.4.1 & 3.4.2. Let us consider the situation of data from 2 receivers to 4 satellites, there is a total of 8 ambiguities (RS) to be estimated. The number that must be held fixed as reference ambiguities is R+S−1 (Table 1), or 5, leaving (R−1)(S−1) , or 3, to be estimated. All the ambiguities from receiver 1 (the base station) are held fixed (that is, n_1^1, n_1^2, n_1^3 & n_1^4) and the ambiguity to the satellite in channel 1 at receiver 2 (the base satellite) is held fixed (that is, n_2^1). The task therefore is to estimate the 3 remaining ambiguities (n_2^2, n_2^3 & n_2^4).

In practice, however, because the basic data input into the solution is the double-differenced observable, the 3 double-differenced (or lumped) ambiguities are estimated instead (see section 3.4.3): k_{12}^{12} (= $n_1^1 - n_1^2 - n_2^1 + n_2^2$), k_{12}^{13} (= $n_1^1 - n_1^3 - n_2^1 + n_2^3$) & k_{12}^{14} (= $n_1^1 - n_1^4 - n_2^1 + n_2^4$).

Note that this is equivalent to the base station / base satellite approach referred to earlier. An inspection of the 3 double-differenced ambiguity terms reveals that 3 of the 4 individual ambiguity parameters n_j^i in each double-differenced ambiguity term are in fact the same parameters that would have been held fixed if the double-difference observable model was formulated with individual ambiguity parameters. Therefore estimating k_{12}^{12}, k_{12}^{13} & k_{12}^{14} is <u>equivalent</u> to estimating the double-differences of the 3 free ambiguity parameters n_2^2, n_2^3 & n_2^4 with respect to the base satellite and base station. If a new satellite rises, and observations commence to it, the new ambiguity parameter k_{12}^{15} (= $n_1^1 - n_1^5 - n_2^1 + n_2^5$) is introduced. Again, 3 of the 4 terms (n_1^1, n_1^5 & n_2^2) can be considered non-estimable if the individual (undifferenced) ambiguity modelling approach were used, while one of the ambiguity parameters (n_2^5) could be estimated.

Because the base satellite is the "pivot" in this double-differencing process, if this satellite sets or observations to this satellite are no longer available, then a new base satellite must be selected. If we assume that the new base satellite is satellite 2, the new double-differences are formed between satellites 2-3 & 2-4. The new ambiguity parameters associated with these observables are k_{12}^{23} (= $n_1^2 - n_1^3 - n_2^2 + n_2^3$) & k_{12}^{24} (= $n_1^2 - n_1^4 - n_2^2 + n_2^4$). None of these differenced ambiguity parameters belong to the set of 3 estimable ambiguity parameters originally defined in relation to satellite 1. However, instead of introducing new ambiguity parameters into the design matrix of the adjustment, each of these is modelled as a combination of 2 previously defined double-differenced ambiguity parameters. That is: $k_{12}^{23} = k_{12}^{13} - k_{12}^{12}$ & $k_{12}^{24} = k_{12}^{14} - k_{12}^{12}$. In this case a pair of double-differenced ambiguity parameters are explicitly included in the design matrix. Such a scheme is an elegant solution to the problem of rising and setting satellites during a session.

The TRIMBLE software and hardware package is designed to give survey accuracy results, and results at the few parts per million precision level are often obtained. While there is some atmospheric delay modelling, there is no option for estimating atmospheric bias parameters. The program possesses no satellite orbit adjustment capability. Furthermore, the mathematical correlations in the double- (and triple-) differences are not modelled, hence the results are not strictly equivalent to an adjustment based on one-way phase data.

4.2 PoPS™

The PoPS™ program was developed to process GPS observations made using the WILD-MAGNAVOX 101 GPS receiver (FREI et al, 1986). Although PoPS™ is a commercial GPS adjustment program like TRIMVEC™, it does have additional features and capabilities that reflect its heritage. The program is based on the algorithms and processing philosophy of the high precision GPS adjustment software of the Astronomical Institute, University of Berne (GURTNER et al, 1985). In particular, it is a multi-station program (that is, it can handle data collected simultaneously from more than two receivers) in which the mathematical correlations introduced through forming the double-differenced observables can be included if so required. In addition, there is considerable flexibility in terms of session and solution definition, as well as a choice of atmospheric delay models. However no atmospheric bias parameters can be estimated, and the program has no satellite orbit adjustment capability.

After synchronising the observation time-tags using the pseudo-range data, the single-differences (between-station differences) are formed. Cycle slip detection and repair is carried out on the single-differenced data directly, as well as with the aid of the double-differences. The actual phase adjustment is carried out using double-differenced observables.

As in the case of the TRIMVEC™ software, we can discuss the phase configuration defect problem in PoPS™ at two levels. If we assume that all the cycle ambiguities are modelled as individual (undifferenced) parameters then there is a rank defect in the double-difference adjustment model of R+S–1. This is the number of ambiguities to be held fixed as reference ambiguities and, as discussed in section 3.4.1, a common technique is to hold the ambiguities from a base station and a base satellite fixed, and estimate the remaining RS–R–S+1 ambiguities (Table 1). If we consider the example of 2 receivers and 4 satellites, the task is one of estimating the 3 ambiguity parameters n_2^2, n_2^3 & n_2^4.

However, where the PoPS™ software differs from TRIMVEC™ is that the basic observable is a single-difference, not a double-difference. The ambiguity parameters modelled are therefore the single-differenced ambiguities. In this example there are 4 independent single-difference parameters: k_{12}^1 (= n_1^1 - n_2^1), k_{12}^2 (= n_1^2 - n_2^2), k_{12}^3 (= n_1^3 - n_2^3) & k_{12}^4 (= n_1^4 - n_2^4). According to section 3.4.3 there is a rank defect of R–1 (= 1), and hence one reference single-differenced ambiguity must be held fixed and the others are to be estimated. This is the between-station ambiguity involving the "reference" satellite, say k_{12}^1. The reference satellite is selected by the program to be the satellite to which the greatest number of observations are made during a session. (In the case of 3 receivers in a session, the reference ambiguities are, for example, k_{12}^1 and k_{13}^1.)

When the double-differences are formed with respect to the reference satellite, each double-difference observable contains the difference between 2 single-differenced ambiguity parameters. That is, from the 4 single-difference observables, 3 double-differences are constructed: $k_{12}^{12} = k_{12}^1 - k_{12}^2$, $k_{12}^{13} = k_{12}^1 - k_{12}^3$ & $k_{12}^{14} = k_{12}^1 - k_{12}^4$. That is, each double-difference observation equation in the adjustment contains two ambiguity terms: one for the reference single-differenced ambiguity which is held fixed (and may not even appear explicitly in the design matrix) and one free single-differenced ambiguity parameter. The 3 estimated (free) and one reference single-difference ambiguity parameters are themselves the between-stations differences involving the 3 free and 5 reference undifferenced ambiguities.

In the event of the reference satellite setting or observations to this satellite no longer being available, a new reference satellite is chosen to control the data differencing and in order to maintain the same reference single-differenced ambiguity (say k_{12}^1) the new double-differenced ambiguity parameter is modelled as a combination of 2 previously defined double-differenced ambiguity parameters. Referring to the example discussed earlier, the new double-differenced ambiguity parameters are defined as: $k_{12}^{23} = k_{12}^{13} - k_{12}^{12}$ ($= k_{12}^2 - k_{12}^3$) & $k_{12}^{24} = k_{12}^{14} - k_{12}^{12}$ ($= k_{12}^2 - k_{12}^4$). In each case, a pair of previously defined single-differenced ambiguities are explicitly included in the design matrix.

4.3 U.S. National Geodetic Survey Program PHASER

The PHASER program package was developed by a government agency (the U.S. National Geodetic Survey - KASS & DULANEY, 1986) to satisfy its requirements for GPS phase reduction software. The operational environment in which the program was intended to be used has strongly influenced its design. The main characteristics of PHASER are:

- It is GPS instrument-independent (that is, able to process input from any type of GPS receiver).
- It has multi-station capability.
- It was a designed to run on a powerful mini-computer, as is usually found in mapping and survey organisations.
- The program system is capable of delivering relative positioning results at the several parts per million accuracy level, a limit defined by the program's inability to estimate atmospheric bias parameters or to adjust GPS satellite orbit(s).

Although the cycle slip repair and pseudo-range solution for observation time-tag / GPST offset procedures are essentially similar to TRIMVEC™, PHASER was the first program to make direct use of the one-way phase data in the phase adjustment step (GOAD, 1985). The basic observable model is a modification of equation (2-13) using a novel technique for handling the clock biases. PHASER uses the base station / base satellite concept to define two types of lumped clock bias parameters:

(1) For any one-way phase observation at epoch e involving the base station r or base satellite s the total clock biases are modelled by a single parameter (see equation (3-9)):

$$B_j^i(t_e) = f_0\varepsilon^i(t_e) - f_0\varepsilon_j(t_e) + n_j^i \tag{3-20}$$

where $i=s$ or $j=r$. There are $(R+S-1)T$ of these B parameters.

(2) For any remaining phase observation involving neither the base station r nor the the base satellite s , the total clock biases are modelled by a linear combination of B_j^i bias parameters and the double-differenced ambiguity parameter (see section 3.4.3):

$$k_{rj}^{si} + B_j^s(t_e) + B_r^i(t_e) - B_r^s(t_e) \tag{3-21}$$

where:

$$\begin{aligned} k_{rj}^{si} &= B_r^s(t_e) - B_r^i(t_e) - B_j^s(t_e) + B_j^i(t_e) \\ &= n_r^s - n_r^i - n_j^s + n_j^i \end{aligned} \tag{3-22}$$

There are $(R-1)(S-1)$ of these k parameters.

Note that k_{rj}^{si} is an integer, and is a constant for a whole observation session (assuming there are no cycle slips), while $B_j^i(t_e)$ has a non-integer value and must be estimated on an epoch-by-epoch basis. The similarity with the procedure used by TRIMVEC^{TM} is obvious, except that the clock bias parameters $B_j^i(t_e)$ have also to be estimated as the clock phase errors (ε^i & ε_j) have not been eliminated through double-differencing. There are no rank defects in these clock bias parameters as the total number of modelled clock biases (= (R–1)(S–1) + (R+S–1)T) is equal to the number of clock biases required (= RS+ST+RT–R–S–T+1), as defined in Table 1 for the undifferenced model. However, no attempt is made to explicitly estimate the clock phase errors.

If we consider the example of 2 receivers and 4 satellites, the clock bias parameters in the 5 one-way phase observations involving either the base station (say receiver 1) or base satellite (say satellite 1) are: $B_1^1(t_e)$, $B_1^2(t_e)$, $B_1^3(t_e)$, $B_1^4(t_e)$ & $B_2^1(t_e)$. The 3 remaining phase observations contain the following clock bias parameters:

1: $k_{12}^{12} + B_2^1(t_e) + B_1^2(t_e) - B_1^1(t_e)$

2: $k_{12}^{13} + B_2^1(t_e) + B_1^3(t_e) - B_1^1(t_e)$

3: $k_{12}^{14} + B_2^1(t_e) + B_1^4(t_e) - B_1^1(t_e)$

The contribution of the 8 observations to the normal equation matrix (NEM) is computed and the 5 B_j^i clock bias parameters are estimated for epoch e and their effect on the common part of the NEM (containing the geodetic parameters of interest and the constant bias parameters k_{12}^{12} , k_{12}^{13} , k_{12}^{14}) is removed using standard partitioning techniques (see equation (2-23)). Once all the observations have been processed, the reduced NEM is inverted and the common parameters solved for. If the estimated values of the k parameters are close to integers (as they should be for short interstation distances), the double-differenced ambiguities can be resolved to their integer values and a subsequent ambiguity-fixed solution attempted.

If observations to a new satellite become available, or from a new station, new k parameters are incorporated into the NEM, but the number of B parameters remains unchanged. In the event of observations involving either the base station or base satellite no longer being available, there will not be sufficient data to perform an epoch solution for the B parameters. This will manifest itself as a singularity when that part of the NEM containing the B parameters is inverted. The remedy in PHASER is to zero out the corresponding nondiagonal row and column elements of the NEM. Such a simple algorithm is equivalent to forming another combination of observations in the double-differences (that is, selecting a new base satellite or station). Alternatively, a new base satellite (or base station) is selected from the remaining satellites (or stations), and no further contribution to the k parameter associated with this formerly non-base satellite (or station) is made. In this, and a number of other respects, an undifferenced approach to carrier phase adjustment involves less "bookkeeping", particularly an advantage when multi-station adjustments are performed. Such an approach also allows the data to remain uncorrelated, hence simplifying the adjustment process. In addition, no data is ever excluded from an adjustment, as can happen in the case when double-differenced observables have to be generated.

4.4 UNSW Program NIBBLE/CRUNCH

The NIBBLE/CRUNCH phase adjustment program was designed for high precision GPS relative positioning applications in which, in addition to being able to process multi-station, multi-instrument (instrument-independent) and multi-session campaigns, a satellite orbit adjustment capability was required (RIZOS & STOLZ, 1988). As in the case of the PHASER program the basic observable in NIBBLE/CRUNCH is one-way carrier beat phase (equation (2-13)). The manner in which it handles the clock biases, and overcomes the rank defect problems associated with them, is however different to that of PHASER. Because PHASER uses a combination of B and k parameters (both of which are of the lumped variety) there are no rank defects. In the case of NIBBLE/CRUNCH all clock bias parameters (ε^i , ε_j & n_j^i) are explicitly modelled. In fact there are four types of clock bias parameters (MASTERS, 1988; private communication):

(1) RT receiver clock phase errors (ε_j)

(2) ST satellite clock phase errors (ε^i)

(3) base station / base satellite ambiguity parameters (n_j^i)

(3) non-base station / non-base satellite lumped ambiguity parameters (k_{rj}^{si})

There are a total of RS ambiguity parameters: R+S−1 of the n_j^i variety for observations involving either a base station r or base satellite s , (R−1)(S−1) of the k_{rj}^{si} variety (section 3.4.3) for those observations involving neither the base station nor the base satellite. This formulation is similar to that of PHASER except that the clock phase errors have not been lumped with the base station / base satellite ambiguities. This has the advantage that the clock errors of receivers and satellites can be separately weighted in the solution.

In the example of 2 receivers and 4 satellites, there are 5 undifferenced ambiguity parameters: n_1^1 , n_1^2 , n_1^3 , n_1^4 & n_2^1 ; each associated with a one-way phase observation from receiver j to satellite i (with i=s=1, or j=r=1). Each of these 5 phase observations also contain 2 epoch parameters: one for the receiver clock error ($\varepsilon_j(t)$) and one for the satellite clock error ($\varepsilon^i(t)$). The remaining 3 phase observations involving neither the base station nor the base satellite contain the following clock bias parameters:

1: $k_{12}^{12} + n_2^1 + n_1^2 - n_1^1 + \varepsilon^2(t) - \varepsilon_2(t)$
2: $k_{12}^{13} + n_2^1 + n_1^3 - n_1^1 + \varepsilon^3(t) - \varepsilon_2(t)$
3: $k_{12}^{14} + n_2^1 + n_1^4 - n_1^1 + \varepsilon^4(t) - \varepsilon_2(t)$

There are a total of RS+ST+RT clock bias parameters, resulting in a rank defect of R+S+T−1 (Table 1, model 1). The method used in NIBBLE/CRUNCH to eliminate the rank defects in the clock biases is model 5, Table 2. The reference clock biases are specified as follows:

- All receiver clock errors fixed for epoch t_1.
- All satellite clock errors fixed for t_1.
- Receiver clock t fixed for all epochs.

Note that all the reference clock biases are clock error parameters. There are no reference ambiguities. The selection of which receiver clock is to be the epoch reference clock is random.

Using the same method as in section 3.4.2, it can be demonstrated that the R+S−1 estimable undifferenced ambiguities are:

$$E(\hat{n}_j^i) = f_0\,[\varepsilon^i(t_1) - \varepsilon_j(t_1)] + n_j^i \tag{3-23}$$

The estimates of the cycle ambiguities will not be integers as they include the clock errors at the reference epoch (t_1). The estimates of ε^i , ε_j & n_j^i are of no interest, and NIBBLE/CRUNCH does not output their values. The parameters of interest are the (R−1)(S−1) k_{rj}^{si} parameters which have integral expectation values, and the expected values are equal to the double-differenced ambiguities involving base station / base satellite ambiguities (equation (3-19)). While the clock phase error parameters are estimated on an epoch-by-epoch basis, and their contribution is removed from the NEM, the k_{rj}^{si} and n_j^i parameters are included in the common parameter set. If the estimated values of the k_{rj}^{si} parameters are close to integers, cycle ambiguity resolution may be attempted in a subsequent step.

As with PHASER, if observations to a new satellite are available, or from a new station, new k_{rj}^{si} parameters are introduced into the NEM, although the number of n_j^i parameters associated with the base station / base satellite remains the same. If there is insufficient data at an epoch to solve for the ε^i & ε_j parameters, the same remedy is used as in the case of PHASER. The appropriate nondiagonal rows and columns of the NEM are zeroed out until the rank defects are overcome. However, in the event of observations involving either the base station or base satellite no longer being available, no special action is taken. Although no further contribution to the NEM is made for the associated n_j^i ambiguity parameters, there is no need to select another base satellite (or base station). Again, the advantage of the undifferenced approach to carrier phase adjustment is that no complex algorithm is needed to account for satellites rising and setting during an observation session.

The NIBBLE/CRUNCH package has since been superseded by the USMASS program.

4.5 UNSW Program USMASS

The UNSW Satellite Measurement Analysis Software System (RIZOS & STOLZ, 1988) is a general-purpose station coordinate and orbit determination program recently developed at the School of Surveying, and is capable of, amongst other things, of processing GPS phase data. As with the NIBBLE/CRUNCH and PHASER programs the basic observable is one-way carrier beat phase data (equation (2-13)), although the manner in which the clock biases are modelled owes more to NIBBLE/CRUNCH than to PHASER. All clock biases are explicitly modelled but the reference clock biases are selected according to model 1 and 2, Table 2. The reference clock biases are specified as in section 3.4.2, that is:

- Ambiguities for receiver r fixed for all satellites.
- Ambiguities involving satellite s fixed for all receivers.
- Receiver or satellite clock t fixed for all epochs.

and the estimable (non-reference) ambiguities are (equation (3-19)):

$$E(\hat{n}_j^i) = n_j^i - n_j^s - n_r^i + n_r^s \qquad (3\text{-}24)$$

The estimable ambiguities are in reality double-differenced ambiguities (with respect to the reference ambiguities n_j^s , n_r^i & n_r^s), and hence have integral expectation values. It is not necessary to have the same reference clock at each epoch, and the selection of which receiver or satellite clock is to be held fixed is purely random. In the example of 2 receivers and 4 satellites, with base station being receiver 1 and base satellite being satellite1, the 5 reference ambiguities are: n_1^1 , n_1^2 , n_1^3 , n_1^4 & n_2^1 , while the 3 estimable ambiguities are: n_2^2 , n_2^3 , n_2^4. As in the case of the PHASER and NIBBLE/CRUNCH programs, situations in which satellites rise and set, and stations start and end observing, at different times are easily handled. No observation differencing takes place. Instead, at worse the set of reference clock biases changes, although often no further action is necessary as the undifferenced approach to phase adjustment simply lets the normal matrix decide which parameters can be estimated and which cannot. Furthermore, the data remains uncorrelated and no special treatment of the correlation problem (as it exists for differenced observables) is required.

5. ACKNOWLEDGEMENTS

The first author (D.B.G.) was supported in his studies at the School of Surveying, University of New South Wales, by a study award from the Dept. of Survey and Land Information, New Zealand.

6. REFERENCES

BEUTLER, G., BAUERSIMA, I., GURTNER, W. & ROTHACHER, M., 1987. Correlations between simultaneous GPS double difference carrier phase observations in the multistation mode: Implementation considerations and first experiences. **Manuscripta Geodaetica**, 12, 40-44.

CROSS, P.A., 1983. Advanced least squares applied to position fixing. Working Paper No.6, Dept Land Surveying, North East London Polytechnic, 205pp.
FREI, E., GOUGH, R.J., & BRUNNER, F.K., 1986. PoPS™: a new generation of GPS post-processing software. Proc. 4th Int. Symp. on Satellite Positioning, Austin, Texas, 28 Apr - 2 May, 1986, 455-473.
GOAD, C.C., 1985. Precise relative position determination using Global Positioning System carrier phase measurements in a non-difference mode. Proc. 1st Int. Symp. Prec. Pos. GPS, U.S. Dept. of Commerce, NOAA, Rockville, Md., 15-19 April, 1985, 347-356.
GRAFAREND, E., & SCHAFFRIN, B., 1976. Equivalence of estimable quantities and invariants in geodetic networks. **Zeitschrift für Vermessungswesen**, 101(11), 485-491.
GURTNER, W., BEUTLER, G., BAUERSIMA, I., & SCHILDKNECHT, T., 1985. Evaluation of GPS carrier difference observations: the bernese second generation software package. Proc. 1st Int. Symp. Prec. Pos. GPS, U.S. Dept. of Commerce, NOAA, Rockville, Md., 15-19 April, 1985, 3463-372.
KASS, W.G., & DULANEY, R.L., 1986. Procedures for processing GPS phase observations at the National Geodetic Survey. Proc. 4th Int. Symp. on Satellite Positioning, Austin, Texas, 28 April - 2 May, 1986, 753-765.
KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A., & COLLINS, J., 1987. **Surveying with GPS**. Ferd. Dümmlers Verlag, Bonn, 128pp.
LIEBELT, P.B., 1967. **An introduction to optimal estimation**. Addison-Wesley, Reading, Mass.
LINDLOHR, W., & WELLS, D.E., 1985. GPS design using undifferenced carrier beat phase observations. **Manuscripta Geodaetica**, 10, 255-295.
REMONDI, B.W., 1985. Global Positioning System carrier phase: description and use. **Bull Geod**, 59, 361-377.
RIZOS, C., & GRANT, D.B., 1990. Time and the Global Positioning System. In Unisurv S-38, "Contributions to GPS Studies", C. Rizos (ed.), report of the School of Surveying, University of N.S.W.
RIZOS, C. & STOLZ, A., 1988. The UNSW satellite measurement analysis software system (USMASS). **Aust.J.Geod.Photogram.Surv.**, 48, 1-28.
RIZOS, C., SUBSUANTAENG, S., & SUBARYA, C., 1990. Developing and testing GPS software: considerations, results and conclusions. Proc. 2nd Int. Symp. Prec. Pos. GPS, Ottawa, Canada, 3-7 September, 1990.
SCHAFFRIN, B., & BOCK, Y., 1988. A unified scheme for processing GPS dual-band phase observations. **Bull.Geod.**, 142-160.
TRIMBLE NAVIGATION LTD, 1986. TRIMVEC™ GPS survey software preliminary user's manual. Trimble Navigation Ltd, Sunnyvale California.
WELLS, D.E., LINDLOHR, W., SCHAFFRIN, B., & GRAFAREND, E., 1987. GPS design: undifferenced carrier beat phase observations and the fundamental differencing theorem. Tech. rept. 116, Dept. of Surveying Engin., University of New Brunswick, 141pp.
WELSCH, W., 1980. Some techniques for monitoring and analysing deformations and control nets. Rept. Universität der Bundeswehr, München, F.R.G., 193pp.

TIME AND THE GLOBAL POSITIONING SYSTEM

C. Rizos
D.B. Grant[1]

School of Surveying
University of New South Wales
P.O.Box 1
Kensington, N.S.W., 2203, Australia

ABSTRACT

Time is a fundamental concept inherent in all satellite measurement and analysis techniques. Although positioning results are invariably expressed in three-dimensions, time plays a number of crucial roles in any geodetic solution scheme involving satellite data. In the first instance time enters through the fact that observing stations attached to the earth's surface and the orbiting satellites to which measurements are made are in relative motion. Furthermore, the satellites are moving with respect to an inertial reference frame, and the station coordinates are known, or required, in an earth-fixed reference system which, because of earth rotation and other perturbations, is moving relative to the inertial frame. In addition, all range, or range-like, measurements in space geodesy rely on clocks, as they are essentially timing differences (between time of transmission and time of arrival) of a particular satellite signal. In this paper we discuss the role of time in the context of the Global Positioning System (GPS) technology.

Time enters GPS data analysis in basically three ways:

- Through the time-tags of the satellite orbits.
- Through the time-tags of the measurements.
- Through the measurements themselves.

Because of the nature of the measurements and the fact that GPS is a multi-satellite, multi-user (and hence multi-clock) system, time plays an even more important role in GPS positioning than in any other space positioning technique.

[1]Now at the Dept. of Survey and Land Information, Wellington, New Zealand.

1. TIME SYSTEMS - SOME DEFINITIONS

In order to fully understand the role of time in GPS data analysis it is necessary to briefly review the various time systems involved, and their associated time scales. Some of these definitions are standard and inherent in all space positioning technologies, while others are particular to the GPS system. All are essential to the subsequent development of the GPS observation equations, to understanding the elimination (or estimation) of GPS clock errors, and to determining the accuracy with which time must be defined at various stages of GPS data processing.

In general there are three different time systems that are used in space geodesy (KING et al, 1987):

- Dynamical time
- Atomic time
- Sidereal time

Dynamical Time is the uniform time scale which governs the motions of bodies in a gravitational field: that is, the independent argument in the equations of motion for a body according to some particular gravitational theory, such as Newtonian mechanics or General Relativity. **Atomic Time** is time defined by atomic clocks, and is the basis of a uniform time scale on the earth. **Sidereal Time** is measured by the earth's rotation about its axis, and although sidereal time was once used as a measure of time it is much too irregular by today's standards. Rather, it is a measure of the angular position of a site on the earth with respect to a 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, or time scales, that are used in space geodesy.

Some time scales have a special importance because they provide the "benchmark" or reference scale within a particular time system. This often occurs by international convention. Often, however, the time scales to which we have access are merely realisations of the "true" or definitive reference time scale (or scales) associated with each time system. In the remainder of this section each of these time systems are briefly described, and the various measures of time are discussed, together with how they are obtained in practice, and their relevance to GPS.

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 (the barycentre). Dynamical time measured in this system is called **Barycentric Dynamical Time** (TDB -- the abbreviation for this and most other time scales reflects the French order of the words). A clock fixed on the earth will exhibit periodic variations as large as 1.6 milliseconds with respect to TDB due to the motion of the earth in the sun's gravitational field. However, in describing the orbital motion of near-earth satellites we need not use TDB, nor account for these relativistic variations, since both the satellite and the earth itself are subject to essentially the same perturbations.

For satellite orbit computations, 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). In the terminology of General Relativity, TDB corresponds to **Coordinate Time**, and TDT to **Proper Time**.

The predecessor of TDB was known as **Ephemeris Time** (ET).

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 des Poids et Mesures (BIPM) in Sévres, France, of data from atomic frequency standards (atomic "clocks") in many countries. Prior to 1 January, 1988, this function was carried out by the Bureau International de l'Heure (BIH). TAI is a continuous time scale and serves as the practical definition of TDT, being related to it by:

$$\text{TDT} = \text{TAI} + 32.184 \text{ seconds} \qquad (1\text{-}1)$$

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

Because TAI is a continuous time scale, it has one fundamental problem in practical use: the earth's rotation with respect to the sun is slowing down by a variable amount which averages, at present, about 1 second per year. Thus TAI would eventually become inconveniently out of synchronisation with the solar day. This problem has been overcome by introducing **Coordinated Universal Time** (UTC), which runs at the same rate as TAI but is incremented by 1 second jumps (so-called "leap seconds") when necessary, normally at the end of June or December of each year. At present (June 1990) one needs to add 25 seconds to UTC clock readings to obtain time expressed in the TAI scale.

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

$$\text{GPST} + 19 \text{ seconds} = \text{TAI} \qquad (1\text{-}2)$$

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 on the Celestial Sphere -- 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 celestial equator and the earth's orbital plane with respect to inertial space, and because of the irregularity of the earth's rotational motion itself. UT1 is derived from the analysis of observations carried out by the International Earth Rotation Service based in Paris, and is reconstructed from published corrections to UTC, that is:

$$UT1 = UTC + DUT1 \qquad (1\text{-}3)$$

2. MEASURES OF TIME IN GPS

Each time scale is defined by the period of the basic oscillation of the frequency-determining element (be it the earth's rotation in the case of Sidereal Time, or the oscillation of atoms in the case of TAI) which is measured, and the origin of the time scale, which may be either arbitrarily defined or agreed upon by international convention. The inverse of the period is the frequency, which is directly related to the rate of the time scale. The rate is given by:

$$R = \frac{f}{f_0} \qquad (1\text{-}4)$$

where f is the "true" frequency of the oscillator (determined by comparison with a perfect oscillator) and f_0 is the nominal frequency used to convert the measured oscillator cycles into units of time (seconds). A perfect oscillator has a rate of unity, which implies that its nominal frequency is "true". We can use the two notions of "origin" (or datum) and "rate" (or scale) to differentiate between various time systems and time scales.

A related concept is the drift of an oscillator relative to a perfect oscillator. This is the difference between its rate and the rate of a perfect oscillator (which is unity as noted above):

$$D = R - 1$$

$$= \frac{(f - f_0)}{f_0} \qquad (1\text{-}5)$$

Figure 2.1 illustrates the relationship between the various time scales in common use (KING et al, 1987). The vertical axis indicates the relative offsets of the origins of the time scales, and the slope of the lines indicate their drift. Note that with the exception of UT1 all time scales (nominally) have zero drift (constant and true frequency) as defined by TAI.

In practice, atomic time is the most accessible of all the time systems, and the most accurate. That is, we can build a clock to maintain an atomic time scale. In addition, with the exception of the Sidereal Time scales, it can be used to define or relate all other time scales. If we define "true" or "perfect" time to be, for example, the fundamental time scale in satellite motion, that is TDB, the time scale provided by such an atomic clock is a realisation of this "true" time. It is not itself a true or perfect time scale, as "true" time could only be maintained by a perfect atomic clock (if such a clock could be built) located at the solar system barycentre (not very accessible!). In the parlance of General Relativity this is referred to as "Coordinate Time". Nevertheless this (unattainable) "true" time is an ideal choice for the basic time system to develop the GPS carrier phase observation equations. The most accurate measure of this "true" time on earth (that is, "Proper Time") is maintained by the Bureau International des Poids et Mesures (BIPM). TAI is related to other useful time scales by offsets defined by international convention (Figure 2.1). Hence the "true" time can be related to these other time scales.

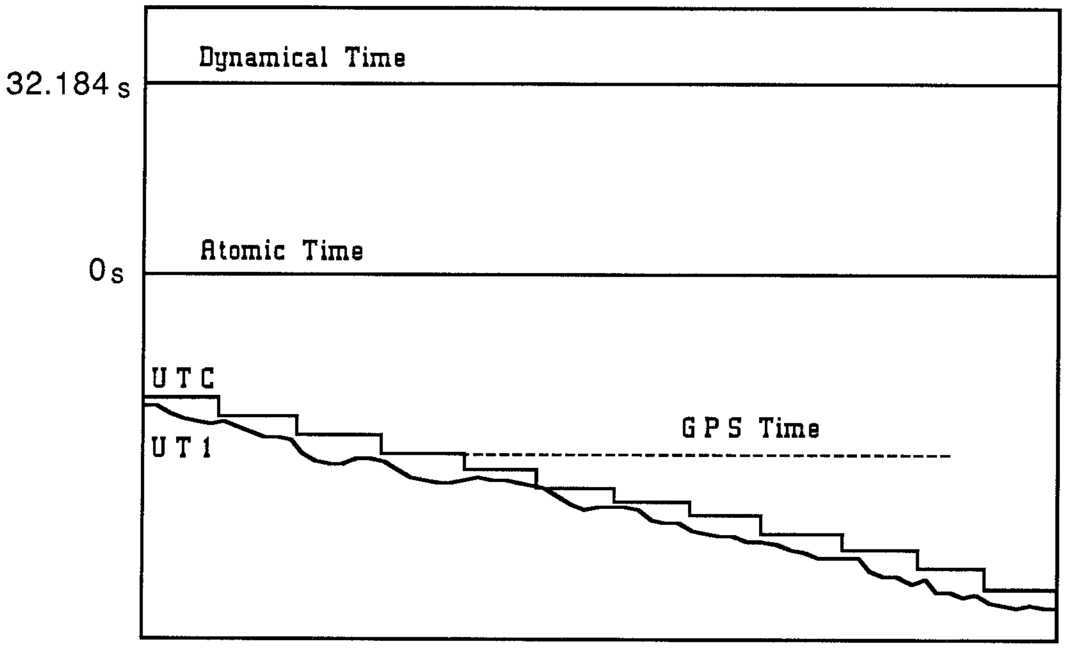


Figure 2.1 Time Standard Relationships
(Adapted from CONLEY, 1984)

In practical terms, each of the world's time centres keeps a local realisation of UTC, the epoch and rate of which relative to UTC(BIPM) are monitored and corrected periodically. UTC(AUS), for example, is maintained by the Australian Survey and Land Information Group (AUSLIG) in Canberra (formerly the Division of National Mapping) in cooperation with most of Australia's time and frequency laboratories. Its relationship to UTC(BIPM) is known to about 5 microseconds. UTC may be accessed by users by tuning in to radio time signals, which have an uncertainty of about 1 millisecond due to changes in the propagation delay through reflection caused by the ionosphere, or to higher accuracies by direct land line or satellite communication with the time-keeping centre. These days the GPS system of satellites can be used to **transfer time**, by permitting easy access to GPS Time (section 5.3).

2.1 Clock Time

Individual atomic clocks, and high precision quartz oscillator clocks, maintain their own time scale, which is therefore a realisation of "true" time if they can be synchronised to TAI, UTC or GPST. We refer to such time scales in this report as **Clock Time**. For example, a different Clock Time is maintained by each of the GPS satellite clocks and each ground receiver clock. The deviation between Clock Time and "true" time may be considered as:

(1) a time dependent offset with respect to TAI, UTC or GPST, or

(2) as clock rate irregularities over a finite time span arising from the imperfect frequency standard used (section 4.1).

For example, the GPS satellites carry both rubidium and cesium frequency standards which may drift off GPST. However their performance (that is, their rate) is continuously monitored by the GPS Control Segment.

2.2 Time Implied by Satellite Motion

Another time scale which is implicit in all measurement systems based on the motion of satellites is the one related to the time-tags of the satellite ephemerides. The motion of a satellite can be considered as defining a unique time scale which is often referred to as the **Satellite Ephemeris Time**, or SET (and should not be confused with the formal Ephemeris Time, which was superseded by international agreement in 1984 by TDT and TDB). Often there is an attempt made to ensure that SET is coincident with some other atomic time scale such as TAI, UTC or GPST. However, just as individual clocks can have differences with respect to "true" time, individual satellite ephemerides can be considered imperfect realisations of SET, and referred to here as **Satellite Orbital Time** (SOT). SOT is an approximate measure of SET, and hence of Dynamical Time, and should not be confused with satellite Clock Time, though both are approximate, but different, measures of "true" time.

2.3 The GPS Measurement in a Two-Clock Ranging System

GPS positioning is therefore concerned with a variety of satellite orbital, satellite clock and receiver clock time scales. Hence, for high precision applications, the relationships between each time scale and "true" time (through one of its common realisations, such as TAI, UTC and GPST) must be established to some degree of accuracy. This is particularly necessary in the modelling of the GPS measurements themselves, which are essentially the difference between the reference phase of a satellite clock at the time of signal transmission and the reference phase of the receiver clock at the time of signal reception. This clock difference comprises a number of components (DELIKARAOGLOU, 1987), some significant and some that can be ignored if certain operational conditions are fulfilled:

(1) Difference between receiver Clock Time and Proper Time at the station (the "receiver clock error").
(2) Difference between Proper Time and Coordinate Time at the station

(General Relativity time effects: TDT - TDB).
(3) Difference between Coordinate Time at signal transmission and signal reception (directly related to the geometrical receiver-satellite range).
(4) Difference between Coordinate Time and Proper Time at the satellite (General Relativity time effects).
(5) Difference between Proper Time and satellite Clock Time at the satellite (the "satellite clock error").

General Relativity enters here because one clock is in motion relative to the other. In our subsequent discussions we will assume that the General Relativity time effects on GPS measurements have been accounted for in a preprocessing stage and will be ignored (for more details the reader is referred to, for example, ASHBY, 1987; ZHU & GROTEN, 1988).

We can therefore identify four basic stages of GPS data processing where time enters in a number of guises:

- Stage 1: The generation of a satellite ephemeris in an inertial reference frame, the orientation of the earth in space and the determination of the satellite's coordinates in the earth-fixed system.

- Stage 2: The differences in satellite clock times and "true" time, how this is monitored by the GPS Control Segment, and the effect of residual clock errors on GPS measurements.

- Stage 3: The signal transmission time -- a measure of the range, and the effect of receiver clock errors on the GPS measurements.

- Stage 4: Accounting for residual satellite and receiver clock errors on the GPS measurements. The differences in receiver clock times and "true" time, the operational requirements for the synchronisation of GPS receivers, and the simultaneity condition for GPS observations, and how these differences can be either determined or at least accounted for in the processing stage. In addition the impact of clock errors on the measurement time-tags, and through them on the quantities used in GPS data processing.

3. TIME, SATELLITE EPHEMERIDES AND EARTH ORIENTATION

Time, as it relates to near-earth satellite motion, has three main manifestations:

(1) As the variable necessary for the computation of the perturbing forces operating on the satellite, and the independent argument in the Equations of Motion,

(2) As the independent variable in the numerous transformations necessary to relate the inertial reference system used for the formulation and solution of the Equations of Motion to the earth-centred / earth-fixed reference system in which the tracking station positions are referred to, and

(3) As a means of tagging the instant of transmission of a signal from a satellite.

For a GPS user who does not compute his own satellite ephemerides, and instead accepts orbital information from another source, point (1) is of little interest. The analyst's main concern is that the ephemeris time-tags (defining the SOT) are sufficiently close to the "true" time required by his data processing software. The SOT (and SET) maybe UTC, IAT, GPST, or some other well-defined time scale.

Further, if the satellite positions are provided in relation to an earth-fixed reference system, point (2) also does not directly concern the average user, unless the satellite reference system is not the same earth-fixed system to which the user's station position solution is to be referred. In such a case, an additional transformation between satellite and tracking station reference systems is required which may or may not be dependent on time. For example, the GPS Broadcast Ephemerides contained in the Navigation Message transmitted by the satellites are provided in an earth-centred / earth-fixed reference system known as WGS84. This is therefore the reference system in which all station solutions are referred if the Broadcast Ephemerides are used.

For high precision geodetic applications of the GPS technology, some form of orbital adjustment is necessary and therefore points (1) and (2) are of critical importance.

Finally, the determination of the coordinates of a satellite at the time of transmission of the signal is a basic operation within all satellite tracking data processing software.

3.1 Satellite Ephemeris Generation - Time Base Considerations

The best analytical theories of orbital motion produce GPS satellite ephemerides with an accuracy of several dekametres, for arc lengths of less than one day (ABBOT et al, 1983). Analytical orbit models therefore are inadequate, both for operational orbit computations (for multi-day arcs) to support GPS surveying and navigation, as well as for geodetic applications where high accuracy orbits are essential. For such high accuracy applications, orbit modelling is based on the Equations of Perturbed Motion (KING et al, 1987, chapter 7), which are numerically integrated in an inertial reference frame, using Dynamical Time (TDT, or a suitable substitute such as TAI) as the independent variable. This inertial reference frame is commonly defined by the true or mean equator and equinox at some (arbitrary) time epoch.

In effect the satellite reference system is "frozen" at that instant. We can relate the coordinates of a satellite, at any instant in time, to different reference systems through a certain combination of transformations described in the next section.

3.2 The Earth's Motion in Space

The modelling of the GPS observable and the estimation of tracking station positions depend upon the precise definitions of the coordinate systems in which the observations are made and subsequently analysed. We therefore need to describe the earth's motion in space, and define the conventional celestial and conventional terrestrial reference systems used.

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 26000 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.

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 epoch are called the mean equator and equinox-of-date. The precession component relates the mean equator and equinox-of-date to the mean equator and equinox of some defined reference epoch. For example, we may define the **Celestial Reference System** (CRS) by the equator and equinox at a particular date (or epoch). The equator and equinox at 12hTDB on 1st January 2000 (Julian Date 2451545.0), designated J2000, defines a conventional CRS (sometimes known as the CCRS , or simply the Conventional Inertial System). (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 1st January 4713BC.) Applying nutation to the mean equator and equinox of a given date yields the **instantaneous or true equator and equinox** of that date.

Two additional transformations are necessary to relate the CCRS to an earth-fixed system. These are the **earth orientation parameters**. 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**. The relationships used for these transformations (precession, nutation, rotation, and polar motion) effectively define the CCRS by describing its relationship to the terrestrial system, which we refer to here as the **Conventional Terrestrial Reference System** (CTRS). For a full treatment of this subject the reader is referred to KOVALEVSKY et al (1989).

The four transformation components that relate celestial and terrestrial reference systems centred at the earth are (see KAPLAN, 1981; CONLEY, 1984; for definition of constants):

- **Precession**: Transformation of mean-of-date coordinates $\mathbf{r}(t_0)$ at one epoch t_0 to mean-of-date coordinates $\mathbf{r}'(t_1)$ at another epoch t_1 is effected by:

  $$\mathbf{r}'(t_1) = \mathbf{P}.\mathbf{r}(t_0)$$

  where

  $$\mathbf{P} = \mathbf{R}_3(-z_A).\mathbf{R}_2(-\theta_A).\mathbf{R}_3(\zeta_A)$$

- **Nutation**: Transformation of mean-of-date coordinates $\mathbf{r}'$ to true-of-date coordinates $\mathbf{r}''$ is effected by:

  $$\mathbf{r}'' = \mathbf{N}.\mathbf{r}'$$

where

$$\mathbf{N} = \mathbf{R}_1(-\varepsilon-\Delta\varepsilon).\mathbf{R}_3(-\Delta\psi).\mathbf{R}_1(\varepsilon)$$

- **Earth Orientation**: Transformation of true-of-date coordinates $\mathbf{r}''$ to earth-fixed coordinates $\mathbf{r}_e$ is effected by:

$$\mathbf{r}_e = \mathbf{R}_2(-x).\mathbf{R}_1(-y).\mathbf{R}_3(\theta).\mathbf{r}''$$

where x , y are the coordinates of the rotation pole relative to the fixed z-axis of the CTRS system, and θ is the Greenwich Apparent Sidereal Time.

3.3 Coordinates of a Satellite at Signal Transmission Time

One of the basic requirements in adjustments of satellite tracking data is to relate the coordinates of the satellite (in the case of an orbit adjustment -- the *a priori* coordinates) to those of the tracking station. The quantity that directly relates the position vector of satellite i to the position vector of station j is the satellite-receiver range at time t :

$$\rho_j^i(t) = \left| \mathbf{r}_c^s(t^s) - \mathbf{r}_c(t) \right| \qquad (3\text{-}1)$$

when expressed in terms of a celestial reference frame (not fixed to the earth). $\mathbf{r}_c^s$ and $\mathbf{r}_c$ are the position vectors in, for example, the CCRS, of the satellite and station, and t^s is the true time of transmission. Alternatively we may use the model:

$$\rho_j^i(t) = \left| \mathbf{r}_e^s(t^s) - \mathbf{r}_e \right| \qquad (3\text{-}2)$$

when expressed in terms of an earth-fixed reference frame. $\mathbf{r}_e^s$ and $\mathbf{r}_e$ are the position vectors in, for example, the CTRS, of the satellite and station respectively (note the absence of a time argument for $\mathbf{r}_e$ in the case of a stationary tracking receiver).

The important point to note in the use of equations (3-1) or (3-2) is that we are calculating the path of a signal travelling at the speed of light. We use a model of geometric range based on spatial coordinates and relate this to the observations of transit time. This leads us into the realm of Einstein's theories of Special and General Relativity. The relativistic corrections which may be applied to GPS observations (if required, see for example ASHBY, 1987) are valid for calculations carried out in a quasi-inertial (non-rotating) reference frame such as the CCRS. Thus equation (3-1) is valid. However, equation (3-2) is not strictly valid because the reference frame has rotated significantly during transit time.

One simple method to overcome this problem is to define a special temporary "computational" reference frame for the purpose of calculating the satellite-receiver range. This frame is quasi-inertial, earth-centred and we define the axes as being coincident with those of the CTRS at the time of reception. The receiver coordinates at the time of reception are thus the same in both frames. The satellite coordinates at the time of transmission differ from the CTRS coordinates primarily because of the rotation between the two frames during the signal transit time. (There is also an effect in the precession, nutation and polar motion during the time interval $t - t^s$, but it is minute in comparison to earth rotation.) If we use $\mathbf{r_t}$ to represent position vectors in this temporary computational frame, we can correct equation (3-2) by writing it as:

$$\rho_j^i(t) = \left| \mathbf{r_t}^{\mathbf{s}}(t^s) - \mathbf{r_t} \right|$$

$$= \left| \mathbf{R}_3(\omega\Delta t).\mathbf{r_c}^{\mathbf{s}}(t^s) - \mathbf{r_e} \right| \qquad (3\text{-}3)$$

where ω is the rate of rotation of the earth. The modelled range is now correctly defined in terms of satellite and receiver coordinates in the CTRS frame. Note that the temporary computational frame serves only to allow us to generate equation (3-3) and avoid violation of Einstein's theories of Special and General Relativity. It has no other function in modelling GPS observations.

How do we obtain the coordinates of the satellite and the receiver in the same reference frame for equations (3-1) or (3-3)? The receiver coordinates are usually given or required in an earth-fixed (CTRS) system. If the Broadcast Ephemeris is used, the resulting satellite coordinates at the time of transmission will also be in terms of a CTRS. If a precise ephemeris is used, the satellite coordinates may be expressed in an inertial or quasi-inertial system such as the CCRS. The complete transformation between the CCRS and the CTRS is:

$$\mathbf{r_e}^{\mathbf{s}} = \mathbf{R}_2(-x).\mathbf{R}_1(-y).\mathbf{R}_3(\theta).\mathbf{N}.\mathbf{P}.\mathbf{r_c}^{\mathbf{s}} \qquad (3\text{-}4)$$

or

$$\mathbf{r_c}^{\mathbf{s}} = \mathbf{P}^{\mathbf{T}}.\mathbf{N}^{\mathbf{T}}.\mathbf{R}_3^{\mathbf{T}}(\theta).\mathbf{R}_1^{\mathbf{T}}(-y).\mathbf{R}_2^{\mathbf{T}}(-x).\mathbf{r_e}^{\mathbf{s}} \qquad (3\text{-}5)$$

Equation (3-4) may be used to convert satellite coordinates in terms of the CCRS to the CTRS for use in equation (3-3). Equation (3-5) may be used to convert receiver coordinates in terms of the CTRS to the CCRS for use in equation (3-1). In practice the situation is even more complicated.

3.3.1 Conventional Terrestrial Reference System

A CTRS is defined by the totality of procedures, models and constants associated with the establishment and maintenance of a primary (global) network of reference stations. Stations in a geodetic 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 10cm or less in most areas of the world. Such a system is possible because of the large number of observations made over the past 10 years by 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 AUSLIG Laser Ranging Station at Orroral Valley (near Canberra, in eastern Australia) and the NASA Laser Ranging Station at Yaragadee (north of Perth, in Western Australia) are known with respect to each other and to laser ranging stations in North America and Europe with an uncertainty of about 5cm (e.g. STOLZ & MASTERS, 1983). From VLBI observations, the coordinates of the NASA Deep Space Network Tracking Station at Tidbinbilla (near Canberra) and the Parkes radio astronomy telescope (north of Canberra, in NSW) are also known with respect to each other and similar radio telescopes in North America and Europe with a similar uncertainty of about 5cm (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 near the SLR and VLBI tracking stations, a network of sites within Australia could be tied into the CTRS. Such a GPS network is being established in Australia and New Zealand, as part of an expanded worldwide CIGNET system to support high precision civilian GPS activities (NEILAN & MELBOURNE, 1989).

3.3.2 Conventional Celestial Reference System

The realisation of a 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 of GPS satellites, so that they provide a relatively 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 (such as the CIGNET network) whose coordinates are known more precisely from other space techniques.

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. Furthermore, as their rate-of-change is measured in terms of tens of arcsecs per year(!), an error in the time argument of over an hour would need to be made before **P** and **N** terms affect the position of the GPS satellite by a metre, relative to the earth's surface. This is unlikely to be the case. (The corrollary of this is that the precession and nutation matrices need not be recomputed at intervals less than an hour.) The effect of UT1 and polar motion is, however, another matter.

3.3.3 UT1 and Polar Motion

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.

Fortunately, "rapid service" values of UT1 and pole position with uncertainties less than 0.02arcsec are available from the IERS and other organisations within 10 days of the time when the observations were made, and predicted values with uncertainties less than 0.05arcsec 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 0.05arcsec produces a baseline error of less than 3 parts in 10^7 (KING et al, 1987), so the accuracy of UT1 and pole values is not a problem for most applications. The analyst 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.

However, in addition to the unpredictability of UT1 and pole position, the apparent error in the orientation of the earth-fixed reference system (and hence of the GPS satellite's position relative to ground receivers) due to an error in the time argument for the computation of UT1 must be taken into account. The error relation is that 1 second of time is equivalent to 15arcseconds. If the observation time-tag were therefore in error by 1 millisecond, the orbit of the GPS satellite would appear to be displaced (rotation about the earth's spin axis) by 15 milliarcseconds, or 1.5m.

Such an error in the time argument (due to the local clock error), see section 5.4.2, is possible <u>if the time of transmission has to be derived indirectly from the time of reception</u> (the observation time-tag). In the context of GPS data processing this is discussed below.

3.3.4 Interpolation of the Broadcast Ephemeris at Signal Transmission Time

The Broadcast Ephemeris is a convenient representation of the satellite orbit, having the following characteristics:

- It is a compact representation based on 14 parameters (VAN DIERENDONCK et al, 1980) which are changed hourly, but which can be validly used for up to a four hour period centred at the epoch known as the "time-of-ephemeris".
- The satellite coordinates are given directly in the WGS84 system (the realisation of the CTRS in the GPS system). Hence no reference system transformations are required to be carried out by the user.
- The computation of the satellite position only requires the input of a time argument (see KING et al, 1987, for the algorithms involved).

The time input into the calculation of the satellite position is the transmission time of the signal in the GPST system. There are essentially two techniques by which the time of transmission can be determined:

(1) One technique makes use of the range measurement itself and the time of reception as noted on the receiver clock. In the case of GPS, the pseudo-range is converted to transit time:

$$\tau_j^i = \rho_j^i / c \qquad (3\text{-}6)$$

where c is the velocity of light. It is then subtracted from the pseudo-range time-tag to give the signal transmission time directly:

$$\tau_j^i = t_j - t_j^{si} \qquad (3\text{-}7)$$

where t_j is the time of reception and t_j^{si} is the time of transmission (which is sought). (Internal to the receiver however, it is the PRN codes modulated on the L-band signals that give the transmission time, and the pseudo-range is determined from this transmission time and the receiver-determined reception time.)

(2) The second technique uses an indirect approach, not requiring any knowledge of the satellite-receiver range. In order to solve equation (3-7) for modelling an observation time-tagged at t_j, we must iterate as both sides of the equation are functions of the transit time. This is done in practice by starting with a reasonable guess for τ, for example 0.07 seconds in the case of GPS satellites. With this initial guess we can derive an estimate of the transit time (equation (3-7)), determine the satellite position vector:

$$\mathbf{r}^{\mathbf{s}}_{\text{iteration}} = \mathbf{r}^{\mathbf{s}}\,(t_j - \tau^i_{j\,\text{iteration}}) \qquad (3\text{-}8)$$

From this a new transit time is computed using this satellite position and the unchanged receiver position (equation (3-1) or (3-3), & (3-6)):

$$\tau^i_{j\,\text{iteration+1}} = \frac{\rho^i_{j\,\text{iteration}}}{c} \qquad (3\text{-}9)$$

With this new transit time we can repeat the process (equations (3-8) & (3-9)) until convergence. About 3 orders of magnitude are gained with each iteration, and the process usually converges in 2-3 iteration steps. (REMONDI, 1984, describes a more efficient algorithm for GPS transit time computations that minimises the number of iterations necessary.)

There are several comments that need to be made with regards to determining GPS satellite position at transmit time:

- As the Broadcast Ephemeris provides the satellite position vector in an earth-fixed reference system, the appropriate range model to use (for a static receiver) is equation (3-3).

- The direct (pseudo-range) technique for determining time of transmission does not require that the receiver clock time by synchronised with "true" time as the measurement (the transit time) itself contains that error. Hence a local receiver clock error does not propagate into the time of transmission. In fact the receiver clock is irrelevant. However, the time of transmission is directly generated by the satellite clock, and any discrepancy between it and the Satellite Ephemeris Time (in which the time-tags of the orbit are defined) will appear as an alongtrack error in the orbit, by all receivers tracking this satellite, at that epoch.

- The indirect technique for determining time of transmission, on the other hand, is independent of the satellite clock (in fact, one need not even exist), but entirely dependent on the receiver clock. If the local receiver clock time is discrepant from SET, the effect is a time-tag error affecting all satellites being tracked at that epoch.

These latter two points are important when we come to define the notion of Computation Time for GPS data adjustment (section 5.4.2). The indirect approach is likely to lead to greater errors in range modelling as receiver clock error is likely to be greater than a GPS satellite clock error (especially if the broadcast *a priori* clock error model is used -- section 4.4.1), unless the receiver clock error is explicitly estimated (section 5.3).

4. CLOCK BEHAVIOUR AND ERROR MODELLING

Individual clocks maintain their own time scale. The aim is to relate this time scale, in as rigorous a manner as possible, to "true" time. We are therefore dealing with the deviations of each clock time scale from some "perfect" time scale. A study of clock behaviour is useful from two points of view:

- in understanding the limits of clock performance, so that valid approximations can be made concerning clock error(s). For example, clock errors can be assumed to be nuisance parameters effectively eliminated if certain operational requirements are met in the case of GPS surveying.

- in defining explicit models for clock behaviour.

We discuss below some measures of clock performance or stability, and relate them to common clock error models used in GPS.

4.1 Frequency, Phase and Time

Today's high precision clocks are all based on some form of frequency standard or oscillator. In the context of the GPS system these belong to one of two classes:

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

(2) The various types of quartz crystal oscillators, of which the oven-controlled variety is the commonest used in GPS receivers.

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

$$\Phi(t_i) - \Phi(t_{0i}) = f_i\,(t_i - t_{0i}) \tag{4-1}$$

The wavelength of the phase cycle is (c / f_i), where c is the velocity of electromagnetic radiation. Substituting the appropriate scaling for the phase change to convert it into a time interval, and defining the origin of the time scale at the arbitrary reference epoch of the cycle counter, we can obtain an expression for a clock reading as (WÜBBENA, 1988):

$$t_i\,(t) - t_{0i} = \frac{1}{f_0} \int_{t_0}^{t} f_i(t)\; dt \tag{4-2}$$

with:

t_0	is the reference epoch
t_{0i}	is the clock reading at the reference epoch
$f_i(t)$	is the frequency of the oscillator
f_0	is the nominal oscillator frequency

A standard model for the frequency of an oscillator is:

$$f_i\,(t) = f_0 + \Delta f + \dot{f}\,(t - t_0) + f_r\,(t) \tag{4-3}$$

with:

Δf	is a frequency bias
$\dot{f}$	is a frequency drift
$f_r(t)$	are unmodelled random frequency errors

Substituting equation (4-3) into (4-2) gives:

$$t_i\,(t) = t_{0i} + (t - t_0) + \frac{\Delta f}{f_0}(t - t_0) + \frac{\dot{f}}{2f_0}(t - t_0)^2 + \frac{1}{f_0} \int_{t_0}^{t} f_r(t)\; dt \tag{4-4}$$

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

$$\begin{aligned} \varepsilon_i\,(t) &= t_i - t \\ &= a_0 + a_1\,(t - t_0) + \frac{a_2}{2}(t - t_0)^2 + \int_{t_0}^{t} y(t)\; dt \end{aligned} \tag{4-5}$$

where:

a_0	is the clock bias term
a_1	is the clock drift term
a_2	is the clock drift-rate, and

$\int_{t_0}^{t} y(t)dt$ is the integrated random fractional frequency error

Expressed as "phase" error this is:

$$\Phi_\varepsilon(t) = f_0\, \varepsilon_i(t) \qquad (4\text{-}6)$$

In the case of an L1 GPS phase measurement, the frequency is 1575.42MHz, and the effective wavelength of the signal is approximately 19cm. One cycle of the signal therefore is tracked by a frequency counter in approximately 0.6 nanoseconds.

Clock error, whether expressed in terms of phase (equation (4-6)), time (equation (4-5)) or frequency (equation (4-3)) instability, consists essentially of two distinct components:

(1) The systematic part which can be determined (and predicted). This is the explicit polynomial part of equations (4-3) and (4-5), although the order of the polynomial may be higher than that indicated.

(2) The random part, which may be significant enough to have to be taken into account.

Therefore in addition to exhibiting **deterministic deviations** from a "true" time scale, they also undergo **stochastic variations** in both time (or phase) and frequency. An example of a realisation of the time difference between a commercial cesium clock and a time scale generated from an ensemble of atomic clocks is shown in Figure 4.1. The top graph shows how the clock has a frequency bias so that the time scale appears to drift linearly away from the "true" time (defined here by the group of atomic clocks). This linear drift (approximately 50 msec/year) does not affect the clock's ability to keep accurate time as long as the rate of the drift (coefficient a_1) is known or can be estimated very well. The middle graph shows the residuals after fitting a straight line to the top graph. The variation is now of the order of 5 msec/year, but the quadratic appearance indicates a higher order effect still present, in this case a significant frequency drift in the clock (the coefficient a_2 representing "ageing"). That is, the frequency of the clock appears to change linearly with time. The bottom graph shows the residuals after removing a quadratic function. These are now primarily stochastic variations (a higher order polynomial could be used to remove residual "systematic" trends, but we will consider here "stochastic" variations as being those that remain after second polynomial modelling).

This random part needs to be taken into account. If this random part is significant (it is of the order of 0.5 msec/year in the example described above) then its behaviour can be characterised as some form of **random process**. The GPS satellite and receiver clocks are discussed in section 4.4., but we first need to define a measure for clock stability that is appropriate for this random signature.

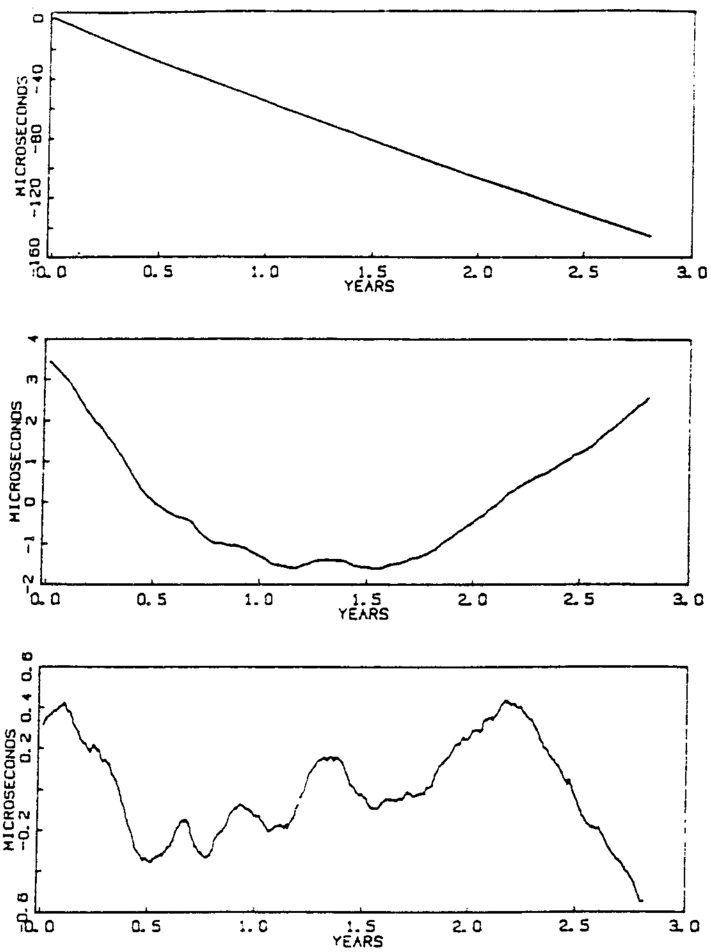


Figure 4.1 Time Difference between a Commercial Atomic Clock and the U.S. Naval Observatory Time Scale (JONES & TRYON, 1987) (The middle graph has linear trend removed, and the bottom graph has a quadratic removed.)

4.2 Frequency and Time Stability

The component of equation (4-3) attributable to random error sources is known as the random fractional frequency deviation, or the integrated random fractional frequency error (equation (4-5)). The total fractional frequency deviation (systematic + random), or just the random part, can be analysed in the time or frequency domain. The standard approach is to deal with the sample variance of only the random fractional frequency fluctuations in the time domain. As we can measure the frequency count or time difference over some elapsed time interval τ we can define the mean value of the fractional frequency deviation (see, for example, WÜBBENA, 1988):

$$y_k = \frac{1}{\tau} \int_{t_k}^{t_k+\tau} y(t)\, dt = \frac{[\Phi(t_k+\tau) - \Phi(t_k)]}{2\pi f_0 \tau} \qquad (4\text{-}7)$$

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

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

$$< \sigma_y^2(N,T,\tau) > = < \frac{1}{N-1} \sum_{n=1}^{N} [y_n - \frac{1}{N}\sum_{k=1}^{N} y_k]^2 > \qquad (4\text{-}8)$$

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

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

$$\sigma_y^2(\tau) = E\{\sigma_y^2(N=2,T=\tau,\tau)\} = E\{\frac{(y_{k+1} - y_k)^2}{2}\} \qquad (4\text{-}9)$$

This two-sample variance of the fractional frequency error is the standard measure of clock stability used. One of its special advantages is its relative simplicity: it is only a function of τ and can be plotted in the form of stability graphs such as those in Figure 4.2 (HELLWIG, 1979). The units for $\sigma_y(t)$ are dimensionless. **Note that in Figure 4.2 the linear drifts of the quartz crystal and rubidium oscillator of 1 part in 10^{10} and 1 part in 10^{12} have been removed.**

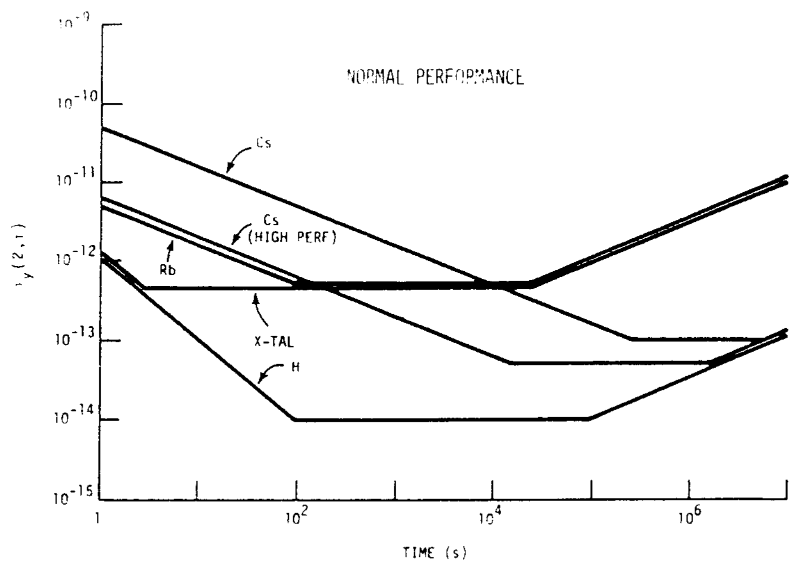


Figure 4.2 Square-Root of the Allan Variance of Typical Oscillators (after removal of linear trend for crystal and rubidium oscillators) (HELLWIG, 1979)

How does one interpret frequency stability plots such as these? The clock stability is portrayed as a function of the time interval between the monitoring of a particular clock. If we assume that at the start of the interval the clock is synchronised with (or has been compared to) "true" time, the amount by which the clock has "deviated" (on average) after a certain time interval τ is given by the square-root of the Allan Variance $\sigma_y(\tau)$ times τ . For example, quartz crystal oscillators are as precise as hydrogen masers if the time intervals are less than approximately 5 seconds. In the short term, up to 10^4 seconds cesium standards fare badly in comparison with other frequency standards. However, their medium to long term performance is superior to all but the hydrogen maser, which it rivals after approximately 10^6 seconds.

A point worth noting with regard to Figure 4.2 is that, with the possible exception of the cesium oscillator, the rubidium, quartz crystal and hydrogen maser oscillators exhibit distinct short-, medium- and long-period stability regimes. These regimes have the following characteristics:

(1) Short-period: the Allan Variance **decreases** as τ increases.

(2) Medium-period: the Allan Variance is a **constant**.

(3) Long-period: the Allan Variance **increases** as τ increases.

4.3 Stochastic Clock Error Models

The Allan Variance is a measure of the performance of a clock. It can be used to predict the behaviour of clocks in the GPS system over time spans ranging from fractions of a second to several hours, and hence the likely build up of clock error (phase, time or range equivalent). This is important for various aspects of GPS observable modelling (section 6 & 7). Usually the clock errors are in fact eliminated through some form of observation differencing. However, there may be instances when a clock error model is explicitly required.

Particular circumstances will influence the form of the clock error model:

- A clock error model involves the parametrisation of the systematic deviations, as well as accounting for the random deviations. Such a model is only possible in the context of an optimal filtering scheme in which the systematic part of the clock error is estimated (usually as polynomial coefficients, as in equation (4-25)) and the clock error residuals are treated as a random process which is the source of some form of "system noise" in the filter (see, for example, LANDAU, 1988; WÜBBENA, 1988; MERMINOD, 1989).

- An *a priori* clock model, on the other hand, such as the one transmitted by the GPS satellites (section 4.4.1), is purely a deterministic model (even though the numerical values of the clock error coefficients themselves may have been determined from a filter solution).

This section briefly describes the random processes appropriate for clock error modelling when the clock error is estimated using a Kalman filter procedure. As a starting point we will use the Square-Root Allan Variance graph (Figure 4.2). The three frequency stability regimes mentioned earlier can be characterised by different random processes. These random processes will be introduced from a purely intuitive point of view, devoid of mathematical details.

First let us summarise some points made earlier (both explicitly and implicitly):

- The instantaneous frequency is defined by equation (4-3). It may also be expressed in terms of the phase output of an oscillator:

$$\Phi = 2\pi f_0 t + \Phi_s(t) + \Phi_r(t) \tag{4-10}$$

  where Φ_s is the systematic part of the deviation and Φ_r is the remaining random part.

- The fractional frequency error is an instantaneous measure of the frequency (or phase) stability, and is defined as:

$$y(t) = \frac{\dot{\Phi}_r(t)}{2\pi f_0} \tag{4-11}$$

  note that $y(t)$ is dimensionless.

- Another parameter is:

$$x(t) = \frac{\Phi_r(t)}{2\pi f_0} \tag{4-12}$$

  which is the oscillator pulse expressed in units of time. $x(t)$ is the instantaneous (random) time error of a clock driven by an oscillator having an instantaneous frequency f.

- $y(t)$ and $x(t)$ are random processes.

- The fractional frequency error can be analysed in the frequency or time domain. The measure of frequency stability in the frequency domain is typically expressed in terms of the one-sided spectral density $S_y(f)$ of $y(t)$. Both the time and frequency domain characterisations can be obtained from each other through simple transformation formulae (see, for example, RUTMAN, 1978).

- A measure of frequency stability in the time domain is convenient as it can be characterised by moment indicators such as mean value, variance and covariance functions. If $y(t)$ is a stationary and ergodic process, we can obtain the statistical mean value and variance of $y(t)$ based on the sampling:

$$M_y = \frac{1}{N} \sum_{n=1}^{N} y(n) \tag{4-12}$$

$$R_0 = \frac{1}{N-1} \sum_{n=1}^{N} [y(n) - M_y]^2 \qquad (4\text{-}13)$$

where R_0 , the true variance can be estimated from the sample variance. That is, $E\{R_0\} = R_0$.

- However, in the time domain, phase, frequency and fractional frequency are all functions of time. In addition it is not possible to measure instantaneous frequency. Therefore these quantities are best characterised by their variation over a specific time interval τ . We define the average fractional frequency error over the time interval τ (equation (4-7):

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

note that as $\tau \to 0$ then $y_k \to y(t)$, y_k is therefore a good approximation to $y(t)$ for short intervals τ .

- We can now analyse y_k in the time domain. The standard variance formula is applied (equation (4-13)) and when $N = 2$, $T = \tau$ we obtain:

$$\sigma_y^2(N=2, T=\tau, \tau) = \frac{(y_{k+1} - y_k)^2}{2} \qquad (4\text{-}14)$$

- The mean value is the Allan Variance or "pair variance" (equation (4-9)):

$$\sigma_y^2(\tau) = E\left\{\frac{(y_{k+1} - y_k)^2}{2}\right\} \approx \frac{1}{M-1} \sum_{k=1}^{M} \frac{[y_{k+1} - y_k]^2}{2} \qquad (4\text{-}15)$$

note that it is a function of τ , and as $\tau \to 0$ and $M \to \infty$, the Allan Variance tends towards the true variance of $y(t)$.

- The Allan Variance is not based on any physical model of the oscillator, instead, the stability graphs are produced from the results of testing actual oscillators. (For the long-period portion of the graphs, long measurement time periods are required, and hence the results are not as reliable as for the short-, medium-period portions.) However, as we shall see below, the Allan Variance computations are not strictly valid for the medium- and long-period sampling intervals.

- For the prediction of clock errors, a simple prediction is:

$$\sigma_x^2(\tau) = \tau^2 . \sigma_y^2(\tau) \qquad (4\text{-}16)$$

The Allan Variance of oscillators is convenient means of characterising frequency stability in the context of qualitative assessments as to whether it is appropriate to model (in the deterministic sense) clock errors in GPS, or whether there are alternative processing procedures that largely eliminate the clock errors (see section 6). Allan Variance is not, however, a convenient expression of clock/frequency stability for stochastic modelling. Attempts are made to characterise the Allan Variance graph as one or more random processes, and to assign to each process a value for the spectral density. All such attempts are approximations (see LANDAU, 1988, for discussion on this problem). "Total" stochastic clock error models appropriate for GPS are discussed in section 4.4. What is presented below is the rationale behind the choice of approximate models based on a number of possible random noise processes.

4.3.1 Short-Period Stability Models

An inspection of Figure 4.2 shows that there is a distinct frequency stability regime where the Allan Variance decreases as the time interval τ increases. This portion of the graph can be expressed as:

$$\sigma_y^2(\tau) = K_1.\tau^{\alpha_1} \tag{4-17}$$

where K_1 for various types of commercially available oscillators is given in Table 4.1. α_1 is positive.

The integral of y(t) is a constant and y(t) is a stationary process, of which the most appropriate model is that of **white noise**. This means that the values of y(t) are uncorrelated over time, so that if y(t) were known at some instant it could not be predicted at a short time later. This would only be true if there was infinite variance on the white noise. Though unrealistic, it is a good approximation of random frequency noise for short elapsed times and has important implications for GPS clock modelling (section 4.4). (White noise on the frequency implies a random walk process for the phase or time error caused by the frequency deviations.)

4.3.2 Medium-Period Stability Models

This frequency stability regime is characterised by a constant Allan Variance. This implies that y(t) increases as τ increases. This portion of the graph can be expressed as:

$$\sigma_y^2(\tau) = \sigma_{yF} = \text{constant} \tag{4-18}$$

where σ_{yF} for various types of commercially available oscillators is given in Table 4.1.

There is no simple random noise model in the time domain that can describe this behaviour. Instead a non-stationary process for the frequency deviations known as **flicker noise** is assumed. Flicker noise is defined in the frequency domain, but its physical cause is uncertain. To be used in filter models for clock error it must be transformed into the time domain, and a discussion of how such a transformation can be approximated is found in LANDAU (1988). Because of its relative complexity, compared with the short-period and long-period random processes, it is often not used (see, for example, BAUSTERT et al, 1989;

MERMINOD, 1989).

4.3.3 Long-Period Stability Models

This frequency stability regime is characterised by the Allan Variance increasing with increasing t . This implies that y(t) increases rapidly as τ increases. This portion of the Allan Variance graph can generally be approximated by:

$$\sigma_y{}^2(\tau) = K_2.\tau^{\alpha_2} \tag{4-19}$$

where K_2 for various types of commercially available oscillators is given in Table 4.1. α_2 is negative.

y(t) is a non-stationary process best approximated by a **random walk**, which itself implies a white noise process on the derivative of the frequency: the frequency drift ($\dot{f}$), and a second order random walk (known also as integrated random walk or random ramp) on the phase or time error (MERMINOD, 1989).

Table 4.1 Typical Performance Data for Commercially Available Oscillators (Adapted from HELLWIG, 1979)

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

4.4 Clock Models Appropriate for GPS

From the discussions in the previous sections, the fractional frequency error y(t) can be represented as a set of three independent noise processes $Z_n(t)$ (LANDAU, 1988):

$$y(t) = Z_{-2}(t) + Z_{-1}(t) + Z_0 \tag{4-20}$$

where Z_0 , Z_{-1} , Z_{-2} represent the white noise, flicker noise and random walk processes respectively. The spectral density of Z_n is given by:

$$S_{Z_n}(f) = h_n\, f^{\,n} \qquad 0 \leq f \leq f_h$$

or

$$S_{Z_n}(f) = 0 \qquad f > f_h\,,\ h = -2, -1, 0 \tag{4-21}$$

where the h_n are constants and f_h is the system noise bandwidth. The one-sided spectral density of y(t) is given by:

$$S_y(f) = h_{-2}\, f^{-2} + h_{-1}\, f^{-1} + h_0 \tag{4-22}$$

The Allan Variance can be obtained by transformation of the spectral densities into the time domain (see, for example, RUTMAN, 1978):

$$\sigma_y^2(\tau) = \int_0^{\infty} S_y(f)\, \frac{2\sin^4 \pi\tau f}{(\pi\tau f)^2}\, df \tag{4-23}$$

The result is of the form:

$$\sigma_y^2(\tau) = \frac{h_{-2}\, 2\pi^2\tau}{3} + h_{-1}\, 2 \ln 2 + \frac{h_0}{2\tau} \tag{4-24}$$

A task is therefore to determine the spectral densities for the various h_n . In general the flicker noise model is not used (see, for example, BAUSTERT et al, 1989).

4.4.1 GPS Satellite Clocks

The GPS satellite component most likely to fail is the clock. Each satellite is launched with several clocks on board. Any one of the oscillators can be commanded by the GPS Control Segment to be the primary spacecraft oscillator. The prototype (Block I) satellites were launched with different oscillator designs in order to test which would function best under space conditions. Some of the satellites are at present operating with cesium clocks, others with rubidium clocks. The Block II GPS satellites are equipped with two cesium oscillators and two rubidium oscillators, and the decision as to which will be the "master" clock is made by the Master Control Station on the basis of the monitored performance of the clocks. The clock bias, drift and drift-rate are explicitly determined in the same procedure as the satellite ephemeris is estimated. The behaviour of each GPS satellite clock is monitored against GPS Time (GPST), which is itself kept synchronised to UTC as defined by the U.S. Naval Observatory.(after taking into account the leap second offsets -- section 1.2). As a consequence, the clock behaviour so determined in used to:

(1) occasionally reset the satellite clock so that it is within 1 millisecond of GPST.

(2) the offset, drift and drift-rate of the satellite clocks are available to all GPS users via clock error coefficients broadcast in the Navigation Message.

The satellite clock error model is a polynomial of the form:

$$\varepsilon = a_0 + a_1\,(t - t_0) + a_2\,(t - t_0)^2 \tag{4-25}$$

where

a_0 is the clock bias term

a_1 is the clock drift term
a_2 is the clock drift-rate

What is available to users is really a prediction of the clock behaviour some time into the future (possibly up to 24 hours ahead). As the random deviations of even cesium and rubidium oscillators are not predictable, such deterministic models of satellite clock error (mis-synchronisation with GPST) are accurate to about 20 nanoseconds, or six metres in equivalent range. The residual satellite clock error can therefore not be neglected.

A number of options are available for handling this error:

- construct a range-like observable from which the satellite clock error ε has been eliminated.

- model the satellite clock error as a random process.

In the context of GPS phase data reduction (section 6.1) the former involves observation differencing, while the latter requires explicit estimation of the clock error on an epoch-by-epoch basis. Although a random walk process for the time error would be most appropriate for short sampling periods (subsecond to several minutes), this could only be implemented within an optimal filtering estimation scheme. In general however, a standard least squares algorithm is used in which the time error is modelled as white noise. This is in fact no model, in the stochastic sense, as no attempt is made to relate the clock error from epoch to epoch.

Stochastic models of GPS satellite clock error are included in the estimation algorithms for satellite orbit determination used by the Control Segment (for example, SWIFT, 1985), as well as those interested in processing undifferenced or single-differenced phase data.

4.4.2 GPS Receiver Clocks

Most GPS receivers are equipped with oven-controlled, quartz crystal oscillators. Their major advantages are that they are small, consume little power and are relatively inexpensive. In addition they have good short-term stability (Figure 4.2). Some receivers are, however, equipped with ports to permit the connection of an external frequency standard such as a cesium, rubidium and even a hydrogen maser.

Although the time defined by individual receiver clocks have essentially arbitrary origins, they can be tied to a well established time scale, such as GPST, in a number of ways. Generally, the time origin of a GPS receiver is set automatically as soon as sufficient satellites are tracked (section 5.3). The subsequent time scale defined by the corrected receiver clock is then nominally that of GPST because:

- the synchronisation at some epoch (that is, defining the time origin) is susceptible to error. Generally, it can be defined only at the 0.1 microsecond level.

- the stability of the scale is directly related\to the quality of oscillator used, and how often the current clock time is compared with GPST through the use of GPS pseudo-range observations.

However, as with satellite clock errors, GPS phase data processing to survey accuracies requires that the clock errors be explicitly estimated, or eliminated, so that no residual clock errors remain.

4.4.3 One-Way Ranging

All GPS measurements involve two clocks: the satellite clock responsible for generating the signal, and the receiver clock, against which the incoming signal is compared. A GPS range is derived from the difference between the receiver clock reading at the time of reception of a signal and the satellite clock reading at time of transmission.

GPST is the intermediate time scale against which all GPS clocks are compared. Hence as GPST is "true" time, the offset of any clock from this particular time scale is generally treated in GPS processing as a clock error. A GPS range therefore is contaminated by two clock errors: it is in fact, for this reason, termed a **pseudo-range**.

The dominant error is, however, the receiver clock error because:

- the satellite clock, being either a rubidium or cesium atomic standard, is of higher quality than the average receiver clock, which is generally a quartz crystal oscillator.

- the satellite clock time is regularly monitored by the GPS Control Segment, and the clock error is modelled by transmitted polynomial coefficients (section 4.4.1).

This is illustrated schematically in Figure 4.3.

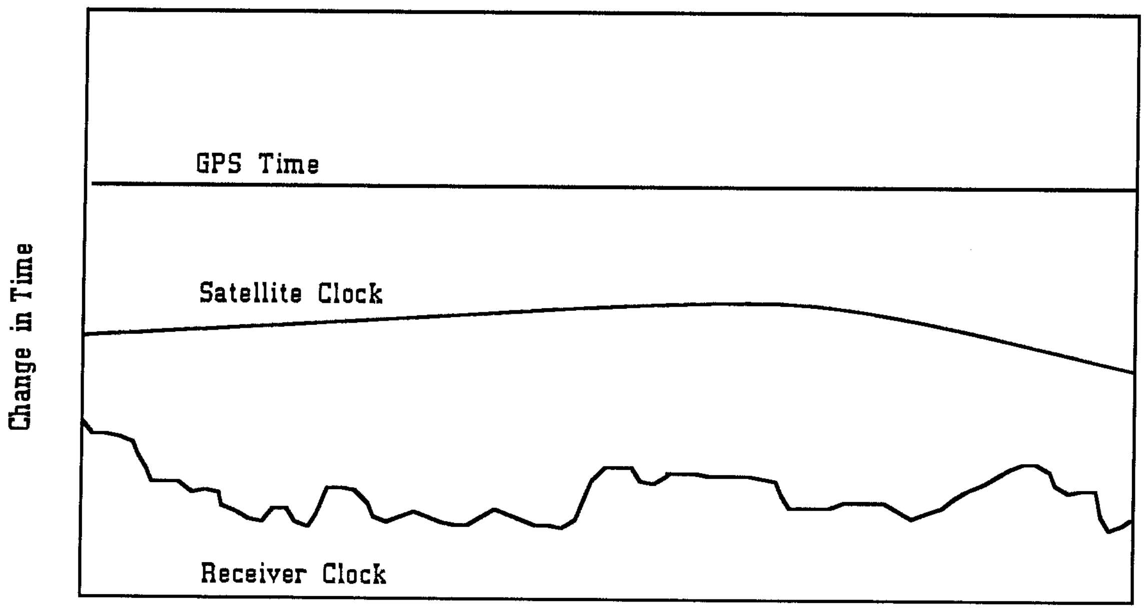


Figure 4.3 Variations in Satellite Clock and Receiver Clock Rate (Adapted from ROCKEN, 1988)

5. THE GPS MEASUREMENT PROCESS

5.1 Defining Clock Error

The clocks used to generate GPS measurements and to time-tag them are imperfect, however the theoretical time frame for modelling the observations and for undertaking the geodetic calculations is "true" time. It is useful to suppose initially that every clock, or oscillator, can be compared directly with a "true" oscillator which has a known and constant frequency f_0 with respect to "true" time.

In the following mathematical development the symbol Φ is used to represent the phase of this "true" oscillator and T to represent the "true" time generated by the oscillator. The behaviour of phase with time obeys the relation (equation (4-1) & (4-2)):

$$\Phi(T) = \Phi(T_0) + \int_{T_0}^{T} f(T)\, dT \qquad (5\text{-}1)$$

where f(T) is the time dependent frequency of the oscillator and T_0 is the time at some arbitrary initial reference epoch. Because the frequency of the "true" oscillator is constant we can write:

$$\Phi(T) = \Phi(T_0) + f_0\,(T - T_0) \qquad (5\text{-}2)$$

The clock time generated by an imperfect clock oscillator is obtained by counting the number of whole and fractional cycles and dividing by the nominal frequency. The frequency of any oscillator is constant with respect to its own clock time. In general however it will not be constant with respect to any another time scale or oscillator. Using ϕ as the phase of an imperfect oscillator and t(T) as the clock time defined by that oscillator at "true" time T we can write:

$$\phi(T) = \phi(T_0) + f_0\,[\, t(T) - t(T_0)\,] \qquad (5\text{-}3)$$

where we have chosen the nominal frequency f_0 of the clock oscillator to be the same as the nominal frequency of the perfect oscillator. The term $[\, t(T) - t(T_0)\,]$ is the number of seconds of clock time that have elapsed in the interval between "true" times T and T_0.

The frequency of this oscillator is not constant with respect to "true" time and thus the clock time will have a non-linear behaviour with respect to "true" time. We can represent the difference between "true" time and clock time by phase errors, by time errors, or by frequency errors. The nominal L1 carrier frequency of GPS satellites is 1575.42MHz and 1 cycle of phase is equivalent to approximately 0.6 nanoseconds. The measured quantity in GPS is phase and these measurements are made at intervals timed by receiver clocks. The most useful representations of clock behaviour are therefore phase errors or clock errors. To relate frequency error to time or phase error we need to know the complete behaviour of the frequency with time so that the integral in the following equation (equation (5-1) can be evaluated:

$$\phi(T) = \phi(T_0) + \int_{T_0}^{T} f(T)\, dT \qquad (5\text{-}4)$$

The phase and clock errors are only required at the discrete epochs when measurements are made, and therefore it is not necessary to know the complete behaviour of the oscillator with time. It is often more efficient to describe clock behaviour in terms of phase or clock errors at specific epochs rather than by a continuous model of frequency error such as a polynomial.

A general model such as a polynomial has been used by some authors to represent the continuous satellite and receiver clock behaviours as they impact on GPS observation modelling (for example, KING et al, 1987). This method is used successfully to represent the satellite oscillators for low accuracy applications (section 4.4.1). A second order polynomial is sufficient for most purposes. However there is a possibility that the stability of the satellite oscillators may be intentionally degraded or "dithered" under the policy of "selective availability". For example, if sudden changes in satellite oscillator frequency occur, these will not be well modelled by a polynomial.

Therefore, for reasons given above, the GPS observation equations will be developed using discrete (epoch-by-epoch) clock errors rather than frequency polynomials. This does not preclude the subsequent development of polynomials to model these clock errors. It does however ensure that the equations developed are general and do not reflect one particular continuous clock modelling strategy.

The clock error at "true" time T is represented by $\varepsilon(T)$ where

$$\varepsilon(T) = t\,(T) - T$$

or

$$t\,(T) = T + \varepsilon(T) \qquad (5\text{-}5)$$

If we substitute equation (5-5) into equation (5-3) we obtain:

$$\phi(T) = \phi(T_0) + f_0\,[\,T - T_0 + \varepsilon(T) - \varepsilon(T_0)\,] \qquad (5\text{-}6)$$

and using equation (5-2) to substitute for $f_0(T - T_0)$ we get:

$$\phi(T) = \Phi(T) + f_0\,\varepsilon(T) + \phi(T_0) - \Phi(T_0) - f_0\,\varepsilon(T_0) \qquad (5\text{-}7)$$

or

$$\phi(T) - \Phi(T) - f_0\,\varepsilon(T) = \phi(T_0) - \Phi(T_0) - f_0\,\varepsilon(T_0)$$
$$= K \qquad (5\text{-}8)$$

The term K is constant in time as the equation is true for any T. The phase difference between two oscillators is closely related to the clock difference. Two clocks reading exactly the same time will have oscillators with exactly the same phase. Applying this to the above equation, if the clock error $\varepsilon(T)$ is zero then K is equal to the difference between $\phi(T)$ and $\Phi(T)$, which we have noted is then also zero. We can therefore write:

$$\phi(T) = \Phi(T) + f_0\,\varepsilon(T) \tag{5-9}$$

This follows from the definition of clock error and the fact that the clock time is generated by the phase of the oscillator, which in effect means that time and phase are completely dependent on each other.

5.2 The GPS Phase Observable

5.2.1 Signal Transit Time

Let the **time of reception** of a signal at receiver j be T_j , and let the **time of transmission** of that signal from satellite i be T_j^i. The **transit time** of the signal from i to j is $\tau_j^i(T_j)$ and is defined by (equation (3-6)):

$$\tau_j^i(T_j) = T_j - T_j^i \tag{5-10}$$

The beat phase formed as an observation in the receiver is the difference between the phase of the received signal and the phase of the local receiver oscillator. The phase of the received signal at the time of reception T_j is equal to the phase of the transmitted signal at the time of transmission T_j^i (this is implicit in the definition of the transmission time). Therefore:

$$\phi_{bj}^i(T_j) = \phi_{rj}^i(T_j) - \phi_{loj}(T_j)$$

$$= \phi^{ti}(T_j^i) - \phi_{loj}(T_j) \tag{5-11}$$

where:

- $\phi_{bj}^i(T_j)$ = the carrier beat phase for receiver j , satellite i , at reception time T_j
- $\phi_{rj}(T_j)$ = the received signal phase from satellite i at receiver j at time T_j
- $\phi_{loj}(T_j)$ = the local oscillator phase of receiver j at time T_j
- $\phi^{ti}(T_j^i)$ = the transmitted signal phase from satellite i at transmission time T_j^i

From equation (5-9) we can write an expression relating phase and clock phase error of the local oscillator:

$$\phi_{loj}(T_j) = \Phi(T_j) + f_0\,\varepsilon_j(T_j) \tag{5-12}$$

where $\varepsilon_j(T_j)$ is the clock phase error of receiver j. Using equations (5-2), (5-9) and (5-10), an expression relating transmitted phase and satellite clock phase error, at reception time can be formed:

$$\begin{aligned} \phi^{ti}(T_j^i) &= \Phi(T_j^i) + f_0\, \varepsilon^i(T_j^i) \\ &= \Phi(T_j) - f_0\, \tau_j^i(T_j) + f_0\, \varepsilon^i(T_j^i) \end{aligned} \qquad (5\text{-}13)$$

where $\varepsilon^i(T)$ is the clock phase error attributable to the oscillator of satellite i. Combining equations (5-11), (5-12) and (5-13) leads to an expression for the **carrier beat phase**:

$$\phi_{bj}^i(T_j) = -\, f_0\, [\, \tau_j^i(T_j) - \varepsilon^i(T_j^i) + \varepsilon_j(T_j)\,] \qquad (5\text{-}14)$$

The transit time component in equation (5-14) is made up of two parts. The main part is derived from the true geometric range ρ_j^i between the satellite at the time of transmission and the receiver at the time of reception. This is determined from the position vectors of the satellite and the receiver, as defined in the same reference system (see section 3.3.4). Given the speed of the signals (c , the speed of light) we can calculate the time taken for the signal to travel this distance. The value we use for the speed of light c is the speed in a vacuum and hence this is the primary means of defining scale in GPS. It is the satellite and receiver position information contained in the geometric range that allows us to use GPS for positioning. The second part of the transit time term accounts for the extra time taken for the signal to travel through the earth's atmosphere. This is caused by a change in velocity as the signal traverses layers of different density, but as the effect is small compared with the geometric range it can be modelled as a time delay, a phase delay or as an increase in range, and can be assumed to not affect GPS scale defined by c . If we incorporate this as a phase correction term ϕ_{atmos}, then it may be further broken down into terms for the different parts of the atmosphere: ϕ_{ion} for the ionosphere component, and ϕ_{trop} for the troposphere component. Hence:

$$f_0\, \tau_j^i(T_j) = (\, f_0 / c\,)\, \rho_j^i(T_j) + \phi_{atmos} \qquad (5\text{-}15)$$

Combining equations (5-14) and (5-15) we obtain a model for carrier beat phase:

$$\phi_{bj}^i(T_j) = -\, (\, f_0 / c\,)\, \rho_j^i(T_j) + f_0\, [\, \varepsilon^i(T_j^i) - \varepsilon_j(T_j)\,] - \phi_{atmos} \qquad (5\text{-}16)$$

5.2.2 Measured Carrier Beat Phase

The <u>measured</u> carrier beat phase differs from the above <u>modelled</u> beat phase in a number of important respects. Firstly, the carrier beat phase measurements will have random noise associated with the measurement process. This may be accommodated by an additional noise term ϕ_{noise} , with a magnitude of only a fraction of a cycle (a few mm in equivalent range). Secondly, the definition of clock phase error in equation (5-9) depends on the integral and fractional part of the carrier phase. When the carrier beat phase is measured in a GPS receiver the measurement is ambiguous with regards to the number of integer cycles. This integer is set in a counter to some arbitrary value when the satellite signal is first acquired (for example zero), see Figure 5.1. There is therefore an unknown integer number of cycles n_j^i difference between the measured carrier beat phase and the model beat phase in equation (5-16). This will be unique to a particular satellite-receiver pair. Furthermore, it will be a constant for as long as the receiver continues to track and count the integer number of cycles from the time the satellite signal is first acquired. At any epoch other than the initial measurement epoch, the instrument measures the fractional phase $\phi_{fj}^i(T_j)$ and, in addition, takes a reading $\phi_{cj}^i(T_j)$ on the cycle counter. This combined fractional and integer phase observation is commonly referred to as **integrated carrier beat phase**, and n_j^i is the so-called **cycle ambiguity**.

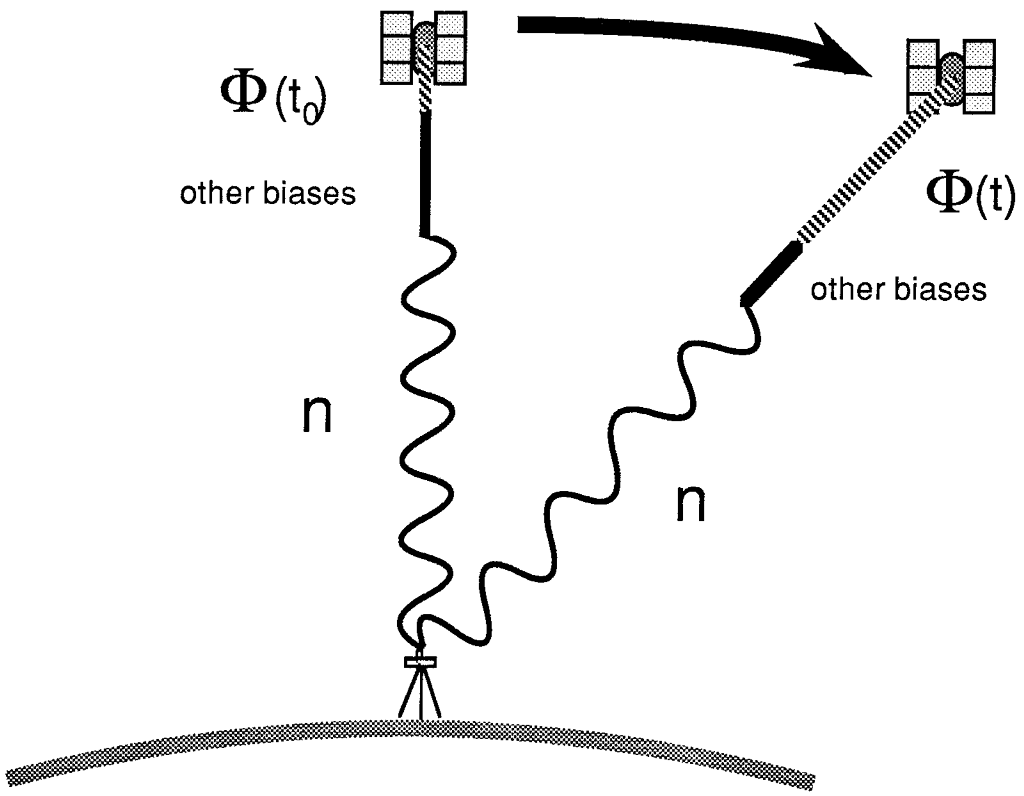


Figure 5.1 Integrated Carrier Beat Phase and the Ambiguity

The model of the measured carrier beat phase (in units of cycles) therefore includes these three additional terms:

$$\phi_{bj}^i(T_j) = \phi_{fj}^i(T_j) + \phi_{cj}^i(T_j)$$

$$= -\,(f_0 / c)\,\rho_j^i(T_j) + f_0\,[\,\varepsilon^i(T_j^i) - \varepsilon_j(T_j)\,]$$

$$+\, n_j^i - \phi_{atmos} + \phi_{noise} \qquad (5\text{-}17)$$

Note that such an observation model is necessary for both L1 and L2 integrated carrier beat phase measurements.

5.3 The GPS Pseudo-Range Observable

The model for the measured pseudo-range follows directly from equation (5-17), after conversion to units of length:

$$\rho_{dj}^i(T_j) = \rho_j^i(T_j) + c\,[\,\varepsilon^i(T_j^i) - \varepsilon_j(T_j)\,] - \rho_{atmos} + \rho_{noise} \qquad (5\text{-}18)$$

The points to note are:

- The absence of the ambiguity term.
- The satellite and receiver clocks terms are identical to those in the carrier phase model (having been generated by the same clocks). This close relationship to carrier phase often leads to pseudo-ranges being referred to as code-phase.
- The noise on the measurement ρ_{noise} is several orders of magnitude larger than for carrier phase, being of the order of metres to dekametres.

The last point in particular is significant. It limits the usefulness of pseudo-range data to relatively low accuracy applications. For these applications it is sufficient to model the satellite clock error $\varepsilon^i(T_j^i)$ by equation (4-25), and to correct the pseudo-range data. This is also done, to some extent, to the atmospheric effect, using simple refraction models. The remaining bias is therefore the receiver clock error:

$$\rho_{dj}^i(T_j) = \rho_j^i(T_j) + c.\varepsilon_j(T_j) + \rho_{noise} \qquad (5\text{-}19)$$

This error is, in effect, the offset of the receiver clock time from that of the corrected satellite clock time. It is a constant error in the ranges to any satellite, made at time $T_j)$. This has two important implications:

(1) If the position of the receiver is known, then (with the satellite position defined by the Broadcast Ephemeris) the range $\rho_j^i(T_j)$ is also known. The clock offset at time $T_j)$ can be determined directly from equation (5-19), with one pseudo-range observation $\rho_{dj}^i(T_j)$. This is a very effective form of **time transfer**, as the derived time offset now defines the relationship between local clock time and a well defined time scale such as GPST (and thence TAI, UTC, etc.).

(2) If the position of the receiver is unknown, then there are four parameters in equation (5-19) which are unknown (the 3 coordinate components

plus the receiver clock error $\varepsilon_j(T_j)$). The measurement of four pseudo-ranges (to four GPS satellites) at the same time permits a direct solution of all parameters to be made. This is the so-called **navigation solution**.

In both cases, given the level of measurement noise, and possible errors in the orbit and satellite clock error model, the relationship between receiver clock time and GPST can probably be defined to better than 0.1 microsecond. This is however not sufficiently accurate for modelling the receiver clock error in the carrier beat phase (equation (5-17)), though it is sufficient for the purpose of correcting the time-tags of the observations (see section 5.4.2).

5.4 Errors in Modelling the Observation

Mismodelling of the integrated carrier phase observation (equation (5-17)) can result from a number of effects. In addition to the atmospheric effects (which we do not wish to comment on in this paper), the main errors arise from:

(1) The measurement noise may not be truly random, but influenced by systematic disturbances external to the receiver circuitry. One effect is **multi-path**, resulting from spurious signal reflections from nearby surfaces. For such effects there is generally no remedy available through improved observation modelling.

(2) The cycle ambiguity n_j^i may not be constant for the entire period that satellite i is tracked by the receiver (j). A jump in the value is known as a **cycle-slip**. This complicates carrier phase processing, and is generally detected and repaired in a preprocessing step (so that in fact the corrected observations do satisfy equation (5-17)).

(3) Errors in the calculation of the **satellite-receiver range** $\rho_j^i(T_j)$. These are discussed below in more detail.

5.4.1 The Satellite-Receiver Range Model

The satellite-receiver range contains all the information required for positioning with GPS. The range is a function of both the position vector of the satellite and the position vector of the receiver. In the least squares adjustment of the carrier phase data this range is modelled using the *a priori* information about the position vectors. This modelled value is affected by errors in the model and in the *a priori* values. If these are significant we should include further parameters in the adjustment, such as corrections or changes to the receiver coordinates or orbital parameters to account for errors in the satellite ephemerides. However, as the satellite-receiver range changes with time, any uncertainty in the time scale used for the phase data calculations will also affect the accuracy of this model, and hence of the adjustment itself.

For the purposes of this discussion we assume that the range modelling is performed in an earth-centred / earth-fixed (ECEF) coordinate frame (section 3.3). We will also assume that the receiver is fixed in this frame (that is, we are dealing with static rather than kinematic positioning), and that effects such as earth tides and tectonic plate motion are insignificant. Hence the relative motion between satellite and receiver is described entirely by the change of the satellite position vector relative to this frame.

We can identify four major sources of error in the computation of the satellite-receiver range:

(1) Errors in the ECEF coordinates of the receiver.
(2) Errors in the satellite ephemeris.
(3) Errors in the transformation between inertial and ECEF frames.
(4) Errors in the time input to the calculation of the range.

Errors (1) and/or (2) may be accounted for by including appropriate parameters in the GPS phase observation model for the data adjustment. If the ephemeris is given in terms of ECEF coordinates (as in the case of the Broadcast Ephemerides) then any errors in the transformation between inertial and ECEF frames (3) will be included with the ephemeris errors, (2). (The most common error is to neglect the incremental rotation of the earth between time of transmission of a signal and time of reception when the range modelling is carried out in the ECEF frame -- section 3.3.) Otherwise the transformation parameters (for precession, nutation, polar motion, earth rotation) are usually considered to be to sufficient accuracy for the purposes of GPS data processing. Error (4) is the remaining influence on the range calculation that needs particular attention.

5.4.2 Time-Tag Error, Computational Time and the Calculation of Model-Range

We need the time of transmission of the signal **with respect to the Satellite Ephemeris Time** in order to calculate the satellite position vector from the ephemeris. A number of techniques can be used (see section 3.3.4) but the accuracy of this calculation depends on whether there is any difference between the time of transmission used in the calculation and the true time of transmission expressed in terms of SET (or its realisation, Satellite Orbital Time - SOT). Here we assume the ephemeris time-tags are expressed in the SOT system.

The time of reception used for this calculation is usually based on the receiver time-tag for the observation (that is, Local Clock Time) but may differ from it if corrections have been applied. For example, the receiver clock may be synchronised with some time standard such as UTC, before, during and/or after field operations. This corrected time of reception can be used, together with the known difference between this time standard and the SOT, to obtain an *a priori* estimate of corrections to be applied when calculating the satellite-receiver range. These corrections can be applied to either the observation time-tags or the ephemeris time-tags, thus accounting for the effect of any time system error in the modelling of the GPS data.

Alternatively, we can include extra parameters in the adjustment and estimate the difference between the time scale of the observation time-tags and the SOT scale using either the pseudo-range data or the phase data itself. This *a posteriori* correction can then be applied in subsequent iterations of the adjustment. (Generally the analysis of pseudo-range data occurs at the *preprocessing stage*, the time-tags then are assumed corrected for the subsequent phase adjustment step.)

The point to note is that the time system used for the calculations of satellite-receiver ranges is not necessarily Receiver Clock Time, Satellite Ephemeris Time, GPST or any of the other well defined time scale. We can refer to this time scale as **Computational Time**. The difference between Computational Time and SET is the difference between the time-tags of the ephemeris and the time-tags of the observations (which may have had some corrections applied). This difference is called the **time-tag error**. An unintended error may be introduced into the geodetic calculations if significant time-tag errors are present, and they are not accounted for. The precise meaning of the word "significant" will be discussed later.

In the above discussion we have assumed that the observation time-tags have been generated by the receiver clock, and then perhaps corrected in some way. The receiver clock is therefore performing two functions: it generates the signal which is used to form the carrier beat phase and it acts as a clock to produce the observation time-tags. The time-tags and the observations are therefore not strictly independent, and an error in one will lead to an error in the other. However, according to our definition of the time-tag error, it may not be completely the result of receiver clock error. It may result, in addition, from:

- errors in the corrections applied to the receiver clock time to correct it (or synchronise it with "true" time) -- see section 5.3, and from

- uncertainties in the realisation of Satellite Ephemeris Time (that is, the SOT).

In GPS data processing the receiver clock errors and the observation time-tag errors are in fact often regarded as independent. The pseudo-ranges, if measured, are affected by the receiver clock error to the same extent that the phases are. We could use them in two ways:

(1) To carry out a pseudo-range "navigation solution" in which the estimated receiver clock error (offset from GPST) is used to define a **corrected Receiver Clock Time**, to which the observation time-tags refer (and hence defines the Computational Time) -- section 5.3.

(2) The pseudo-range, together with the observation time-tag, can be used to calculate the time of transmission of the signal in the time scale of the satellite clock (section 3.3.4). After applying the *a priori* satellite clock error model transmitted in the Navigation Message, this would result in an observation time-tag independent of the receiver clock. The Computational Time would in effect be a **corrected Satellite Clock Time.**

As we shall see, the difference between either definition of Computational Time and the GPS Satellite Ephemeris Time is insignificant.

In this discussion we will regard time-tag errors as being independent of receiver clock errors (affecting the observations themselves), while still remembering that in practice they are often linked. We need to introduce a set of additional parameters to account for the time-tag error, denoted by $te_j(T_j)$, for receiver j at time T_j .

We can now consider the effect of this time-tag error on the calculation of the satellite-receiver range. The effect on the calculated range of a small discrepancy between the time of transmission expressed in the Receiver Clock Time system and the equivalent event in the SET system can be represented by a Taylor's series expansion:

$$\rho_j^i(T_j + te_j(T_j)) = \rho_j^i(T_j) + te_j(T_j)\, \dot{\rho}_j^i(T_j) + \tfrac{1}{2}\,(\, te_j(T_j)\,)^2\, \ddot{\rho}_j^i(T_j) + \ldots \qquad (5\text{-}20)$$

where $\dot{\rho}$ is the rate of change of range (range-rate) and $\ddot{\rho}$ is the second derivative of range with respect to time (range acceleration). The maximum range-rate for GPS satellites (taking earth rotation into account) is approximately 760 m/sec and the maximum range acceleration is approximately 0.14 m/sec^2.

If we seek to keep the error in the satellite-receiver range to below 1mm (equivalent to 0.005 cycles or 3 picoseconds at the GPS L1 carrier frequency of 1575.42MHz) then the second order term in equation (5-20) is only significant if the time-tag error is greater than approximately 0.1 seconds. The first order term will be significant if the time-tag error is greater than 1 microsecond. In GPS adjustments, the Computational Time based on either corrected Satellite Clock Time or corrected Receiver Clock Time, the time-tag errors are less than 1 microsecond and both of these terms may be ignored.

5.4.3 Modelling Time-Tag Error in Phase Observations

However, if Computational Time has not been defined, with the aid of pseudo-range data, to be close to SET, and the time-tag error is indeed large, it would need to be estimated as an additional parameter during the analysis of the phase data alone. If we substitute the zero and first order terms of equation (5-20) for $\rho_j^i(T_j)$ in equation (5-17) we obtain a more complete expression for the modelled carrier beat phase:

$$\begin{aligned}\phi_{bj}^i(T_j) &= -\,(\, f_0 / c\,)\,[\, \rho_j^i(T_j) - te_j(T_j)\, \dot{\rho}_j^i(T_j)\,] \\ &\quad + f_0\,[\, \varepsilon^i(T_j^i) - \varepsilon_j(T_j)\,] \\ &\quad + n_j^i - \phi_{atmos} + \phi_{noise} \qquad (5\text{-}21)\end{aligned}$$

In equation (5-21) we have a different time-tag error term for every receiver and for every epoch. It should not be necessary to estimate all of these. Over an observation period of say 3 hours (approximately 10^4 seconds) a quartz crystal receiver clock with a drift of 10^{-10} sec/sec will not change its error by more than 1 microsecond, and thus a single error term per receiver for all epochs should be sufficient. In any case, 2 terms per receiver (time-tag offset and drift-rate) should be sufficient to ensure that all time-tag errors greater than 1 microsecond are accounted for.

If the receivers have been synchronised to each other in some way (as the early Macrometer "codeless" receivers had to be -- KING et al, 1987), but it is suspected that they are not synchronised with respect to SET, then it may be sufficient to have one time-tag error parameter which describes the average error of all receivers.

6. DIFFERENCED GPS PHASE OBSERVABLES

6.1 The Elimination of Clock Errors by Differencing

In developing the GPS carrier beat phase observations in the previous section, we have briefly considered the effect of systematic time errors on an individual observation. This is however not the whole picture. We are more interested in how these observations are brought together in an adjustment and how errors affect the adjusted parameters.

When we form a least squares model using undifferenced phase observations (equation (5-17) or (5-21)), we find ourselves making assumptions about the parameters, for example, that two observations have a common clock error. The same assumptions are made when differences are formed between observations to eliminate common errors. In GRANT et al (1990) the authors discuss the **Fundamental Differencing Theorem**, which may be summarised as follows: regardless of whether we eliminate parameters by differencing or estimate them in an undifferenced model, the results for the other adjusted parameters will be the same. From this we can deduce that assumptions that are made explicitly in forming differenced observation equations are also made implicitly in formulating the undifferenced model. The derivation of the mathematical models for the various observation differences is therefore a useful means of examining these assumptions. (We also do so because in practice this is the most common way of dealing with clock biases. Note here that we are using the term "clock bias" as defined by GRANT et al, 1990, to refer to the linear biases of receiver and satellite clocks, $\varepsilon_j(T)$ and $\varepsilon^i(T)$, as well as the integer ambiguity term n_j^i.)

The clock biases are significant error sources and must be either estimated or eliminated in some way if survey or geodetic accuracies are to be achieved. The clock difference between satellite and receiver affects the phase observation directly and an error of 3 picoseconds in this clock difference will affect the GPS L1 carrier beat phase by 1mm (0.005 cycles). We do not have *a priori* information about the clock differences to this level of accuracy so we cannot eliminate the clock errors by means of an *a priori* model. If all the clock bias terms are included in the adjustment model to be estimated, the number of parameters will become unwieldy as the number of observation epochs increases. These are generally nuisance parameters as we are usually not interested in the clock behaviour itself but are only concerned with ensuring that it does not affect the geodetic parameters such as receiver positions or orbit parameters. A common technique in GPS adjustment therefore is to form differences of the observations to eliminate some or all of the unwanted clock biases.

Given that clock errors may be eliminated by differencing, it may be asked why spend so much time and energy defining and describing these errors? The reason is that on closer inspection we find that these errors are not always completely eliminated by differencing. If certain assumptions are made about the clock errors, based on our knowledge about the field procedures, the office processing, the receiver operation and clock stability, then the clock errors do effectively cancel. If these assumptions are not valid then the effect of the clock errors on the adjustment is not as well understood. For example, a breakdown in procedures or hardware may result in significant, non-cancelling, clock errors. In section 7.2 we describe several scenarios which may lead to the assumptions being invalid.

We also find that the time-tag errors in equation (5-21) are not eliminated by differencing. In many cases they will be known *a priori* to an accuracy sufficient to allow them to be eliminated from the model. Otherwise they must be estimated if the highest accuracy is to be achieved.

A close study of the assumptions often made is rewarding even though they are, in most cases, accepted as being true without question. It gives a greater insight to the processing options available for GPS observations. It provides an understanding of the reasoning behind recommended field and office procedures. It helps the analyst to determine what may have happened when things go wrong. In these cases it also allows the analyst to make the appropriate corrections to the data processing to eliminate or minimise the effect of the additional errors.

6.1.1 Forming Observation Differences: the Assumptions

The assumptions which are often made are:

(1) It is assumed that the observations from one receiver to different satellites at a measurement epoch were all made at the same time.

(2) It is assumed that the receiver clock is used to determine the time of observation and that all receivers have been set so as to take observations at the same nominal observation times. In other words, it is assumed that where the observations are not simultaneous, this is entirely due to receiver clock errors. This may not be the case when GPS receivers of different makes are combined in a campaign.

(3) It is assumed that the receivers have been somehow synchronised with each other so that the "true" observation times are close for all receivers. "Close" will be defined later in section 6.3.2.

(4) It is assumed that the receivers have been synchronised with respect to some standard time scale (or that the lack of synchronisation has been measured) so that the relationship between the observation time-tags, the ephemeris time-tags, and the time-tags of the earth rotation parameters can be specified in the adjustment. That is, it is assumed that the time-tag errors are so small that they have no effect on the adjustment, or that they have been largely eliminated by appropriate transformation of the Computational Time (as discussed in section 5.4.2).

(5) It is assumed that the satellite clocks are so stable that the clock errors do not change significantly between the various transmission times for signals which are all received at the same nominal observation time. This is discussed in section 6.3.1.

(6) It is assumed that the receivers have maintained lock on the satellite

signals and tracked the number of integer cycles so that the integer ambiguity is a constant for all observation epochs. That is, there are no cycle-slips.

These assumptions will be investigated in detail as they arise in the formulation of the mathematical models for the various differenced carrier beat phase observables. As well as identifying the assumptions we will try to identify the circumstances for which they are not valid. To do this we need to set a criterion for "validity". We will define an assumption as being invalid if the error in the observation which results from the assumption is a significant proportion of the measurement noise. Assuming that the measurement noise is several millimetres we have chosen 1mm to be a "significant proportion" of the noise. In calculating the significance of systematic errors we will consider the worst case scenario in order that we remain conservative when deciding whether the assumptions are valid.

6.2 Between-Satellite Differences

It is a well known that the receiver clock errors may be eliminated by forming the difference between phase observations from one receiver to two satellites at the same time. The operator ∇ indicates a **between-satellites difference.** Using Equation (5-19) with satellites 1 and 2 and receiver j we can write:

$$\nabla\phi_j^{12}(T_j) = \phi_{bj}^{1}(T_j) - \phi_{bj}^{2}(T_j) \quad (6\text{-}1)$$

where we have made the assumption (number (1) in section 6.1.1), that the two observations have been made at the same time T_j. We can write the observation based on this assumption as:

$$\begin{aligned}\nabla\phi_j^{12}(T_j) = &-(f_0/c)\,[\,\rho_j^{1}(t_j) - \rho_j^{2}(t_j) - te_j(t_j)\,(\,\dot{\rho}_j^{1}(t_j) - \dot{\rho}_j^{2}(t_j)\,)\,] \\ &+ f_0\,[\,\varepsilon^1(T_j^1) - \varepsilon^2(T_j^2)\,] \\ &+ n_j^1 - n_j^2 - \nabla\phi_{atmos} \end{aligned} \quad (6\text{-}2)$$

Note that the receiver clock error has been eliminated but that the time-tag error of the observation has not. The maximum possible value for the range-rate difference in equation (6-2) is twice that of a single range-rate (that is, 1520 m/sec, assuming range-rates of +760 m/sec and –760 m/sec, respectively for the two satellites in question). If we now assume that the time-tag error is less than 0.7 microseconds then it will have no significant effect on the differenced phase. This is assumption (4) of section 6.1.1, and leads to the familiar between-satellite difference observable:

$$\begin{aligned}\nabla\phi_j^{12}(T_j) = &-(f_0/c)\,[\,\rho_j^{1}(t_j) - \rho_j^{2}(t_j)\,] \\ &+ f_0\,[\,\varepsilon^1(T_j^1) - \varepsilon^2(T_j^2)\,] \\ &+ n_j^1 - n_j^2 - \nabla\phi_{atmos} \end{aligned} \quad (6\text{-}3)$$

We will now consider further our initial assumption. While it may be the intention of a receiver manufacturer to ensure that the observations to different satellites are all made at the same time, it may not always be possible in practice. For example, it has sometimes been found with certain GPS receivers, such as the Trimble 4000 series, that occasionally an observation to one satellite is made at a time which differs by 1 millisecond from the other observations. This is also true of WM101 observations made in the "raw" data format (MERMINOD, 1989). If it is not known that such time-tag offsets have occurred at each measurement "epoch", there will be an error introduced in the between-satellite difference which may need to be modelled.

When we introduced time-tag errors in section 5.4.1, we assigned an error to every receiver. We now see that it is possible to have an error for every observation. This error may have two effects. Firstly we can no longer use a single time-tag error, as in equation (6-2), secondly we have to look at whether the receiver clock errors at the different observation times do cancel.

The size of the first part of this error depends on the range-rate of the observed satellite. In the case of the "millisecond slip" alluded to above, the error could range from –0.76 to +0.76 metres, or approximately ±4 phase cycles. This is a significant error and needs correcting. Any difference between observation times less than 1.3 microseconds will be insignificant. If an error is known to have occurred we can correct the time-tag for that individual observation and use this corrected time-tag to calculate the satellite receiver range. Therefore the "millisecond slips" only present a problem if they have not been identified, or if the processing software used does not allow for a different time-tag to be assigned to every observation.

The size of the second part of the error depends on the stability of the receiver clock. Given a receiver clock stability of 1 part in 10^{10}, the observation times would need to differ by 30 milliseconds before a significant non-cancelling receiver clock error was introduced.

6.3 Between-Station Differences

The satellite clock errors may be almost eliminated by forming the difference between phase observation from two receivers to one satellite at the same time. The operator indicates a **between-station difference**. The observation times are given by the receiver clocks and include receiver clock errors. Therefore, although the nominal observation or reception times of the two measurements may be the same, the "true" time of measurement will differ from one receiver to the next.

If the nominal reception time is t then the true times of reception, T_1 and T_2, at receivers 1 and 2 are:

$$t = t_1 = t_2$$
$$T_1 = t - \varepsilon_1(T_1)$$
$$T_2 = t - \varepsilon_2(T_2)$$

and therefore

$$T_1 - T_2 = \varepsilon_2(T_2) - \varepsilon_1(T_1) \qquad (6\text{-}4)$$

The observation time difference (for observations that are supposed to be simultaneous) is thus equal to the difference in receiver clock errors. (This is assumption number (2) of section 6.1.1.) Care must though be taken to ensure that this is true when mixing GPS receivers of different brands.

We can now develop the observation equation for between-receiver differences. Using equation (5-21) with receivers 1 and 2 and satellite i we can write:

$$\Delta\phi_{12}{}^{i}(t) = \phi_{b1}{}^{i}(T_1) - \phi_{b2}{}^{i}(T_2) \qquad (6\text{-}5)$$

and thus

$$\Delta\phi_{12}^{i}(t) = -(f_0/c)\,[\,\rho_1^{i}(t_j) - \rho_2^{i}(t_j) - te_1(t_1)\,\dot{\rho}_1^{i}(t_1) + te_2(t_2)\,\dot{\rho}_2^{i}(t_2)\,]$$
$$+ f_0\,[\,\varepsilon^{i}(T_1^{i}) - \varepsilon^{i}(T_2^{i})\,]$$
$$- f_0\,[\,\varepsilon_1(T_1) - \varepsilon_2(T_2)\,]$$
$$+ n_1^{i} - n_2^{i} - \Delta\phi_{atmos} \qquad (6\text{-}6)$$

6.3.1 Satellite Transmission Times

The satellite clock errors still appear in this equation because the two observed signals were transmitted at different times. The term on the second line containing these clock errors describes the change in satellite clock error between the two transmission times. It is usually considered to be insignificant but we need to examine this (assumption number (5) of section 6.1.1).

Using equations (5-10) & (6-4) we can determine the relationship between the two transmission times:

$$T_1^{i} - T_2^{i} = T_1 - T_2 + \tau_2^{i}(T_2) - \tau_1^{i}(T_1)$$
$$= \varepsilon_2(T_2) - \varepsilon_1(T_1) + \tau_2^{i}(T_2) - \tau_1^{i}(T_1) \qquad (6\text{-}7)$$

The relationship between these times is illustrated in Figure 6.1, in which the more general situation of two receivers and two satellites is shown.

For simplicity in examining the likely magnitude of errors under this assumption we will represent the change in the satellite clock error between T_1^{i} and T_2^{i} by a linear drift (rate of change of clock error with time $\dot{\varepsilon}^{i}(T)$).

Note that this drift could include a jump or step in the satellite clock error between the two transmission times:

$$\varepsilon^{i}(T_1^{i}) - \varepsilon^{i}(T_2^{i}) = \dot{\varepsilon}^{i}(T)\,(\,T_1^{i} - T_2^{i}\,)$$
$$= \dot{\varepsilon}^{i}(T)\,(\,\tau_2^{i}(T_2) - \tau_1^{i}(T_1)\,)$$
$$+ \dot{\varepsilon}^{i}(T)\,(\,T_2 - T_1\,) \qquad (6\text{-}8)$$

The significance of this depends on two factors:

- the difference in transit times, and
- the difference in observation times.

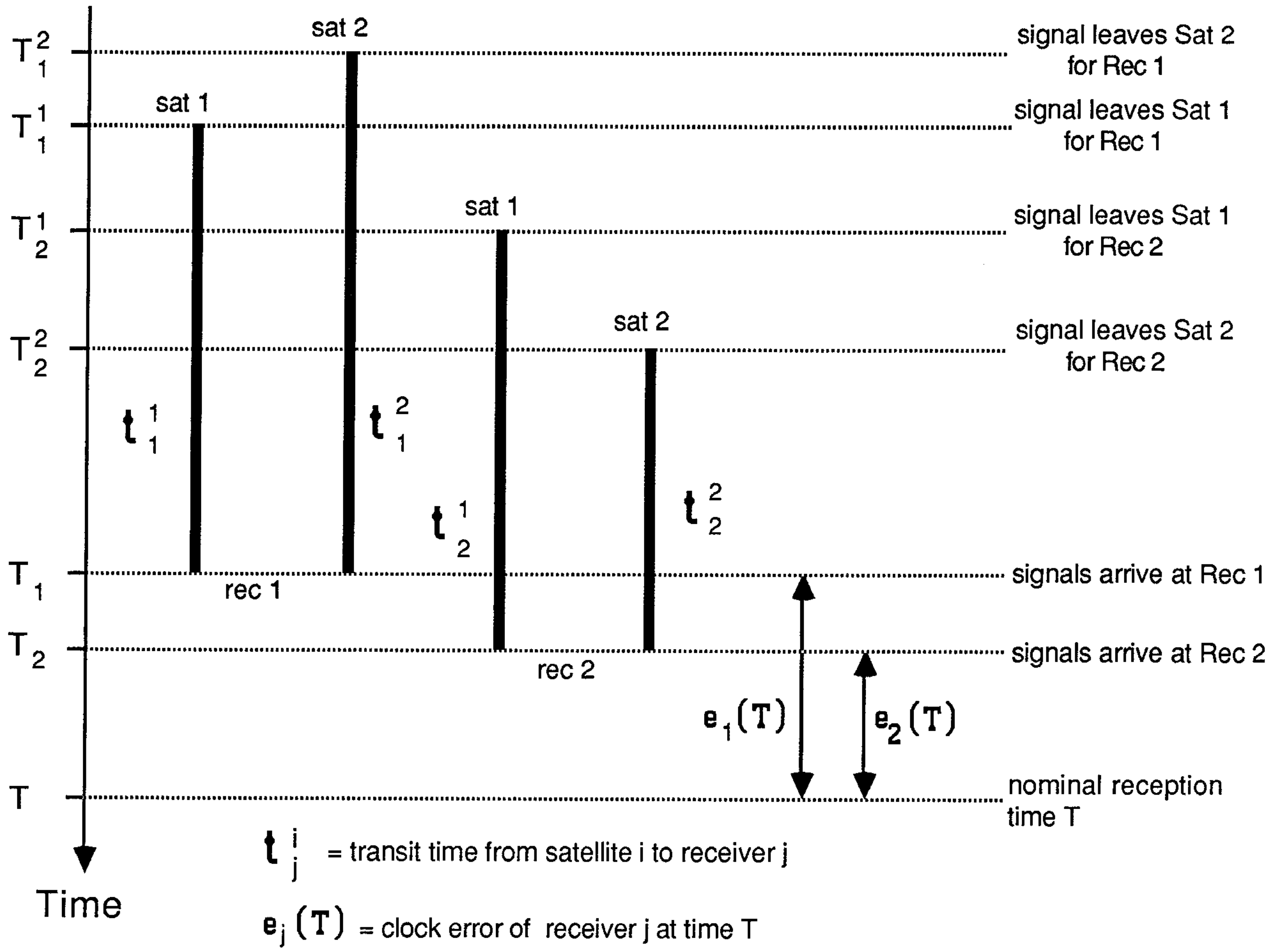


Figure 6.1 Relationships between Clock Errors, Time of Transmission and Time of Reception

The difference in transit times depends on satellite-receiver geometry. For a baseline of 300km in length this will be less than a millisecond (and will be even less for shorter baselines). If the cesium clocks in the satellites are behaving to specifications, we can expect the short term stability over a millisecond to be at least 1 part in 10^{10}. The error over a 1 millisecond interval will therefore be of the order of 0.1 picosecond, and thus insignificant. Unless we have reason to suspect high frequency noise in the satellite clocks with clock error changes of several picoseconds within a period of less than a millisecond (short term stability of only 1 part in 10^8) or a step in the clock error of several picoseconds then we may ignore the effect of transit time differences on the satellite clock error. We can thus approximate equation (6-8) by:

$$\varepsilon^i(T_1^i) - \varepsilon^i(T_2^i) \approx \dot{\varepsilon}^i(T)\,(T_1 - T_2)$$
$$\approx \dot{\varepsilon}^i(T)\,(\varepsilon_2(T_2) - \varepsilon_1(T_1)) \qquad (6\text{-}9)$$

We may now write the observation equation as:

$$\begin{aligned}\Delta\phi_{12}{}^{i}(t) = &-(f_0/c)\,[\,\rho_1{}^{i}(t_j) - \rho_2{}^{i}(t_j) - te_1(t_1)\,\dot{\rho}_1{}^{i}(t_1) + te_2(t_2)\,\dot{\rho}_2{}^{i}(t_2)\,] \\ &+ f_0\,\dot{\varepsilon}^{i}(T)\,(\,\varepsilon_2(T_2) - \varepsilon_1(T_1)\,) \\ &- f_0\,[\,\varepsilon_1(T_1) - \varepsilon_2(T_2)\,] \\ &+ n_1{}^{i} - n_2{}^{i} - \Delta\phi_{atmos} \qquad (6\text{-}10)\end{aligned}$$

If the observation time difference between receivers $(T_1 - T_2)$ or the clock error difference $(\,\varepsilon_2(T_2) - \varepsilon_1(T_1)\,)$ is less than 30 milliseconds then the satellite clock terms of equation (6-6) may be ignored if we assume a satellite clock short term stability of 1 part in 10^{10}. This is assumption (3) of section 6.1.1.

If the receivers have been synchronised accurately then observations at different receivers will be virtually simultaneous and we can use this to cancel some of the common errors such as those of satellite clocks. If, through faulty hardware or field procedures, the receivers make observations at significantly different times then this assumption must be re-examined. The required accuracy of synchronisation between receiver clocks of 30 milliseconds (resulting in an error of 1mm in the observation due to non-cancellation of satellite clock errors) should not be difficult to maintain in practice.

If the pseudo-range navigation solution is used to synchronise the receiver clocks (as is usually the case for code-correlating receivers), then we would expect that the clock error would be less than 1 microsecond (section 5.2). Under the policy of "selective availability", the system is degraded for non-authorised users, and the positional accuracy is expected to be ±100 metres. This still corresponds to a time error of the order of only 0.3 microseconds.

If the observation time differences are greater than 30 milliseconds, then either a higher noise level must be accepted or the satellite clock errors must be modelled in some way. One possibility is to use polynomials to model the satellite clocks. The use of polynomials has been discussed earlier and there are potential difficulties with this method. A polynomial which models the clock behaviour between measurement epochs for the whole of the observation session (up to an hour or more) may not adequately represent the short term behaviour between transmit times (microseconds to seconds). An alternative would be to use Doppler observations, if available, as they are direct observations of the instantaneous rate of change of phase and can be used to calculate what the phase observation would have been at a different time. However, higher order satellite clock error terms such as those introduced by Selective Availability will not be accounted for. The validity of this approach therefore depends on the accuracy of the Doppler observations, the magnitude of non-linear terms in the satellite clock error and the size of the time correction. If we assume that the satellite clock errors cancel (assumption number (5) of section 6.1.1) we obtain equation (6-11) below.

Eliminating the satellite clock errors we now have:

$$\Delta\phi_{12}^{i}(t) = -(f_0/c)\,[\,\rho_1^{i}(t_j) - \rho_2^{i}(t_j) - te_1(t_1)\,\dot{\rho}_1^{i}(t_1) + te_2(t_2)\,\dot{\rho}_2^{i}(t_2)\,]$$
$$- f_0\,[\,\varepsilon_1(T_1) - \varepsilon_2(T_2)\,]$$
$$+ n_1^{i} - n_2^{i} - \Delta\phi_{atmos} \qquad (6\text{-}11)$$

6.3.2 Time-Tag Errors

Let us now examine more closely the time-tag error terms in the above observation equation. We can write:

$$te_1\,\dot{\rho}_1 - te_2\,\dot{\rho}_2 = \tfrac{1}{2}(\,te_1 + te_2\,)(\,\dot{\rho}_1 - \dot{\rho}_2\,) + \tfrac{1}{2}(\,\dot{\rho}_1 + \dot{\rho}_2\,)(\,te_1 - te_2\,)$$

$$= te_{mean}\,(\,\dot{\rho}_1 - \dot{\rho}_2\,) + \rho_{mean}\,(\,te_1 - te_2\,) \qquad (6\text{-}12)$$

where we have simplified the notation. Here we have partitioned the time-tag error into two parts:

- a **common** time-tag error given by te_{mean} , and
- a **relative** time-tag error given by $(te_1 - te_2)$

The effect of these terms will be different. The range-rate will be similar at both receivers and so the mean range-rate will be much larger than the difference of the range-rates. The mean range-rate cannot be larger than the maximum range-rate of 760 m/sec. The magnitude of the range-rate difference has been found to be:

$$(\,\dot{\rho}_1 - \dot{\rho}_2\,) < L \times 1.9 \times 10^{-4}\ \text{m/sec}$$

where L is the length of the baseline in metres. For a 10km baseline the range-rate difference is therefore less than 2 m/sec.

The observation may therefore be nearly 400 times more sensitive to relative time-tag differences than to a common time-tag error, for a 10km baseline. While a relative time-tag difference of 1.3 microseconds may cause an observation error of 1mm, it will take a common time-tag error of at least 500 microseconds (0.5 milliseconds) to produce the same observation error. (This is dependent on the length of the baseline. A common time-tag error of 50 microseconds may cause a 1mm error in the observation of a 100km baseline.)

Figures 6.2 and 6.3 illustrate the effects of relative and common time-tag errors. In these diagrams we have assumed that the observations from the two receivers have the same time-tags and, for simplicity, that the transit times are the same from the satellite to the two receivers. Although, this last assumption is not strictly correct, we can accurately model the motion of the satellite during the time of transit, and it allows for a simplification of the diagrams. We also assume that the satellite ephemeris is free of error.

The points T1 and T2 in these diagrams refer to the true satellite positions at the time of transmission to receivers 1 and 2. These are both at the same point due to the assumption made above. C1 and C2 are the respective calculated positions.

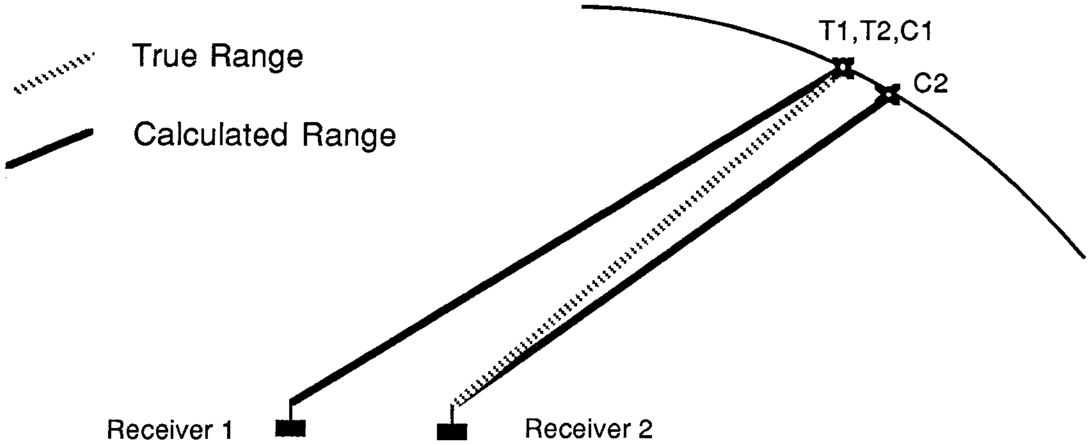


Figure 6.2 Relative Time-Tag Error

In Figure 6.2 we assume that there is no time-tag error at receiver 1 and a significant error at receiver 2. The calculated and true ranges from receiver 2 are different from each other, whereas those from receiver 1 are the same. This may arise from a poorly defined Computational Time based on corrected receiver clock time (section 5.4.2), or because the two receivers were not synchronised with each other.

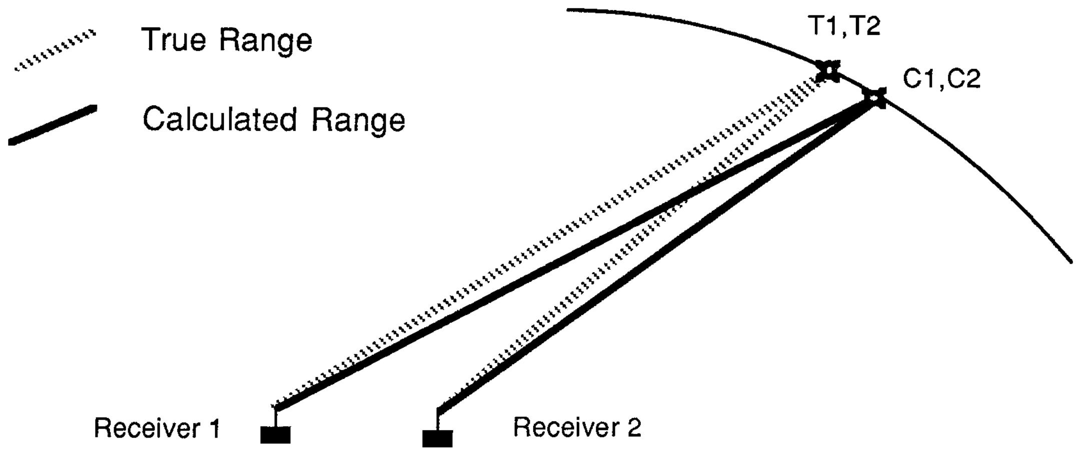


Figure 6.3 Common Time-Tag Error

Figure 6.3 illustrates the situation when there is an equal time-tag error at both receivers. There is a difference between calculated and true ranges for both receivers, this difference being nearly the same for each receiver.

There is another way of assessing the effect of a common time-tag error. It can be seen from Figure 6.3 that a time-tag error which is common to all receivers has the same effect as an alongtrack ephemeris error (this may arise from a poorly defined Computational Time based on corrected satellite clock time -- section 5.4.2). It is possible to have a time-tag error that is common to all receivers and all satellites.

We can use the rule-of-thumb commonly quoted to estimate the effect of ephemeris errors on adjusted baselines. If we divide the baseline length by the satellite altitude we get the ratio that propagates ephemeris error into the adjusted baseline. For example, if the ephemeris error is 20 metres (and given the satellite altitude of 20×10^6 metres), a 1 part per million (ppm) error will be introduced into the baseline. The satellites move 20 metres in approximately 5 milliseconds. If we are using an ephemeris with errors of 20 metres then a common error in the receiver time-tags of 1 or 2 milliseconds will not add substantially to the systematic errors which have already been introduced by an inaccurate ephemeris.

(It follows from this that it makes no sense to estimate a common time-tag error if orbital parameters are to be estimated in the adjustment. Any common time-tag error will be absorbed by the orbital parameters as an along-track shift in the satellite positions.)

When we difference between stations the intention is to eliminate the effect of satellite clock errors. We see above that this differencing also reduces the effect of common time-tag errors by about 2 or 3 orders of magnitude compared with the effect of the individual time-tag errors in the undifferenced model.

We can therefore deduce that when the satellite clock parameters are estimated, the clock error estimates also absorb most of the common part of the time-tag errors. Hence eliminating the satellite clock error parameters on an epoch-by-epoch basis eliminates this part of the time-tag errors. This is because according to the Fundamental Differencing Theorem, we may choose whether to reduce the effect of these common time-tag errors by differencing or by allowing the satellite clock error parameters to absorb their effects. Either way the adjustment is relatively insensitive to these errors insofar as they affect the estimates of the non-bias parameters.

If we assume that the common and relative time-tag errors are so small that they have no significant effect on the observation (assumption number (4) of section 6.1.1) then we can write the familiar simplified between-station observation equation:

$$\begin{aligned} \Delta\phi_{12}{}^i(t) = & -(f_0/c)\,(\,\rho_1{}^i(T_1) - \rho_2{}^i(T_2)\,) \\ & - f_0\,[\,\varepsilon_1(T_1) - \varepsilon_2(T_2)\,] \\ & + n_1{}^i - n_2{}^i - \Delta\phi_{atmos} \end{aligned} \qquad (6\text{-}13)$$

In this equation the satellite errors and time-tag errors have been completely removed and the receiver clock errors appear as linear biases.

6.4 Between-Epoch Differences

Parameters which are constant with time may be eliminated by forming the **between-epoch difference**. The operator δ is used to represent this operation. This observation is formed by taking the difference between observations made from a receiver j and a satellite i at two different epochs. It is not generally used in GPS data processing because the number of parameters that are removed (the ambiguities) is relatively small while the number of clock biases remaining (the clock errors) is large. The change of carrier beat phase with time is however equivalent to the doppler frequency which is the observable used in TRANSIT data processing.

If we assume that the integer ambiguity parameters are independent of time then they cancel out when we form this differenced observation. T_1 and T_2 are the two "true" times of reception at receiver j , and $T_1{}^i$ and $T_2{}^i$ are the two transmission times from satellite i :

$$\delta\phi_j{}^i(T_{12}) = \phi_{bj}{}^i(T_1) - \phi_{bj}{}^i(T_2)$$

$$\begin{aligned} &= -(f_0/c)\,[\,\rho_j{}^i(t_1) - \rho_j{}^i(t_2)\,] \\ &\quad + (f_0/c)\,[\,te_j(t_1)\,\dot{\rho}_j{}^i(t_1) - te_j(t_2)\,\dot{\rho}_j{}^i(t_2)\,] \\ &\quad + f_0\,[\,\varepsilon^i(T_1{}^i) - \varepsilon^i(T_2{}^i) - \varepsilon_j(T_1) + \varepsilon_j(T_2)\,] \\ &\quad - \delta\phi_{atmos} \end{aligned} \qquad (6\text{-}14)$$

If the receiver has not maintained lock on the satellite signal there may be different integer ambiguity terms for following observations. Our ability to detect this cycle-slip depends on how well we can model the change in the other terms between epochs.

The effect of the time-tag errors will depend on how much they and the range-rates have changed between epochs. Which itself will depend in part on the length of time between epochs used to form the difference. The length of time between epochs could be equal to the length of the entire observing session, up to several hours! Using the technique of equation (6-12), we find that the total effect is equal to the mean time-tag error multiplied by the change in range-rate **plus** the mean range-rate multiplied by the change in time-tag error.

The maximum possible range-rate is 760 m/sec and the maximum possible change in range-rate is 1520 m/sec if the satellite is tracked from rise to set. Under these circumstances the magnitude of time-tag errors that can be therefore considered insignificant are those less than 0.7 microseconds (average) and 1.3 microseconds (change in time-tag error).

6.5 Double-Differences

This is the term usually applied to the observation which has been formed by differencing between both satellites and receivers. This double-difference phase may be created by either forming the between-satellite difference phase and then differencing between stations, or by first forming the between-station difference and then differencing between satellites:

$$\begin{aligned}\phi_{DD}(t) &= \nabla\Delta\phi_{12}{}^{12}(t) = \Delta\nabla\phi_{12}{}^{12}(t) \\ &= \nabla\phi_1{}^{12}(T_1) - \nabla\phi_2{}^{12}(T_2) \\ &= \Delta\phi_{12}{}^1(t) - \Delta\phi_{12}{}^2(t) \qquad (6\text{-}15)\end{aligned}$$

We have derived a variety of between-stations difference and between-satellites difference equations that can be used to form this double-difference observation equation, depending on the assumptions made. For example, we can use equations (6-6), (6-10), (6-11), or (6-13) from the between-station difference equations. We will start from equation (6-6):

$$\begin{aligned}\phi_{DD} = &- (f_0/c)\,[\,\rho_1{}^1(t) - \rho_1{}^2(t) - \rho_2{}^1(t) + \rho_2{}^2(t)\,] \\ &+ (f_0/c)\,[\,te_1(T_1)\,(\,\dot{\rho}_1{}^1(t) - \dot{\rho}_1{}^2(t)\,) - te_2(T_2)\,(\,\dot{\rho}_2{}^1(t) - \dot{\rho}_2{}^2(t)\,) \\ &+ f_0\,[\,\varepsilon^1(T_1{}^1) - \varepsilon^1(T_2{}^1) - \varepsilon^2(T_1{}^2) + \varepsilon^2(T_2{}^2)\,] \\ &+ n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2 - \nabla\Delta\phi_{atmos} \qquad (6\text{-}16)\end{aligned}$$

We have assumed once again that for each receiver the observations to the two satellites are made at the same time. If this assumption is incorrect then the observation may be affected in the same way as was found with the between-satellite difference observation equation developed in section 6.2.

If the criteria given in section 6.3.1 for the assumption that satellite clock errors cancel are satisfied then we may use equation (6-11) to form the double-difference observable:

$$\begin{aligned}\phi_{DD} = &- (f_0/c)\,[\,\rho_1{}^1(t) - \rho_1{}^2(t) - \rho_2{}^1(t) + \rho_2{}^2(t)\,] \\ &+ (f_0/c)\,[\,te_1(T_1)\,(\,\dot{\rho}_1{}^1(t) - \dot{\rho}_1{}^2(t)\,) - te_2(T_2)\,(\,\dot{\rho}_2{}^1(t) - \dot{\rho}_2{}^2(t)\,) \\ &+ n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2 - \nabla\Delta\phi_{atmos} \qquad (6\text{-}17)\end{aligned}$$

We next make the assumption that time-tag errors have an insignificant effect on the observation by heeding the requirements of section 6.3.2, namely that relative time-tag errors are less than 1 microsecond and that common time-tag errors are less than 1 or 2 milliseconds. We can then use equation (6-13) to form the familiar simplified double-difference observation equation:

$$\phi_{DD} = -(f_0/c)\,[\,\rho_1{}^1(t) - \rho_1{}^2(t) - \rho_2{}^1(t) + \rho_2{}^2(t)\,] + n_1{}^1 - n_1{}^2 - n_2{}^1 + n_2{}^2 - \nabla\Delta\phi_{atmos} \qquad (6\text{-}18)$$

The only clock biases remaining in this equation are the integer ambiguities. The validity of this equation is therefore dependent on how close to reality are the assumptions (1) to (5) of section 6.1.1.

6.6 Triple-Differences

This observation could also be referred to as a double-difference rate. It is the difference between double-differences formed at two epochs.

We could use equations (6-16), (6-17) or (6-18) to form this triple-difference observation equation. The principles behind differencing and the effect of the various assumptions made have been covered above, so we will use equation (6-18) together with the common assumption that the integer ambiguities are independent of time to derive the simplified triple-difference observation equation:

$$\begin{aligned}\phi_{TD}(t_{ab}) &= \delta\Delta\nabla\phi_{12}{}^{12}(t_{ab}) \\ &= \phi_{DD}(t_a) - \phi_{DD}(t_b) \\ &= -(f_0/c)\,(\,\rho_1{}^1(T_{a1}) - \rho_1{}^2(T_{a1}) - \rho_2{}^1(T_{a2}) + \rho_2{}^2(T_{a2})\,) \\ &\quad + (f_0/c)\,(\,\rho_1{}^1(T_{b1}) - \rho_1{}^2(T_{b1}) - \rho_2{}^1(T_{b2}) + \rho_2{}^2(T_{b2})\,) \\ &\quad - \delta\Delta\nabla\phi_{atmos} \qquad (6\text{-}19)\end{aligned}$$

This equation is based on assumptions (1) to (6) of section 6.1.1. Note that all the clock bias terms, including the integer ambiguities, have been eliminated.

If our assumption that the integer ambiguity is constant in time is false then this observation will be affected. The triple-difference observation is especially useful in this regard as not only have the unknown clock biases been removed, but also the effect of errors in the other terms such as atmospheric delay, ephemeris error, receiver coordinate error, are substantially reduced by triple-differencing. Especially if the epochs are only a few seconds or minutes apart.

The potential error of several cycles caused by the "millisecond slip" (section 6.2) may have the same appearance as a cycle-slip except that the error will not be an integer.

7. SYSTEMATIC CLOCK ERRORS AND THEIR IMPACT ON GPS PHASE DATA SOLUTIONS

In the preceding sections we have examined more closely the assumptions made in section 6.1.1 and have investigated the conditions required for these assumptions to be valid. We noted earlier that the Fundamental Differencing Theorem tells us that the results from a triple-difference adjustment, where all the linear clock biases have been eliminated, is the same as that from undifferenced phase solution in which they are all explicitly estimated.

In other words, the assumptions which were required to form the triple-difference equation (6-19) are made implicitly when estimating the parameters using the undifferenced phase observation. To summarise the results of the previous section we will state these assumptions again here and define the conditions that are required for us to be sure that the assumptions are valid. The criteria by which we judge whether assumptions are valid is if they introduce less than 1mm error in the observation modelling.

7.1 Clock Errors and GPS Phase Observation Modelling

Assumption1

We have assumed that all observations to the satellites from a receiver were made at the same time. If the observations are made within 1 microsecond of each other then we can regard them as being simultaneous. If the difference in time between observations is greater than 1 microsecond then we need to specify a different time-tag for every observation rather than a common time-tag for all of them.

If we do have a different time-tag for each observation then we can model the range correctly. However, if the times of observation differ by more than 30 milliseconds, and the receiver clock is only stable to 1 part in 10^{10}, then it may not be valid to assume one receiver clock error in the observation model that is common to all the observations. In other words, between-satellite differencing may not completely cancel the receiver clock error.

Assumptions 2 and 3

We have assumed that the receiver clock determines the observation time and that all receivers have been set by the operators to take observations the same rate. In other words we assume that the nominal observation times are identical and that the actual observation times differ only by the receiver clock errors. The most significant timing problem to be faced by an analyst these days is likely to arise from mixing data from different brands of receivers, where the nominal observation time-tags (as opposed to observation time-tags corrupted by receiver clock error) may be significantly different.

We have also assumed that the receiver clocks have been synchronised so that the differences in receiver clock error are sufficiently small for us to be able to say that observations at different receivers occur at the same time. This assumption is tied up with assumption 2 above. (Note, however, that due to field procedure problems, different receivers may make observations at different times even though their clocks are synchronised!)

If there is a difference in observation times greater than 30 milliseconds then the satellite clock errors may not be the same for two observations to the same satellite, at the same nominal observation time.

Assumption 4

We have assumed that the receiver clocks, which have been synchronised with each other, are also linked in some way with a standard time scale so that we can remove the time-tag errors between the receiver clocks and satellite ephemeris time.

If there is a difference in time-tag errors between receivers greater than 1 microsecond then it may be necessary to estimate this time-tag error difference in the adjustment.

The limitation on common time-tag error which meets the same standard as the other assumptions here (introducing less than 1mm error in the observation) depends on the baseline length. It is, for example, 0.5 milliseconds for a 10km line. However we have also noted that the effect of this error on the baseline is likely to be only 1ppm for 4 or 5 milliseconds in the common time-tag error.

Assumption 5

It is assumed that the satellite clocks are stable to at least 1 part in 10^{10} so that the satellite clock errors are the same at the various transmission times that make up one epoch (as defined by the various ground receivers operating).

This assumption is tied up with assumptions 2 and 3. The figure of 30 milliseconds between observation times is based on the assumed stability of the satellite clock. It may need to be reconsidered if "satellite clock dithering" occurs under "selective availability" conditions.

Assumption 6

We have assumed that the receivers maintain lock on the signal and that the integer ambiguities are independent of time. This can be tested by forming differences between epochs and comparing the observed difference with the modelled value. Triple-differencing is especially useful here as the other systematic errors in the model are minimised or eliminated by the triple-differencing operation.

7.2 Practical Considerations for GPS Surveys

We have seen that while the observations are sensitive to differences in the clock errors between satellite and receiver of several picoseconds, they are only sensitive to microsecond errors in the time-tag difference between receivers and millisecond time-tag errors that are common to all receivers.

The velocity of the carrier signal between satellite and receiver is the speed of light, approximately 300000km/sec, whereas the maximum velocity of the satellite towards the receiver is approximately 0.7km/sec. This is the reason for the difference of 5 to 6 orders of magnitude between the sensitivity of the observations to clock errors and time-tag errors. Also the adjustment of GPS carrier phase observations gives results for receiver coordinates which have relative precisions about 3 orders of magnitude greater than the absolute precisions (depending on the length of the line). This is also true of the estimates of the time-tag errors. The relative time-tag errors may be determined to a higher precision than the absolute or common time-tag error. The difference between absolute and relative precision comes from the geometry. The altitude of the satellites is about 3 orders of magnitude greater than the length of a typical GPS baseline (20km).

KING et al (1987) state that receiver clocks should be synchronised to within about 1 microsecond of each other and a few milliseconds relative to some time standard such as UTC. We have demonstrated in section 6, the reasoning behind these recommendations.

For GPS adjustments of the highest possible accuracy, we will extend the recommendations as follows:

(1) The observations at different receivers should be taken within 30 milliseconds of each other to ensure that satellite clock errors can be cancelled (or estimated directly without the need for polynomials). This will be satisfied if the receiver clock differences are kept to less than 30 milliseconds. If a survey is to be conducted using receivers of different manufacture then a problem could arise with different observation times even where the clock differences (synchronisation errors) are small.

(2) The differences between the clocks generating the time-tags (receiver clocks) should be known to better than 1 microsecond to ensure that time-tags can be adjusted, to keep the relative time-tag error at less than 1 microsecond. If it is known, or suspected, that the relative time-tag errors may be larger than this then they should be estimated in the adjustment.

(3) That part of the time-tag error that is common to all receivers should be kept to within a few milliseconds to keep the baseline errors to less than 1 ppm. This will usually involve measuring the difference with respect to some standard time scale such as UTC or GPST and then ascertaining the ephemeris time scale. If it is known, or suspected, that the difference is larger than this then it should be estimated (though not if the orbits are also adjusted).

If the pseudo-range solution is used to correct the receiver clocks (section 5.3) and the broadcast ephemeris is used (as will usually be the case with code-correlating receivers) then the above conditions should be satisfied without further action. Even in this case it is still possible that a time-related error may occur. Five scenarios are described below where the analyst may need to consider the effect of clock errors, even though they have been supposedly eliminated by double-differencing.

<u>Scenario 1</u>
A receiver is used in which the pseudo-range solution is used to correct the receiver clock. The survey begins when only two or three satellites are visible. (These observations can still contribute to the solution. Four satellites are only needed for an instantaneous pseudo-range solution.) The operator in entering the station coordinates makes a large error, say 1°, in the latitude. There is not enough information to calculate position and time from two or three satellites and so the pseudo-range solution is constrained by the false station coordinates resulting in a false clock offset solution. The calculated clock offsets, which are used to "correct" the receiver clock, change as the geometry changes and only

jump to the correct value when four satellites become visible. This problem can be corrected by post-processing software provided that the software has been written to cover such a contingency and provided the observation times are within 30 milliseconds of each other.

Scenario 2
The receiver clocks have all been correctly synchronised to GPST by one means or other and the analyst, rather than using the broadcast ephemeris, obtains an ephemeris from some other agency without realising that it is in terms of some other time scale such as UTC or TAI. The resulting common time-tag error of several seconds will distort the adjustment.

Scenario 3
The receiver, due to a quirk in the internal signal processing, occasionally makes observations to different satellites at slightly different times without attaching the correct time-tag to this observation. If it is not known that this has occurred, an error is introduced into the observation which depends on the range-rate of the satellite times the error. It is not removed by differencing. In the case of the "millisecond" slip this error can be up to ±4 cycles. This problem can be overcome if the software is written to handle it.

Scenario 4
A "codeless" receiver is used and the receiver clocks are synchronised to each other and to a standard time scale such as UTC. The differences are later measured to determine the clock drift but an error is made in the sign of the differences and they are applied the wrong way in the data processing, resulting in a significant time-tag error. If the time-tag error is estimated the mistake can be identified and corrected.

Scenario 5
Receivers of different makes which observe at different nominal times are used in a GPS survey. The analyst must account for the possibility that relative time-tag errors may be of the order of a fraction of a second (or several seconds in the case of observation epochs taken at intervals of greater than a second). It may be possible to "correct" the data to the relevant epochs (if the interval is short and Doppler measurements are available). More commonly, a relative time-tag correction would need to be estimated as part of the adjustment process.

8. ACKNOWLEDGEMENTS

The second author (D.B.G.) was supported in his PhD studies at the School of Surveying, University of New South Wales, by a study award from the Department of Survey and Land Information, New Zealand. The authors wish to acknowledge the important contributions made by Wan Xuan Fu and Pen-Shan Hung on the topic of clock "performance".

9. REFERENCES

ABBOT, R.I., CEFULA, P.J., & TSE, S.F., 1983. Satellite orbital theory for a small computer. **Astrodynamics, 54**, 543-572.

ASHBY, N., 1987. Relativistic effects in the global positioning system. Proc. IAG Symposia: Relativistic Effects in Geodesy, Advances in Gravity Field Modelling, and Analysis of Satellite Altimetry, Vancouver, Canada, Aug. 10-22, 1987, 41-50.

BAUSTERT, G., HEIN, G.W. & LANDAU, H., 1989. On the use of GPS in airborne photogrammetry, hydrographic applications, and kinematic surveying. Proc. 5th Int. Geod. Symp. Sat. Doppler Pos., Las Cruces, New Mexico, Mar. 13-17, 1989, 1029-1040.

CONLEY, R., 1984. Reference frames, earth models and GPS. GPS JPO, Directorate of Systems Engin., YEEE 84-040, 137pp.

DELIKARAOGLOU, D., 1987. On principles, methods and recent advances in studies towards a GPS - based control system for geodesy and geodynamics. Preprint, NASA Tech. Memo. 100716, Goddard Space Flight Center, Greenbelt, Md., U.S.A., 105pp.

GRANT, D.B., RIZOS, C., & STOLZ, A., 1990. Dealing with GPS biases: some theoretical and software considerations. In Unisurv S-38, report of the School of Surveying, University of N.S.W.

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**, 39-51.

HELLWIG, H., 1979. Microwave time and frequency standards. **Radio Science, 14(4)**, 561-572.

JONES, R.H., & TRYON, P.V., 1987. Continuous time series models for unequally spaced data applied to modeling atomic clocks. **SIAM J. Sci. Stat. Comput., 8(1)**, 71-81.

KAPLAN, 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.

KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A., & COLLINS, J., 1987. **Surveying with GPS**. Dümmler, Bonn, 128pp.

KOVALEVSKY, J., MUELLER, I.I., & KOLACZEK, B., (eds.), 1989. **Reference Frames in Astronomy and Geophysics**. Kluwer Academic Publishers, 474pp.

LANDAU, H., 1988. Zur Nutzung des Global Positioning Systems in GeodLANDAU, H., 1988. Zur Nutzung des Global Positioning Systems in Geodäsie und Geodynamik: Modellbildung, Softwareentwicklung und Analyse. Rept. 36, Inst. of Astro. & Physical Geodesy, Univ. FAF, Munich, F.R.G., 284pp.

MERMINOD, B., 1989. The use of Kalman filters in GPS navigation. Unisurv S-35, report of the School of Surveying, University of N.S.W., 203pp.

NEILAN, R.E., & MELBOURNE, W.G., 1989. GPS global tracking system in support of missions to the planet earth. CSTG Bull. 11, Int. Coord. of Space Techniques for Geodesy & Geophysics, 129-140.

REMONDI, B.W., 1984. Using the Global Positioning System (GPS) phase observable for relative geodesy: modeling, processing, and results. Rept. CSR-84-2, Center for Space Research, University of Texas at Austin. 360pp.

ROCKEN, C., 1988. The Global Positioning System: a new tool for tectonic studies. PhD thesis Univ. of Colorado, 265pp.

RUTMAN, J., 1978. Characterization of phase and frequency instabilities in precision frequency sources: fifteen years of progress. Procs. IEEE, 66(9),

1048-1075.
STOLZ, A., & MASTERS, E.G., 1983. Satellite laser ranging measurements of the 3200km Orroral-Yaragadee baseline. **Aust.Surveyor**, **31**, 557-562.
SWIFT, E.R., 1985. NSWC's GPS orbit/clock determination system. Proc. 1st Int. Symp. Prec. Pos. GPS, U.S. Dept. of Commerce, NOAA, Rockville, Md., 15-19 April, 1985, 51-62.
VAN DIERENDONCK, A.J., RUSSELL, S.S., KOPITZKE, E.R., & BIRNAUM, M., 1980. The GPS navigation message. Special issue of **Navigation** on GPS, 55-73.
WÜBBENA, G., 1988. GPS carrier phases and clock modeling. In **GPS-Techniques Applied to Geodesy and Surveying**, E.Groten & R.Strauss (eds.), Lecture notes in Earth Sciences, 19, Springer-Verlag, 381-392.
ZHU, S.Y. & GROTEN, E., 1988. Relativistic effects in GPS. In **GPS-Techniques Applied to Geodesy and Surveying**, E.Groten & R.Strauss (eds.), Lecture notes in Earth Sciences, 19, Springer-Verlag, 41-46.

RESOLUTION OF THE CYCLE AMBIGUITIES

B. Merminod[1]

School of Surveying
University of New South Wales
P.O.Box 1
Kensington, N.S.W., 2203, Australia

ABSTRACT

For short baselines, the reliable resolution of the integer cycle ambiguities appears to be the key to precise relative GPS positioning. However, in order to maximise the conditions for this to occur, phase observations should be made during sessions that involve significant changes in the receiver-satellite geometry. This condition implies that longer observing sessions are preferable, which, in turn, can be a strong economic limitation in the use of GPS satellite surveying technology.

How is receiver-satellite geometry quantified, and hence the conditions for ambiguity resolution? At present, the satellite geometry indicators commonly available - the well known DOP factors (Dilution Of Precision) - refer to the instantaneous pseudo-range solution for the receiver coordinates and reflect the processes involved in the use of the carrier phase only in a very limited manner. This report describes a set of alternative precision indicators closely related to the geometrical variations of the receiver-satellite geometry and the special properties of phase measurements.

These indicators, also presented in the form of DOP factors, refer to the three following cases:

- precision of the coordinates, when cycle ambiguities are also present as parameters (in ambiguity-free solutions), or when they are absent altogether (as in the case of triple differences solutions.
- precision of the determination of the ambiguities.
- precision of the coordinates, once the ambiguities are fixed to integer values (as in ambiguity-fixed solutions).

This report describes the characteristics of these BDOP (Bias DOP) factors and discusses how they can be used to assist in survey planning: the selection of observation window, session length and the satellites to be tracked.

[1]Now at the Dept. of Land Survey and Physical Planning, Maseru, Lesotho.

1. INTRODUCTION

1.1 Background Remarks

The optimum selection of satellites to be tracked has been addressed in many papers about GPS. The techniques are all derived from the strength-of-solution analyses of GPS pseudo-range navigation. However, the systematic treatment of the cycle ambiguities (and hence GPS phase observations) has tended to be avoided for two main reasons:

- Most GPS users are interested in the navigation capabilities of the system. Phase measurements and hence cycle ambiguities are completely irrelevant to them.

- Geodesists are interested in the best possible accuracy in the coordinates. Thus, resolving the cycle ambiguities is only an unfortunate duty (one more) on the way to satisfying their primary goal.

In general, even if geodetic applications are considered, the usual approach is to assume that the cycle ambiguities have somehow already been resolved to their integer values (and hence transforming ambiguous phase observations into unambiguous range observations). Attention is then concentrated on studying factors limiting accuracy, such as atmospheric refraction, multipath and ephemeris errors (see, for example, LANDAU & EISSFELLER, 1985).

Occasionally apparently good observation sessions, with reasonable satellite coverage and of respectable duration, lead to poor results because the cycle ambiguities cannot be resolved. In the absence of other information, the standard response is to recommend longer observation sessions, in the hope that there will be enough data to improve the quality of the cycle ambiguity estimates, and hence optimise the chances of resolving them.

For survey management authorities, the situation is worse. As soon as they try to edict recommendations concerning the duration of GPS sessions, to satisfy some prescribed survey accuracy standard, there arises a case where the suggested requirements were fulfilled and unsatisfactory results were still obtained. Consequently, the recommended practices and specifications for GPS surveys are made even more conservative.

If the duration of the observations is to be kept as short as possible, resolving the cycle ambiguities is the most stringent limitation in the use of GPS for precise small scale surveys. There are techniques to evaluate the suitability of a given satellite coverage and the probability of resolving the cycle ambiguities. The aim of this report is to describe the derivation of indicators of precision suitable for GPS surveying. These indicators have been developed with the following guidelines in mind:

- They should represent the true geometric strength of the observations performed in as "condensed" a form as possible (preferably as one number).

- Their expression must remain general, valid for all commonly available GPS measuring systems.
- The algorithms should be portable, and suitable for implementation on PC computers, as used for routine GPS data processing.

These indicators can then assist in the selection of observing time and session duration.

1.2 Cycle Ambiguities in GPS Processing

1.2.1 Why Cycle Ambiguities Occur

The GPS carrier phase observation (on L1 or L2 wavelengths) contain, in effect, a fraction of the true range between the satellite and the receiver (KING et al, 1987).

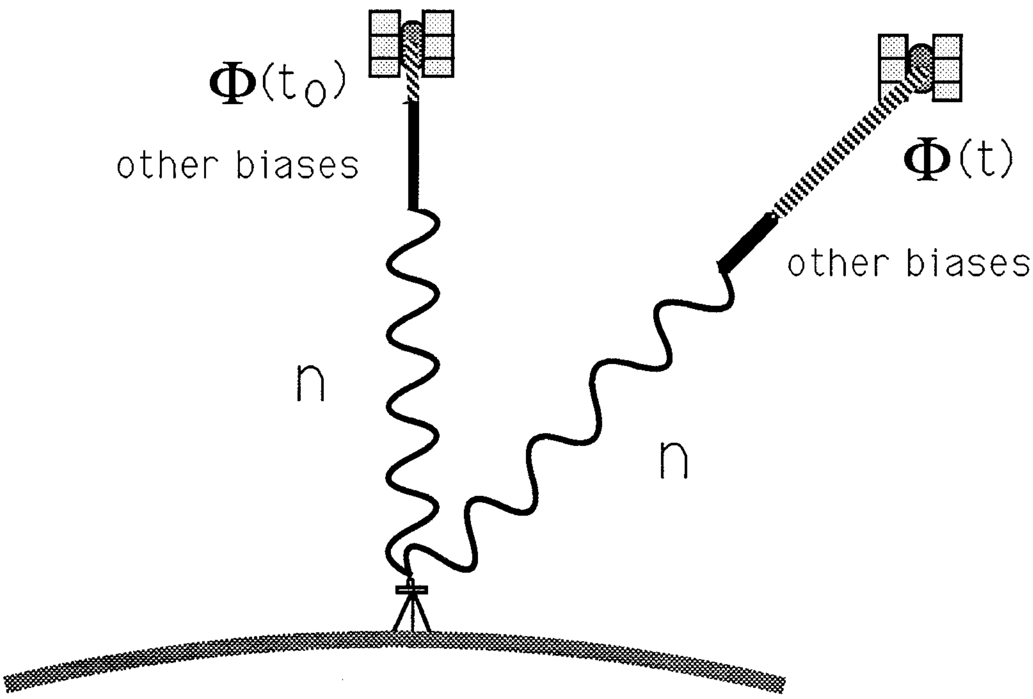


Figure 1: Integrated carrier beat phase observations

The difference between the observed carrier phase and the receiver-satellite range is affected by several **biases** (IBID, 1987). In particular, for GPS surveying instrumentation, the initial integer number of wavelengths ("n") is unknown because its value cannot be measured directly. Such observations are known as "integrated carrier beat phase" measurements (Figure 1). In other words: changes in range (or range differences) and not ranges are observed. The phase measurements are therefore "**ambiguous**" ranges. Note that if carrier beat phase observations were not "accumulated" or "integrated", by measuring the change in phase (or range) from the epoch of initial signal lockon, the value of n would change with each epoch. This would be a next to useless observable type. The value of this unknown initial (integer) bias is an important task of the GPS data processing software. Such a solution is known as an "ambiguity-free" solution.

At this stage, no use of the integer nature of n can be made as it is indistinguishable from other, non-integer biases such as satellite and receiver oscillator instabilities and atmospheric refraction, and in fact is contaminated by them. Thus, bias "fixing", or "ambiguity resolution" as it is more correctly known, is only possible after all biases are eliminated or otherwise accounted for.

1.2.2 Why Cycle Ambiguities Must be Resolved

Phase observations can be modelled as (in units of cycles):

$$\Phi(t) = (f_0 / c) \cdot \rho(s,r,t) - n \qquad (1\text{-}1)$$

where f_0 is the nominal frequency of the signal, c is the velocity of electromagnetic radiation, and $\rho(s,r,t)$ is a function of time, satellite and receiver positions. $\rho(s,r,t)$ is a range, which may be contaminated by a number of biases. If this observation is related with a previous one at time t_0, a range difference is obtained. For our purpose, the main interest of the model is the ability to isolate the integer unknown n with a sufficient number of observations. According to HATCH (1987), resolving the ambiguity in effect means shifting the integer term to the left hand side of the above equation. Thus, it becomes an *a priori* known part of the observation:

$$\Phi(t) + n = (f_0 / c) \cdot \rho(s,r,t) \qquad (1\text{-}2)$$

This is nothing other than a range observation equation. The number of equations is the same as before, but the number of unknowns is smaller: one less bias (n) for each observing satellite-station combination. These two types of observations (equations (1-1) and (1-2)) have quite different properties. For short observing sessions, the geometric strength of the second type of observation (and hence the solution derived from it) is greater.

1.2.3 Why New Indicators of Precision are Needed

"Indicators of precision" are commonly computed by the field planning components of most commercial GPS software packages. However, the indicators presently used were developed for use with instantaneous measurements of the time delay of a signal. This time delay, scaled by the velocity of light, gives a measure of the range between the satellite and the receiver, biased by possible clock errors: the **pseudo-range**. These indicators, principally GDOP, PDOP and TDOP (their derivation is given in, for example, LANDAU & EISSFELLER, 1985) are not suitable for describing the precision of coordinates determined from phase observations, with or without cycle ambiguity resolution, because they refer to (MERMINOD et al, 1990):

- Range measurements to the satellites. However the determination of the cycle ambiguities depends upon range difference measurements. (There would be no ambiguities if range observations were available!)

- The instantaneous solution for positioning. However the resolution of the ambiguities requires observations over some period of time, as the ambiguous range must reflect varying receiver-satellite geometry.

- Single receiver operation, as is the norm for GPS navigation, as opposed to the multi-receiver mode of operation in GPS surveying.

NORTON (1987) reports on a test of cycle ambiguity fixing for different GPS sessions and tried to relate it to the GDOP factor. It was found that apparently a reliable solution for the ambiguities is obtained when the GDOP is either constantly low or rapidly changing, irrespective of its value. This indicated that there was no general correspondence between GDOP and the conditions for reliable estimation of the ambiguities.

Before proceeding to derive alternative DOP factors, we need to investigate the nature of the mathematical model of phase observations, and various linear combinations commonly used in GPS data reduction algorithms. This is essential as DOP factors based on phase observations must simulate the outcome of a GPS phase data reduction.

2. MODEL OF THE OBSERVATIONS

2.1 Integrated Carrier Beat Phase

2.1.1 Functional Model

For the following discussions, the classical observation equation of the integrated carrier beat phase is taken as the starting point (GRANT et al, 1990):

$$\phi_r^s(t) = \frac{1}{\lambda}\sqrt{[\mathbf{X}^s(t) - \mathbf{X}_r]^2} + f_0\,[\varepsilon_r(t) - \varepsilon^s(t)] - n_r^s + (\phi_{atmos})_r^s \qquad (2\text{-}1)$$

where f_0 is the nominal frequency of the carrier wave (L1=1575.42 MHz, L2=1227.60 MHz). $\mathbf{X}$ is the position vector of either the satellite or the receiver (depending on the superscript or subscript). The inverse of the wavelength $1/\lambda$ converts the range between satellite and receiver, usually expressed in metres, to cycles of the carrier wave. All time dependent terms in the equation are expressed in the same time frame, referred to here as true time. $\varepsilon^s(t)$ and $\varepsilon_r(t)$ are, respectively, the satellite and receiver oscillator (clock) errors at each epoch (observed - true time, here in seconds). n_r^s is the initial number of cycles of the carrier phase and $(\phi_{atmos})_r^s$ is the phase delay due to atmospheric refraction. No cycle slips are assumed to be present.

To isolate n_r^s and fix it to an integer value, more assumptions are required:

1. The receiver and satellite clock errors are somehow accounted for (at better than one wavelength: L1 is approximately 19cm, L2 approximately 26cm), or entirely eliminated by data differencing.

2. "Short" baselines are considered, so that environmental influences affect the measurements almost equally at two GPS receivers. For any satellite, the difference between $(\phi_{atmos})_r^s$ at both receiver sites is a small fraction of a wavelength. Thus, the atmospheric term is independent of the receiver site and, in effect, becomes indistinguishable from the satellite clock error. The effect of systematic errors in $\mathbf{X^s}$ can also be minimised in such a case.

The latter assumption is particularly critical. From experience with single-frequency GPS surveys, it is known that it is usually not possible to resolve ambiguities for baselines over about 20-30km, irrespective of the session length, without appropriate modelling of the atmospheric refraction. With the use of dual-frequency instrumentation the ionospheric delay can be measured and hence the residual atmospheric refraction contains only the tropospheric component. Use of dual-frequency observations promises ambiguity resolution for shorter observing sessions, and even in circumstances when long series of single-frequency observations do not permit ambiguity resolution.

Invoking the "short" baseline assumption and using the wavelength of the carrier wave as the basic length unit (for phases, coordinates and biases), a simplified functional model can be expressed as (note the dropping of the subscript in the atmospheric bias term):

$$\phi_r^s(t) = \rho_r^s(t) + f_0\,[\varepsilon_r(t) - \varepsilon^s(t)] - n_r^s + (\phi_{atmos})^s \qquad (2\text{-}2)$$

The range term ρ is receiver dependent. Two cases are possible, one for each terminal station of the baseline: the fixed "control" site, and the free "remote" site. As the position vector of both the control site $\mathbf{X_{ro}}$ and the satellite $\mathbf{X^s}(t)$ are assumed known, the range between them is assumed known and hence free of errors. Grouping the known terms into the constant k, the model for the phase observation at the control site is simply:

$$\phi_{ro}^s(t) = f_0\,[\varepsilon_{ro}(t) - \varepsilon^s(t)] - n_{ro}^s + (\phi_{atmos})^s + k_{ro}^s(t) \qquad (2\text{-}3)$$

For the range between the remote site and the satellite, only an approximate position of the site is available. Thus, the functional model for observations from the remote site contains the partial derivatives of the range from the satellite, with respect to the coordinates of the remote site, to complement the approximate range contained in the constant k. (Additionally, for a short baseline, the coordinates of the control site can be used to compute these partials.) The linearised model therefore is:

$$\phi_r^s(t) = \mathbf{R^s}\cdot \mathbf{dX_r} + f_0\,[\varepsilon_r(t) - \varepsilon^s(t)] - n_r^s + (\phi_{atmos})^s + k_r^s(t) \qquad (2\text{-}4)$$

with $\mathbf{dX_r}$ the (increments of the) coordinates of the remote site and $\mathbf{R^s}$. is the transpose of the unit-vector from the satellite in question to the control site:

2.1.2 Stochastic Model

We assume that all one-way carrier phases are independent and have the same variance. As we are interested in error propagation rather than in absolute values, this variance is set to unity. Hence the covariance matrix of one-way observations is the identity matrix.

The sample rate of the observations also influences the results. One set of phase measurements per minute is assumed. The use of other measurement rates can easily be accounted for: it suffices to set the variance of the phase observations to the inverse of the rate expressed in minutes (see Appendix E.4).

Cancellation (or modelling) of environmental effects is likely to be incomplete for longer baselines, mainly due to different environmental conditions at both stations. To account for this, terms which are functions of the baseline length could be introduced into the covariance matrix of the observations (see, for example, BOCK et al, 1986). Nevertheless, deciding upon the dependency of the observation precision on the baseline length at the present stage would restrict the generality of the approach. For practical use, the value of an indicator of precision must be multiplied by some factor, which is best determined by experience. At this level only, the inclusion of length dependent and/or independent terms might be appropriate.

2.2 Single-Differences

Building differences between simultaneous one-way phase observables at receiver 1 and receiver 2 to satellite i , can be regarded as a premultiplication of the vector of one-way phases by an appropriate matrix operator:

$$\Delta\phi_{12}{}^{i}(t) = \mathbf{D}_{\Delta} \cdot [\ \phi_1{}^{i}(t)\ ,\ \phi_2{}^{i}(t)\ ] \tag{2-5}$$

or expressed in matrix form, and dropping super- and subscripts:

$$\Delta\Phi = \mathbf{D}_{\Delta} \cdot \Phi$$

Explicitly, with 4 satellites:

$$\begin{bmatrix} \phi^1 \\ \phi^2 \\ \phi^3 \\ \phi^4 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \phi_1{}^1 \\ \phi_1{}^2 \\ \phi_1{}^3 \\ \phi_1{}^4 \\ \phi_2{}^1 \\ \phi_2{}^2 \\ \phi_2{}^3 \\ \phi_2{}^4 \end{bmatrix} \tag{2-6}$$

If we replace the one-way phases by their functional model, the between-receiver single-differences have the form:

$$\Delta\phi^s(t) = \Delta\rho^s(t) + f_0\Delta\varepsilon_r(t) - \Delta n^s \tag{2-7}$$

The satellite clock error terms have cancelled. $\varepsilon_r(t)$ is the difference in receiver clock errors, and $\rho^s(t)$ is the difference in geometric range from the two receivers to the satellite in question. With the "short" baseline assumption, the atmospheric delay terms have also vanished. According to the linearised model of the one-way phase observations (equations (2-3) and (2-4)), the between-receiver single-differences can be written as:

$$\Delta\phi^s(t) = \mathbf{R^s} \cdot \mathbf{dX_r} + f_0\Delta\varepsilon_r(t) - \Delta n^s + \Delta k^s(t) \tag{2-8}$$

Thus, at the single-difference level, the design matrix for range parameters remains unchanged. The role of the control station is only to eliminate the satellite clock errors and reduce the effect of other error sources (for example, atmospheric refraction and satellite orbit errors). For a "short" baseline, it can be shown (see Appendix E) that the azimuth, the slope and the length of the baseline are irrelevant as far as the precision of the determination of the coordinates of the remote site is concerned.

According to the assumption $\mathbf{C}_\Phi = I_8$ for the one-way phases, the covariance of the single-differences becomes:

$$\mathbf{C}_{\Delta\Phi} = \mathbf{D}_\Delta \cdot I_8 \cdot \mathbf{D}_\Delta^{\mathbf{T}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = 2 \cdot I_4 \tag{2-9}$$

From a computational point of view, differencing between 2 receivers reduces both the number of observations and the number of unknowns in the functional model, without introducing correlations. However, it should be noted that not all single-differences can be independent if more than 2 receivers are considered, as data from some receivers must be used several times to build a set of single-differences that includes all receivers. Furthermore, the variance of the single-differences is twice as large as that of one-way phases (equation (2-9)). The differencing operator **D** can be used at all epochs, but its dimension may change according to the number of satellites being tracked.

2.3 Double-Differences

At each epoch, the single-differences can then be differenced between-satellites. More than 2 satellites are generally involved and therefore double-differences are correlated. The differencing matrix operator can be defined as:

$$\nabla\Delta\Phi = \mathbf{D}_{\nabla}\cdot \Delta\Phi \tag{2-10}$$

The optimal choice of an operator is not straightforward and several alternatives have been used:

a) Fixed reference: for example the reference satellite could be the one with the lowest PRN number.
b) Sequential: with respect to the previous satellite.
c) Sequential and uncorrelated.
d) Reduction by the mean.

These options are investigated in Appendix A and the covariance matrix of the differenced observations is explicitly derived for each option. For each operator proposed, if we replace the single-differences with their functional model and drop all super- and subscripts, the double-differences have the form:

$$\nabla\Delta\phi(t) = \nabla\Delta\rho(t) + \nabla\Delta n \tag{2-11}$$

The receiver clock terms, independent of the satellites, have cancelled. Building double-differences is therefore a convenient way to avoid modelling oscillator clock errors, and is the standard observable used in most GPS data processing software.

2.4 Triple-Differences

A simple technique for computing differences between-epochs is to always consider only two successive epochs. Thus, the end epoch of a differenced observation is also the start epoch of the next one. In this case the triple-differenced observations are correlated pairwise. Typically, for e epochs, we have:

$$\begin{bmatrix} \delta\nabla\phi(t_1) \\ \delta\nabla\phi(t_2) \\ \cdots \\ \cdots \\ \delta\nabla\phi(t_e) \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & \ldots & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & \ldots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & \ldots & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \nabla\phi(t_0) \\ \nabla\phi(t_1) \\ \nabla\phi(t_2) \\ \cdots \\ \cdots \\ \nabla\phi(t_e) \end{bmatrix} \tag{2-12}$$

The differencing operator acting on the double-differences is identical to that used to create double-differences by the sequential differencing option (see Appendix A.2). There are, of course, other ways of building triple-differences, such as using the first epoch as the fixed reference epoch. However, the sequential approach presents a definite advantage: if lock on a satellite is lost between one epoch and the next, only the corresponding triple-difference will be affected.

If we replace the double-differences with their functional model, the triple-differences have the form:

$$\delta \nabla \phi(t) = \delta \nabla \rho(t) \qquad (2\text{-}13)$$

The integer ambiguity terms, which were invariant with time, have cancelled. Often in GPS data processing, triple-differences are used in order to avoid dealing with the integer ambiguities. A discussion from a more geometrical perspective is given in Appendix E.

3. ELIMINATION OF INTEGER CYCLE AMBIGUITIES

The approach is based on an explicit removal of the cycle ambiguities from the set of solve-for parameters, using the fact that they are constant with respect to time. This is realised by differencing the observations between epochs. In this context, the only purpose of the other intermediate differencing steps - between receivers and satellites - is to eliminate the other (mainly clock) biases.

Until the ambiguities have been fixed to integer values, measurements of the carrier phase are essentially **range difference** observations. An intuitive understanding of this type of observation is difficult to obtain. A geometrically based analysis, closely related to satellite kinematics, is presented in Appendix E, where the concept of **paired-satellites** provides an interpretation of the process involved. However, for the actual derivation of the proposed indicator of precision, we simply consider the computation of triple-differences as described in section 2.4.

3.1 Global Solution

Let us consider a session with e epochs and m possible satellites. All satellites are not necessarily observed at all epochs. Starting with the model of the single-differences described in section 2.2 (equation (2-8)), the triple-differenced observations for the entire session can be expressed as a vector equation:

$$\delta\nabla\Delta\Phi = \mathbf{D}_\delta \cdot \mathbf{D}_\nabla \cdot \mathbf{R} \cdot \mathbf{dX_r} + \delta\nabla\Delta K \qquad (3\text{-}1)$$

where $\mathbf{dX_r}$ contains the 3 unknown coordinate component increments of the remote site. The matrices premultiplying $\mathbf{dX_r}$ contain several differencing operators $\mathbf{D}_\delta$ and $\mathbf{D}_\nabla$. $\mathbf{R}$ contains the partial derivatives associated with all satellites observed and has the maximum dimension: (e x m) x 3. $\mathbf{D}_\nabla$ is block diagonal and contains e elements, each operating on m observations. $\mathbf{D}_\delta$ contains up to (m-1) elements, each operating on the time series of up to e double-differences. Hence the design matrix premultiplying $\mathbf{dX_r}$ has maximum dimension [(e-1) x (m-1)] x 3. $\delta\nabla K$ is the vector of *a priori* information. The covariance matrix of the triple-differenced observations is $\{\mathbf{D}_\delta \cdot \mathbf{D}_\nabla)\} \cdot 2 \cdot I_{e.m} \cdot \{\mathbf{D}_\delta \cdot \mathbf{D}_\nabla\}^\mathbf{T}$. The computation of the normal matrix is then straightforward.

This approach has the merit of indicating the amount of data to be dealt with, but obviously the storage requirements are very large and hence impose a limit to the number of epochs per session. Fortunately, there are ways to divide the global matrix operations into smaller ones. Triple-differences at different epochs are correlated with each other. Hence, the basic assumption for direct accumulation of data in the normal matrix, as demonstrated in Appendix C, is not fulfilled. The simplest way to overcome this problem would be to ignore the correlation between successive triple-differences. However, a strict solution requires this correlation to be taken into account. Let us consider both options.

3.2 Correlated Triple-Differences

Neglecting the correlation between epochs leads to a much weaker solution. This is demonstrated in Appendix E.4, by inspection of the observation weights for both alternatives. On the other hand, the solution is made less sensitive to errors in a particular measurement, because an error in one triple-difference is not propagated to the following ones. This is why correlated triple-differences are commonly used to identify cycle slips in GPS data processing. Strictly speaking, the triple-differenced observations only offer this interesting feature when they are not rigorously computed!

3.3 Uncorrelated Triple-Differences

An attractive method to account for pairwise correlation between epochs is to decorrelate the observations with the recursive algorithm presented in Appendix B.2. By doing so, an independent set of observations is obtained at each new epoch and the normal matrix can be incremented directly (Appendix C). However, this procedure is only valid for pairwise correlated observations and triple-differences are correlated pairwise only if double-differences are uncorrelated. This influences the choice of the operator when differencing between satellites: it is preferable to use the sequential operator to build the double-differences because it too correlates them pairwise, as with triple-differences at successive epochs. The same decorrelating algorithm can then be used!

However, an important drawback with this approach is the difficulty in dealing with changes in the satellite configuration during a session. A decorrelated double-differenced observation may contain data from several satellites. When changes occur in the satellite coverage, relations between observations at successive epochs may be broken. Thus, to obtain correct results, the same satellites must be observed at every epoch during the entire session.

The 3 x 3 normal matrix is inverted when all desired triple-differenced observations have been processed. Once inverted the covariance matrix of the remote site coordinates is obtained.

4. EXPLICIT RESOLUTION OF INTEGER CYCLE AMBIGUITIES

This approach is based on double-differenced observations, where the integer ambiguities remain as unknown parameters in the observation model.

4.1 Single-Differences

From section 2.2, for an epoch with q observed satellites, we have q single-differenced observations of the form:

$$\Delta\phi^s(t) = \mathbf{R}^s \cdot \mathbf{dX_r} + f_0 \Delta\varepsilon_r(t) - \Delta n^s + \Delta k^s(t) \tag{2-8}$$

One of the main problems with forming between-receiver single-differences arises from changes in the satellite constellation during an observation session. It may not be possible to know at the beginning of a session how many and which satellites will be observed, and consequently used to build single-differences. Therefore, all available satellites, say m, must be considered as potential sources of unknown ambiguities. For this purpose, one column of the design matrix should be dedicated to each available satellite. Furthermore as we are only interested in the design matrix, the k^s terms are not required. The receiver clock terms are also omitted, as they will cancel later anyway, when differencing between satellites. In matrix equation form, dropping the super- and subscripts, the single-differenced observations can be partitioned as follows:

$$\Phi = \mathbf{A} \cdot \mathbf{X} = \begin{bmatrix} \mathbf{R} & \mathbf{J} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{dX_r} \\ \mathbf{n} \end{bmatrix} \tag{4-1}$$

The vector of unknowns **X** consists of the 3 coordinate component increments $\mathbf{dX_r}$ for the remote site and m single-differenced cycle ambiguities **n**. The design matrix **A** therefore comprises a part, **R**, containing range partials and an operator **J**, which selects the q observed satellites out of the m possible ones. For an epoch with 4 satellites observed out of a possible 5, the selection operator would be:

$$\mathbf{J} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{4-2}$$

4.2 Double-Differences

4.2.1 Choice of Unknowns

The basic model for double-differences follows from the single-difference model developed above:

$$\nabla\Phi = \mathbf{D}_\nabla \cdot \mathbf{A} \cdot \mathbf{X} = \mathbf{D}_\nabla \cdot [\,\mathbf{R} \;\; \mathbf{J}\,] \cdot \begin{bmatrix} \mathbf{dX_r} \\ \mathbf{n} \end{bmatrix} \tag{4-3}$$

At this point there are a number of ways of defining the unknown ambiguities. Two approaches based on "lumped" ambiguity terms (GRANT et al, 1990) are considered here:

1. Single-differenced cycle ambiguities are chosen as parameters, and their number is equal to the number of satellites available: m.

2. Double-differenced cycle ambiguities are considered, and their number is equal to m-1.

These options can be compared with those for the treatment of a levelling network without fixed origin. In the former, the height of each point is considered unknown. In the latter, we determine m-1 height differences joining all points.

4.2.2 Single-Differenced Ambiguities

Design matrix: we build q-1 double-differenced observations between satellites i and j :

$$\nabla\Delta\phi(t) = \nabla\Delta\rho(t) - \Delta n^j + \Delta n^i \tag{4-4}$$

with $1 \leq i, j \leq m$. In matrix form the design matrix is simply:

$$\mathbf{D}_\nabla \cdot \mathbf{A} \cdot \mathbf{X} = \mathbf{D}_\nabla \cdot \mathbf{R} \cdot \mathbf{dX_r} + \mathbf{J} \cdot \mathbf{n} \tag{4-5}$$

Normal matrix: It is best obtained not by using the above mentioned design matrix, but by using the kernel approach described in Appendix A.2.d. This involves "reduction by the mean", which is independent of any differencing operator:

$$\mathbf{N} = \mathbf{A}^\mathbf{T} \cdot \mathbf{K}_\mathbf{n} \cdot \mathbf{A} \tag{4-6}$$

Accumulation of epochs: The storage of the normal matrix is designed so that both a column and a row are dedicated to each unknown. Hence, its dimension is (3+m) x (3+m) , where m is the maximum number of satellites available.

Elimination of missing satellites: Up to this stage, everything has been designed for the maximum number of satellites available. In the general case however, out of the m possible satellites only a certain number q' is observed at least once during a session. Thus, when all the data for a session has been accumulated in the normal matrix, the rows and columns corresponding to satellites which have not been observed are removed and the size of the matrix reduces accordingly to $(3+q') \times (3+q')$. This does not affect the 3×3 coordinate submatrix.

The solution of the normal equations is not unique. The matrix $\mathbf{N}$ has a rank defect of one because no reference ambiguity has been defined (GRANT et al, 1990). This compares with the levelling network example mentioned earlier: it is impossible to solve for heights using only height difference observations. The rank defect can be overcome by introducing a constraint on the unknowns. Our choice is that the part of the trace of the covariance matrix referring to the ambiguities be a minimum. A principal component analysis of the normal matrix shows that the eigenvector associated with the zero eigenvalue is $\mathbf{s} = (0,0,0,1, . . . ,1)^{\mathbf{T}}$. In other words, if $\mathbf{x}$ is a solution, then $\mathbf{x} + k\mathbf{s}$ is also a solution, and it therefore only suffices to add the same quantity to all the ambiguities to obtain another solution. The elements of $\mathbf{s}$ referring to the coordinate submatrix are all zero, and consequently the minimum partial trace condition as presented above is equivalent to the minimum (full) trace condition. Using such a condition avoids the choice of one satellite ambiguity having to be held fixed. (CASPARY, 1987, is a good textbook on these topics, which are not specific to satellite geodesy problems.)

Covariance matrix: The generalised inverse of the accumulated normal matrix is the covariance matrix of the unknowns:

$$\mathbf{Q_{xx}} = (\mathbf{N}+\mathbf{ss}^{\mathbf{T}})^{-1} \cdot \mathbf{N} \cdot (\mathbf{N}+\mathbf{ss}^{\mathbf{T}})^{-1} \qquad (4\text{-}7)$$

For the case of interest, $\mathbf{N}+\mathbf{ss}^{\mathbf{T}}$ is simply obtained by adding a constant term to all elements of the ambiguity submatrix of $\mathbf{N}$ (dimension $q' \times q'$). It is worthwhile to consider separately the components of the covariance matrix referring to the coordinates (dimension 3×3) and the ambiguities.

Coordinate submatrix: It is independent of the reference ambiguity selected. This, again, is related to the fact that the eigenvector of the normal matrix associated with the zero eigenvalue is $(0,0,0,1, . . . ,1)^{\mathbf{T}}$.

Ambiguity submatrix: It indicates the precision of each estimated ambiguity with respect to the reference ambiguity (or, in the case here, the reference combination of ambiguities). This submatrix depends on the selection of the reference ambiguity. The minimum trace criteria used here is equivalent to fixing the integer ambiguity to a satellite that would be at the centroid of the visible satellite constellation. The relative precision of the ambiguities is not affected by the choice of the reference ambiguity, but unfortunately absolute measures of precision of ambiguities are not unique.

4.2.3 Doubly-Differenced Ambiguities

Design matrix: at each epoch, q-1 equations are formed:

$$\nabla\Delta\phi(t) = \nabla\Delta\rho(t) + \nabla\Delta n \tag{4-8}$$

In matrix form this is:

$$\mathbf{D}_\nabla \cdot \mathbf{A} \cdot \mathbf{X} = \mathbf{D}_\nabla \cdot \mathbf{R} \cdot \mathbf{dX_r} + \mathbf{D}_\nabla \cdot \mathbf{J} \cdot \mathbf{n} \tag{4-9}$$

The doubly-differenced ambiguities depend on $\mathbf{D}_\nabla$ and $\mathbf{J}$, both of which can change between epochs. We may instead replace:

$$\mathbf{D}_\nabla \cdot \mathbf{J} \cdot \mathbf{n} \quad \text{by} \quad \mathbf{J}' \cdot (\mathbf{D}_\nabla' \cdot \mathbf{n})$$

where $\mathbf{D}_\nabla'$ is a predefined differencing operator of dimension $(m-1) \times m$, constant for the whole observation session. The intention is to concentrate the variations between epochs in matrix $\mathbf{J}'$, a selection operator for the appropriate double-differenced ambiguities.

The desired replacement is only possible if $(\mathbf{D}_\nabla \cdot \mathbf{J}) = (\mathbf{J}' \cdot \mathbf{D}_\nabla')$. Both sides can be postmultiplied by the same term:

$$\mathbf{J}' \cdot \mathbf{D}_\nabla' \cdot \mathbf{D}_\nabla'^{T}(\mathbf{D}_\nabla'\mathbf{D}_\nabla'^{T})^{-1} = \mathbf{D}_\nabla \cdot \mathbf{J} \cdot \mathbf{D}_\nabla'^{T}(\mathbf{D}_\nabla'\mathbf{D}_\nabla'^{T})^{-1} \tag{4-10}$$

By definition, $\mathbf{D}_\nabla'$ is constructed from m-1 combinations. Due to its square matrix structure, $\mathbf{D}_\nabla'\mathbf{D}_\nabla'^{T}$ is positive definite of rank m-1 and has a regular inverse. The new selection matrix therefore becomes:

$$\mathbf{J}' = \mathbf{D}_\nabla \cdot \mathbf{J} \cdot \mathbf{D}_\nabla'^{T}(\mathbf{D}_\nabla'\mathbf{D}_\nabla'^{T})^{-1} \tag{4-11}$$

Thus, we can replace the $\mathbf{J}'$ matrix in the model (equation (4-9)):

$$\mathbf{D}_\nabla \cdot \mathbf{A} \cdot \mathbf{X} = \mathbf{D}_\nabla \cdot \mathbf{R} \cdot \mathbf{dX_r} + \mathbf{D}_\nabla \cdot \mathbf{J} \cdot \mathbf{D}_\nabla'^{T}(\mathbf{D}_\nabla'\mathbf{D}_\nabla'^{T})^{-1} \cdot (\mathbf{D}_\nabla' \cdot \mathbf{n}) \tag{4-12}$$

In explicit matrix form, the double-difference model becomes:

$$\nabla \Phi = \mathbf{D}_\nabla \cdot \mathbf{A} \cdot \mathbf{X} = \mathbf{D}_\nabla \cdot \begin{bmatrix} \mathbf{R} & \mathbf{J} \cdot \mathbf{D}_\nabla'^{T}(\mathbf{D}_\nabla'\mathbf{D}_\nabla'^{T})^{-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{dX_r} \\ \nabla\, \mathbf{n} \end{bmatrix} \tag{4-13}$$

where $\nabla \mathbf{n} = \mathbf{D}_\nabla' \cdot \mathbf{n}$.

Example: let us consider two different sets of doubly-differenced ambiguities: $\mathbf{D}_{\nabla f}'$ a fixed-reference differencing operator and $\mathbf{D}_{\nabla s}'$ a sequential-differencing operator. The new unknowns $\nabla \mathbf{n}$ are:

$$\begin{aligned} &\text{with } \mathbf{D}_{\nabla f}': \quad \Delta n^2 - \Delta n^1,\ \Delta n^3 - \Delta n^1,\ \Delta n^4 - \Delta n^1,\ \Delta n^5 - \Delta n^1 \\ &\text{with } \mathbf{D}_{\nabla s}': \quad \Delta n^2 - \Delta n^1,\ \Delta n^3 - \Delta n^2,\ \Delta n^4 - \Delta n^3,\ \Delta n^5 - \Delta n^4 \end{aligned} \tag{4-14}$$

For the matrix $\mathbf{J}'$ defined above, $\mathbf{D}_{\nabla f}'$ leads to:

$$\mathbf{J} \cdot \mathbf{D}_{\nabla}'^{T}(\mathbf{D}_{\nabla}'\mathbf{D}_{\nabla}'^{T})^{-1} = \frac{1}{5} \cdot \begin{bmatrix} -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 4 \\ -1 & 4 & -1 & -1 \end{bmatrix} \tag{4-15}$$

and $\mathbf{D}_{\nabla s}'$ leads to:

$$\mathbf{J} \cdot \mathbf{D}_{\nabla}'^{T}(\mathbf{D}_{\nabla}'\mathbf{D}_{\nabla}'^{T})^{-1} = \frac{1}{5} \cdot \begin{bmatrix} 1 & 2 & 3 & -1 \\ -4 & -3 & -2 & -1 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & -2 & -1 \end{bmatrix} \tag{4-16}$$

These matrices form the second part of the design matrix. They obviously depend on the choice of $\mathbf{D}_{\nabla}'$.

Normal matrix: It is not necessary to compute $\mathbf{J}'$ explicitly. As with the singly-differenced ambiguity approach, it is preferable to use the "reduction by the mean" approach to construct $\mathbf{N}$. The accumulated normal matrix depends on $\mathbf{D}_{\nabla}'$. This is logical as the unknowns to which $\mathbf{N}$ refers are defined by $\mathbf{D}_{\nabla}'$. $\mathbf{N}$ has full rank and can be inverted directly. In fact the choice of $\mathbf{D}_{\nabla}'$ can be considered a datum selection. However, as before, the coordinate submatrix is completely independent of the datum: it is the sum of the products $\mathbf{R}^T \cdot \mathbf{K}_n \cdot \mathbf{R}$ from all epochs, which only depend on the number and the positions of the satellites observed.

Covariance matrix: It has a smaller size: $(3+q'-1) \times (3+q'-1)$, but exhibits the same features: the coordinate submatrix is the same as before and the ambiguity submatrix depends on the datum selection $\mathbf{D}_{\nabla}'$. The Bias Dilution of Precision 1 (BDOP1) can also be extracted, as before, by taking the trace of the coordinate covariance submatrix. BDOP2 is defined as the trace of the ambiguity submatrix and, like the submatrix itself, depends on the choice of the reference ambiguity.

4.2.4 Comparison of the Approaches

The main difference between using singly-differenced ambiguities and doubly-differenced ambiguities is the stage at which the reference (singly-differenced) ambiguity is selected. The computation burden for BDOP2 is very similar and changes in the satellite coverage can be overcome with the same flexibility. When it comes to comparing the prediction and results of a measurement campaign, one should keep in mind that the selection of the reference ambiguity term in the GPS post-processing software may be different to the one used to compute BDOP2. (For example, quite often the singly-differenced ambiguity associated with the satellite observed at most epochs is held fixed.) This can result in some apparent discrepancies in results for the ambiguities (and their precisions) between different software packages.

4.3 BDOP2 versus BDOP1

It should be emphasised that the coordinate covariance submatrix, and accordingly its eigenvalues and eigenvectors, are exactly the same as those resulting from the approach using uncorrelated triple-differences. Hence, all the information extracted from the computations for the BDOP1 factor using the triple-difference approach is obtained here as part of the solution for BDOP2. Furthermore, the problem of dealing with a changing satellite configuration is more easily overcome using the double-differences approach than using triple-differenced observations.

An inspection of some computation examples (see section 7), indicates that BDOP1 and BDOP2 have similar signatures. Two reasons can explain this behaviour:

1. They are extracted from the same matrix, in which coordinates and ambiguities are strongly interrelated.

2. They describe similar properties of the satellite configuration. If it is easy to compute precise coordinates when ambiguities are eliminated, it must also be easy to explicitly solve for them.

Due to its invariance with respect to the computational approach - differencing level and selection of the reference ambiguity - BDOP1 therefore represents a more general concept in GPS data processing.

5. ONCE CYCLE AMBIGUITIES ARE ESTIMATED AND RESOLVED

5.1 Range Observations

In this last case, it is assumed that the ambiguities have already been fixed to their integer values. Thus, the relevant observations are now unbiased ranges. For each observation session, we only consider the (3 x 3) part of the accumulated normal matrix corresponding to the coordinates, formed from double-differenced observations. It has already been pointed out that this submatrix is unaffected by the approach taken for ambiguity modelling (singly- or doubly-differenced parameters), or whether they are eliminated altogether, as in the (uncorrelated) triple-differenced approach. Its inverse is the covariance matrix of the coordinates of the remote site.

5.2 BDOP3: Extraction and Features

We define the Bias Dilution of Precision 3 indicator (BDOP3) as the trace of the 3 x 3 covariance matrix. This factor can be considered as being equivalent to an "accumulated PDOP", which simply reflects the fact that the differential positioning is not an instantaneous process, but involves observations at several epochs. The peaks of the BDOP3 curve are smoothed, even for relatively short sessions with high PDOP values at individual epochs. The magnitude of this indicator of precision, less than the corresponding BDOP1 (or BDOP2), at least for short sessions, indicates that in general the accuracy of the coordinates is higher once the ambiguities are (correctly) resolved than if they were not resolved to their integer values.

6. IMPLEMENTATION OF COMPUTATIONAL PROCEDURE

6.1 Quick Mathematical Reference

The normal matrix contains the contribution of all observations collected during a survey session. Its dimension is 3 + the number of satellites.

$$\mathbf{N_{xx}} = \begin{bmatrix} \mathbf{N_{cc}} & \mathbf{N_{cb}} \\ \mathbf{N_{bc}} & \mathbf{N_{bb}} \end{bmatrix} \tag{6-1}$$

where the coordinate only part is contained in $\mathbf{N_{cc}}$, the cycle ambiguity part is $\mathbf{N_{bb}}$, while $\mathbf{N_{cb}}$ and $\mathbf{N_{bc}}$ are the submatrices formed from the cross multiplication of coordinate and ambiguity terms. The cofactor matrix $\mathbf{Q_{xx}}$ is obtained by the generalised inversion (with minimal trace) of $\mathbf{N_{xx}}$:

$$\mathbf{Q_{xx}} = \begin{bmatrix} \mathbf{Q_{cc}} & \mathbf{Q_{cb}} \\ \mathbf{Q_{bc}} & \mathbf{Q_{bb}} \end{bmatrix} \tag{6-2}$$

The covariance matrices and the corresponding factor of Dilution Of Precision factors for the different cases are:

$$\mathbf{BDOP1} = \sqrt{\text{trace}\left[\mathbf{Q_{cc}}\right]} \tag{6-3}$$

$$\mathbf{BDOP2} = \sqrt{\text{trace}\left[\mathbf{Q_{bb}}\right]} \tag{6-4}$$

$$BDOP3 = \sqrt{\mathrm{trace}\left[N_{cc}^{-1}\right]} \qquad (6\text{-}5)$$

6.2 The "PREDICT" Program

At the School of Surveying, University of New South Wales, a satellite prediction software program, known as "PREDICT", has been developed by Dr. C.Rizos. The principal aim of this program is to produce satellite alert information, sky-plots and DOP factor-plots. Recorded Broadcast Ephemeris messages as well as formatted files containing the Keplerian elements of simulated orbits can be used as input. The algorithms for estimation of the ability to solve for the ambiguities processing double- and triple-differenced phase observations has been implemented as an additional task within "PREDICT".

6.2.1 Features

The computations can be performed for any observation session length, independent of the specified output rate. This makes it possible, for example, to plot estimators for one hour sessions every ten minutes. The end product is a plot of the proposed BDOP indicators. Except for the approach using decorrelated triple-differences (see Appendix A.2.c), it is not necessary that the number of satellites observed remains constant during a session.

6.2.2 Observations

For each epoch (for example, each minute) the algorithm determines if a satellite is visible, according to the selected elevation cut-off angle. When all visible satellites are found, the design matrix for that observation epoch is computed. The differenced design matrix (correlated or decorrelated), and the associated covariance matrix, is derived.

6.2.3 Accumulation Process

The "observations" are accumulated in two steps:

1. The first step is to add the normal matrices obtained at each epoch, until an output is required (according to some selected rate).

2. The second step is a "rolling" process. The normal matrices obtained above are added until the selected session length is reached. (The session length can be any multiple of the output rate.) The freshly computed first step normal matrix replaces the oldest one and the process repeated.

6.2.4 Extraction of the DOP Factor

At each output period, the accumulated normal matrix is inverted, giving the covariance matrix of the unknowns. The DOP factor is then computed and directed to the printer and/or plotter.

6.3 Examples to Illustrate the Characteristics of the BDOP Factors

6.3.1 Four Satellite Coverage

An example of a four satellite constellation (satellites 6, 9, 11, 13), as viewed from Sydney, on the 9th September, 1987, for a 6 hour period from 20h 00m (local time) to 2h 00m on the following day, is considered. Figure 2 is a sky-plot showing the paths of the satellites in relation to a polar plot centred at the Sydney site's zenith. The position and motion of the associated paired-satellites (see Appendix E) are also shown at 22h 40m.

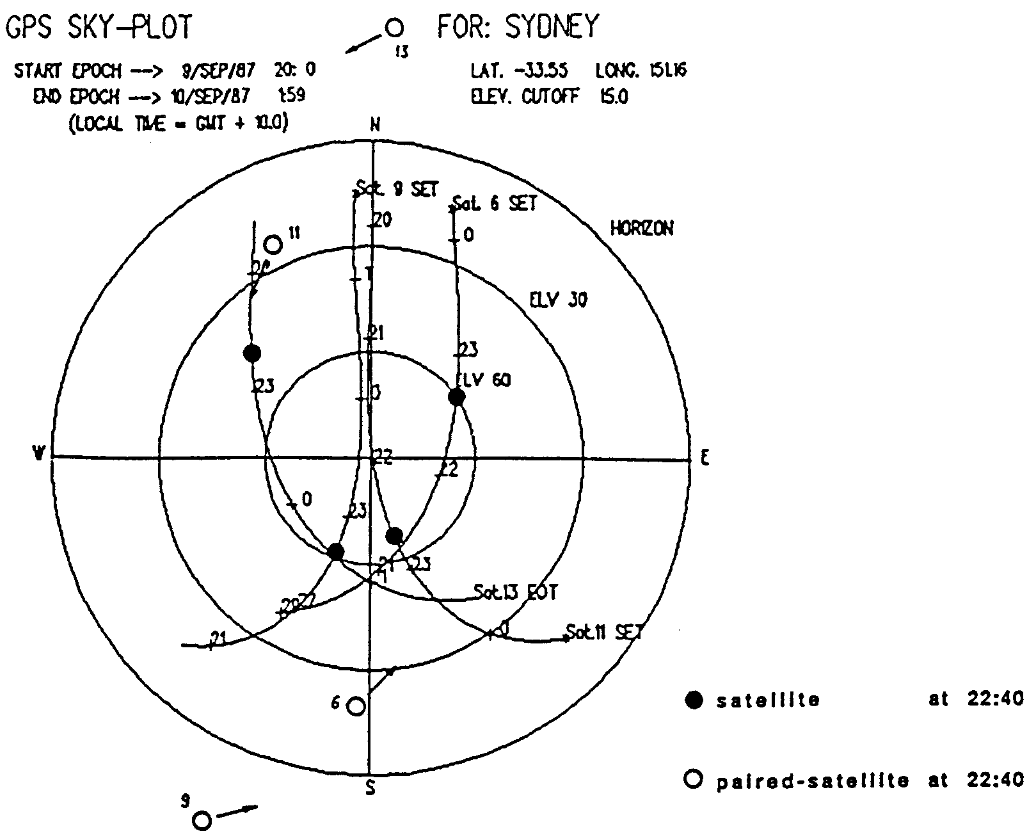


Figure 2: A sky-plot centred at Sydney, Australia for the 9th-10th September, 1987, showing a four satellite constellation and the associated locations of the paired-satellites.

Figure 3 is a plot of the Geometric Dilution of Precision (GDOP) for this period. There is a large increase around 22h 40m, even though there is one satellite in each quadrant. At this time, all the satellites have approximately the same elevation angle and it is not possible to separate the receiver clock error from the height of the station in the solution. (The PDOP would identical behaviour.) On the basis of studying such a plot, the surveyor could draw the conclusion that the constellation is not favourable for GPS positioning.

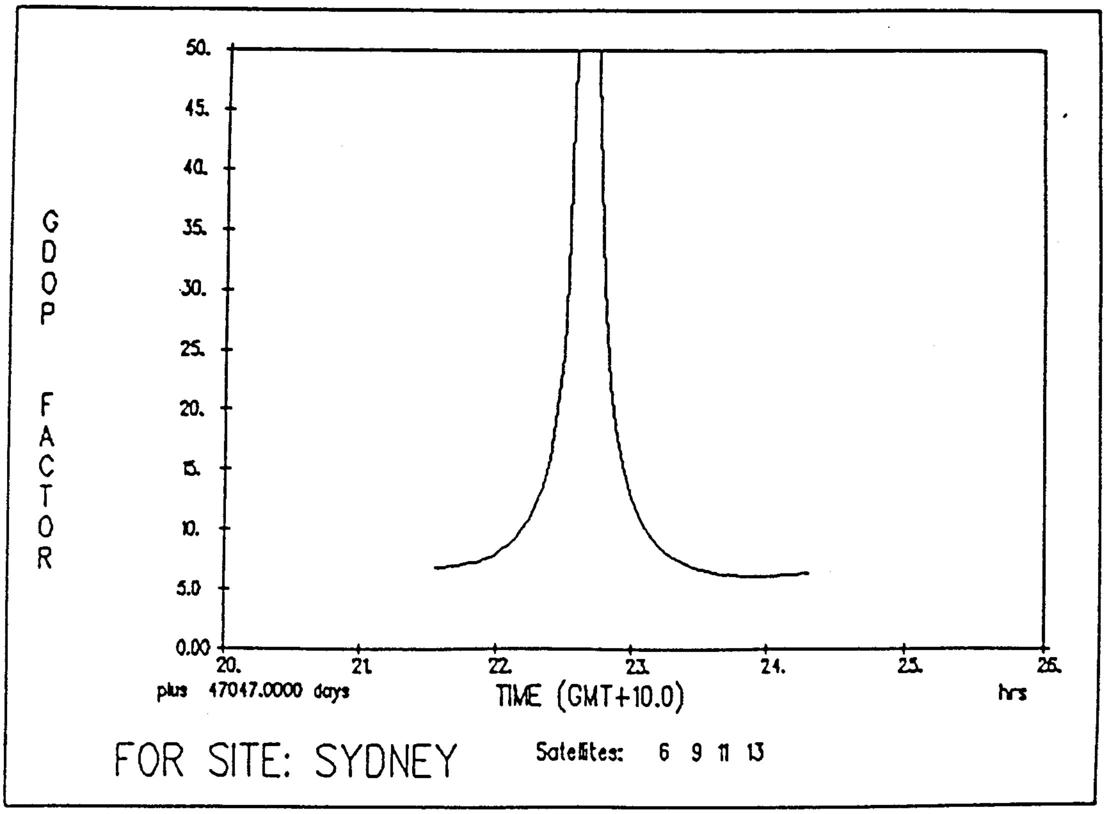


Figure 3: The GDOP factor at Sydney, Australia for the 9th-10th September, 1987, for a four satellite constellation.

The BDOP1, BDOP2 and BDOP3 factors have been calculated for observation session lengths of 15, 30 and 60 minutes and plotted in Figures 4, 5 and 6 respectively. (The value indicated on these plots is for the end of the observation session.) The BDOP1 and BDOP2 indicators do not have a peak as in the case of the GDOP/PDOP, but the BDOP3 plot does show a small peak. Nevertheless the peak is strongly reduced when averaged over a session. An extension of the session length, from 15 to 30 and 60 minutes, causes a strong decrease in the value of the BDOP1 and BDOP2 factors, indicating a strengthening of the phase solution with increasing session length. However, the decrease in the value of the BDOP3 indicator with an extension of the observation session length is not as pronounced as for the BDOP1/BDOP2 factors, indicating that ambiguity-fixed solutions do not exhibit a strengthening of the solution to the same degree that ambiguity-free (and especially the transition from ambiguity-free to ambiguity-fixed) solutions exhibit.

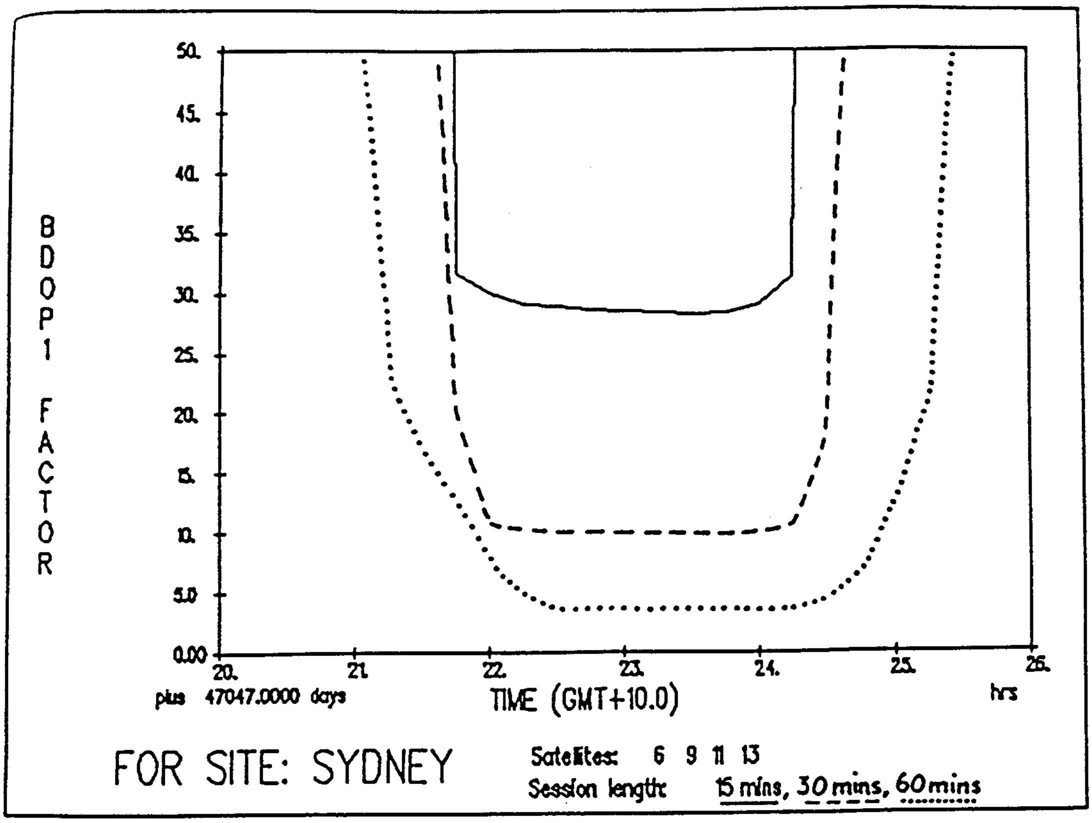


Figure 4: The BDOP1 factor at Sydney, Australia for the 9th-10th September, 1987, for a four satellite constellation and observing sessions of 15, 30 and 60 minutes.

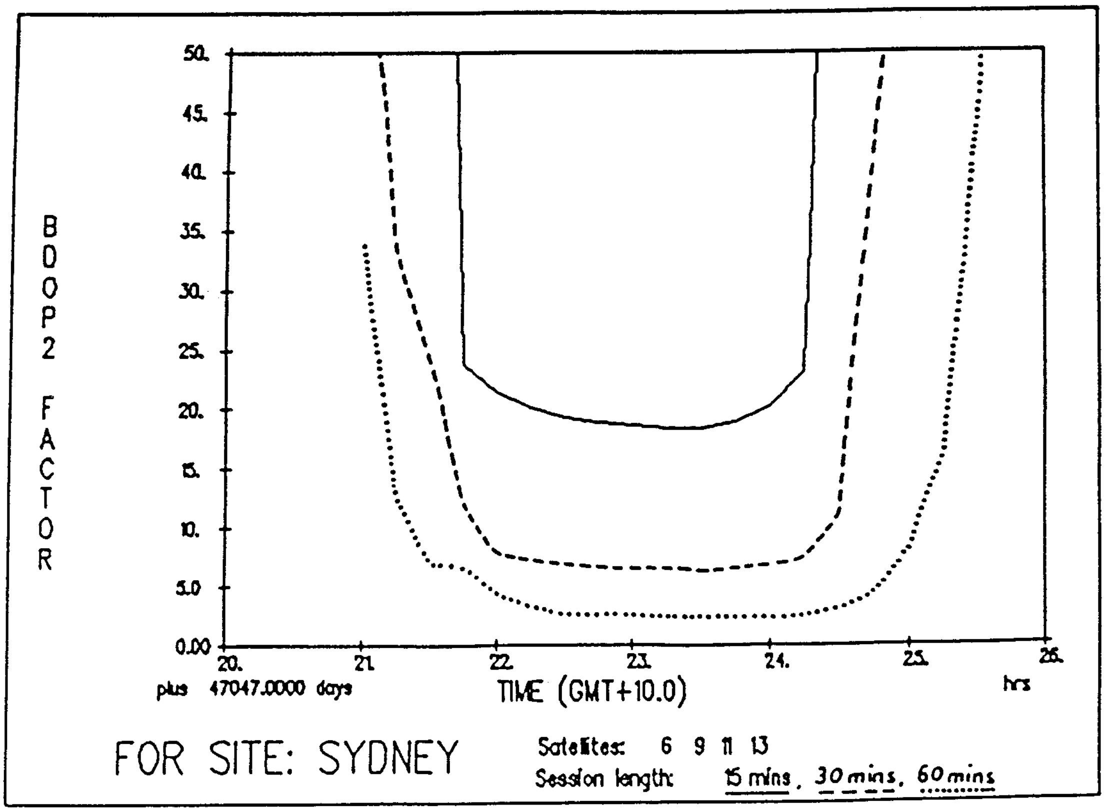


Figure 5: The BDOP2 factor at Sydney, Australia for the 9th-10th September, 1987, for a four satellite constellation and observing sessions of 15, 30 and 60 minutes.

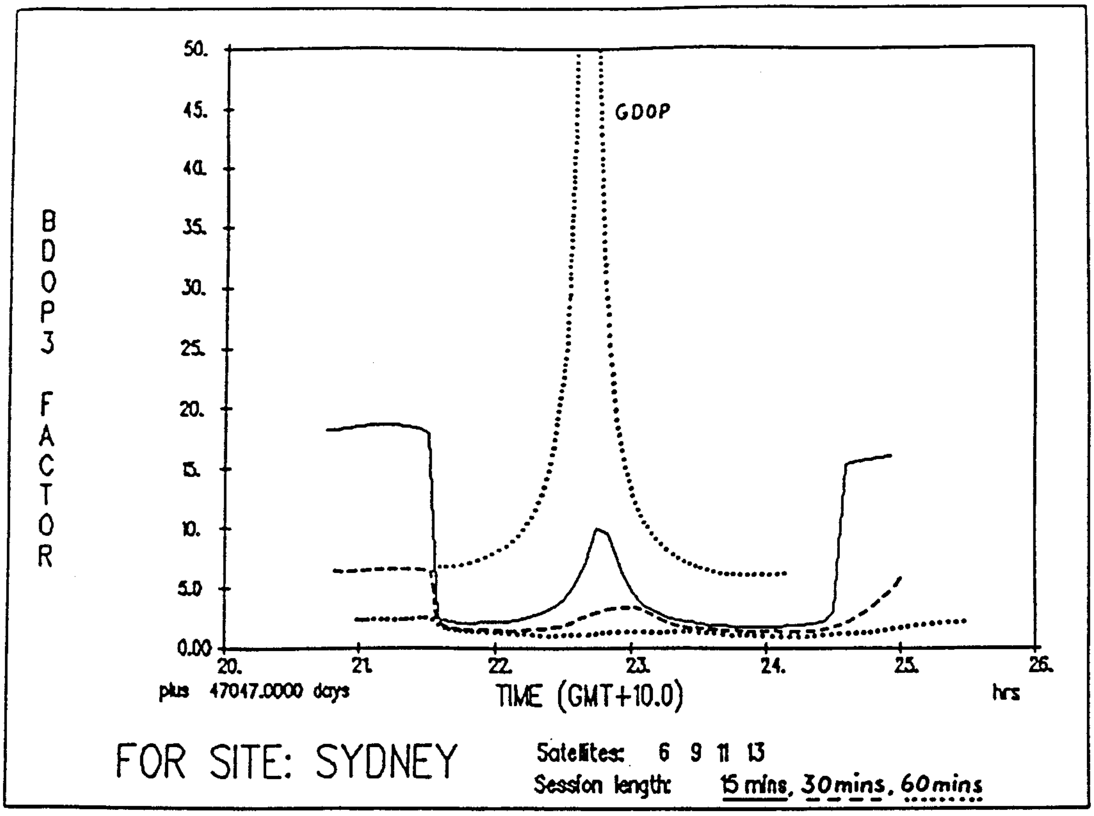


Figure 6: The BDOP3 factor at Sydney, Australia for the 9th-10th September, 1987, for a four satellite constellation and observing sessions of 15, 30 and 60 minutes.

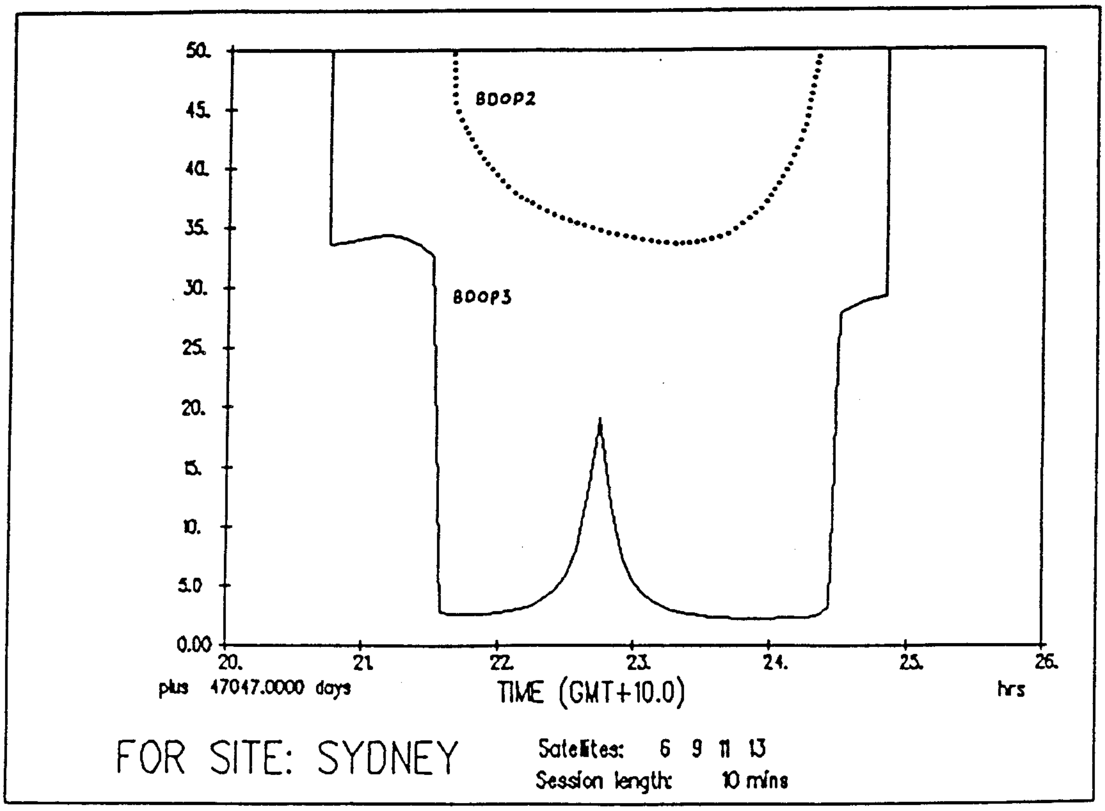


Figure 7: The BDOP2 and BDOP3 factor at Sydney, Australia for the 9th-10th September, 1987, for a four satellite constellation and an observing session of 10 minutes.

In comparing the relative magnitudes of BDOP2 and BDOP3 we see from Figure 7 that even for observation sessions as short as 10 minutes, the peak of BDOP3 is already smaller than BDOP2. Thus, a resolution of the ambiguities would already improve the final coordinates for even such short session.

6.3.2 Three Satellite Coverage

We now consider a reduced constellation of three satellites by eliminating satellite 13 from the earlier example.

BDOP2 and BDOP3 have been calculated for a 30 minute session and plotted in Figure 8. For half hour sessions, the precision of the estimated integer cycle ambiguities is reasonable only at the beginning of the observing window, for sessions ending between 21h 30m and 22h 00m. Figure 9 conveys the same information but for 60 minute observation sessions. For one hour sessions, the probability of resolving the integer cycle ambiguities appears to be reasonable throughout the observing window, between 21h 15m and 24h 15m (end of session).

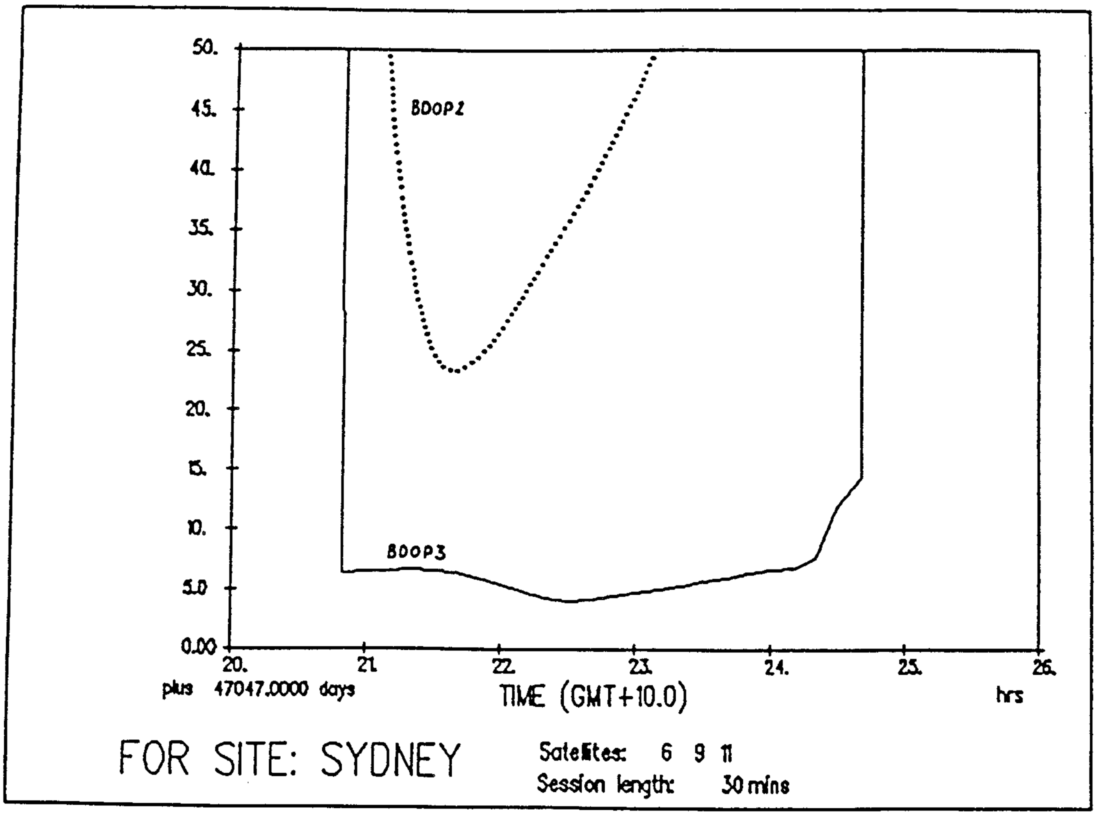


Figure 8: The BDOP2 and BDOP3 factor at Sydney, Australia for the 9th-10th September, 1987, for a three satellite constellation and an observing session of 30 minutes.

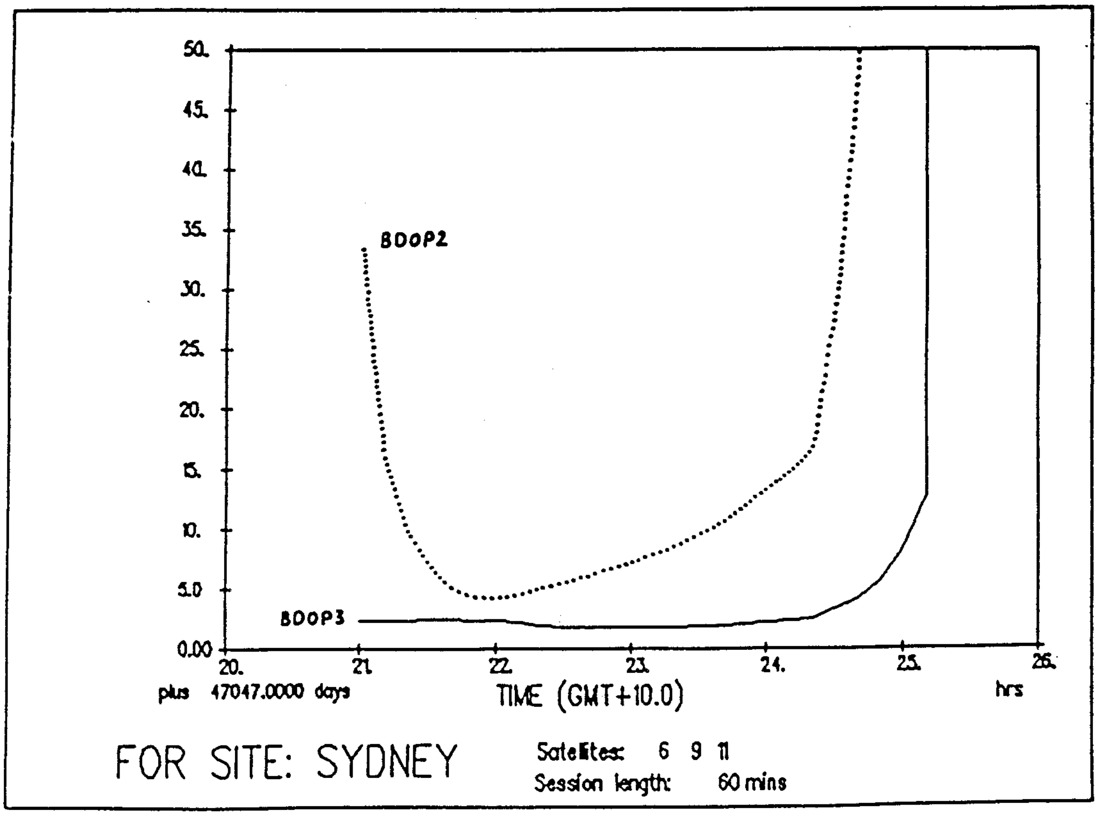


Figure 9: The BDOP2 and BDOP3 factor at Sydney, Australia for the 9th-10th September, 1987, for a three satellite constellation and an observing session of 60 minutes.

6.3.3 Block I Satellite Coverage

Using the (1988) 7 Block I satellite GPS constellation, PDOP (Figure 10) and the 3 BDOP factors for a 60 minute session length (Figure 11) were computed for Sydney, Australia, for a 5 hour window on the 1st November, 1988. A minimum elevation cutoff of 15° was applied. The PDOP was calculated for all 4 satellite combinations visible and the **lowest** of these values is plotted in Figure 10. Note that PDOP falls below 5 only after about 18h 05m. BDOP1 has been calculated for observation session lengths of 30, 60, 90 and 120 minutes and plotted in Figure 12. (In these plots all BDOP values have been plotted against the mid point of the simulated observing session, that is, the plotted BDOP1 at 19h 30m for a 120 minute observing session is that for an observing session running from 18h 30m to 20h 30m. In addition, all visible satellites were assumed to have been tracked.) The relationship between troughs and peaks in the PDOP plot and in the BDOP plots is not clear as BDOP continues to drop even when PDOP reaches a plateau after 18h 05m because of rising satellites (from 3 to 7) during the next hour or two. Nevertheless low BDOP values are associated with observing sessions that include the period of rapidly falling PDOP.

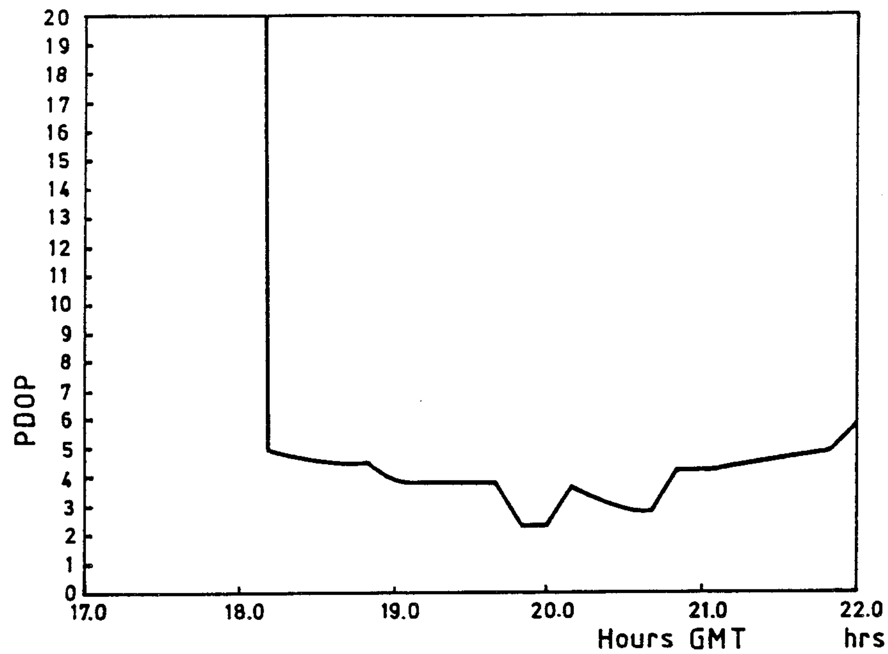


Figure 10: The PDOP factor at Sydney, Australia for 1st November, 1988, for the (1988) 7 Block I satellite constellation.

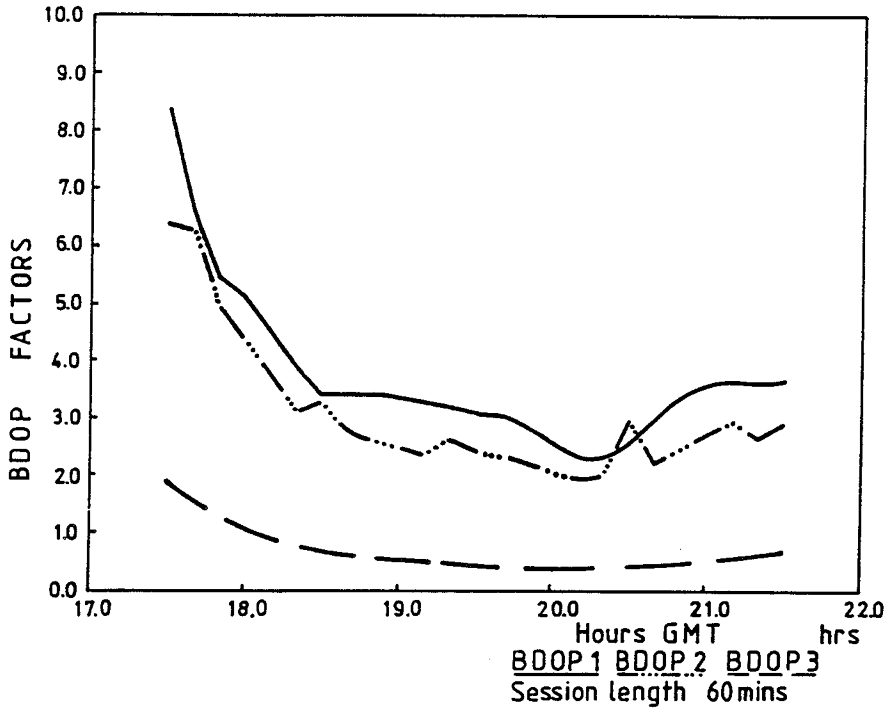


Figure 11: BDOP1, BDOP2 and BDOP3 indicators for Sydney, Australia for 1st November, 1988, for the (1988) 7 Block I satellite constellation and an observing session of 60 minutes.

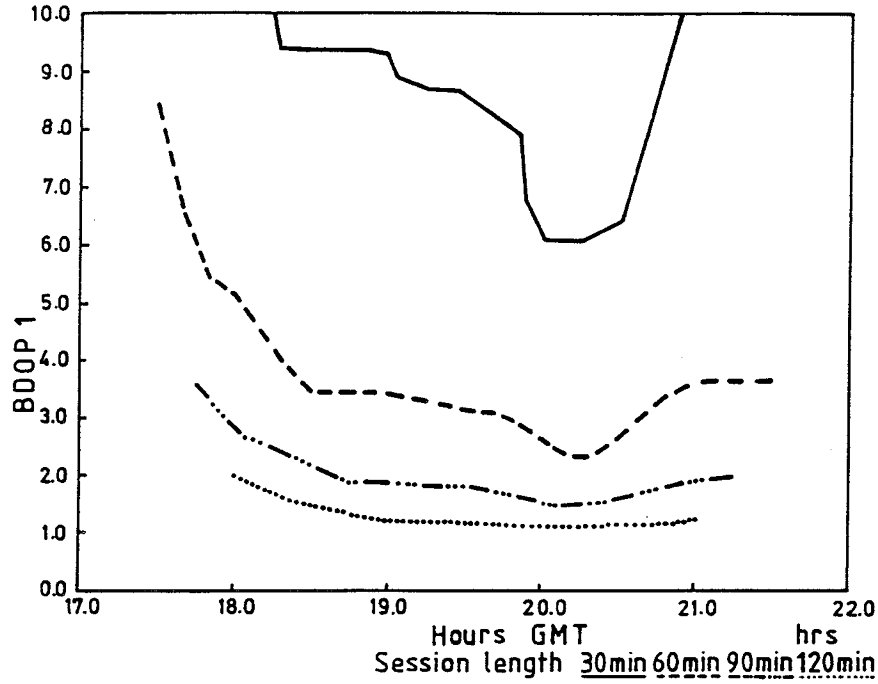


Figure 12: The BDOP1 factor at Sydney, Australia for 1st November, 1988, for the (1988) 7 Block I satellite constellation and observing sessions of 30, 60, 90 and 120 minutes.

6.3.4 Block II Satellite Coverage

To study the relationship between PDOP and BDOP factors further a set of simulated orbital elements for an earlier proposed GPS 21 (18 + 3 active spares) satellite constellation was used to calculate PDOP and BDOP1 for a survey planned for the 1990's. (It is now well known that the operational system will comprise a 21 + 3 satellite constellation.) The receiver position used to calculate these is for Wellington, New Zealand. A minimum elevation cutoff of 15° was again applied and the number of satellites visible at any instant varied from between 4 and 7. As in the case of Figure 10, the PDOP was calculated for all 4 satellite combinations visible and the **lowest** of these values is plotted in Figure 13. The BDOP values were computed using all visible satellites of the nominal 18 satellite constellation. (It was assumed the 3 spare satellites were not available for tracking.) In the case of this satellite constellation the satellite geometry repeats every 11 hours 58 minutes (with different satellites) and thus the DOP factors were only plotted over a 12 hour period. Note in Figure 13 the peaks in the PDOP plot at 0h 20m, 7h 00m and 9h 40m.

BDOP1 has been calculated for observation session lengths of 30, 60 and 120 minutes. The BDOP1 based on all satellites visible in an observing session are plotted in Figure 14 for various session lengths. Note the minima at 0h 00m, 3h 45m, 7h 15m, and 9h 45m, and that 3 of these 4 minima correspond to the 3 PDOP maxima. The fourth BDOP1 minimum occurs at an extended period of low PDOP. This is consistent with the findings of NORTON (1987), that is, it appears that sharp peaks in PDOP correspond to good observing sessions. Hence the commonly accepted belief that low PDOP (or GDOP) indicates good geometry for GPS carrier phase adjustment may be true in some circumstances, but it is more often wrong than right. The observing windows of all the BDOP1 maxima (centred at 1h 45m, 5h 30m, 8h 30m and 10h 45m) correspond to periods of relatively good (that is, small) PDOP!

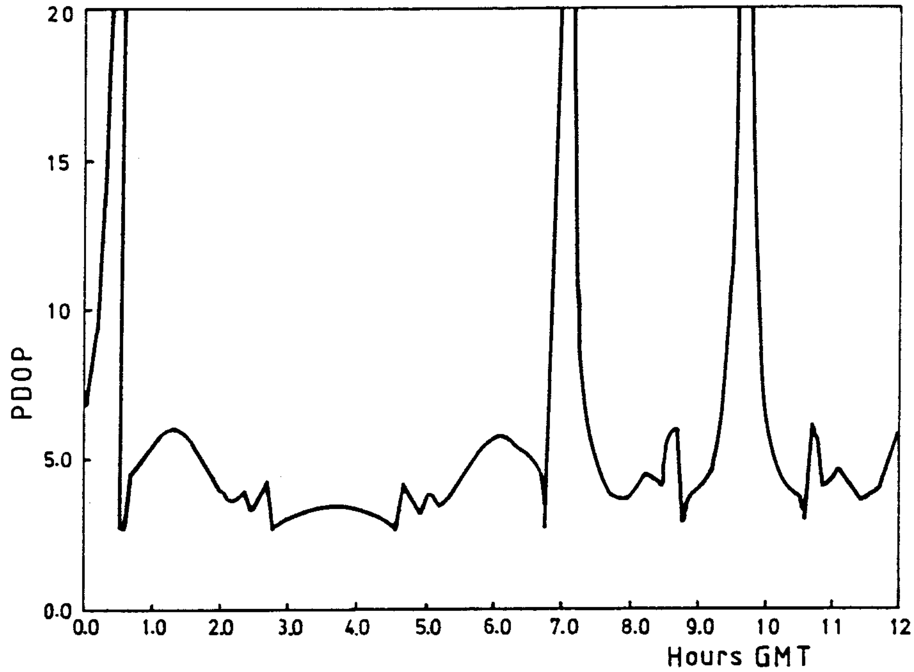


Figure 13: The PDOP factor at Wellington, New Zealand, assuming the proposed nominal 18 satellite constellation.

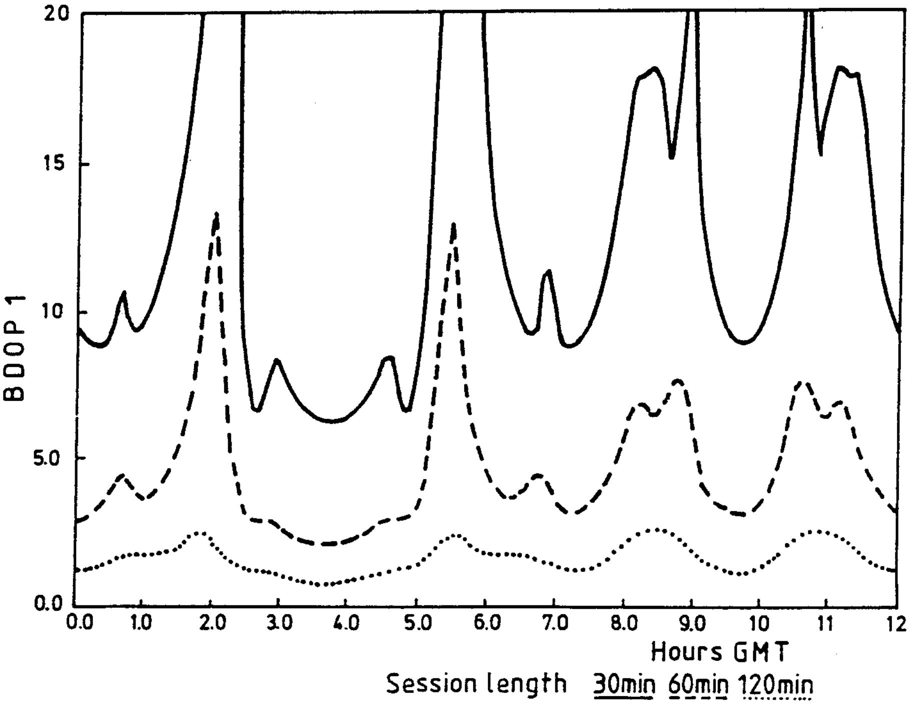


Figure 14: The BDOP1 factor at Wellington, New Zealand, assuming the proposed nominal 18 satellite constellation and observing sessions of 30, 60 and 120 minutes.

6.3.5 Characteristics of the Precision Indicators

The BDOP factors have, in general, the following characteristics:

- Indicators can be computed for as few as two satellites. This is in accordance with reality: contrary to common belief, GPS surveys do not require the simultaneous availability of 4 satellites. However, to obtain similar quality results with less satellites, longer sessions are required.

- Indicators can be computed for observing sessions during which satellites are rising and setting.

- Note in Figures 12 and 14 the dramatic improvement and smoothing of BDOP1 that occurs for longer session lengths.

- In Figures 12 and 14, the best observing time with a 30 minute session is significantly better than the worst observing time with a 60 minute session. The timing of the start of an observing session is therefore more critical for short sessions than for long ones.

- Another factor that influences the magnitude and variability of BDOP1 is the number of satellites observed. Note in the examples in section 6.3.3 & 6.3.4 that all satellites that were visible were assumed to have been tracked.

- The variations in the value of BDOP1 for different sessions during the day are significant, even for the comparatively long 120 minute observing sessions. Even for a 120 minute observing session, the ratio of best to worst BDOP1 is more than 2 to 1.

- Despite the 24 hour coverage of the 21 (18 + 3) satellite constellation there will be some observing times that are significantly better than others. This may not necessarily be the case for a 24 (21 + 3) satellite or 21 + remaining Block I satellite constellation.

- The behaviour of BDOP1 and BDOP2 (Figure 11) is similar, as they describe similar properties of the satellite configuration. That is, if it is easy to compute precise coordinates when ambiguities are also estimated, it must also be comparatively easy to solve explicitly for them. Interest in the BDOP2 factor is therefore rather limited.

- Furthermore, the exact definition of BDOP2 depends on the method used to overcome the rank defects in the solved-for ambiguity parameters (GRANT et al, 1989), and is therefore not uniquely defined.

- On the other hand, BDOP1 and BDOP3 describe the precision of the coordinates before and after ambiguity resolution and are not affected by the selection strategy used for the solve-for ambiguity parameters. Their definition is unique and they are therefore best suited for general use.

- BDOP3 is equivalent to an accumulated PDOP over an observing session and the peaks of the instantaneous PDOP indicator are smoothed, even for relatively short sessions of 30 minutes.

- The magnitude of BDOP3 is always smaller than the corresponding BDOP1 (or BDOP2), see Figure 11. However, the decrease of BDOP3 with an extension of the observing session length is not as pronounced as for BDOP1 (and BDOP2). This indicates that it is less important to resolve the ambiguities for longer sessions. In any case, it is usually BDOP1 that will distinguish a "good" observing session from a "bad" one.

- For some special purposes, the horizontal and vertical components of BDOP1 and BDOP3 can be extracted, just as HDOP and VDOP are derived from the cofactor matrix of the navigation fix (section 7.3).

- The BDOP units in the plots are dimensionless. They can be scaled into estimates of precision (or accuracy) for the estimated coordinates or ambiguities by multiplying them by the measurement precision (or accuracy). For carrier phase observations, the precision is of the order of several millimetres. However this multiplying factor could be proportional to the baseline length, in order to account for systematic errors.

7. SUMMARY AND CONCLUSIONS

7.1 Summary

GPS surveying is essentially a two step procedure:

(1) the tracking of a number of GPS satellites simultaneously from two or more ground sites, over a finite observing session, and
(2) the post-processing in one adjustment of the recorded carrier phase data.

For optimal planning of GPS surveys it is necessary to have an objective measure of the precision attainable for position determination using GPS surveying techniques, for different observation scenarios. At present, the "measures of precision", or "precision indicators", commonly available are based on GPS navigation principles, and are only of limited use for GPS surveying. For example, the Position Dilution Of Precision (PDOP) is a shorthand means of quantifying the influence of satellite(s)-receiver geometry for a navigation fix based on simultaneous pseudo-ranges to 3 or more GPS satellites. However the use of DOP factors derived from GPS navigation considerations is inappropriate for designing the observation scenario for two important reasons:

(a) The navigation solution is based on the **instantaneous** satellite-receiver geometry, whereas the GPS carrier phase adjustment is governed by the continually changing geometry of an observing session which may last for several hours.

(b) The pseudo-ranges are considered to be biased only by the errors of receiver oscillator (which changes from epoch to epoch and needs to be estimated or eliminated). In a GPS carrier phase adjustment the double-differenced observation (or equivalent) is predominantly biased by the integer cycle ambiguities (other biases still exist, but they are not related to satellite and measurement geometry). The geometric strength of an **ambiguous range** (integrated carrier phase) is quite different to that of a **pseudo-range**.

Furthermore, for short baselines, the reliable estimation of the cycle ambiguities and their subsequent resolution to their true integer values in an "ambiguity-fixed" solution, leads to an improvement in the quality of the results over and above that available from the initial "ambiguity-free" solution. However a reliable estimate of the integer biases is only possible from a strong ambiguity-free solution. The authors have therefore proposed an alternate set of precision indicators, referred to as BDOP (Bias Dilution Of Precision) factors:

- **BDOP1** for the precision of the coordinates before resolution of the cycle ambiguities,
- **BDOP2** for the precision of the determination of the cycle ambiguities,
- **BDOP3** for the precision of the coordinates after resolution of the cycle ambiguities.

The following characteristics of the BDOP factors were noted:

(1) The BDOP factors appear to be indicators of **accuracy** as well as **precision**. That is, periods of low BDOP indicate periods of lower sensitivity to (most) systematic errors.
(2) Times of low BDOP values do not coincide with times of low PDOP or GDOP values.
(3) There is a considerable variation in BDOP through the day, even for the operational GPS constellation.
(4) The longer the observing session, the less variation in BDOP through the day.
(5) Observing less than the full set of visible GPS satellites is suboptimal, and increasing the number of satellites tracked is more efficient than lengthening observing sessions.

The BDOP factors can be used for two purposes: (a) as an absolute measure of the likely accuracy of the GPS phase adjustment, or (b) to assist in GPS survey planning by providing a basis for judging the relative quality of different observation scenarios. With regard to the former, if the measurement precision (or accuracy) is known, the product of the measurement standard deviation (or User Equivalent Range Error) and the BDOP factors gives a measure of the precision (or accuracy) of GPS surveys:

(a) For long baselines, where the ambiguities cannot be resolved, BDOP1 is the appropriate indicator for the precision of the receiver coordinates and BDOP2 for the ambiguity parameters.

(b) For short baselines, where the ambiguities can be assumed to have been resolved, BDOP3 is the appropriate indicator for the precision of the receiver coordinates.

However, as GPS surveying is largely influenced by unmodelled systematic errors it is difficult to quantify the magnitude of the measurement errors, and hence the absolute coordinate accuracy to be expected or indicate whether ambiguities can be resolved or not. Therefore, precision indicators are unlikely to be used in this manner except perhaps as guidelines for "GPS standards and specifications". Rather, the BDOP factor of most interest for GPS survey planning is BDOP1.

An analysis of the relative magnitude of BDOP1 when: (a) the satellites to be tracked, (b) the observation start time, and (c) the observing session length are varied can assist in selecting the observation scenario that leads to the strongest ambiguity-free solution. Whereas varying the observation scenario in order to minimise the value of BDOP3 may result in more precise coordinates from an ambiguity-fixed solution, if BDOP1 is so bad then the resolution of cycle ambiguities is unreliable. In other words: **ambiguity resolution is not a miraculous way of improving a poor ambiguity-free solution because ambiguity resolution will always be unreliable in such a case.** BDOP1 has the same function as the fuel indicator of a car: more fuel will not make the car go faster, but with a low tank it might not be possible to reach the destination!

However, the decrease in the magnitude of BDOP3 with an extension of observing session length is not as pronounced as for BDOP1, hence it is less important to resolve ambiguities for longer sessions than for short ones. Furthermore, during periods of low BDOP1 the difference in the precision of an ambiguity-free and an ambiguity-fixed solution is a minimum, hence little is gained by resolving ambiguities under such circumstances.

Our suggestion for the use of such indicators is to therefore concentrate on the variability of BDOP1, and to use it as the single criterion for choosing the satellites to be tracked, the observation start time and the session length, by proceeding in the following manner:

(1) Decide on the satellites to be tracked. Usually this will be all that are visible, although if the GPS receiver can only track 4 or 5, then the following procedure would need to be repeated with all combinations of satellite subsets.

(2) By varying both the session length and the observation start time, compute the BDOP1 factor, and display the results as BDOP1 "contour" on an X-Y plot (with X-axis corresponding to observation start time, and the Y-axis corresponding to the session length), as suggested by HATCH & AVERY (1988).

(3) Select the observation scenario that gives the lowest BDOP1 subject to two planning constraints (for example, satellites tracked + observation start time, or satellites tracked + session length, or least likely, observation start time + session length).

However, it must be emphasised that **cycle ambiguities cannot be assured of being resolved simply by selecting the observation scenario with the lowest BDOP1**. It merely indicates which observation scenario has the best geometry for coordinate determination, and other factors such as the atmospheric refraction effect, multipath, orbit error, and, most importantly, the baseline length come into consideration and will affect the magnitude of the UERE (for example, UERE will increase with increasing baseline length). To assist the GPS survey planning process, it is therefore recommended that:

(a) The appropriate software for computing BDOP1 be included in the field preparation software, as is presently done for PDOP and GDOP.

(b) Government survey and mapping agencies draft guidelines for maximum BDOP, for cases of both short and long baselines, for each class of survey, and to delete all references to PDOP or GDOP factors.
(c) Further covariance studies be carried out in order to ascertain how the product of BDOP1 and UERE varies with increasing baseline length, and whether simultaneous orbit adjustment affects this, and hence develop a planning tool for simulating the conditions for cycle ambiguity resolution.

7.2 Computational Procedure

To compute the proposed indicators in one run, the following procedure is recommended:

1. Assume a stochastic model with independent one-way phases and one observation per minute.
2. Compute the complete accumulated normal matrix, using the singly-differenced cycle ambiguity approach. The dimension is $(3+m) \times (3+m)$, where m is the maximum number of satellites available.
3. Accumulate the normal matrix over the planned observation session duration.
4. Reduce the size of this matrix according to the actual number q' of satellites observed during the considered session. This gives $\mathbf{N}$, of dimension $(3+q') \times (3+q')$.
5. Invert $\mathbf{N}$, introducing the minimal trace condition. This gives the covariance matrix of the unknowns: $\mathbf{Q_{XX}}$.
6. The indicator of the precision for the coordinate determination is BDOP1: the trace of the 3×3 coordinate submatrix $\mathbf{Q_{CC}}$ of $\mathbf{Q_{XX}}$.
7. Isolate $\mathbf{N_{CC}}$ the coordinate submatrix (3×3, full rank) of $\mathbf{N}$.
8. The trace of the inverse of $\mathbf{N_{CC}}$ is BDOP3, the indicator of precision of the coordinates, under the assumption that all ambiguities have been resolved.

It should be emphasised that there is no need to decorrelate the observations in this approach. Furthermore no differencing operator needs to be explicitly defined. All correlations can be taken into account by using the operator for "reduction by the mean" to compute the normal matrix of the double-differenced observations and its associated covariance matrix (see Appendix A.2).

A valuable strategy for GPS planning is to study both at BDOP1 and BDOP3. If integer ambiguities can be resolved, an improvement of BDOP3 should result in more precise coordinates. BDOP1 need not be optimised but an upper limit to its value should be set, according to the performance of the equipment and field experience. BDOP1 has the same function as the fuel gauge of a car: more petrol will not help it go faster, but with a low tank, it might be simply impossible to arrive at the destination.

7.3 Customised Indicators

In some cases, other indicators may be more suitable. Generally, they can be extracted from the same covariance matrices, according to the guidelines given in Appendix D. As an example, a refined procedure is given, in which BDOP1,2 and 3 are replaced respectively by:

(a) Precision of the remote site with range difference measurements. A horizontal error ellipse $\sigma_A, \sigma_B, \theta$ (BHDOP1) and a vertical mean error σ_H (BVDOP1) are extracted from $\mathbf{Q_{cc}}$. This precision should be obtained even if no cycle ambiguities can be resolved.

(b) Mean error of the worst combination of integer ambiguities. The highest eigenvalue λ_{max} is extracted from $\mathbf{Q_{cc}}$. This value is independent of the selection of the reference ambiguity and the weakest linear combination of coordinates cannot be much different from the weakest linear combination of integer ambiguities. This indicator can be used for the choice of the session duration.

(c) Precision of the remote site with range measurements. A horizontal error ellipse $\sigma_A, \sigma_B, \theta$ (BHDOP3) and a vertical mean error σ_H (BVDOP3) are extracted from $\mathbf{N_{cc}}^{-1}$. This precision should be obtained if all cycle ambiguities can be resolved.

8. ACKNOWLEDGEMENTS

Comparisons with the work conducted simultaneously by Ron Hatch have provided an appreciated feedback. Bob King made useful suggestions for the improvement of this report. Chris Rizos provided support in integrating the BDOP computations into his PREDICT software. The University of New South Wales provided the computing facilities. The Swiss Science Foundation is gratefully acknowledged for the financial support of the first author and his family while studying in Australia.

9. REFERENCES

BIERMAN, G. J., 1977. **Factorization methods for discrete sequential estimation**. Academic Press, New York, 241pp.

BOCK, Y., GOUREVITCH, S.A., COUNSELMAN III, C.C., KING, R.W., & ABBOT, R.I., 1986. Interferometric analysis of GPS phase observations. **Manuscripta Geodaetica**, 11, 282-288.

CASPARY, W. F., 1987. Concepts of Networks and Deformation Analysis. Monograph No. 11, School of Surveying, University of N.S.W., 183pp.

DUPRAZ, H., 1979. La géodésie et les satellites. Institut de Géodésie et Mensuration, Ecole Polytechnique fédérale de Lausanne, publication no.18, 8pp.

GEIGER, A., & KAHLE, H.-G., 1982. Zum Dopplerverfahren in der Satellitengeodäsie: ein Überblick. Vermessung, Photogrammetrie, Kulturtechnik 6/82, 181-191.

GRANT, D.B., RIZOS, C., & STOLZ, A., 1990. Dealing with GPS baises: some theoretical and software considerations. In Unisurv S-38, "Contributions to GPS Studies", C. Rizos (ed.), report of the School of Surveying, University of N.S.W.

HATCH, R. R., 1987. Geometric Satellite Geodesy. A report of the Magnavox Advanced Products and Systems Company, 28pp.

HATCH, R.R. & AVERY, E.V., 1988. A strategic planning tool for GPS surveys. Pres. at ASCE Specialty Conference: GPS-88 Engineering Applications of GPS Satellite Surveying Technology, May 11-14, 1988, Nashville, Tennesse, U.S.A.

KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A., & COLLINS, J., 1987. **Surveying with GPS**. Ferd. Dümmlers Verlag, Bonn, 128pp.

LANDAU, H., & EISSFELLER, B., 1985. Optimization of GPS satellite selection for high precision differential positioning. Report 19, GPS research 1985 at Inst. of Astr. and Phy. Geodesy, Univ. FAF, Munich, F.R. Germany, 65-105.

LINDLOHR, W., & WELLS, D.E., 1985. GPS design using undifferenced carrier beat phase observations. **Manuscripta Geodaetica**, 10, 255-295.

MERMINOD, B., GRANT, D.B., & RIZOS, C., 1990. Planning GPS surveys - using appropriate precision indicators. submitted **Can.Surv.**

NORTON, T., 1987. Monitoring the precision of relative GPS positioning. Procs. of the Centenary GPS Conf., Dept. of Land Information, Royal Melbourne Institute of Technology.

APPENDIX A. DIFFERENCING GPS OBSERVATIONS

A.1 General Case

The theoretical presentation here is intended to introduce the reader to some of the data manipulation issues involved in the differencing of carrier phase data. For a more details the reader is referred to, for example, BOCK et al (1986), and LINDLOHR & WELLS (1985).

A set of observations is considered, represented by the linear functional model:

$$\mathbf{l} + \mathbf{v} = \mathbf{A} \cdot \mathbf{x} \tag{A-1}$$

and the stochastic model by the covariance matrix of the observations $\mathbf{C_l}$

A differencing operator $\mathbf{D}$ is applied to the system:

$$\mathbf{D} \cdot \mathbf{l} + \mathbf{D} \cdot \mathbf{v} = \mathbf{D} \cdot \mathbf{A} \cdot \mathbf{x} \tag{A-2}$$

leading to the new system of equations:

$$\mathbf{l'} + \mathbf{v'} = \mathbf{A'} \cdot \mathbf{x} \tag{A-3}$$

The new observations therefore are $\mathbf{l'} = \mathbf{D} \cdot \mathbf{l}$. The covariance of the new differenced observations is:

$$\mathbf{C_{l'}} = \mathbf{D} \cdot \mathbf{C_l} \cdot \mathbf{D}^T \tag{A-4}$$

Application of the minimum quadratic norm, $[\mathbf{v'}^T\mathbf{v'}]$ = minimum, leads to the normal matrix:

$$\mathbf{N'} = \mathbf{A'}^T\mathbf{P'A'} \qquad \text{where } \mathbf{P'} = (\mathbf{DC_lD}^T)^{-1} \tag{A-5}$$

A.2 Some Operators

Let us consider a sample of phase data involving 4 singly-differenced (between-receiver) observations involving a baseline. The covariance matrix of these differenced observations is (equation (2-9) $2 \cdot I_4$. We wish to form double-differences using various between-satellite differencing operators:

(a) The operator for fixed reference satellite differencing has the form:

$$\mathbf{D}_{\nabla a} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \tag{A-6}$$

and the covariance of the resultant double-differences is:

$$C_{\nabla\Delta\Phi} = D_{\nabla a}\, C_{\Delta\Phi}\, D_{\nabla a}^T = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} = 4 \cdot \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \tag{A-7}$$

(b) The operator for sequential satellite differencing has the form:

$$D_{\nabla b} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \tag{A-8}$$

and the covariance of the resultant double-differences is:

$$C_{\nabla\Delta\Phi} = D_{\nabla b}\, C_{\Delta\Phi}\, D_{\nabla b}^T = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix} = 4 \cdot \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{bmatrix} \tag{A-9}$$

(c) In this case, we can consider the differencing operator as being the product of the sequential operator and the appropriate decorrelating operator, as described in Appendix B.2:

$$D_{\nabla c} = L^{-1} \cdot D_{\nabla b} \tag{A-10}$$

According to the law of propagation of variances and the definition of L (given in Appendix B.1), the covariance matrix of this new combined operator is:

$$\begin{aligned} C_{D_{\nabla c}} &= (L^{-1} \cdot D) \cdot I \cdot (L^{-1} \cdot D)^T \\ &= L^{-1} \cdot (D \cdot D^T) \cdot (L^{-1})^T \\ &= L^{-1} \cdot (L \cdot L^T) \cdot (L^{-1})^T = I \end{aligned} \tag{A-11}$$

Similar results can be obtained with other differencing operators, but a different premultiplier L^{-1} is required.

(d) The "reduction by the mean" for n observations is defined by the operator:

$$K_n = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \dots & -\frac{1}{n} \\ \dots & \dots & \dots & \dots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1-\frac{1}{n} \end{bmatrix} \qquad \text{(A-12)}$$

It does not reduce the number of observations and introduces correlations, but it has some properties that makes it very useful compared to the other differencing operators:

(1) It is symmetric: $\mathbf{K_n}^T = \mathbf{K_n}$

(2) It is idempotent: $\mathbf{K_n} \cdot \mathbf{K_n} = \mathbf{K_n}$

Let us consider $\mathbf{D}$, a differencing operator of dimension $(n-1) \times n$, we have, in addition:

(3) $\mathbf{D} \cdot \mathbf{K_n} = \mathbf{D}$

(4) $\mathbf{D}^T \cdot (\mathbf{D} \cdot \mathbf{D}^T)^{-1} \cdot \mathbf{D} = \mathbf{K_n}$

These properties (3 & 4) in particular apply to the operators (a) to (c) described earlier. The product in (4) is called the kernel of the operator. For a set of uncorrelated original observations, it is easy to derive the normal matrix of the differenced observations in the following manner:

$$\begin{aligned} \mathbf{N'} = \mathbf{A'^T P' A'} &= (\mathbf{DA})^T \cdot (\mathbf{D} \cdot \mathrm{I} \cdot \mathbf{D}^T)^{-1} \cdot (\mathbf{DA}) \\ &= \mathbf{A}^T \cdot \mathbf{D}^T (\mathbf{D} \cdot \mathbf{D}^T)^{-1} \mathbf{D} \cdot \mathbf{A} \\ &= \mathbf{A}^T \cdot \mathbf{K_n} \cdot \mathbf{A} \end{aligned} \qquad \text{(A-13)}$$

Thus, the normal matrix does not depend on the choice of the operator **D** !

Applying **D** to observations already reduced by the mean highlights the properties of **K** :

$$\begin{aligned} \mathbf{N'''} &= (\mathbf{DKA})^T \cdot ((\mathbf{DK}) \cdot \mathrm{I} \cdot (\mathbf{DK})^T)^{-1} \cdot (\mathbf{DKA}) \\ &= \mathbf{A}^T\mathbf{K}^T\mathbf{D}^T \cdot (\mathbf{DKK}^T\mathbf{D}^T)^{-1} \cdot \mathbf{DKA} \\ &= \mathbf{A}^T\mathbf{KD}^T \cdot (\mathbf{DKKD}^T)^{-1} \cdot \mathbf{DKA} && \text{symmetry} \\ &= \mathbf{A}^T\mathbf{KD}^T \cdot (\mathbf{DKD}^T)^{-1} \cdot \mathbf{DKA} && \text{idempotence} \\ &= \mathbf{A}^T\mathbf{K} \cdot \mathbf{D}^T(\mathbf{DD}^T)^{-1}\mathbf{D} \cdot \mathbf{KA} && \text{neutral for differencing} \\ &= \mathbf{A}^T\mathbf{K} \cdot \mathbf{K} \cdot \mathbf{KA} && \text{kernel} \\ &= \mathbf{A}^T\mathbf{KA} \qquad = \mathbf{N'} && \text{(A-14)} \end{aligned}$$

More interesting however is the product:

$$\mathbf{N''} = \mathbf{A'}^T\mathbf{A'} = (\mathbf{KA})^T\cdot(\mathbf{KA}) = \mathbf{A}^T\mathbf{K}\cdot\mathbf{KA} = \mathbf{A}^T\mathbf{KA} = \mathbf{N'} \qquad \text{(A-15)}$$

This, again leads to the same normal matrix. The simple expression for the normal matrix is only due to the fact that **K** is symmetric and idempotent. Though it suffices to premultiply the design matrix by its transpose to obtain the normal matrix, the covariance matrix of the observations reduced by the mean is not the identity matrix, but $\mathbf{K}\cdot\mathrm{I}\cdot\mathbf{K}^T = \mathbf{K}$.

APPENDIX B. CORRELATED OBSERVATIONS

B.1 General Case

We consider the functional model: $\mathbf{l} + \mathbf{v} = \mathbf{A} \cdot \mathbf{x}$. According to BIERMAN (1977), if the stochastic model is defined by a positive definite covariance matrix $\mathbf{C_l}$, there exists an operator $\mathbf{L}$ such that:

$$\mathbf{L} \cdot \mathbf{L}^T = \mathbf{C_l} \tag{B-1}$$

and the modified observations $\mathbf{L}^T \cdot \mathbf{l}$ are uncorrelated. We can write a new system of observation equations:

$$\mathbf{L}^T \cdot \mathbf{l} + \mathbf{L}^T \cdot \mathbf{v} = \mathbf{L}^T \cdot \mathbf{A} \cdot \mathbf{x} \tag{B-2}$$

From the law of propagation of variances, we can compute the covariance matrix of these new observations:

$$\mathbf{C}_{\mathbf{L}^{-1}\mathbf{l}} = \mathbf{L}^{-1} \cdot \mathbf{C_l} \cdot (\mathbf{L}^{-1})^T = \mathbf{L}^{-1} \cdot (\mathbf{L} \cdot \mathbf{L}^T) \cdot (\mathbf{L}^{-1})^T = \mathrm{I} \tag{B-3}$$

Considering the observations sequentially, the problem is to extract from each new observation information that is not already contained in the previous ones. Such a procedure is also known as Gram-Schmidt orthogonalisation. However, a disadvantage of the resultant uncorrelated observations is that it is often difficult to obtain a geometric insight.

B.2 Pairwise Correlated Observations

Sequential differencing leads to the following covariance matrix, which has a particular significance in GPS data processing (equation (A-9)):

$$\mathbf{C_l} = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 & \cdots \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \cdots \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \cdots \\ 0 & 0 & -\frac{1}{2} & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{B-4}$$

A suitable choice for the operator matrix $\mathbf{L}$ is given by the algorithm presented in BIERMAN (1977), and referred to in HATCH (1987):

$$L_{1,1} = 1$$

$$L_{i+1,j} = \frac{-1}{2 \cdot L_{i,j}}$$

(B-5)

$$L_{i+1,j+1} = \sqrt{1 - \frac{1}{[2 \cdot L_{i,j}]^2}}$$

Explicitly, this leads to the matrix:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ -\frac{1}{2} & \sqrt{\frac{3}{4}} & 0 & 0 & \dots \\ 0 & -\frac{1}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 0 & \dots \\ 0 & 0 & -\sqrt{\frac{3}{8}} & \sqrt{\frac{5}{8}} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \qquad \text{(B-6)}$$

In the case of 4 observations, the inverse of **L** is the triangular matrix:

$$\mathbf{L^{-1}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}} & \sqrt{\frac{4}{3}} & 0 & 0 \\ \frac{1}{\sqrt{6}} & \sqrt{\frac{4}{6}} & \sqrt{\frac{9}{6}} & 0 \\ \frac{1}{\sqrt{10}} & \sqrt{\frac{4}{10}} & \sqrt{\frac{9}{10}} & \sqrt{\frac{15}{10}} \end{bmatrix} \qquad \text{(B-7)}$$

The fractions are intentionally left unreduced in order to make the systematic structure of the matrix more apparent.

An important feature of $\mathbf{L^{-1}}$ is its ability to be computed recursively using the following algorithm:

$$(L^{-1})_{1,1} = 1$$

$$(L^{-1})_{i+1,j} = \frac{1}{\sqrt{\left(\frac{2}{(L^{-1})_{i,j}}\right)^2 - 1}} \cdot (L^{-1})_{i,j} \qquad \text{for } j \leq i$$

(B-8)

$$(L^{-1})_{i+1,j+1} = \frac{1}{\sqrt{1 - \left(\frac{(L^{-1})_{i,j}}{2}\right)^2}}$$

Thus, it is possible to compute a new uncorrelated observation and a new row of the design matrix **A** , with only the new observation and the last uncorrelated one.

APPENDIX C. ACCUMULATING OBSERVATIONS

C.1 Principle

Adding an independent group of new observations to the system of observation equations is equivalent to adding the corresponding new normal equations to the system of normal equations. The advantage of the second method is that it requires the storage of a smaller amount of data.

C.2 Example with 2 Epochs

We consider the two independent (that is, uncorrelated) sets of observation equations:

$\mathbf{l_1} + \mathbf{v_1} = \mathbf{A_1} \cdot \mathbf{x}$ with an associated weight matrix $\mathbf{P_1}$

$\mathbf{l_2} + \mathbf{v_2} = \mathbf{A_2} \cdot \mathbf{x}$ with an associated weight matrix $\mathbf{P_2}$

Making one system out of these two, we obtain the new design matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_1} \\ \mathbf{A_2} \end{bmatrix} \qquad \text{(C-1)}$$

For independent sets of observations, the global weight matrix is then:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P_2} \end{bmatrix} \qquad \text{(C-2)}$$

The global normal matrix is easily obtained from:

$$\mathbf{N} = \mathbf{A^T} \cdot \mathbf{P} \cdot \mathbf{A} = \mathbf{A_1}^\mathbf{T} \cdot \mathbf{P} \cdot \mathbf{A_1} + \mathbf{A_2}^\mathbf{T} \cdot \mathbf{P} \cdot \mathbf{A_2} = \mathbf{N_1} + \mathbf{N_2} \qquad \text{(C-3)}$$

It is therefore the sum of the individual normal matrices.

APPENDIX D. INDICATORS OF PRECISION

D.1 Covariance Matrix

From least squares adjustments, the precision of the solution is given by the product of the covariance matrix of the unknowns (be they receiver coordinates or ambiguities) and the *a priori* precision of the observations (mean variance of the unit of weight). The covariance matrix has the dimension $u \times u$, where u is the number of unknowns:

$$\Sigma_{\mathbf{xx}} = \mathbf{Q}_{\mathbf{xx}} \cdot \sigma^2 = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1u} \\ q_{21} & q_{22} & \cdots & q_{2u} \\ \cdots & \cdots & \cdots & \cdots \\ q_{u1} & q_{u2} & \cdots & q_{uu} \end{bmatrix} \tag{D-1}$$

A matrix is not very suitable tool for a quick and simple evaluation of the quality of a least squares solution. There needs to be some "compression" of the information in the matrix elements into a simple expression, preferably a single number.

D.2 Principal Component Analysis (PCA)

As discussed in CASPARY (1987), it is possible to decompose the covariance matrix of u unknowns into **eigenvalues** and **eigenvectors** so that:

$$\mathbf{Q}_{\mathbf{xx}} := \begin{bmatrix} s_1 & s_2 & \cdots & s_u \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \dots & \lambda_u \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ s_2 \\ \cdots \\ s_u \end{bmatrix} \tag{D-2}$$

This process is also called "diagonalisation". An eigenvector indicates the direction of an axis of the mean error ellipsoid. If the eigenvectors are normalised, then each associated eigenvalue is the square of the length of the semi-axis.

D.2.1 Largest Eigenvalue

The largest eigenvalue is simply the variance of the worst linear combination of unknowns. In geometric terms, it is the square of the largest semi-axis of the mean error ellipsoid. Thus, it is an appropriate "quality indicator" for the precision of the determination of a combination of unknowns (coordinates and/or ambiguities).

D.2.2 Separation into Horizontal and Vertical Components

If the coordinate components unknowns had been selected to be the geographic triad (latitude, longitude and height) representation of position rather than the 3-D geocentric Cartesian coordinate component representation, the 2 x 2 submatrix of the horizontal components can be isolated and transformed into a (horizontal) error ellipse. This is simply a two-dimensional PCA. On the other hand, the square root of the diagonal term corresponding to the height coordinate component in the solution gives the vertical mean error of the solution. The sum of the squares of the individual components σ_A , σ_B and σ_H is still equal to the sum of the eigenvalues.

D.3 Dilution Of Precision (DOP)

The DOP is such a condensed form of quality indicator derivable from the covariance matrix of the solution. In the context of linear algebra, it is the root of the trace of the matrix. If we now define a global precision scalar for the unknowns:

$$\sigma_x^2 = \left[q_{11} + q_{22} + \ldots + q_{uu} \right] \cdot \sigma^2 \qquad \text{(D-3)}$$

and the dilution of precision factor:

$$DOP = \sqrt{q_{11} + q_{22} + \cdots + q_{uu}} \qquad \text{(D-4)}$$

We therefore have: $\sigma_x = DOP \cdot \sigma$

Note that the DOP is dimensionless and independent of the precision of the observations.

With normalised eigenvectors, the trace of the original covariance matrix remains unaffected by a PCA. Thus, the trace of the covariance matrix is equal to the sum of the eigenvalues. As all eigenvalues are positive, the trace is always larger than the largest eigenvalue. A geometric representation of the root of the trace is straightforward: it is the radius of a sphere that always contains the mean error ellipsoid. It therefore follows that the DOP factor is rather conservative (and consequently safe) "summary" of the information contained within the covariance matrix. However, compared with the largest eigenvalue, the main advantage of the trace as a quality indicator is that no PCA is required.

Example: the precision of the original observations, the carrier phases, is typically 1% of a cycle (or 2mm in terms of wavelength). If the DOP factor is 10, the precision of the unknown parameter estimate is 10% of a cycle (or 2cm). In this case, with a mean error of 0.1 cycle and if the estimated parameter is an ambiguity, the probability of "resolving" the ambiguity to its correct integer value is very high.

APPENDIX E. THE "PAIRED-SATELLITE" CONCEPT

E.1 Introduction

In GPS receivers designed for survey operation (as opposed to pseudo-range navigation applications) the variation of the measured carrier phase from epoch-to-epoch is known as the **accumulated** or **integrated carrier beat phase** (GRANT et al, 1990). It is generated by comparing the incoming (Doppler shifted) carrier signal with a signal generated by the receiver's internal oscillator. This beat frequency is therefore sometimes also referred to as the **Doppler frequency**, and its integral over a certain time period is the **Doppler count**. This observable gives information on the variation with time of the (ambiguous) range between satellite and receiver. The geometric content of this observable is quite different from that of range measurements themselves. In order to assist with an understanding of **range differences**, the concept of a **"paired-satellite"** is first established on the basis of purely geometric considerations. The paired-satellite is imaginary, but can always be associated with the real situation. The equivalence of range difference measurements from one satellite and range observations from its paired-satellite can then be established algebraically.

E.2 Geometric Derivation

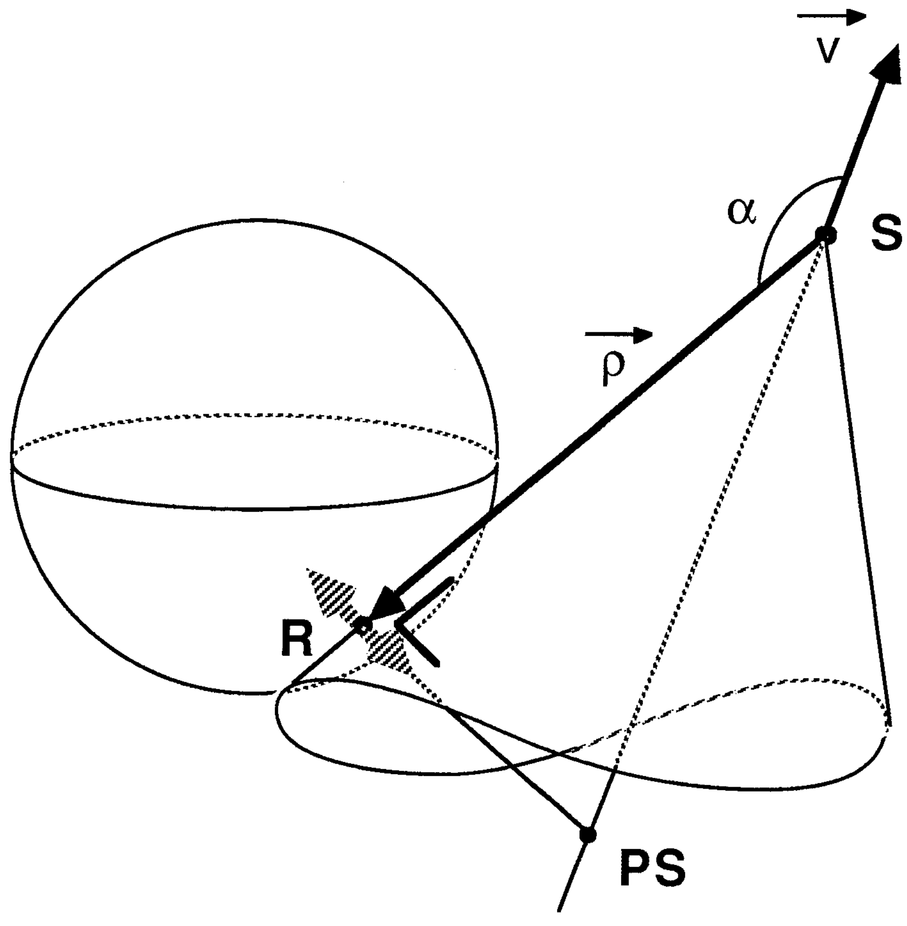


Figure E.1: Geometric relationship between the Paired-Satellites

DUPRAZ (1979) and GEIGER et al. (1982) have shown that the locus of points where a certain Doppler frequency can be observed is a hyperboloid, whose axis of revolution is defined by the velocity of the satellite with respect to an earth-fixed reference frame. For a large distance between the satellite and the receiver, this hyperboloid can be approximated by its asymptotic cone. It therefore follows that the positioning information obtained from the Doppler frequency observable is that in the direction of the perpendicular to the surface of the cone, thus in the direction PS $\leftrightarrow$ R, as indicated by the dashed arrow in Figure E.1. The same information is contained in the range measurement from PS to R.

E.3 Algebraic Derivation

The following development is based on **delta single-differences** $\delta\Delta\Phi(t)$. This phase observable is the variation with time of single-differenced observations. Therefore, it best represents the movement of a satellite with respect to a baseline, which is in this context a very convenient earth-fixed reference (Figure E.2).

The use of paired-satellites can be considered as a permutation of the differencing operations used to generate triple-differences. The normal differencing sequence $\delta\nabla\Delta\Phi(t)$ is replaced by $\nabla\delta\Delta\Phi(t)$. (This of course does not affect the final result.) The main advantage of this permutation is it permits the dynamics of one satellite to be studied before differencing between satellites. Time variations of double-differences $\delta\nabla\Delta\Phi(t)$ would be difficult to deal with because two satellites are involved, confusing the geometry of the situation.

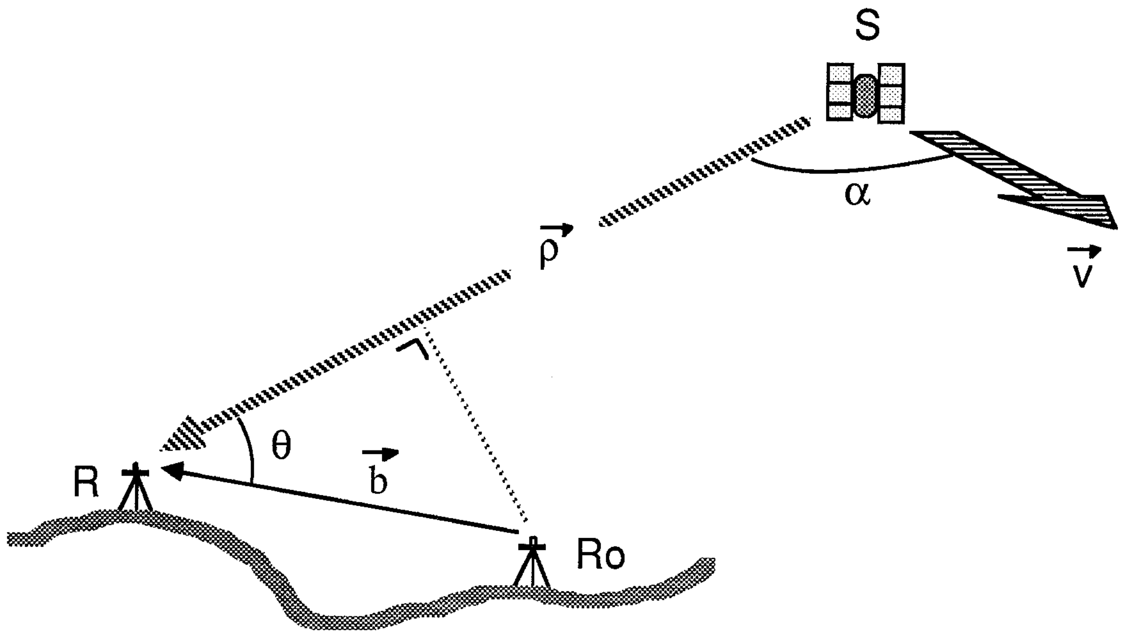


Figure E.2: Baseline Approach

In delta single-differences, satellite clock terms and cycle ambiguities have been eliminated by differencing between receivers and epochs respectively.

However, receiver clock terms are still present and the phase observation model therefore is:

$$\delta\Delta\phi^s(t) = (f_0 / c)\cdot\delta\Delta\rho^s(t) + f_0\cdot\delta\Delta\varepsilon_r(t) \tag{E-1}$$

Let us first consider the range term. For a short baseline, the difference between receivers of the ranges to the same satellite can be expressed as:

$$\Delta\rho^s = \rho_R{}^s - \rho_{Ro}{}^s = b \cdot \cos\theta \tag{E-2}$$

This is of course not exactly true, but the error introduced by this approximation is small for $b \ll \rho^s$ and for small variations in θ. The cosine term is the scalar product ("•") of the following unit vectors:

$$\cos\theta = \frac{\mathbf{b} \bullet \boldsymbol{\rho}}{b \cdot \rho} \tag{E-3}$$

where **b** and b are the baseline vector and baseline length respectively, $\boldsymbol{\rho}$ and ρ are the remote site-satellite vector and its magnitude. A delta single-differenced range over a short time period δt can be approximated as:

$$\delta\Delta\rho = \frac{\partial(\Delta\rho)}{\partial t}\cdot \delta t = b \cdot \frac{\partial(\cos\theta)}{\partial t}\cdot \delta t \tag{E-4}$$

The explicit differentiation with respect to time gives:

$$\frac{\partial(\cos\theta)}{\partial t} = \frac{\mathbf{b}}{b} \bullet \frac{\partial}{\partial t}\left[\frac{\boldsymbol{\rho}}{\rho}\right] \tag{E-5}$$

where

$$\frac{\partial}{\partial t}\left[\frac{\boldsymbol{\rho}}{\rho}\right] = \frac{\mathbf{v}\cdot\rho - \boldsymbol{\rho}\cdot\dot{\rho}}{\rho^2} \tag{E-6}$$

The range-rate $\dot{\rho}$ can be expressed as $v\cdot\cos\alpha$, where v is the norm of the velocity vector (see Figure E.2). Substituting this into equation (E-5) gives:

$$\frac{\partial(\cos\theta)}{\partial t} = \frac{v}{\rho}\cdot\frac{\mathbf{b}}{b} \bullet \left[\frac{\mathbf{v}}{v} - \frac{\boldsymbol{\rho}}{\rho}\cdot \cos\alpha\right] \tag{E-7}$$

The expression in brackets on the right hand side is a vector. Its direction corresponds to that of PS $\rightarrow$ R , as indicated in Figure E.1. At this stage, we refer to it as the PS range vector ρ^{PS} . Some of its properties will be described later. The model for the delta single-differenced range therefore becomes:

$$\delta\Delta\rho = \frac{v}{\rho} \cdot \mathbf{b} \bullet \rho^{PS} \cdot \delta t = \frac{v \cdot \delta t}{\rho} \cdot \rho^{PS} \bullet \mathbf{b} \tag{E-8}$$

This is a linear expression involving the baseline vector **b** . To write a linear adjustment model for the position of the remote site, the baseline is expressed as the difference in coordinates between the control site (subscript "o") and the remote site (superscript "~" for approximate coordinates and prefix "d" for the coordinate increments):

$$\mathbf{b} = \begin{bmatrix} \tilde{X} + dX \\ \tilde{Y} + dY \\ \tilde{Z} + dZ \end{bmatrix} - \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} = \begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} + \begin{bmatrix} \tilde{X} - X_o \\ \tilde{Y} - Y_o \\ \tilde{Z} - Z_o \end{bmatrix} \tag{E-9}$$

This leads to:

$$\delta\Delta\rho = \frac{v \cdot \delta t}{\rho} \cdot \rho^{PS} \bullet \begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} + (\tilde{k} - k_o) \tag{E-10}$$

where the term containing the *a priori* station coordinate information is:

$$(\tilde{k} - k_o) = \frac{v \cdot \delta t}{\rho} \cdot \rho^{PS} \bullet \begin{bmatrix} \tilde{X} - X_o \\ \tilde{Y} - Y_o \\ \tilde{Z} - Z_o \end{bmatrix} \tag{E-11}$$

The coordinate of the remote site can be computed in any desired coordinate system. It suffices only to express the PS range vector in that system. It is generally more convenient to use a local topocentric system, and the vectors for the range and the velocity of the satellite are transformed accordingly, that is the position vector $(dX, dY, dZ)^T$ can be replaced by $(dE, dN, dH)^T$.

Finally, the relation between measurements and unknown coordinates and biases is defined. For each satellite and combination of two epochs, we can write the complete phase observation equation (equation (E-1)), in units of range, as $\text{residual} = v = A \cdot x - l$, that is:

$$\nu = \frac{v \cdot \delta t}{\rho}\left[\frac{\mathbf{v}}{v} - \frac{\rho}{\rho}\cos\alpha\right] \cdot \begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} + f_0 \cdot \delta\Delta\varepsilon_r - \left[\frac{f_0}{c}\delta\Delta\phi - (\tilde{k} - k_o)\right] \quad \text{(E-12)}$$

The only term dependent on the position of the remote site is in the constant part (last bracketed term). As the coefficients of the unknowns are free of baseline components, neither the azimuth, nor the slope, nor the distance between the sites have an influence on the precision of the unknowns (they will of course affect the accuracy). It may be possible to explicitly model the receiver clocks accurately (for example, if an external atomic clock is used to make the phase measurements). In such a case, the clock errors enter into the constant term as well.

E.4 Equivalent Range Observations

A row of the design matrix for the functional model of range difference derived above has the form:

$$A_i = \frac{v \cdot \delta t}{\rho} \cdot [\, \rho^{PS} \,]^T \quad \text{(E-13)}$$

The corresponding observations are considered to be independent and their weight matrix is of the form $\mathbf{P} = I$ (or a multiple of the identity matrix).

Now let us consider a range to the paired-satellite, from the original expression of the PS range vector, its norm is easily obtained:

$$\rho^{PS} = \sqrt{1 - (\cos\alpha)^2} \quad \text{(E-14)}$$

The unit PS range vector can be computed and the design matrix $\mathbf{A'}$ can be defined by rows of the form:

$$A'_i = \frac{[\, \rho^{PS} \,]^T}{\rho^{PS}} \quad \text{(E-15)}$$

The corresponding weight matrix $\mathbf{P'}$ has the structure:

$$P'_{ij} = \left(\frac{v \cdot \delta t \cdot \rho^{PS}}{\rho}\right)^2 \text{ if } i=j \quad \text{and} \quad 0 \text{ otherwise} \quad \text{(E-16)}$$

Each diagonal term is associated with one observation i and $\mathbf{P'}$ is generally not a multiple of the identity matrix.

It can be easily shown that:

$$\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A} \; = \; \mathbf{A'}^T \cdot \mathbf{P'} \cdot \mathbf{A'} \qquad \text{(E-17)}$$

Hence both models lead to the same normal matrix. The advantage of the second approach is that **A'** has the form of a design matrix for range measurements. This therefore proves that a range-rate measurement to one satellite has the same effect as measuring a range in the direction of the corresponding PS range vector. A consequence of this is that we can consider that the PS range vector indicates the location of the paired-satellite, to which range measurements are assumed to have been made.

The position of the paired-satellite ρ^{PS} is a linear combination of the position and the velocity of the satellite. The paired-satellite is in the plane containing both the vector from the receiver to the satellite and the instantaneous velocity vector of the satellite. For GPS, the angle between the position and the velocity of the satellite is generally close to a right angle and, consequently, the corresponding cosine term is small. It follows that the location of the paired-satellite is roughly 90 degrees behind the satellite, or delayed by a quarter of a revolution. Note that a paired-satellite may well be below an observer's horizon, though of course the actual satellite cannot be.

To visualise the situation, it is helpful to indicate the tracks of the paired-satellites on a sky-plot. The suitability of a particular satellite configuration for GPS phase (or range difference) positioning can be easily evaluated: it necessary only to study the configuration of the paired-satellites and consider the strength of the resection of range measurements from them.

The weight of the range measurement to the paired-satellite depends strongly on the specifications of the satellite system, as a quick comparison of GPS and TRANSIT for zenithal observations over one minute shows (equation (E-16)):

(a) GPS: speed = 3.8 km/s, range = 20200 km --> $p' = (0.01)^2$

(b) TRANSIT: speed = 7.2 km/s, range = 1100 km --> $p' = (0.4)^2$

Considering only one satellite, GPS is therefore not very appropriate for range difference measurements, as opposed to TRANSIT. This drawback is more than compensated by the simultaneous availability of several GPS satellites, allowing for the elimination of receiver clock errors.

An extension of the observing period has a different effect on the weighting of range and range-rate measurements. If p_i the weight of a single range measurement, the weight of an accumulation of observations increases in proportion to the number of epochs e (and hence observations):

$$p_\Sigma \; = \; e \cdot p \qquad \text{(E-18)}$$

In the case of the weight matrix **P'** , it appears that accumulating range difference observations makes the global weight increase in proportion to the square of the length of the observing session. If additional measurements are made in between (that is, increase the observation density), they will also improve the weight. For observations at regularly spaced epochs and if p' is the weight of a range difference measurement between two successive epochs:

$$p'_{\Sigma} = \sum_{i=1}^{e-1} (i^2 \cdot p') \qquad \text{(E-19)}$$

This result is more obvious if each new measurement is considered with respect to the first measurement, adding information on the change in range during the time elapsed since the first epoch. The weight propagation shows the importance of the correlations between range difference observations at successive epochs. This also explains why solving for ambiguities is less important for long sessions, since the increase of the geometric strength of range-difference measurements with time is greater than for range measurements.

E.5 Indicators Based on Paired-Satellites

It is not always possible to remove the receiver clock error from the functional model. In such a case the model of the delta single-differences for satellite i reads:

$$\delta\Delta\phi^i(t) = \begin{bmatrix} \sqrt{p'^i(t)} \cdot [\,\rho^{PS}(t)\,]^T & 1 \end{bmatrix} \cdot \begin{bmatrix} dX \\ dY \\ dZ \\ \delta\Delta\varepsilon_r(t) \end{bmatrix} + (\tilde{k}^i - k_o^i) \qquad \text{(E-20)}$$

In the design matrix, the partial derivative for a coordinate is the product of the square root of the weight by the corresponding component of the unit-vector from the paired-satellite to the control station. For GPS, due to the high orbits, values of p' for different satellites and epochs have typically a small range. Thus, the system of equations can be expressed by:

$$\begin{bmatrix} \delta\Delta\phi^1(t) \\ \cdots \\ \delta\Delta\phi^1(t) \end{bmatrix} = \sqrt{p'} \cdot \begin{bmatrix} \rho^{PS1}_{uE} & \rho^{PS1}_{uN} & \rho^{PS1}_{uH} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ \rho^{PSn}_{uE} & \rho^{PSn}_{uN} & \rho^{PSn}_{uH} & 1 \end{bmatrix} \cdot \begin{bmatrix} dX \\ dY \\ dZ \\ \frac{\delta\Delta\varepsilon_r(t)}{\sqrt{p'}} \end{bmatrix} + (\tilde{\mathbf{k}} - \mathbf{k}_o) \qquad \text{(E-21)}$$

Range terms also change at each epoch, but for convenience, this is not shown explicitly. The design matrix has exactly the same structure as that for the single point positioning (navigation) solution using pseudo-ranges. For the range components, the actual satellites are replaced by their corresponding paired-satellites. The receiver clock unknown is simply divided by the square root of the weight.

Hence, the relation between the actual and fictitious satellites established in the previous sections changes slightly if the receiver clock errors are taken into account: measuring pseudo-range differences from a satellite is equivalent to observing pseudo-ranges from its paired-satellite. DOP factors based on paired-satellites can be computed, describing the instantaneous geometry of range-rate measurements, analogous to navigation PDOP and GDOP factors. We will refer to these as PS-DOP factors.

DOP factors can be computed for the paired-satellites (PS-DOP) using the classical algorithms (for example, LANDAU & EISSFELLER, 1985). However, as they are related to an instantaneous solution (based on range-rate observations as opposed to range measurements) they still have some of the drawbacks of usual DOP factors:

- 4 satellites are required unless some unknowns are constrained (for example by fixing height), though it is possible to compute a position completely with less satellites (but more epochs).
- the effect of data accumulation is not taken into account. As pointed out several times, this effect is particularly strong for range difference observations.

To summarise, there is a similar relation between PS-PDOP and BDOP1 as there is between PDOP and BDOP3.

A STUDY OF THE VARIABILITY AND PRECISION OF GPS BASELINE MEASUREMENTS

C.C. Mazur[1]

School of Surveying
University of New South Wales
P.O.Box 1
Kensington, N.S.W., 2203, Australia

ABSTRACT

The length of observations required by GPS receivers within a network is generally a function of:

a. satellite geometry,
b. receiver / hardware characteristics,
c. location on the earth's surface,
d. the conditions prevailing at each survey site at the time of observations, and
e. the prevailing atmospheric conditions at the time.

The indicators of precision developed by MERMINOD (1990) have been designed for use in precise GPS surveys. These indicators have been empirically evaluated in the field and are discussed in this report.

To examine the relationship between real data and the precision estimates determined during the survey planning process, a survey campaign was conducted using a baseline adjacent to the University of New South Wales. Up to three WM102 receivers were used throughout the campaign. After several days of observations, four days of WM102 data was selected for processing. This report discusses:

a. the conduct of the campaign,
b. a description of the WM102 GPS receiver and the PoPS™ software,
c. the results obtained,
d. problems encountered, and
e. areas requiring additional research.

[1]Now at the School of Military Survey, Bonegilla, Victoria, Australia.

1. INTRODUCTION

There are a number of important factors to be considered prior to the conduct of any Global Positioning System (GPS) geodetic survey. These factors include the length of time that each station in the network is going to be occupied, the satellites to be observed, the type of receiver to be used, and the start time of observation. Once the satellites to be observed have been selected, a simulation of the campaign may be conducted using this information, together with the approximate coordinates of the stations to be occupied. From these simulations an estimate of the dilution of precision (DOP) is calculated which can be used to estimate the accuracy of the relative station positions. Thus the strength of solution for the survey may be determined during the survey planning stage. Work done by HATCH & AVERY (1988) and MERMINOD et al (1990) has resulted in the development of precision indicators. These indicators: Hatch's Differential GDOP (DGDOP) and Merminod's Bias DOP (BDOP) have been designed for use in the survey planning stage to determine (in part) the observation schema of the GPS campaign. Merminod's BDOP precision indicators were used exclusively throughout the GPS campaign described in this report.

The present constellation (July 1990) offers Australian surveyors an observation window of approximately 12 hours during which observations to four or more satellites may be made. However the satellite-receiver geometry varies considerably within this window. Surveyors are currently collecting one hour of data for baselines of the order 10 to 30 km with the understanding that given a reasonable satellite coverage they should be able to resolve the integer cycle ambiguities and thus achieve good results. As surveyors are unable to determine in the field when they have sufficient data for this to occur, session lengths tend to be on the conservative side.

For surveys requiring high accuracy, if the integer cycle ambiguities are not all resolved for a given observation session duration, generally the surveyor will repeat the survey using a longer observation period. At present, the duration of the observation window is increasing as each new Block II satellite is launched, and with the full 24 satellite constellation being planned for completion by 1993-94, conducting GPS surveys in this manner is not the most economical use of resources. Once the full constellation has been deployed and surveyors have 24 hour GPS survey capability, session times will need to be planned so that they are sufficiently long enough to result in a solution which is of the required quality with the minimum amount of field time.

1.1 Project Aims

In defining the aims of this project, it is important that they are simple and that most of them are achievable. Whilst this is important for all projects, when working with GPS, unless the aims of any research conducted are simple, then very quickly the interpretation of results obtained can become very difficult. This is due to the number and complexity of the influences affecting GPS survey measurements.

The aims of this project therefore were to:

(1) Confirm a correlation between the precision indicators and the final precision obtained from field measurements.

(2) Investigate the precision of coordinates obtained using static relative position survey techniques.

(3) Attempt to determine prudent observation session lengths given that the relationship between precision indicators and precision values of field measurements has been established.

(4) Investigate ambiguity resolution and the conditions for it to occur.

The research successfully demonstrated that Merminod's BDOP precision indicators determined prior to survey can be a good estimate of the actual precision of station coordinates computed from GPS data recorded during an observation session. The difference in precision values between "predicted" precision estimates and the precision of real data were caused by unmodelled systematic errors which corrupted the quality of our measurements. These are not taken into account in Merminod's indicators. The most significant effect on our measurements is caused by the propagation medium which limits the attainable precision of geodetic GPS measurements.

The software used throughout the campaign, PoPS™ (FREI et al, 1986), was unable to consistently resolve the integer cycle ambiguities in approximately 30% of the observation sessions. This coupled with the unmodelled systematic errors affecting our attainable precision prevented an investigation being undertaken into the determination of optimal session durations. Further investigation will become feasible once more sophisticated modelling techniques are developed which better account for the systematic errors encountered throughout GPS campaigns. It is understood that WILD is due to release an improved version of their PoPS™ software which will use an "enhanced ambiguity resolution" technique (HOAR, 1989). Initial tests conducted by WILD employing this technique claim to have achieved reliable cycle ambiguity resolution with only several minutes of data.

2. PRECISION INDICATORS

There are three design parameters which appropriate "precision indicators" could make a contribution to in GPS survey planning:

- Determining the number of satellites to be observed within a session.
- Determining the optimum time for the commencement of observations.
- Determining the length of observation session commensurate with the accuracy sought.

The number of satellites which can be observed simultaneously varies with the receiver type. For example, the TI-4100 GPS receiver can only observe four satellites simultaneously, whilst the WM102 receiver is capable of tracking up to six satellites simultaneously. Receivers such as the ASHTECH XII and the TRIMBLE 4000ST are able to observe all the satellites "in-view".

The present window in which GPS surveys can be conducted is approximately 12 hours. Even with 24 hour coverage some observation times may be better than others. However survey productivity will be dependent on matching the length of the observation sessions with the receiver-satellite geometry experienced for that session and the survey accuracies sought. (Hence some sessions may be required to be longer than others.) The accuracy of GPS results are generally a function of:

(a) the unmodelled systematic errors influencing the observations,
(b) inaccurate or incomplete observation modelling,
(c) the manner in which the data is processed, and
(d) the receiver-satellite geometry prevailing at the time of the survey.

2.1 Errors and Biases in GPS Observations

The precision attainable from geodetic GPS measurements is degraded by the presence of systematic errors or biases. Errors and biases may be divided into three categories: satellite-dependent, observation-dependent and station-dependent errors.

2.1.1 Satellite-Dependent Errors

These errors are principally:

- Errors in the broadcast ephemeris (and hence in the location of the satellite relative to the ground stations).
- Satellite clock error is a (time dependent) offset existing between a regular time scale such as GPS time and the time kept by the individual satellites.
- Internal delays within a satellite that would influence all the signals transmitted by that satellite.

2.1.2 Observation-Dependent Errors

These errors are generally caused by the propagation medium. The troposphere and ionosphere impede the travel of a satellite signal, and hence cause the modelled range to deviate significantly from the observation. The integer cycle ambiguity is also an observation dependent error (more correctly a "bias") which is characteristic of carrier beat phase observations by a receiver to a specific satellite.

2.1.3 Station-Dependent Errors

Examples of these errors are:

- Receiver multipath disturbance.
- Antenna phase-centre movement.
- Centring error.
- random measurement error
- Receiver clock error.
- Cycle slips in the carrier beat phase observations.

GPS data processing attempts to minimise the influence of these errors through a combination of modelling, estimation and elimination techniques. However, imperfections in this process may result in residual biases. All of these errors and residual biases may be modelled by an "average" standard deviation of uncorrelated "user equivalent range error" (UERE). UERE is the satellite-receiver range error resulting from the total combination of biases affecting GPS measurements.

2.2 User Equivalent Range Error (UERE)

The magnitude of the UERE is determined from the root-sum-square (RSS) of both the estimated GPS measurement noise and systematic errors (or biases). WELLS et al (1986) gives the following estimates of the magnitude of the primary components of the UERE:

(a) The satellite clock bias can contribute anything from hundreds of kilometres to a minimum of about a dekametre if the broadcast clock correction model is used. The receiver clock bias is highly variable and is dependent on such factors as the quality of the clock and how recently it has been synchronised to GPS Time.
(b) The satellite orbital bias is of the order of 10-100 m.
(c) The ionospheric delay bias is a function of zenith angle, being a minimum at the zenith (about 10 m) and a maximum at the horizon.
(d) The tropospheric delay bias is a minimum at the zenith (about 2 m) and approximately 20 m at 10^{o} elevation.

2.3 Dealing With GPS Biases

As already stated, GPS processing techniques attempt to overcome the effects of biases by modelling, estimation and elimination (differencing) techniques. In the differencing process, the carrier phase observations are differenced **between satellites** to minimise receiver dependent systematic errors (GRANT et al, 1990). Differencing **between receivers** minimises the influence of satellite systematic errors. The resulting double-differenced observations, in GPS survey processing, are then modelled in terms of satellite and receiver coordinates and (double-differenced) cycle ambiguity parameters.

The precision with which the station coordinates and ambiguity parameters can be determined is a function of the number of observations taken, the receiver-satellite geometry (influencing the "strength" of the solution), and any residual biases still present in the double-differenced observations (influencing the noise and/or systematic errors in the solution). It would be of great benefit to a surveyor to be able to determine prior to a survey the precision that can be attained for a planned GPS survey. There are two options (MERMINOD, 1990):

(a) A full simulation of phase observation reduction, or

(b) Enhancement of existing approximate quality "indicators" developed for use in GPS navigation.

The latter approach is the preferred option as it does not require complex procedures or software.

2.4 GPS Navigation Precision Indicators

In essence, GPS navigation precision is influenced by many of the same factors that influence GPS surveying, however, the conduct of the GPS survey campaign is quite different:

- In GPS navigation a single mobile receiver is normally used whereas GPS surveying techniques employ several receivers to derive relative position.
- The position determined by a GPS receiver used in navigation mode is an instantaneous one determined using the **pseudo-range observation** (a range contaminated mainly by the receiver clock error). In GPS surveying, **integrated carrier beat phase measurements** (see Appendix 3) collected over an observation session are processed.
- The carrier beat phase observation is obtained through "beating" the incoming carrier phase with a signal produced by the receiver oscillator at the same nominal frequency. If this phase can be measured with an accuracy of 1% of a cycle, the measurement precision is approximately 2 mm which is two orders of magnitude better than ranging with the P code and three orders of magnitude better than the C/A code ranges (WELLS et al, 1986). The carrier beat phase is biased by errors in the receiver and satellite oscillators and also includes the additional bias of the integer cycle ambiguity.

The pseudo-ranges measured by a receiver are calculated from the time interval for a signal to travel from a satellite to a receiver, scaled by the speed of light. There is no synchronisation between the clocks maintained onboard the GPS satellites and the clock in a GPS receiver. As a consequence, the range determined between a satellite and a receiver is contaminated by a receiver-satellite clock offset, which is why this receiver-satellite range is referred to as a pseudo-range. This pseudo-range measurement is also biased by all other systematic errors influencing the GPS survey. The pseudo-range measurement is however very noisy and although systematic errors can be eliminated / minimised, the noise from the measurement remains a problem.

The use of the pseudo-range is fundamental to the determination of an instantaneous position solution in GPS navigation. Using four satellites, the three dimensional receiver coordinates are determined plus the receiver-satellite clock offset. If the receiver is used at sea, the height can be constrained and so a minimum of three satellites can be simultaneously observed to obtain an instantaneous navigation fix. With more than three or four satellites being observed (depending on the mode used), the additional redundancy obtained allows a least squares determination of the receiver coordinates and the clock offset. This increased redundancy improves the precision of the coordinates of the occupied stations. The cofactor matrix (inverse of the normal matrix) (equation 2-2) can be used to study the precision of the coordinates (from the diagonal elements), and information on the correlations between parameters (from the off-diagonal elements).

2.4.1 PDOP and GDOP

A number of precision indicators have been developed for use with GPS navigation. The traditional error ellipse (2 dimensional) or error ellipsoid (3 dimensional) is approximated in GPS navigation by the error sphere whose radius is the root-sum-squared (RSS) of the ellipsoidal axes. This is the **Position Dilution of Precision (PDOP)**. The magnitude of PDOP is the radius of this error sphere. The precision of a GPS navigation position is now estimated by one value instead of the six required to define an error ellipsoid (one defining the length of each axis, and the three orientations). The PDOP indicator is calculated from the instantaneous relative geometry of four or more satellites and the point whose coordinates are to be determined.

A modification of PDOP is known as the **Geometrical Dilution of Precision** (GDOP) as it takes into account the precision of the receiver-clock offset. In the navigation mode, if a GPS receiver can only track a certain number of satellites out of all that may be visible, then the best satellites to track will be those which have the lowest GDOP.

PDOP and GDOP vary in the same manner over an observational session as the precision of a station coordinates will be poor if the receiver-satellite clock offset is also poorly estimated. Because of the close relationship between the two DOP's, often only PDOP needs to be evaluated. This similarity can be seen in examples of PDOP and GDOP for 8th June 1989 and 12th September 1989 given in Appendix 1. For GPS navigation, a PDOP of five or less is desirable. PDOP or GDOP can be pre-determined in the planning phase, and is constant from one day to the next for the same set of satellites and at a time 4 minutes earlier than the previous day. Multiplying PDOP by the UERE will provide an indication of the three dimensional positioning accuracy. For example, for a given navigation fix, a UERE of 20m and a PDOP of four (4) will lead to the navigational position fix having an estimated accuracy of 80m. Whilst PDOP provides an estimate of the accuracy of an instantaneous navigation point position it is unsuited for use with GPS surveying as survey campaigns may last several hours and involve phase tracking by several receivers.

2.5 GPS Survey Precision Indicators

MERMINOD (1990) has developed GPS survey DOP's which are derived from simulations of carrier phase adjustments. The basic double-differenced model reduces the satellite and receiver dependent biases. The resultant model consists of a geodetic part containing the relative coordinate components between two receivers and the integer cycle ambiguity part (there is one cycle ambiguity term for each two satellite-two receiver combination). The resultant normal matrix which contains all observations obtained during a session as:

$$\mathbf{N}_{xx} = \begin{bmatrix} \mathbf{N}_{cc} & \mathbf{N}_{cb} \\ \mathbf{N}_{bc} & \mathbf{N}_{bb} \end{bmatrix} \qquad (2\text{-}1)$$

Where: $\mathbf{N}_{xx}$: singular matrix with rank defect = 1

$\mathbf{N}_{cc}$: coordinates of stations

$\mathbf{N}_{bb}$: cycle ambiguities

N_{cb} , N_{bc} are cross-products of N_{cc} , N_{bb}

Equation (2-1) contains the contributions of all the measurements which were made during the session. If we neglect components for one station (to overcome the rank defect) the reduced normal matrix is of dimension (number of receivers -1) + (number of tracked satellites -1). Inverting this produces the cofactor matrix which can be partitioned as shown below:

$$Q_{xx} = \begin{bmatrix} Q_{cc} & Q_{cb} \\ Q_{bc} & Q_{bb} \end{bmatrix} \tag{2-2}$$

This solution is the **ambiguity-free solution**. If sufficient observations have been made (with sufficient solution strength), the environmental conditions benign, and the baseline short, the estimated ambiguity terms may be close to their integer values. If resolved correctly to the nearest integer the processing can be repeated, removing the resolved ambiguities from the parameter set, to obtain an **ambiguity-fixed solution**. This solution will always be more precise than the ambiguity-free solution as it essentially involves the analysis of unambiguous phase measurements with millimetre precision.

2.5.1 Planning a GPS Survey

When planning a GPS survey, the only observational parameter that the surveyor can generally alter is the start time and duration of the observation session. All other parameters (satellite constellation, satellite-receiver geometry and receiver type) are essentially beyond his control. Currently, GPS surveys of short to medium baselines are conducted with 60 minute observation session lengths to maximise the chances of a ambiguity-fixed solution being obtained. Where the ambiguities have not been resolved, sometimes the session is repeated, possibly with an extension being made to the observation duration.

The BDOP precision indicators may be able to provide, at the survey planning stage, an estimate of the precision which can be anticipated when processing carrier beat phase observations for either an ambiguity-free or ambiguity-fixed solution. This provides an objective basis by which the relative quality of different observation scenarios can be assessed. The three precision indicators provide the relative precision of:

(1) coordinates in an ambiguity-free solution (BDOP1),
(2) ambiguity parameters in an ambiguity-free solution (BDOP2), and
(3) coordinates in an ambiguity-fixed solution (BDOP3).

The absolute coordinate accuracy is difficult to ascertain due to the presence of unmodelled systematic errors. (Note we are referring to the absolute accuracy of the relative position of one end of a baseline with respect to the other.) At present, our lack of understanding of these errors makes the forecasting of an observation session duration sufficiently long enough to resolve the cycle ambiguity virtually impossible. This is particularly true under the effect of "selective Availability", which is a deliberate error introduced by the GPS system controllers. Of the three indicators, BDOP1 is of greatest interest in the survey planning process (MERMINOD et al, 1990).

2.5.2 BDOP Units

The units of the BDOP indicators are dimensionless. Multiplying these values by the UERE, can provide an estimate of the precision of the estimated coordinates. For carrier phase observations, the precision is in the order of several millimetres. But this does not include any systematic errors (or "bias"). This multiplying factor could be proportional to baseline length to account for unmodelled systematic errors. Values for BDOP indicators can be determined from only two satellites, this being the minimum number of satellites required (to form a double-difference) for GPS surveying. The major advantage in using these indicators is in comparing one session with another in the planning stage, using only the relative values of the indicators.

2.5.3 BDOP as an Estimate of Accuracy

BDOP indicators can be used to provide an estimate of the accuracy of the GPS phase adjustment or as a measure of relative quality between a number of alternate scenarios. BDOP factors scaled by the UERE will result in an estimate of the accuracy of a GPS survey. For a long baseline, where the ambiguities cannot be resolved, BDOP1 provides the precision of the receiver coordinates from an ambiguity-free solution. Where the ambiguities can be assumed to have been resolved, BDOP3 will indicate the accuracy of the receiver coordinates from this ambiguity-fixed solution.

2.6 Computing BDOP Indicators (MERMINOD et al. 1990)

The design matrix for the carrier phase adjustment is required for the calculation of the BDOP indicators. The design matrix is computed using approximate positions of the satellites and receivers. Three important assumptions are made:

(1) we only consider short baselines,
(2) all carrier phase measurements are assumed independent and have the same precision, and
(3) the length and orientation of the baseline is irrelevant for relative positioning.

Only the coordinates of one station in the network is required: the indicator being valid for all short baselines in the vicinity. All three BDOP indicators may be calculated by determining the contribution of the double-differenced observations at each epoch and accumulating these over a finite period into the normal matrix. Different session lengths may be evaluated by repeating this procedure and "sliding" the simulated observation period across satellite visibility window. The resultant precision indicators, BDOP1, BDOP2, and BDOP3 are defined as:

BDOP1 Indicates the precision of the coordinates in an ambiguity-free solution. It is defined as the square root of the trace of the coordinate sub-matrix (3 x 3). The magnitude of BDOP1 is influenced by the number of satellites observed. In the calculation of this indicator the assumption is made that all visible satellites are tracked.

$$BDOP1 = SQRT\{TRACE\ [\mathbf{Q_{cc}}]\}$$

BDOP2 Indicates the precision of the ambiguity parameters in an ambiguity-free solution. It is a square matrix with dimension (number of satellites – 1).

$$BDOP2 = SQRT \{TRACE [\mathbf{Q_{bb}}]\}$$

BDOP3 Indicates the precision of the coordinates after all cycle ambiguities have been resolved. The submatrix of the coordinates of the accumulated normal matrix (3 x 3) is extracted and inverted separately, excluding the cycle ambiguities from the set of solve-for parameters.

$$BDOP3 = SQRT \{TRACE [\mathbf{N^{-1}_{cc}}]\}$$

2.6.1 Characteristics of BDOP Indicators

Both BDOP1 and BDOP2 precision indicators have similar characteristics. Their relationship is apparent if we consider the case when the cycle ambiguities are estimated and precise station coordinates are computed. If it is easy to compute precise station coordinates when cycle ambiguities are estimated then it follows that the cycle ambiguities themselves must be easy to solve. However, as there are a number of strategies which can be used to overcome the rank defects in the ambiguity parameters (GRANT et al, 1990), it is not uniquely defined. Consequently the precision indicator BDOP2 is of limited use.

BDOP1 and BDOP3 describe the precision of the coordinates before and after cycle ambiguity resolution, and their definition is unique and independent of the processing strategy used. With both BDOP1 and BDOP3, as the session duration increases so the variation in BDOP values decreases (BDOP3 values will always be less than BDOP1). The BDOP3 precision indicators are appropriate for use with very small baselines due to the high probability of an ambiguity-fixed solution occurring. Under these circumstances, BDOP3 would be a better indicator to use than BDOP1. Examples of BDOP plots for different observation durations are shown at Appendix 2. Note also there is no uniform correlation between low PDOP or GDOP values and low BDOP values. This can be seen by comparing PDOP and GDOP plots in Appendix 1, with BDOP plots in Appendix 2.

2.6.2 Using BDOP Indicators

The strongest ambiguity-free solution will be one derived from an examination of the relative values of BDOP1 obtained by varying the satellites to be tracked, the observation start time, and observation session length. An alternative approach (especially for short baselines) is to vary the above parameters to find the minimum BDOP3 value. This may result in more precise coordinates from an ambiguity-fixed solution, however, if the BDOP1 value is large, the cycle ambiguity may have been incorrectly determined. BDOP3's decrease in magnitude with increasing observation length is not as marked as the decrease in the value of BDOP1 (see Appendix 2). This implies that it is less important to resolve ambiguities for longer sessions than it is for short sessions. During periods of low BDOP1, the difference between an ambiguity-free solution and an ambiguity-fixed solution is a minimum, and hence any gains in precision in going from an ambiguity-free to an ambiguity-fixed solution will be minimal.

When using BDOP indicators in the survey planning phase it would appear that it is best to concentrate on the variability of BDOP1 and to use it as the single criterion on which to select start times, satellites to be tracked, and the length of the observation session.

Various values of BDOP1 can be computed by varying the planning restraints, i.e. session length, satellites tracked and/ or the observation session start times. Plotting these BDOP1 values either individually or on a multiple plot, will allow the best scenario to be chosen by taking two planning constraints into account. The two planning constraints may be:

- satellites tracked and observation start time,
- satellites tracked and session length, or
- observation start time and session length.

Even though this process will result in an observation scenario being chosen with the lowest BDOP1 values, there is no guarantee that the cycle ambiguities will be resolved. The ability to resolve the cycle ambiguities is (in part) a function of changing satellite geometry, and therefore, the length of the observation session. GRANT (1990) states that the ambiguities which are estimated are in fact double-differenced with respect to the reference ambiguities of the base receiver and the base satellite. Their estimation is affected by those errors which are not eliminated through double-differencing. The most significant of these errors are:

- tropospheric delay,
- ionospheric delay,
- satellite orbit errors, and
- errors in the origin station coordinates.

The effect of these errors on the estimated ambiguities (for a baseline) increases as the baseline length increases. The tropospheric delay and ionospheric delay are discussed in Appendix 5.

3. THE SHORT BASELINE CAMPAIGN

To study the variability in the precision of coordinates obtained using GPS surveying techniques, a campaign was conducted to collect data for analysis. The two stations which define the baseline lie approximately 8 km apart and are located at the northern and southern ends of the Inertial Positioning System Test Range. This test range, adjacent to the University of New South Wales, was initially established for the purpose of testing inertial surveying technology (RÜEGER, 1984). The northern stations used throughout the survey are permanent concrete pillars 2 and 3, located on the roof of the Geography and Surveying Building (GAS2 and GAS3). The pillar GAS2 (the north end of the baseline) was used as the fixed station throughout the campaign. The pillar GAS3 is located approximately 2m to the north of pillar GAS2. At the southern end of the baseline the State Permanent Mark SPM8293 was used. This ground mark is grouted into solid sandstone bedrock and is situated on the shoreline of

Botany Bay, at La Perouse. The location and orientation of the baseline is shown below in Figure 3.1.

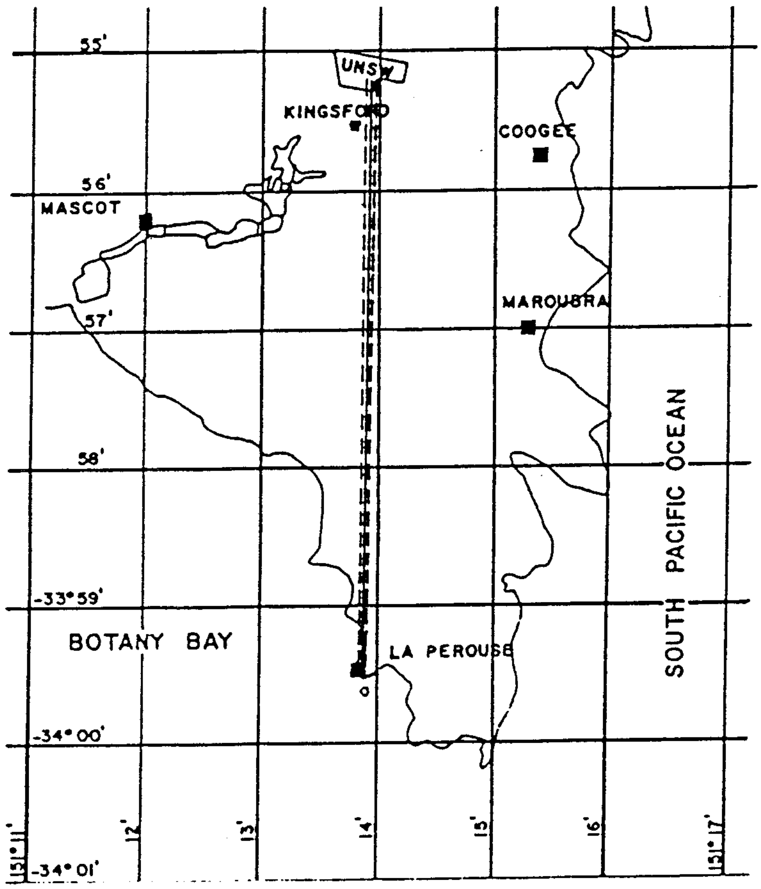


Figure 3.1 Baseline Location (modified from RÜEGER, 1984)

Ideally, a number of baselines with different orientations and lengths should have been investigated. However, given the large amount of data processing required for just one baseline, the observation of additional baselines was not attempted.

Throughout the campaign both stations were free of obstructions above 10 degrees elevation and no obvious physical changes occurred at either sites which could have influenced the signals being received either during or between observation sessions.

3.1 <u>The Campaign</u>

In the majority of GPS surveys (with short to medium baseline lengths) 60 minutes of data is generally sufficient to ensure that the integer cycle ambiguities can be resolved during the data processing stage. Over a small baseline of 8km, it was anticipated that an ambiguity-fixed solution would be obtained with 30 minutes of data or less, given a constellation of five or six satellites and favourable geometry. (Following processing, only for 62 of the 90 possible 30 minute sessions was an ambiguity-fixed solution possible.)

Dual-frequency observations were made throughout the campaign. After each days observations were completed, the data was downloaded to a PC and a backup made. The data from each session was then divided into 30 minute sub-sessions, each of which were then processed separately.

3.1.1 Campaign Timetable

The baseline stations were occupied on the 8^{th} and 9^{th} of June, and on the 12^{th} and 13^{th} of September, 1989. Data was collected in a routine manner for both nights of the June and September stages of the campaign. During the data processing it became apparent that for the baseline GAS3 to SPM8293 on the 12^{th} September, only L1 data was able to be processed, as the L2 data was intermittent and incompletely captured / recorded in the session. (Both receivers worked correctly on the second night of observations. The receivers were not moved between these two sessions.) The aim of this second WM102 campaign was to collect additional data so that a comparison could be made with the data collected in June.

3.1.2 The GAS3 - SPM8293 Baseline

At the completion of the campaign, the three dimensional position offsets were determined for GAS3 with respect to GAS2. Thus, following processing of the baseline GAS3 - SPM8293, the baseline was reduced to the baseline GAS2 - SPM8293. The results obtained were then able to be directly compared with those obtained directly for the baseline GAS2 - SPM8293, observed at the same time for the same session duration. The accuracy of the offsets determined were tested by then processing the baseline GAS2 - GAS3. Pillar GAS3 was given the same *a priori* coordinates as GAS2, and the offsets calculated were entered in the header record in the data file prior to data processing. If the offsets were correct and there was no error in the coordinates of the pillars, a baseline length close to zero would be expected. A zero baseline distance was not expected due to the presence of unmodelled systematic errors and the presence of receiver noise. The resultant baseline length determined by processing data for the baseline GAS2 - GAS3 (giving GAS3 the same coordinates as GAS2 and applying the offsets between the two pillars) was 3.5mm, which was considered acceptable. The most important factor being examined was the relationship of one day's solution to another day's, and the behaviour of the solution over time within the observation session and between the co-observed baselines. The GAS3 offset from GAS2 was computed as being:

ΔE	1.983m
ΔN	-0.307m
Height of Instrument	0.141m

These offsets were used during the September 1989 phase of the campaign.

The session times and type of solution obtained after data processing is given in section 5.

3.2 The WILD-MAGNAVOX GPS Receiver: WM102

The WM102 GPS receiver is essentially an upgraded model of the earlier WM101 receiver which has been given a dual-frequency capability. This data is then able to be processed using either the single frequency observations alone, or using a combination of both the L1 and L2 observations. The WM102 operates with a total of eight channels which function as follows:

Channel 1	Records ephemeris data.
Channel 2	Tracks either one or two satellites. If two satellites are tracked, the receiver switches from one satellite to the other every two seconds.
Channel 3 - 6	One satellite on each channel is tracked continuously on the L1 frequency only.
Channel 7	Used for inter-channel calibration values. This is achieved by tracking up to six satellites using the L1 frequency only and switching between them at two second intervals.
Channel 8	Used to record L2 data by switching sequentially from one satellite to another.

3.2.1 Data Recording

The spacing of each observation epoch is user definable. In our tests observations 60 seconds apart were recorded, on both frequencies:

- The L1 frequency is sampled at a rate of 10 times per second. If a particular channel is tracking two satellites, then two seconds of data are alternatively collected from each satellite. This two-second data sample is compacted and represents the WM102's fundamental data sample. This data sample can either be recorded in this "raw" form, or used to obtain the pre-determined epoch spacing (60 second in our case) "compacted" data observation.

- The compacted L2 data records contain the satellite number, phase, pseudo-range and phase-rate. The observation rate of the L2 data is the same as for L1 data, however Channel 8 is the only channel that records L2 data. The L2 signal acquired by the WM102 is of the order of 300 times stronger than the signal acquired using the alternate squaring technique (WELLS et al, 1986). The WM102 uses the coherent P code where possible and automatically employs a squaring technique only in the event that the P code is encrypted ("anti-spoofing" option active).

3.2.2 Data Session Definition and Observation

The start and end times for each session throughout the campaign were determined with the aid of the School of Surveying's PREDICT program, developed by C.Rizos and modified by B. Merminod and others (MERMINOD, 1990). Sessions were defined by the period that four or more satellites were visible at an elevation of greater than 15^{o} above the horizon.

For each session the WM102's were setup and initialised 15 minutes prior to the commencement of the session. Continuous observations were recorded at each site for the duration of the window after which time the receivers were either manually shutdown or observations automatically halted at the sites operating in the unattended mode (GAS2 and GAS3).

4. GPS DATA PROCESSING WITH THE PoPS™ SOFTWARE

Data collected by the WM102 can be processed directly using PoPS™ software. The software is based on GPS software package developed for scientific purposes at the University of Berne. Some of its capabilities are (FREI et al, 1986):

- handles simultaneous data from several receivers,
- include additional information from previously processed data,
- determine the unknown integer cycle ambiguity where possible (and as required),
- provide results in a manner which can be used by other software,
- provide a field planning component,
- use either the broadcast and precise ephemerides (if available), and
- perform coordinate transformations and map projection calculations.

Data was recorded on standard data cassettes and was downloaded at the end of each session via either the MEMTEC reader or by connecting the WM102 receiver directly to a PC.

Prior to commencing the pre-processing, the data from the multi-hour observation session was divided into 30 minute sub-sessions (see section 5).

4.1 Pre-Processing

The pre-processing stage was divided into four individual steps:

a. Determination of "standard orbit parameters" for each satellite,
b. Calculate single point positions using pseudo-range data,
c. Create single-differences for each site-pair in the network, and
d. Screen poor observations and cycle-slips.

4.1.1 Determination of "Standard Orbits"

Before any phase processing can be carried out using the PoPS™ software, an accurate position of each satellite observed during the observation session is required. Either the broadcast or precise ephemeris can be used. The software determines the positions of all satellites observed during a session from, for example, the ephemeris broadcast by each satellite. This process is performed automatically in the software whenever a new campaign is defined, or whenever additional data is added to an existing campaign. This process does not need to be repeated unless re-processing of the data is performed using, for example, the precise ephemeris.

4.1.2 Single Point Positioning (& Receiver Clock Offset Correction)

The single point positioning process determines the unknown receiver clock offset (with respect to GPS Time), as well as the position of the receiver on the earth's surface to within about 30m using the pseudo-ranges in a quasinavigation solution. Within PoPS™ the clock behaviour can be modelled by polynomials of up to degree six and the technique of differencing compares the phase of the signal from the satellite between one receiver and another. In order for the differencing technique to work, accurate signal reception times are required. To achieve this, a standard time system is adopted and the offset of each satellite and receiver clock from this time (GPS time) must be known to some level (RIZOS & GRANT, 1990). The offset drift and drift-rate of the relatively unstable receiver (quartz crystal) oscillator are modelled with the following accuracies (PoPS™ 3.02):

Offset:	1×10^{-8} sec
Drift:	1×10^{-11} sec/sec
Frequency Drift:	1×10^{-15} sec^2/sec

The data screening routine within this point position stage screens out observations made to satellites which are below the user defined mask elevation angle. Poor observations at this stage can be flagged and excluded from further computations.

4.1.3 The Single Point Positioning Logfile

The single point position logfile provides a summary of the processing options used to determine the point position of all stations in the network, as well as a summary of the data itself:

a. The type of pseudo-ranges processed (whether C/A or P code).
b. The ephemeris used (whether the broadcast or precise).
c. The unknown parameters (clocks and coordinates).
d. The ionospheric and tropospheric models used.
e. The minimum satellite elevation angle used.
f. The observation sample rate.
g. Whether range accuracy weighting is used (using the transmitted UERE).
h. For each site, the satellites observed and number of epochs observed to each satellite are tabulated. A table showing the number of bad observations to each satellite is also provided and was used for data quality assessment.
i. The RMS value of the point position for each site. Throughout the campaign this value was typically less than 2.5m.
j. The three clock parameters are displayed and RMS values for each satellite are listed. This RMS value is useful as an indicator of satellite health.

4.1.4 Data Differencing

PoPS™ creates "between-receiver" single-differences for each site-pair, and stores them in the database (GRANT et al, 1990). This single-difference is the instantaneous difference in phase of a satellite's signal recorded by two receivers observing the satellite at the same (reception) time. Satellite clock bias is effectively eliminated through the use of the single-difference. Double-differencing eliminates both satellite and receiver clock biases, whilst the triple-differencing (used in the initial screening process) are between-epoch differences of the double-differenced data (see Appendix 3).

4.1.5 Data Screening

Both the L1 and L2 frequency data are screened for full and half-cycle slips (first for full cycle slips and then repeated for half-cycle slips). The following dual-frequency combinations can be used during screening:

L1 frequency:	(wavelength approximately 19cm)
L2 frequency:	(wavelength approximately 26cm)
L3 frequency:	This is the standard ionosphere-free combination of L1 and L2 data.
L5 frequency:	The L5 frequency has a wavelength of approximately 86cm and is the beat frequency formed from the difference between L1 and L2 measurements. This is the "widelane" combination.

After data screening and cycle slip editing has been completed, the data is checked for poor quality observations. Data quality can be affected by a malfunctioning satellite or faulty receiver / antenna equipment. Changes in the observation environment can also result in poor data. A check of the data quality can be made by examining the triple-differenced residuals. These values, once computed, can be stored in the data screening logfile. Acceptable triple-difference residuals should be small in magnitude (less than 20mm) and of random sign. Residuals greater than this in magnitude can indicate poor data. An examination of residuals can also reveal residual half or full cycle slips, as shown below in Table 4.1.

Table 4.1 Magnitude of residuals indicating either full or half-cycle slips.

RESIDUAL	CYCLE	FREQUENCY
0.095m	HALF	L1
0.130m	HALF	L2
0.190m	FULL	L1
0.260m	FULL	L2

4.2 Data Computation

After pre-processing, the baseline site-pairs were selected and the computation commenced. For each 30 minute session, a results summary file was produced. The contents of the results summary file includes:

a. Cartesian coordinates of the unknown stations presented in the WGS84 system, as well as the equivalent geographical coordinates (latitude, longitude and ellipsoidal height), along with their standard deviations.
b. The double-differenced ambiguities (plus standard deviations).
c. Inter-station slope distance along with the standard deviation.
d. Residuals from the double-differencing for each epoch in the session.

The standard deviations (or RMS values as they are expressed in the results output) of the coordinates of the unknown stations both before and after ambiguity resolution are a measure of internal precision. In practice, ten times this value is a good measure of accuracy, due to the influence of unmodelled systematic errors on our observations.

4.2.1 Baseline Classification

PoPS™ classifies observed baselines according to the distance between stations:

Table 4.2 PoPS™ Baseline Classification

CATEGORY	L1	L1 + L2
SHORT	<50 km	< 20km
MEDIUM		20-50km
LONG	>50km	> 50km

4.2.2 Single-Frequency Processing Strategy

If single-frequency data is being processed, the computation strategy used in PoPS™ is to process the short lines first and then resolve the integer cycle ambiguities wherever possible. All baselines are then re-processed simultaneously, holding fixed the ambiguities resolved in the initial phase. Ambiguities for long lines may not be resolved due to the possibility of the wrong ambiguities being selected.

4.2.3 Dual-Frequency Processing Strategy

When dual-frequency observations are processed, the initial processing solves for the coordinates of the stations for the short baselines. Wherever possible the cycle ambiguities are resolved. The L5 (beat frequency) is used to solve for the medium baselines after which all baselines are reprocessed holding the resolved ambiguities fixed. The L1 and L2 frequencies are used for the short baselines and L3 is used for both the medium and long baselines. The classification of baselines into one of the three categories within the software may be modified at any stage by the user.

4.2.4 Correlations

As mentioned earlier, correlations are introduced into the data processing through the use of differencing techniques. Correlations in GPS measurements is discussed at Appendix 4. In PoPS™, prior to computing the baseline solution, there are three options available to the user to account for the correlations which exist in the observations:

(1) account for the correlations which occur in forming the double-differences from the single-differences,
(2) account for all correlations, or
(3) account for no correlations at all.

As the campaign was observing data from two stations, only the individual single-differences are uncorrelated. All three options were used during trial data processing, however, only the correlations associated with forming the double-differences from the single-differenced observations were considered at the data processing stage.

4.3 Assessment of Results

The assessment of reliability of the solution computed by PoPS™ was effected through an analysis of:

- the RMS values of the double-differences,
- the estimated ambiguities, and
- the double-difference residuals.

WILD (PoPS™ Manual) recommends that the magnitude of the RMS of the double-differences for baseline lengths less than 20km should be less than 10mm and RMS values for lines between 20 and 100km should be less than 100mm. Typical double-difference RMS values for the short baseline campaign were of the order of 5 to 8mm.

The list of estimated ambiguities can assist in the detection of both full and half-cycle slips. The possibility of half-cycle slips can occur wherever the estimated ambiguities have a decimal component close to 0.5 with an RMS value of less than 20mm. In addition, residuals with large magnitude and/or fluctuating sign are a good indicator of poor data.

5. CAMPAIGN RESULTS

The five multi-hour sessions observed during the campaign were divided into ninety (90), 30 minute sessions. In the processing phase, four days of data were selected for analysis: the 8th and 9th of June, 1989, and the 12th and 13th of September, 1989.

The information which was of primary importance to this study was the manner in which precision estimates obtained in the office prior to the survey compared with those obtained from measurements made in the field. For this GPS campaign, the BDOP values were computed using the ephemeris file collected during each day's observations. This eliminated the possibility of the satellite constellation observed being different to the one used to predict the BDOP precision indicators.

5.1 Single Frequency Data Processing (GAS2 - SPM8293)

A summary of the subsessions with an indication as to whether an ambiguity-fixed or ambiguity-free solution was obtained after processing the single frequency observations is given in Table 5.1 and 5.2. Note that the local time given for each session is the mid-point of a 30 minute session.

Four plots (Figures 5.1, 5.3, 5.5 and 5.7) show the change in the Root Sum Squared (RSS) values of the three dimensional standard deviations for 30 minute observing sessions (L1 data obtained from baseline GAS2 - SPM8293). Each of these plots is followed by the corresponding BDOP1 plot for that day.

Table 5.1 Single Frequency Results for Subsessions (GAS2 - SPM8293)

June						September			
8		9		26		12		13	
Local Time	Ambig Resol.	Local Time	Ambig Resol.	Local Time	Ambig Resol.	Local Time	Ambig Resol.	Local Time	Ambig Resol.
0228	N	0224	Y	0121	N	2031	N	2031	Y
0258	Y	0254	Y	0151	Y	2101	Y	2101	Y
0328	Y	0234	Y	0221	Y	2131	Y	2131	N
0358	N	0354	N	2251	N	2201	Y	2201	N
0428	Y	0424	Y	0321	Y	2231	Y	2231	Y
0458	Y	0454	Y	0351	Y	2301	Y	2301	N
0528	Y	0524	N	0421	Y	2331	Y	2331	N
0558	Y	0554	Y	0451	N	0001	N	0001	Y
0628	N	0624	Y	0521	Y				

Table 5.2 Single Frequency Results for Subsessions (GAS3 - SPM8293)

September			
12		13	
Local Time	Ambig Resol.	Local Time	Ambig Resol.
2031	N	2031	Y
2101	Y	2101	Y
2131	Y	2131	N
2201	Y	2201	N
2231	Y	2231	Y
2301	Y	2301	N
2331	Y	2331	N

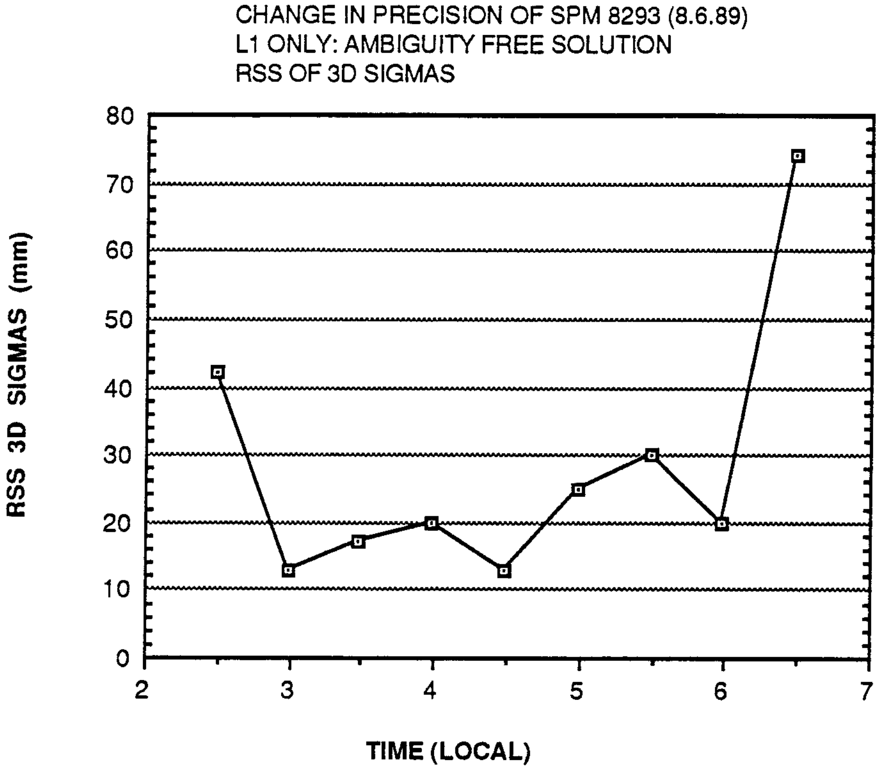


Figure 5.1 RSS of 3D sigmas 8th June 1989 (SPM8293).

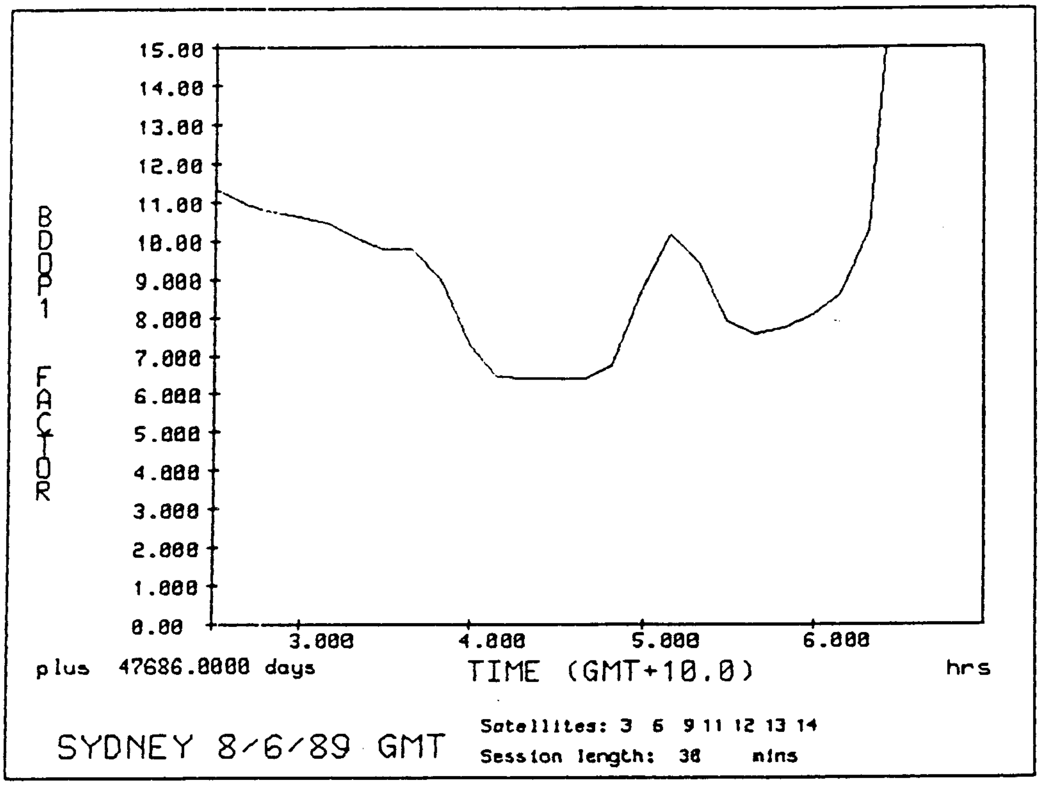


Figure 5.2 BDOP1 plot 8th June 1989 (30 minute session).

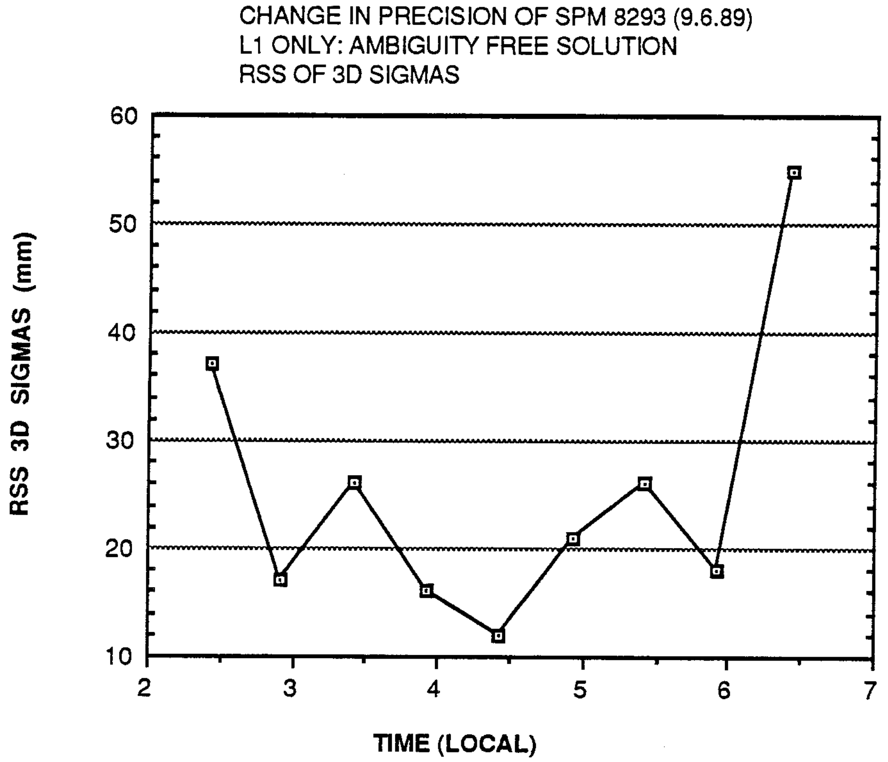


Figure 5.3 RSS of 3D sigmas 9th June 1989 (SPM8293).

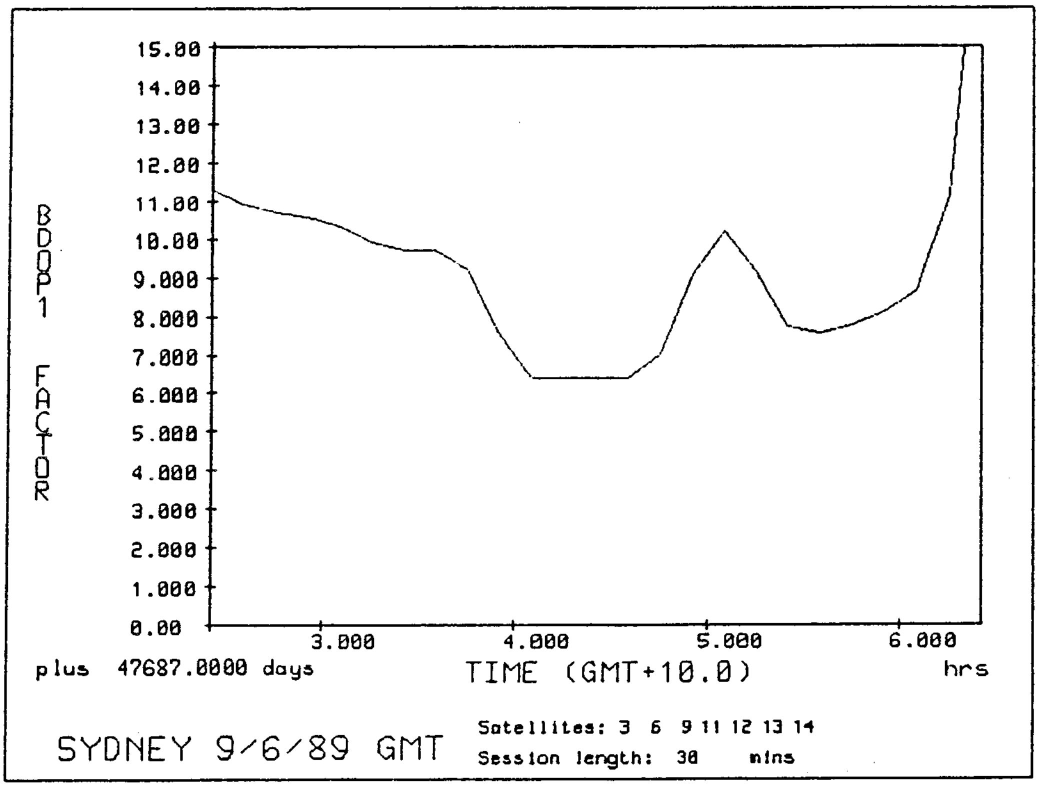


Figure 5.4 BDOP1 plot 9th June 1989 (30 minute session).

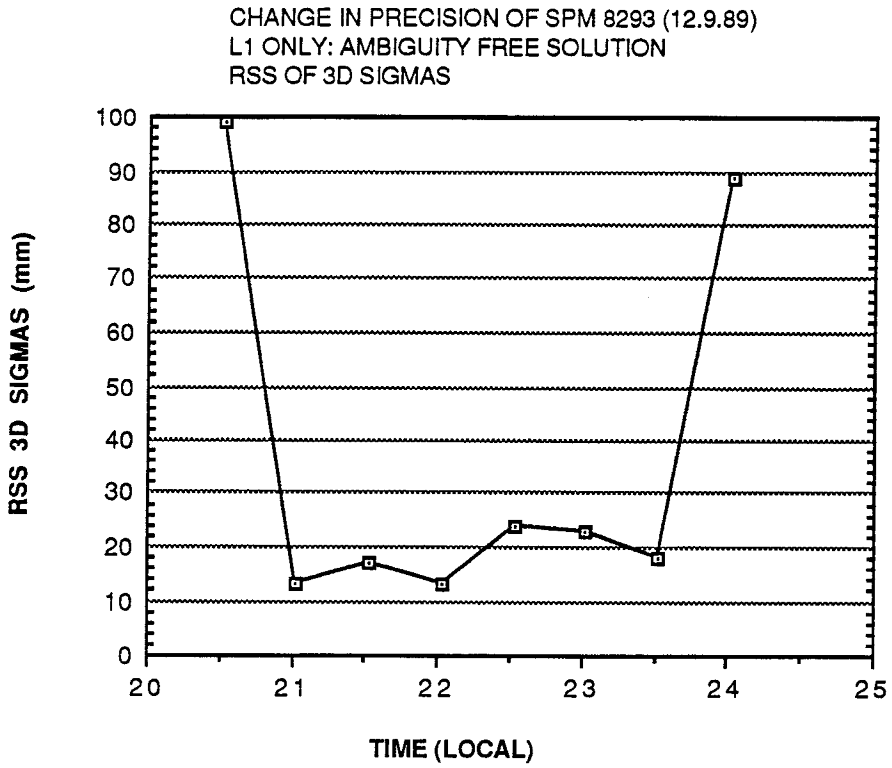


Figure 5.5 RSS of 3D sigmas 12th September 1989 (SPM8293).

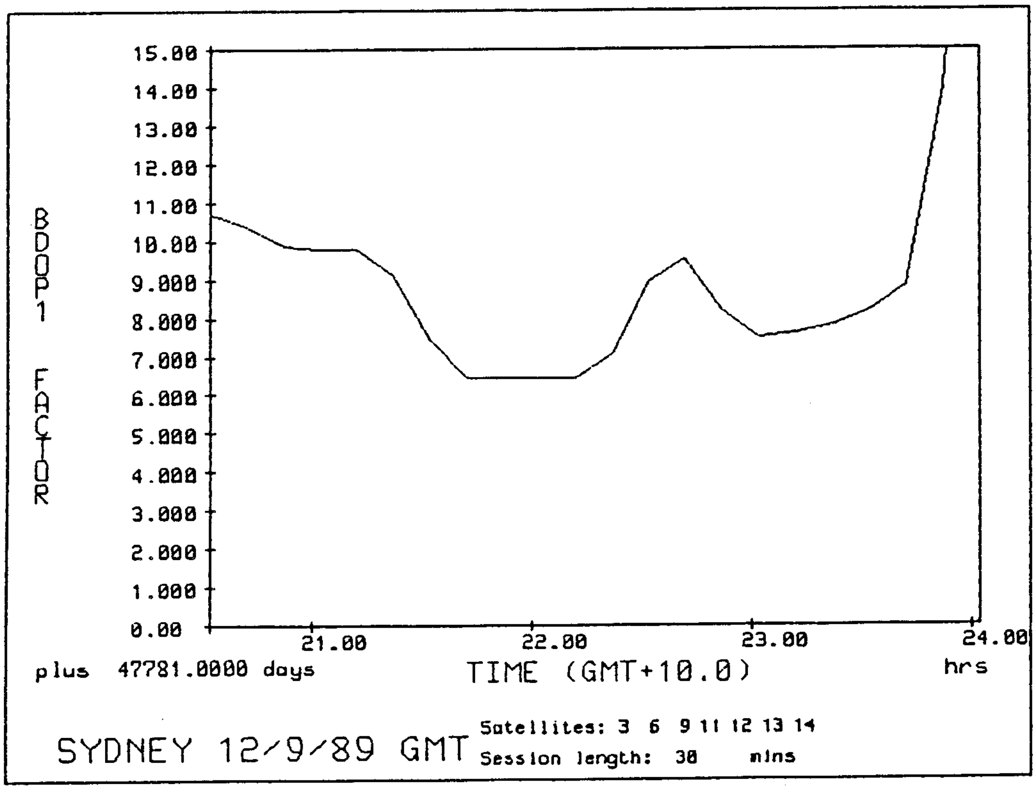


Figure 5.6 BDOP1 plot 12th September 1989 (30 minute session).

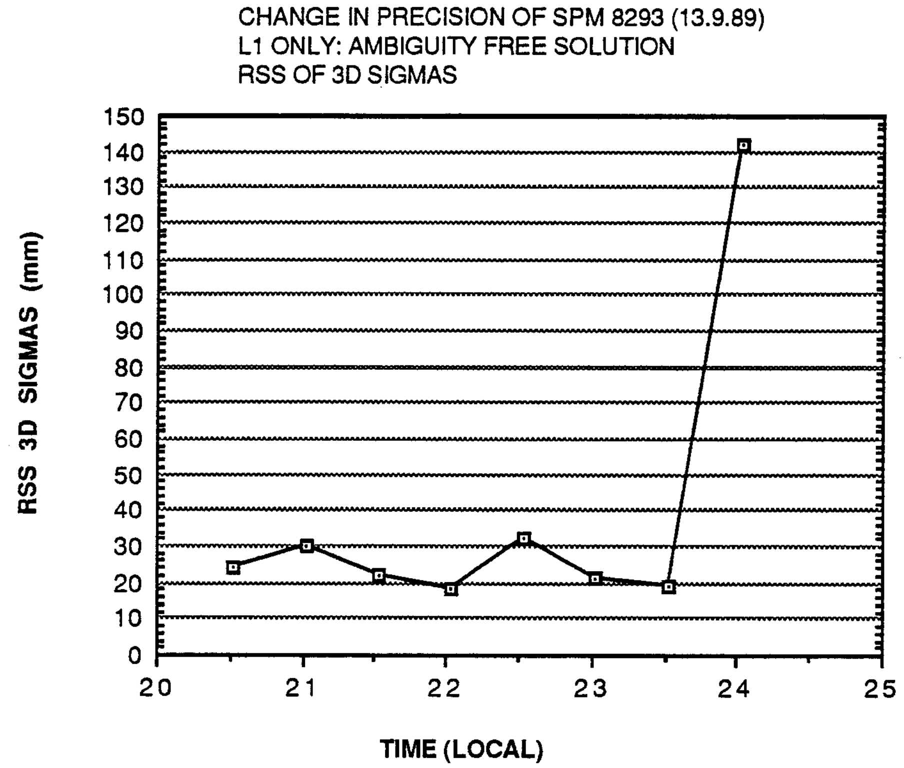


Figure 5.7 RSS of 3D sigmas 13th September 1989 (SPM8293).

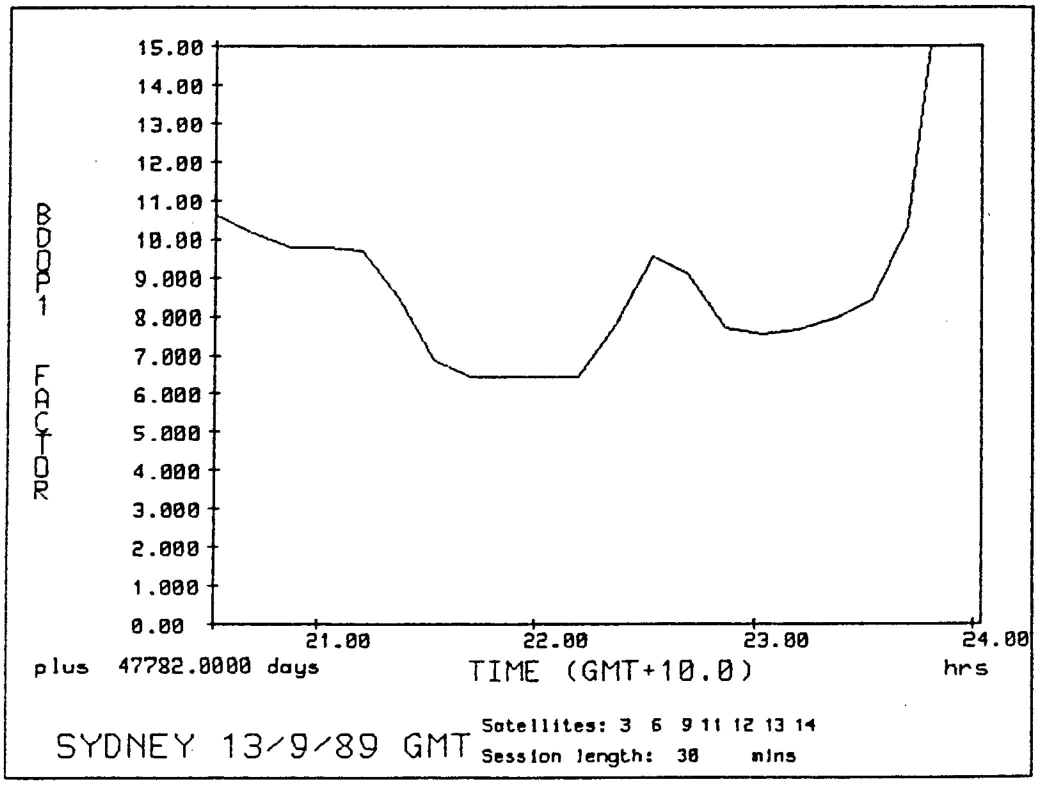


Figure 5.8 BDOP1 plot 13th September 1989 (30 minute session).

As expected, the BDOP1 indicators for the 8th and 9th June are almost identical as are the BDOP1 plots for 12th and 13th September. However, the relationship between the precision indicator(s) and the RSS values for the SPM8293 solutions is not evident. For data processed from the 8th and 9th June, particularly at the beginning of the observation period, there does not seem to be a correlation between the two. For the remaining plots, the general slope of the BDOP1 plots is reflected in the RSS plots. The precision of a station's coordinates which have been determined using GPS relative positioning surveying techniques is affected by all unmodelled systematic effects as well as those residual effects from incompletely modelled errors, whereas BDOP plots only reflect the theoretical precision of the survey measurements with no influence of unmodelled systematic errors.

If a systematic error(s) affects both stations of a given baseline in a uniform manner for the duration of the campaign, this will result in a scaling effect on the 3D RSS precision values and thus should not affect the the "BDOP1 signature" (for that session) in the RSS precision values determined from GPS measurements. This uniform influence on a baseline is unlikely to occur in practice. The influence of systematic errors on GPS measurements is dynamic in nature. This will have an instantaneous scaling effect on the precision values obtained from measurement and as a result, the BDOP signature which we could expect to see under ideal observing conditions will be distorted. This is our explanation for the poor correlation between the plots of our precision indicators and the precision values determined from our single frequency baseline measurements (GAS2 - SPM8293).

To ensure that there were no external factors influencing the data recorded throughout the campaign, the Royal Australian Survey Corps's weekly GPS notices were consulted. From these notices, changes to satellite health and GPS system status can be analysed for the period of the campaign. After examination of these messages, there did not appear to be any major problems with any of the satellites during the course of the campaign which could have adversely affected the data.

5.2 Dual-Frequency Data Processing (GAS2 - SPM8293)

A summary of the subsessions with an indication as to whether an ambiguity-fixed or ambiguity-free solution was obtained after processing the dual-frequency observations is given in Table 5.3 and 5.4. (Note that, as with Table 5.1 and 5.2, the local time given for each session is the mid-point of a 30 minute session.)

The data was re-processed using dual-frequency data to examine the effects of using observations on both L1 and L2, on the precision of SPM8293 coordinates within a multi-hour observation session. The 30 minute session definition remained unchanged from the earlier single-frequency processing. The data was re-screened for full cycle-slips followed by re-screening for half-cycle slips using the L3 triple-difference observable. Once completed, a data screening logfile was produced for each of the 30 minute sessions. From this the quality of the data was assessed as the logfile recorded the number of cycle-slips detected, the RMS of the triple-differences and the largest remaining L1 and L2 residuals.

Table 5.3 Dual-Frequency Results for Subsessions (GAS2 - SPM8293)

June						September			
8		9		26		12		13	
Local Time	Ambig Resol.	Local Time	Ambig Resol.	Local Time	Ambig Resol.	Local Time	Ambig Resol.	Local Time	Ambig Resol.
0228	N	0224	Y	0121	N	2031	YY	2031	Y
0258	Y	0254	Y	0151	Y	2101	Y	2101	Y
0328	Y	0234	Y	0221	Y	2131	Y	2131	N
0358	N	0354	N	2251	N	2201	Y	2201	YY
0428	Y	0424	Y	0321	Y	2231	Y	2231	Y
0458	Y	0454	Y	0351	Y	2301	Y	2301	N
0528	Y	0524	N	0421	Y	2331	Y	2331	YY
0558	Y	0554	Y	0451	N	0001	N	0001	Y
0628	N	0624	Y	0521	Y				

Table 5.4 Dual-Frequency Results for Subsessions (GAS3 - SPM8293)

September	
13	
Local Time	Ambig Resol.
2031	Y
2101	Y
2131	N
2201	YY
2231	Y
2301	N
2331	N
0001	NN

(Note: in Table 5.3 and 5.4 the dual-frequency sessions which resulted in an ambiguity-free or ambiguity-fixed solution different from the single frequency solution have been depicted by NN or YY.)

Figures 5.9 to 5.12 below are similar to Figures 5.1 to 5.8, however dual-frequency observations were processed.

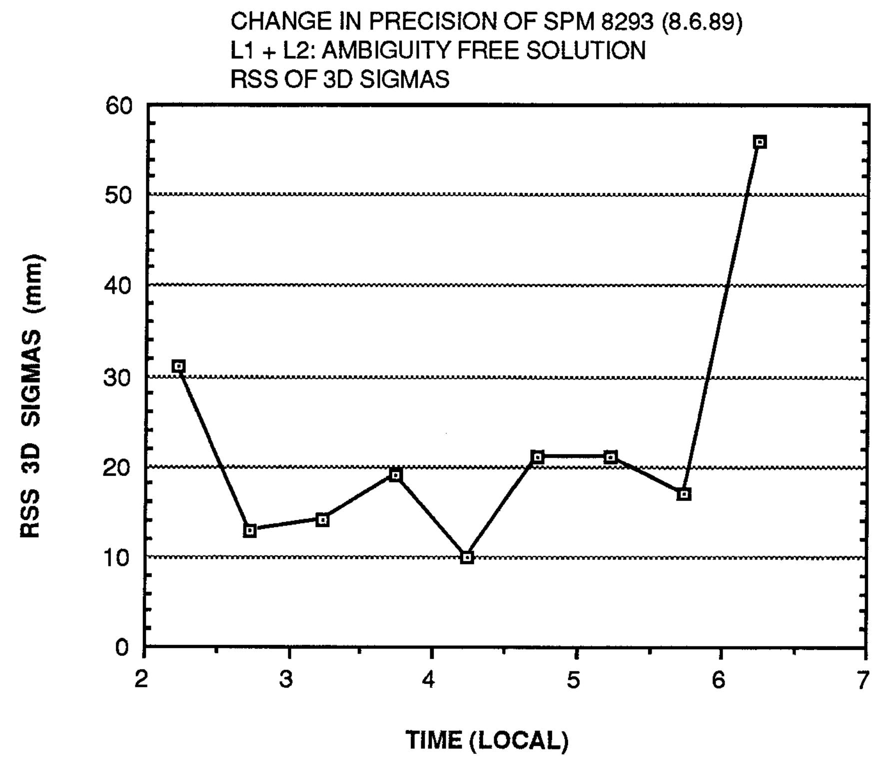


Figure 5.9 RSS of 3D sigmas 8th June 1989 (SPM8293).

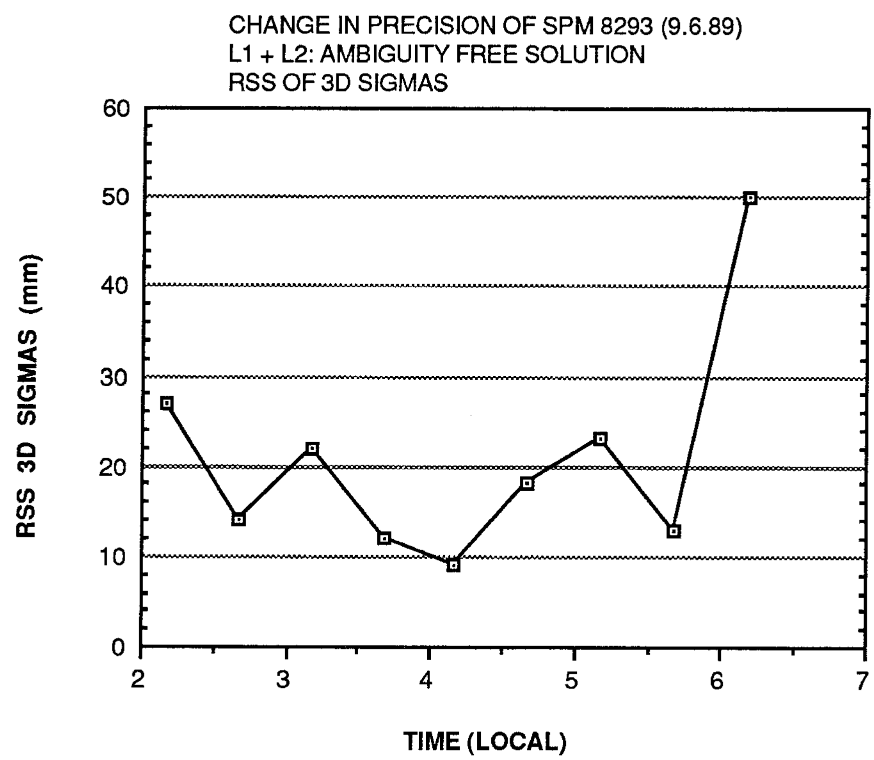


Figure 5.10 RSS of 3D sigmas 9th June 1989 (SPM8293).

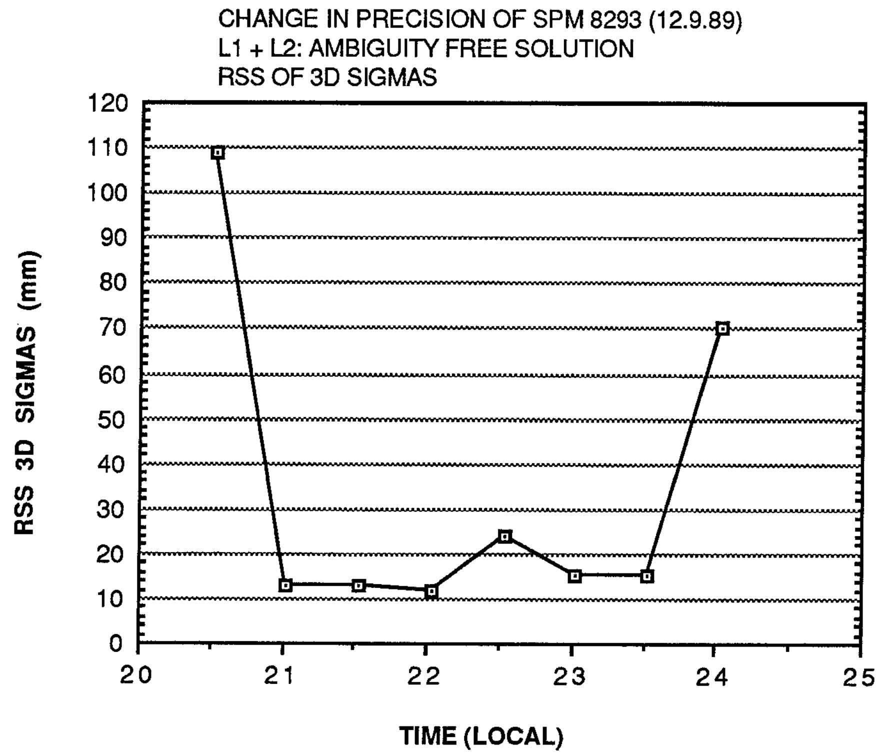


Figure 5.11 RSS of 3D sigmas 12th September 1989 (SPM8293).

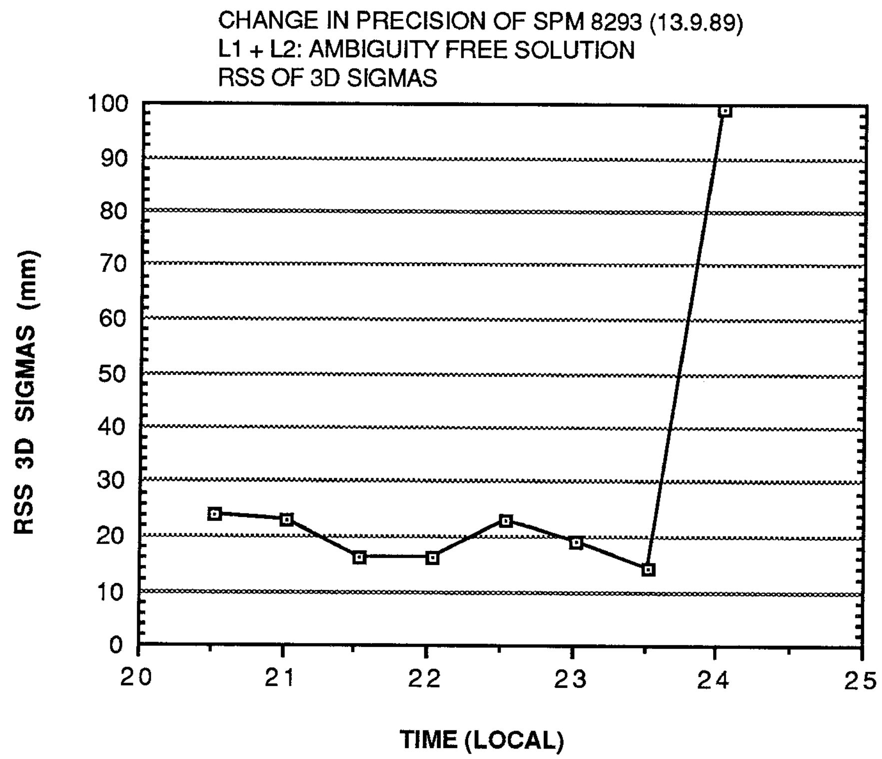


Figure 5.12 RSS of 3D sigmas 13th September 1989 (SPM8293).

From an examination of the plots there is generally good agreement between the BDOP values and the empirical RSS precision values, except for the first 60 to 90 minutes of each multi-hour session (with the exception of the 13th September). The graph of the 13th September 3D RSS precision values shows very close agreement to the BDOP1 plot for that day. This graph demonstrates that given GPS survey conditions which do not depart markedly from our modelled environment, it can be concluded that Merminod's precision indicator BDOP1 could determine the times at which we can expect to obtain results of the best quality (that is, with the highest precision). Even though the other 3D RSS plots exhibit this correlation to a lesser extent, this does not alter the conclusion that using such precision indicators for planning GPS field surveys is a worthwhile exercise.

The precision of the solution obtained from processing data recorded by two receivers sited together (GAS2 and GAS3) and a third receiver (SPM8293) should be identical given that both receivers were recording similar data. Following data processing, this was seen to be the case.

7. CONCLUSIONS AND RECOMMENDATIONS

It has been established now both mathematically and empirically that there appears to be a strong correlation between the BDOP indicators of precision and the precisions gained from GPS field survey. However, there will always be some unmodelled systematic errors present throughout a GPS survey campaign regardless of the time of commencement of the survey and the results shown above are an indicator of this.

The graphs of the RSS of the three dimensional standard deviations of SPM8293 coordinates show that unmodelled systematic errors do noticeably degrade the quality of our measurements, and hence the associated results obtained from a conventional double-differenced phase solution. The presence of these errors severely mares the usefulness, at present, of BDOP1 as a precision indicator and until these errors are better modelled, any possibility of using them to determine optimum observation times will be limited. The main sources of error which have the largest influence on the precision of our measurements are the meteorological conditions prevailing throughout the campaign, and the effect of the propagating medium on the transmission of the satellite's signal to the receiver. For longer baselines, ionospheric activity will introduce errors into GPS measurements obtained with single-frequency instrumentation. The magnitude of these errors will increase with the approaching period of maximum solar activity (HIRMAN et al, 1989).

Before we are able to predict optimum session observation durations these systematical effects on our measurements must be better understood and models derived which are both more appropriate to the observation environment and are capable of reflecting the changes which occur in that environment during the course of the campaign.

For a surveyor who wishes to determine the optimum time to conduct a GPS campaign, the BDOP indicator still provides this information, regardless of the unmodelled systematic errors which may be present at the time of survey. Co-variance analyses of GPS carrier phase adjustments by GRANT (1990) have demonstrated that an improvement in satellite geometry (indicated by BDOP1) also corresponds with a lower sensitivity to systematic errors. In this manner, a benefit of using BDOP precision indicators is in the capability of comparing one observation session with another in the field planning stage.

During periods of low BDOP1 values, the difference between an ambiguity-free solution and an ambiguity-fixed solution is a minimum which suggests any gains in precision from obtaining an ambiguity-fixed solution will be minimal. Due to the manner in which BDOP is derived, it clearly has more relevance as a precision indicator to GPS relative position surveys than either PDOP or GDOP, which are currently being used extensively.

Additional research is required to refine the meteorological models used during data processing. As demonstrated by BEUTLER et al (1989) for small network GPS surveys, results obtained after using meteorological data in post-processing are most probably going to be of poorer quality than those obtained using a "standard" meteorological model (and no meteorological data measurements at all).

8. ACKNOWLEDGEMENTS

The author wishes to acknowledge the following for providing valuable assistance to this project: my supervisor Dr C. Rizos for his guidance and encouragement; the Director of Survey - Army, for sponsoring my studies at the University; Mr R. Eckels, Product Specialist GPS (WM102), Wild-Leitz Australia, Mr R. MacLeod, Project Manager GPS (ASHTECH), SAGEM Australasia Pty Ltd; the Australian Antarctic Division for providing GPS equipment for the second phase of the survey campaign, and Mr B. Hirsch for his computer hardware and software support.

9. REFERENCES

BAUERSIMA, I., 1983. NAVSTAR / Global Positioning System (GPS). Mitteilungen der Satellitenbeobachtungsstation Zimmerwald, Nr. 10, Astronomical Institute, University of Berne.

BEUTLER, G., BAUERSIMA, I., GURTNER, W., ROTHACHER, M., SCHILDKNECHT, T., & GEIGER, A., 1989. Atmospheric refraction and other important biases in GPS carrier phase observations. Monograph 12, "Atmospheric Effects on Geodetic Space Measurements", edited by F.K.Brunner, School of Surveying,University of New South Wales, 15-44.

FREI, E., GOUGH, R., & BRUNNER, F.K., 1986. PoPS™: a new generation of GPS post-processing software. Proc. 4th Int. Symp. on Satellite Positioning, Austin, Texas, 28 Apr - 2 May, 1986, 455-473.

GOAD, C.C., 1985. Precise relative position determination using Global Positioning System carrier phase measurements in a non-difference mode. Proc. 1st Int. Symp. Prec. Pos. GPS, U.S. Dept. of Commerce, NOAA, Rockville, Md., 15-19 April, 1985, 347-356.

GRAFAREND, E.W., LINDLOHR, W., & WELLS,D.E.,1985. GPS redundancy design using the undifferenced phase observations approach. Proc. of the Second SATRAPE Meeting, Saint-Mande, France.

GRANT, D.B., 1990. Combination of terrestial and GPS data in earth deformation studies in New Zealand. Unisurv S-32, report of the School of Surveying, University of N.S.W., 285pp.

GRANT, D.B., RIZOS, C., & STOLZ, A., 1990. Dealing with GPS baises: some theoretical and software considerations. In Unisurv S-38, "Contributions to GPS Studies", C. Rizos (ed.), report of the School of Surveying, University of N.S.W.

HATCH, R.R. & AVERY, E.V., 1988. A strategic planning tool for GPS surveys. Pres. at ASCE Specialty Conference: GPS-88 Engineering Applications of GPS Satellite Surveying Technology, May 11-14, 1988, Nashville, Tennesse, U.S.A.

HIRMAN, J.W., HECKMAN, G.R., GREAR, M.S., & SMITH, J.R., 1989. Solar cycle 22 continues strong climb. EOS, 20(26), 27 June 1989

HOAR, G.J., 1989. Private Communication, MAGNAVOX Advanced Products and Systems Company.

KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A., & COLLINS, J., 1987. **Surveying with GPS**. Ferd. Dümmlers Verlag, Bonn, 128pp.

LINDLOHR, W., & WELLS, D.E., 1985. GPS design using undifferenced carrier beat phase observations. **Manuscripta Geodaetica**, 10, 255-295.

MacLEOD, R.T., & RIZOS, C., 1988. Correlation in GPS surveying: a cause for concern?. Pres. at the Int. Symp. on Global Positioning Systems, Queensland, October, 1988.

MERMINOD, B., GRANT, D.B., & RIZOS, C., 1990. Planning GPS surveys - using appropriate precision indicators. submitted **Can.Surv.**

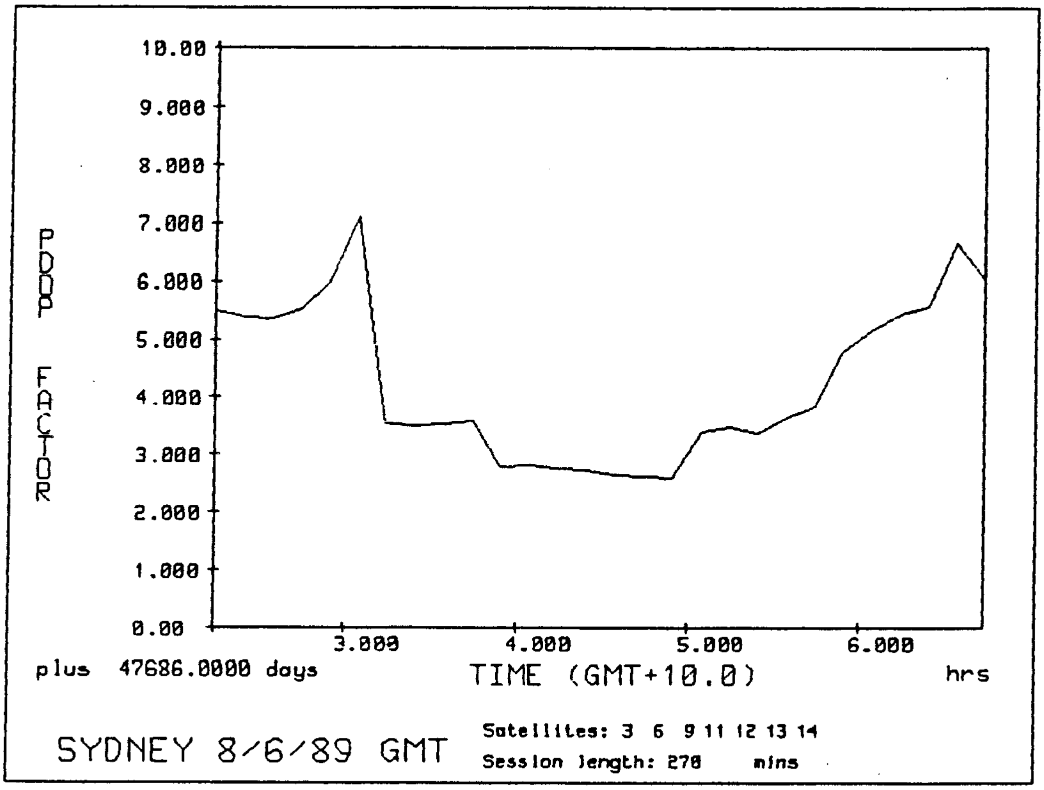
MERMINOD, B., 1990. Resolution of the Cycle Ambiguities. In Unisurv S-38, "Contributions to GPS Studies", C. Rizos (ed.), report of the School of Surveying, University of N.S.W.

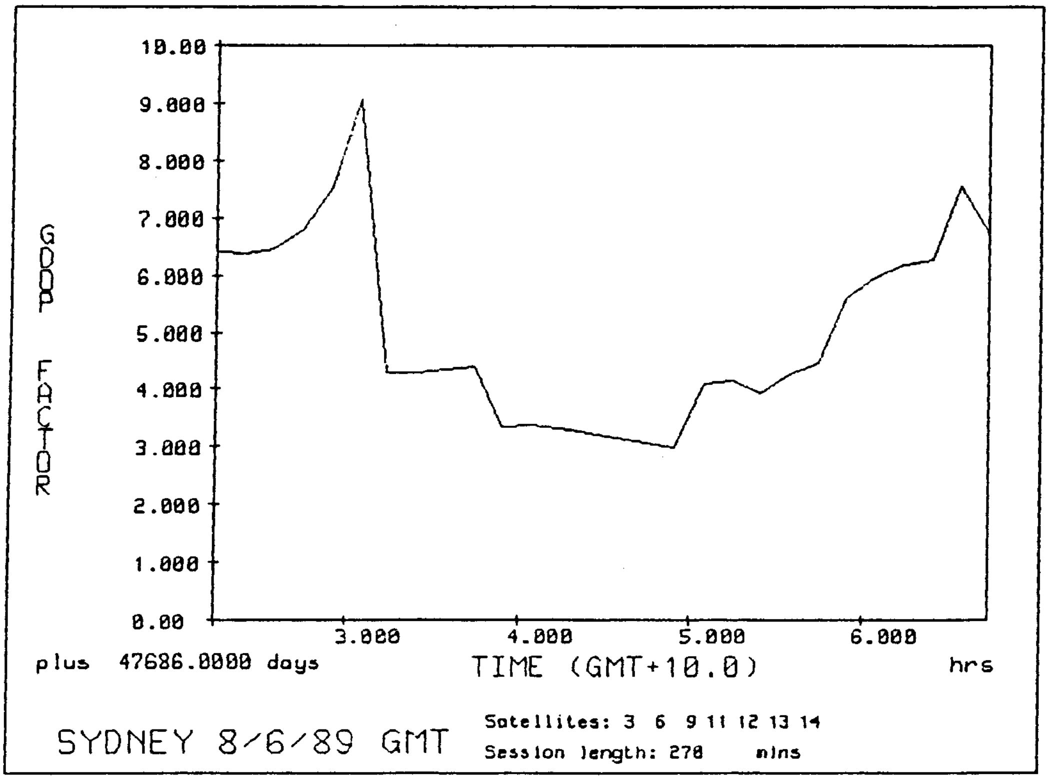
RIZOS, C., & GRANT, D.B., 1990. Time and the Global Positioning System. In Unisurv S-38, "Contributions to GPS Studies", C. Rizos (ed.), report of the School of Surveying, University of N.S.W.

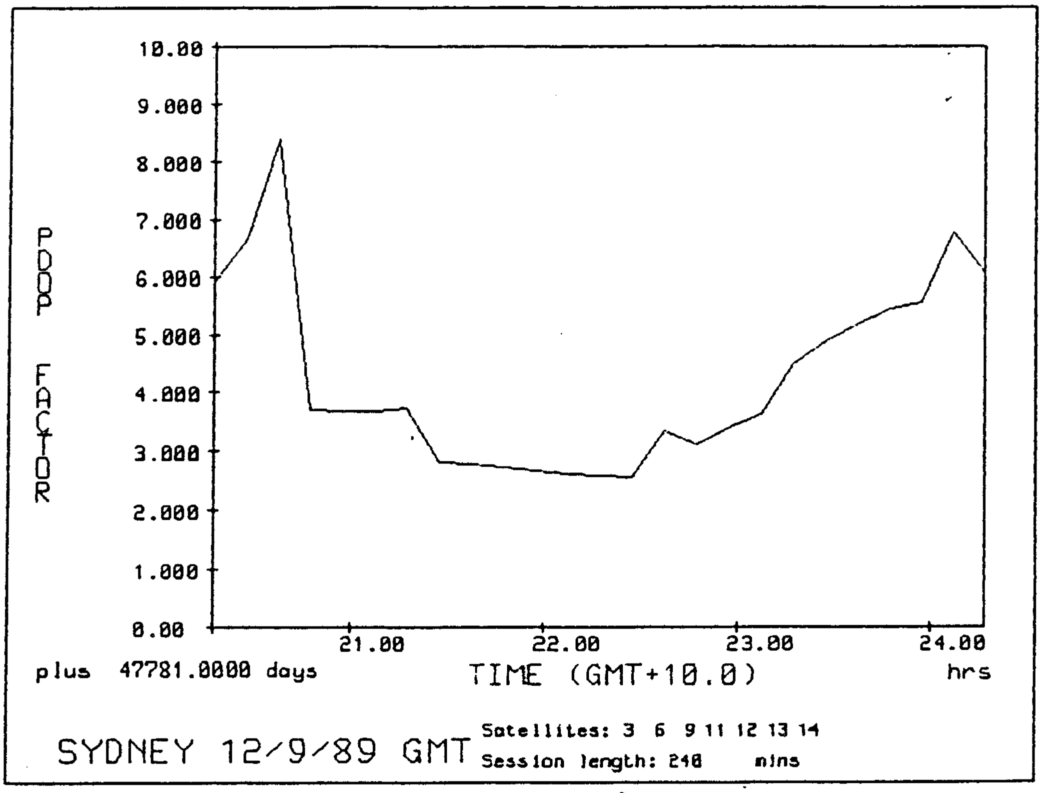
RÜEGER, J.M., 1984. Evaluation of an inertial surveying system. **Aust. Surveyor**, 32(2), 78-98.

WELLS, D.E., BECK, N., DELIKARAOGLOU, D., KLEUSBERG, A., KRAKIWSKY, E.J., LACHAPELLE, G., LANGLEY, R.B., NAKIBOGLU, M., SCHWARZ, K.P., TRANQUILLA, J.M., & VANICEK, P., 1986. Guide to GPS Positioning. Canadian GPS Associates, Fredericton, N.B., Canada.

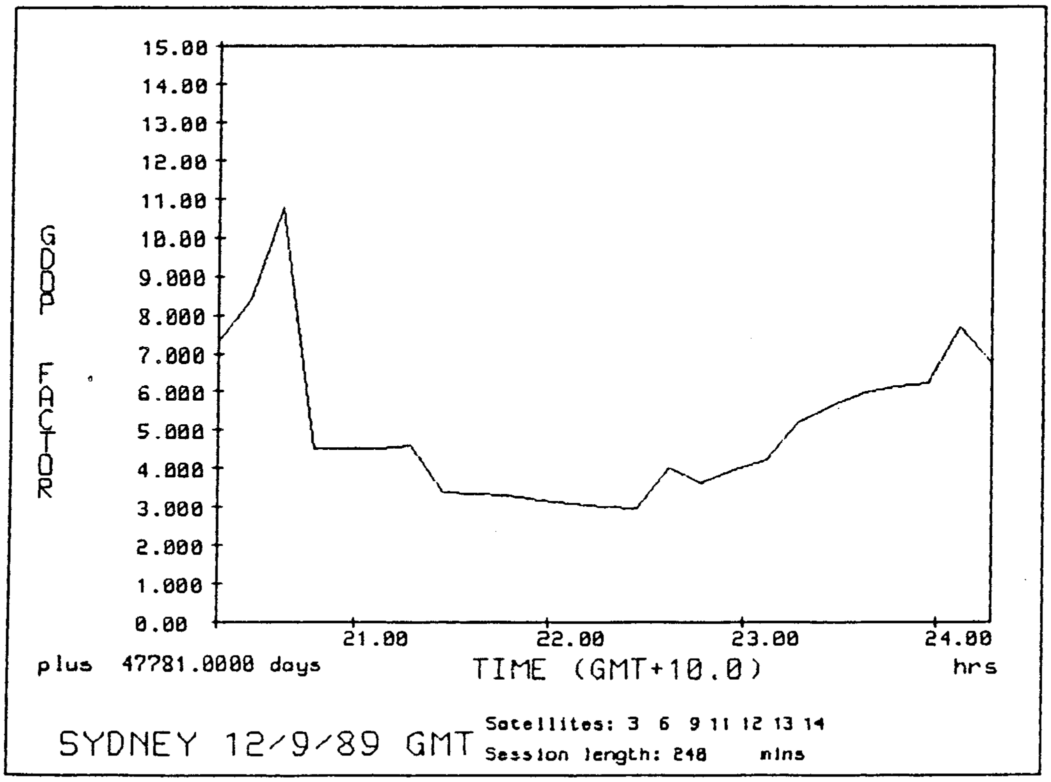
APPENDIX 1. PDOP / GDOP PLOTS FOR 8 JUNE 1989 AND 12 SEPTEMBER 1989





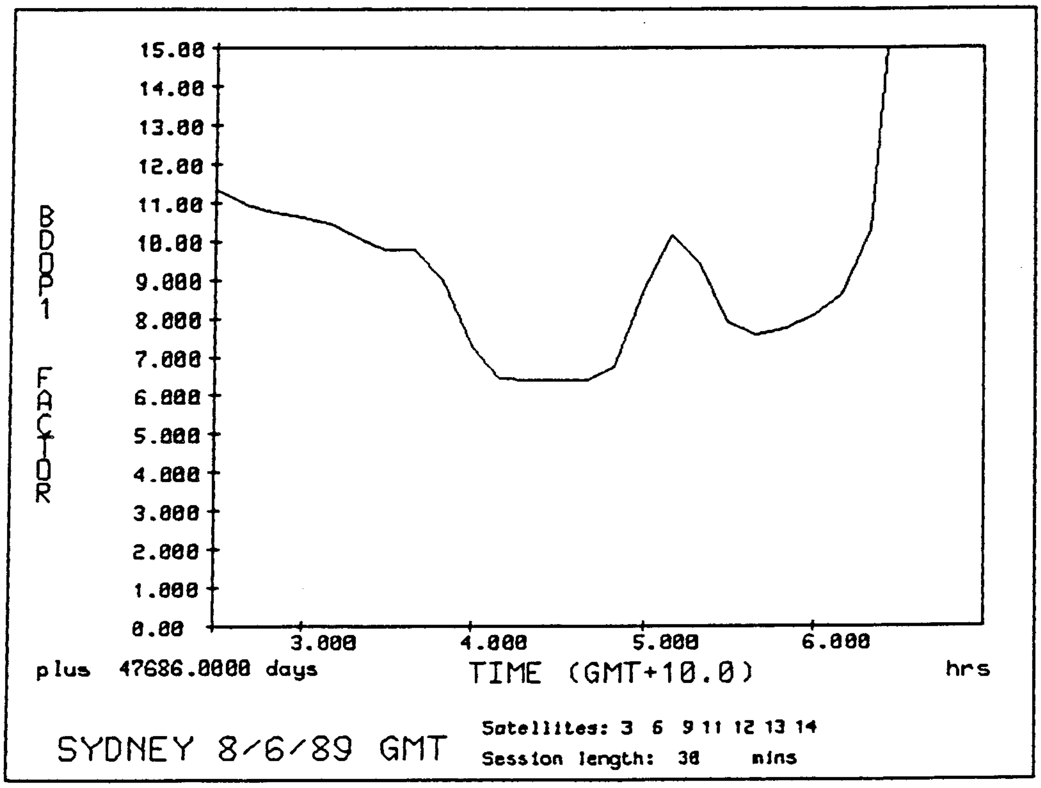


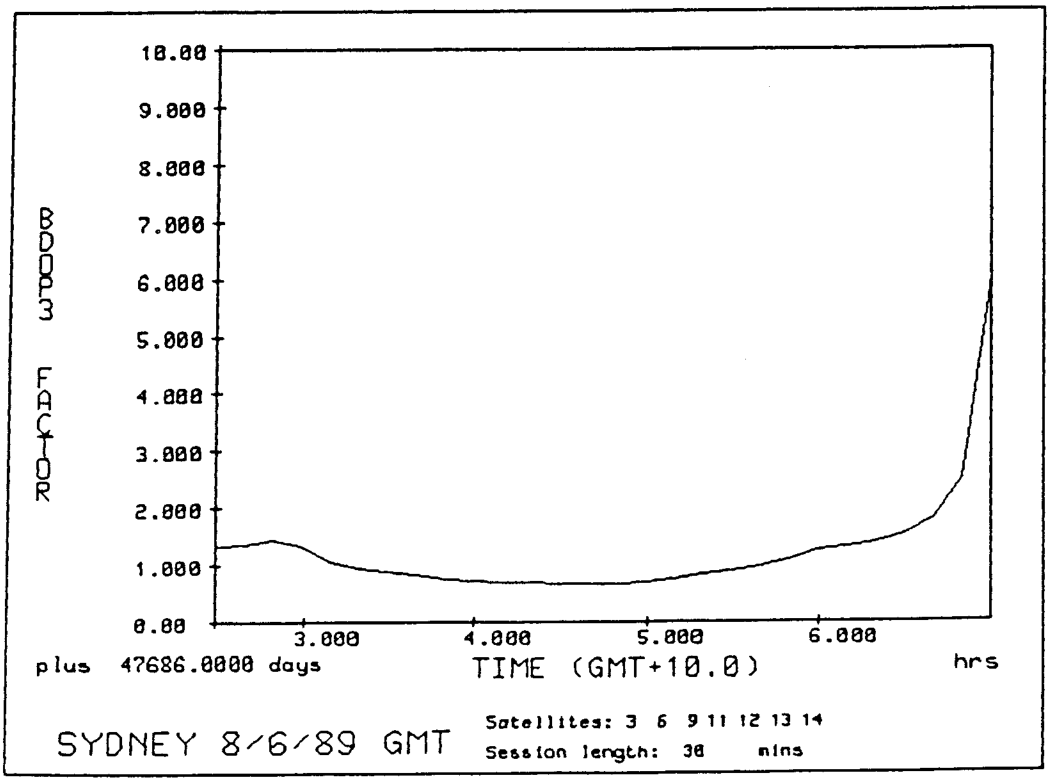
10.00
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.00
PDOP FACTOR
21.00
22.00
23.00
24.00
plus 47781.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 12/9/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 240 mins

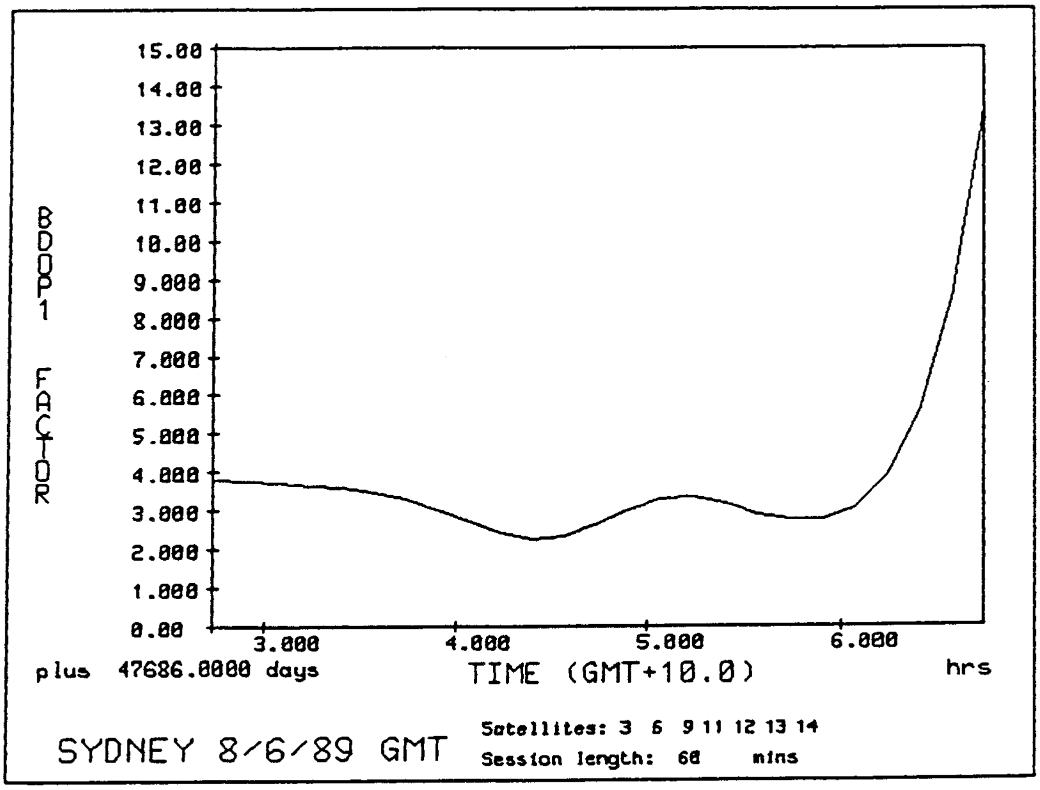


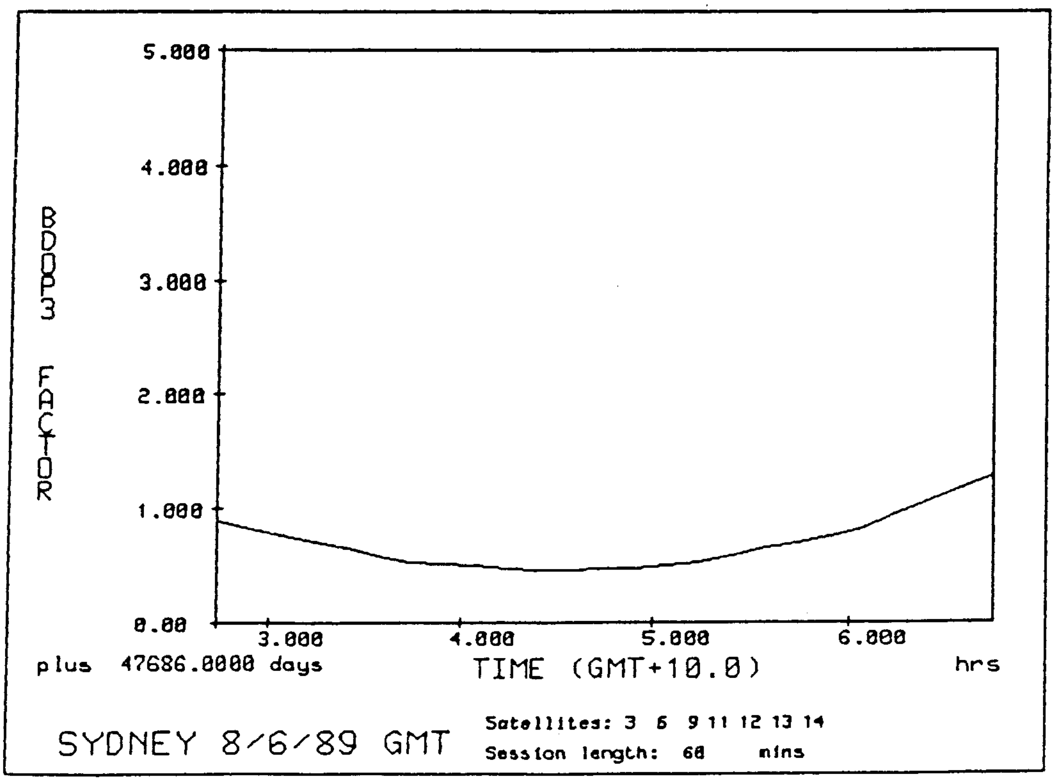
15.00
14.00
13.00
12.00
11.00
10.00
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.00
GDOP FACTOR
21.00
22.00
23.00
24.00
plus 47781.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 12/9/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 240 mins

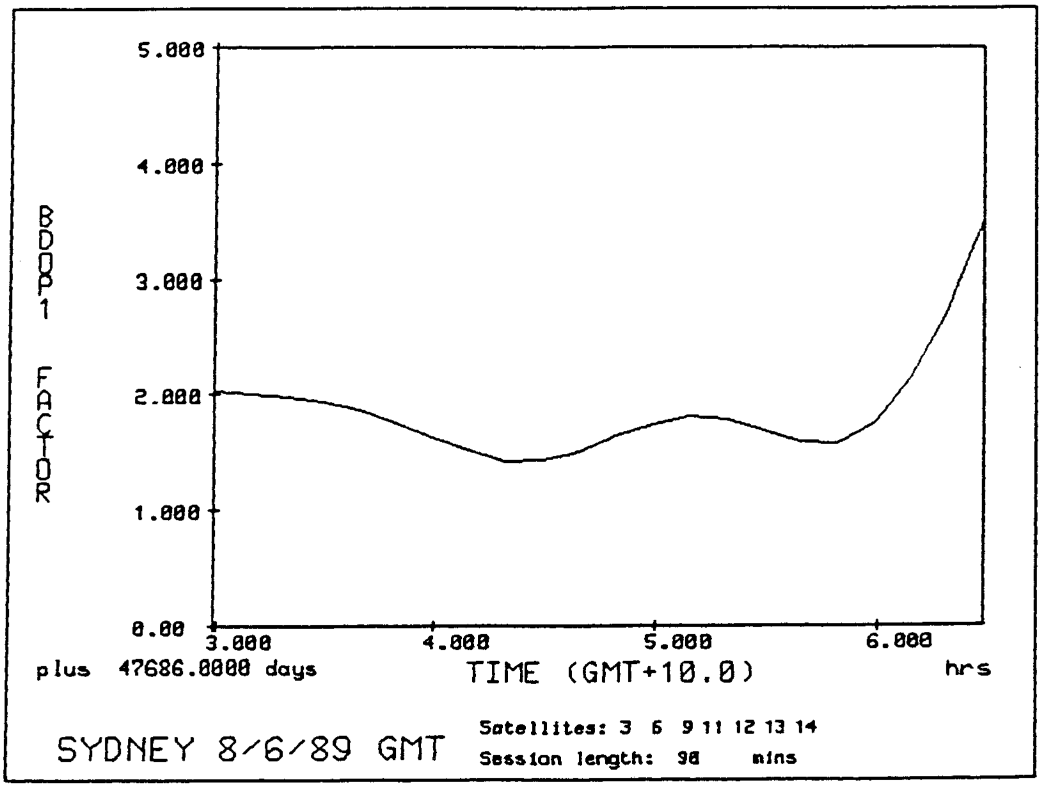
APPENDIX 2. BDOP1 / BDOP3 PLOTS FOR 8 JUNE 1989 FOR 30, 60, 90, 120 MINUTE DURATION OBSERVATION SESSIONS

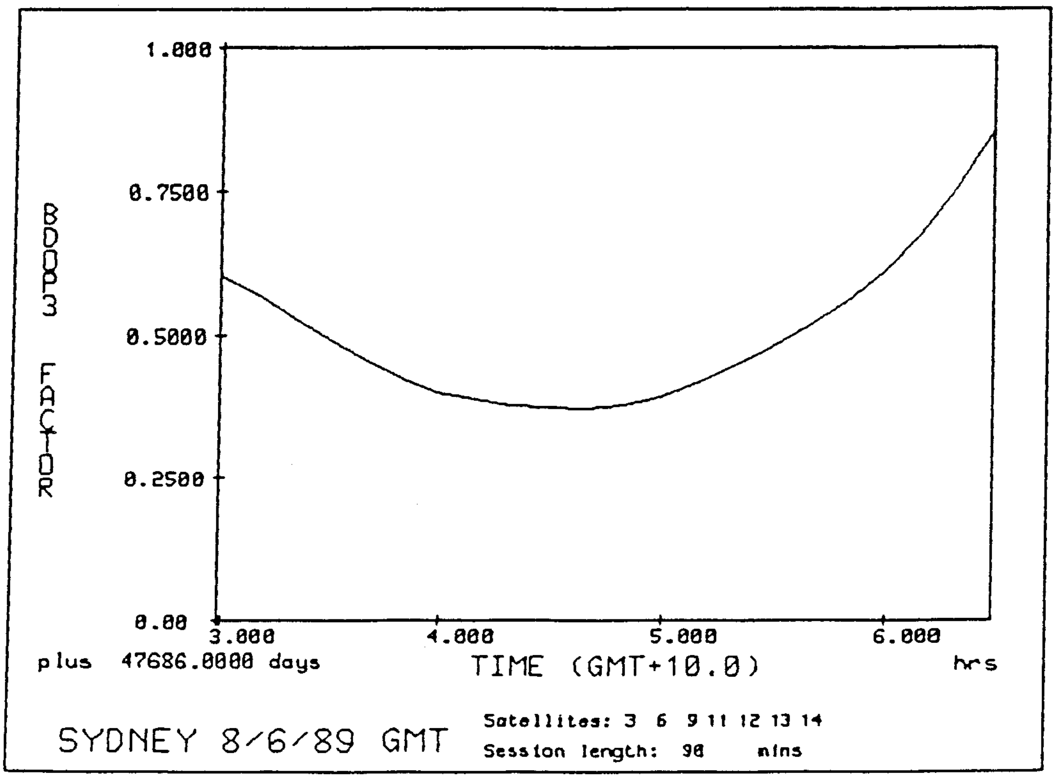


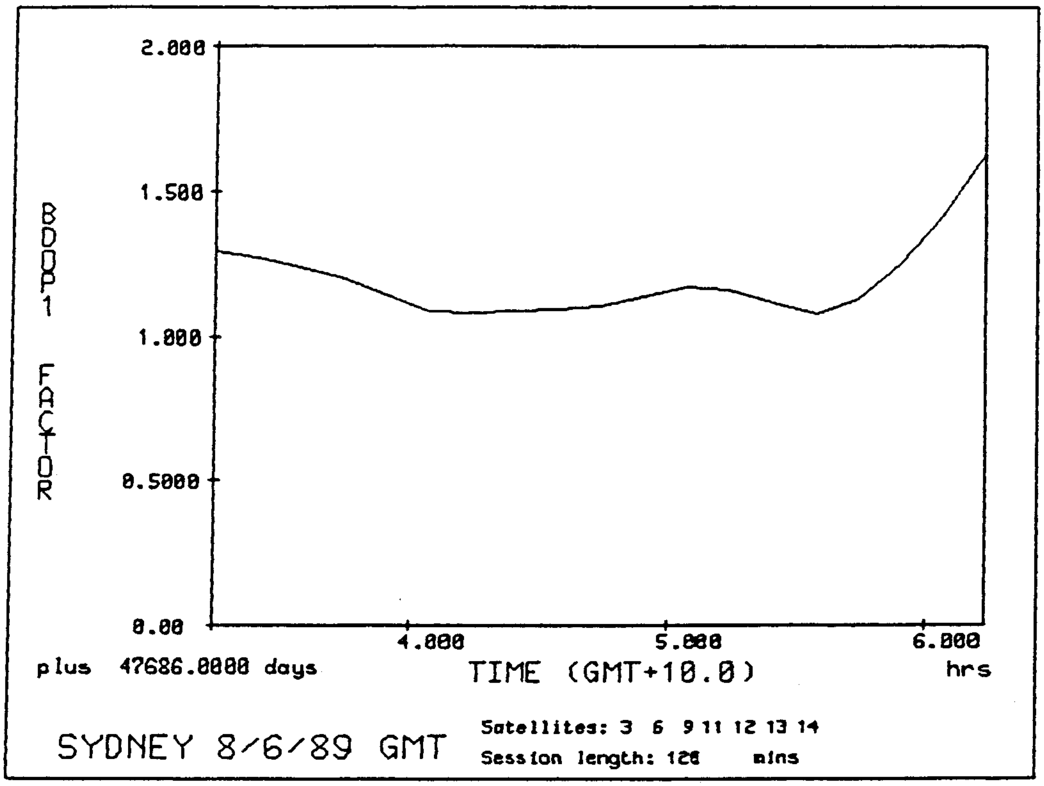



15.00
14.00
13.00
12.00
11.00
10.00
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.00
BDOP1 FACTOR
3.000
4.000
5.000
6.000
plus 47686.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 8/6/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 60 mins

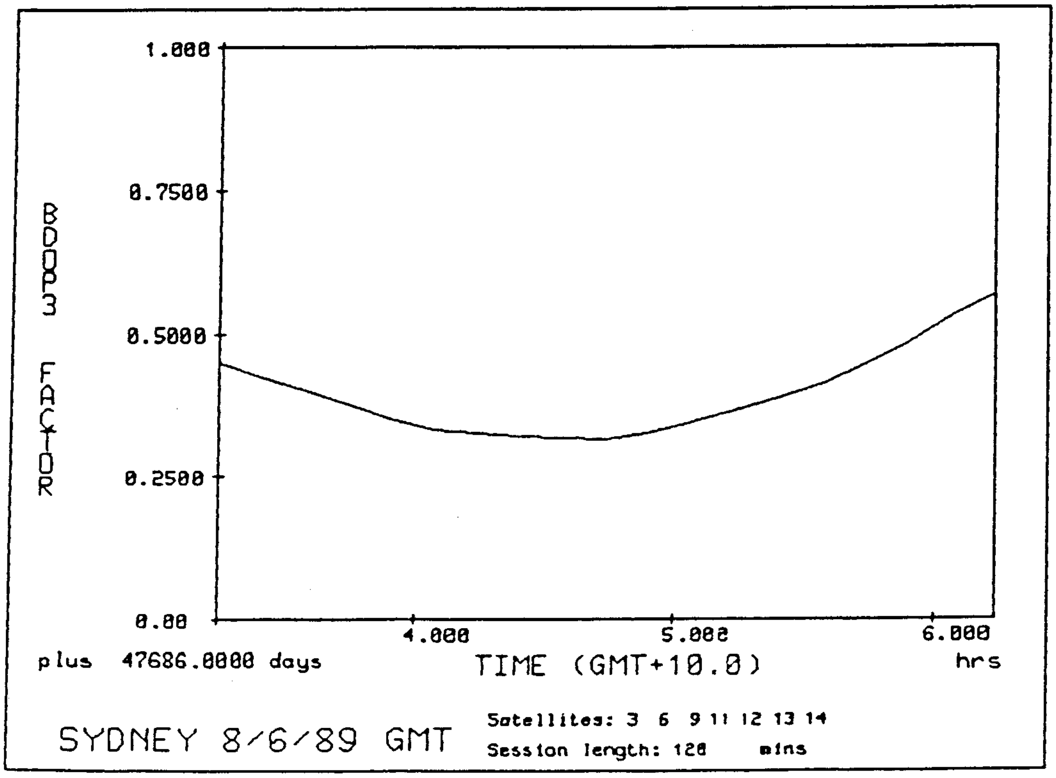

5.000
4.000
3.000
2.000
1.000
0.00
BDOP3 FACTOR
3.000
4.000
5.000
6.000
plus 47686.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 8/6/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 60 mins


5.000
4.000
3.000
2.000
1.000
0.00
BDOP1 FACTOR
3.000
4.000
5.000
6.000
plus 47686.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 8/6/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 90 mins


1.000
0.7500
0.5000
0.2500
0.00
BDOP3 FACTOR
3.000
4.000
5.000
6.000
plus 47686.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 8/6/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 90 mins



2.000
1.500
1.000
0.5000
0.00
BDOP1 FACTOR
4.000
5.000
6.000
plus 47686.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 8/6/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 120 mins



1.000
0.7500
0.5000
0.2500
0.00
BDOP3 FACTOR
4.000
5.000
6.000
plus 47686.0000 days
TIME (GMT+10.0)
hrs
SYDNEY 8/6/89 GMT
Satellites: 3 6 9 11 12 13 14
Session length: 120 mins

APPENDIX 3. THE CARRIER BEAT PHASE OBSERVABLE AND DIFFERENCING TECHNIQUES

Combinations of linear observations are often used in processing GPS data as they reduce or eliminate the effects of some of the linear biases in the observation equations: those due to receiver and satellite clock offsets and the initial (unknown) cycle ambiguities (GRANT et al, 1990). Differenced observations also tend to reduce the effects of orbit errors and the effects of unmodelled tropospheric and ionospheric delays. This reduction is more effective the less the separation of the simultaneously observing receivers.

A3.1 The Carrier Beat Phase Observable

The difference between the carrier signal generated by a receiver's internal oscillator and the carrier signal coming in from a satellite is the GPS observable. This is the phase of the signal which remains when the incoming Doppler shifted satellite signal is beat with the nominally constant reference frequency generated in the receiver. i.e.

$$\phi = \phi_i^k = \phi^k(t) - \phi_i(t) \qquad \text{(A3-1)}$$

Where: ϕ^k = phase of the signal transmitted by the k^{th} satellite at time t
ϕ_i = phase at the i^{th} receiver at reception time T

The carrier beat phase observable can be expressed using two relations:

(a) the phase / frequency relationship which is valid for highly stable oscillators over a short time interval:

$$\phi(t + \delta t) = \phi(t) + f.\delta t \qquad \text{(A3-2)}$$

Now, letting $\delta t = T - t$ we can set $\phi_i(T) = \phi^k(t) + f.(T - t)$ and obtain:

$$\phi = \phi^k(t) - \phi_i(T) = -f.(T - t) \qquad \text{(A3-3)}$$

(b) the fundamental approximation relating to the transmission and receiption times:

$$t + dt + (\rho - d_{ion} + d_{trop})/c = T + dt \qquad \text{(A3-4)}$$

and then

$$T - t = dt - dT + [(\rho - d_{ion} + d_{trop})/c] \qquad \text{(A3-5)}$$

and the carrier beat phase model is:

$$\phi = -(f/c).\rho - f.(dt - dT) - (f/c).(d_{ion} + d_{trop}) \qquad \text{(A3-6)}$$

In practice, the carrier beat phase measurement at some epoch t is based on a phase alignment of the receiver with the incoming carrier signal without the knowledge of which cycle would represent perfect cycle synchronisation. Hence:

$$\phi_{total} = Fr(\phi) + Int(\phi{:}t_0,t) + N(t_0) \qquad \text{(A3-7)}$$

Where: $Fr(\phi)$ = measured fractional phase part
$Int(\phi)$ = integer count of phase cycles from initial epoch to epoch t
N = unknown integer number of cycles at the initial epoch t_0

N is the **cycle ambiguity**, analoguous to the orientation unknown in horizontal direction measurements. As long as the receiver maintains continuous phase lock during an observing session, there is only one ambiguity per satellite-receiver pair. If there are breaks in this lock, cycle slips are said to have occurred. The receiver outputs $\phi_{measured} = Fr(\phi) + Int(\phi;t_0,t)$, therefore:

$$\phi_{total} = \phi_{measured} + N(t_0) \qquad \text{(A3-8)}$$

The equation for the instantaneous carrier beat phase observable for one satellite, one receiver, and at one epoch may now be written as:

$$\phi_{total} = (-f/c).\rho - f.(dt - dT) - (f/c).(-d_{ion} + d_{trop}) + N \qquad \text{(A3-9)}$$

The phase difference between the satellite and the receiver oscillators is now implied by the term $f.(dt - dT)$. If we multiply by the wavelength $\lambda=c/f$ and define $\phi = -\lambda.\phi_{measured}$, we get the corresponding carrier phase equation (in length units):

$$\phi = \rho + c.(dt - dT) + \lambda.N - d_{ion} + d_{trop} \qquad \text{(A3-10)}$$

There are a number of assumptions (LINDLOHR & WELLS, 1985) that need to be made concerning the observables before they can be used in a phase adjustment:

(a) that there are no problems with data simultaneity (ensuring that the relationship between measurement epochs at different receivers is taken into account, see RIZOS & GRANT, 1990),

(b) that there are no malfunctions, that is all receivers obtain data from each satellite at each epoch (to form double-differences), and

(c) the measurements have been pre-processed to remove cycle slips.

During an observation session involving R receivers, S satellites, and T epochs, the number of observed quantities ϕ is $n=RST$.

A3.2 Linear Combinations of Observations (Differencing)

Depending on the type of application and level of accuracy required, there are a number of advantages and disadvantages in forming certain linear combinations of the basic pseudo-range or carrier beat phase observables.

The fundamental aim of the differencing process is to take advantage of correlations to improve the accuracy of relative positions. Using the code or carrier phase measurements for relative positioning usually involves taking differences between measurements. In doing so, the effects of various errors common to the measurements being differenced are either removed or greatly reduced. There are a number of ways in which GPS measurements may be differenced. Differences may be performed between receivers, satellites, epochs, or any combination of these. (The most common solution is to difference in the above order.)

The notation that is used is shown below:

a. Δ denotes a between-receiver differencing of data to the same satellite,

b. ∇ denotes a between-satellite differencing of data at the same receiver, and

c. δ denotes a between-epoch data differencing operation.

A3.2.1 The Between-Satellite Single-Difference

An observation may be formed by differencing the observations to two satellites as simultaneously recorded at a single station (GRANT, 1990). The mathematical model of this between-satellite single-difference range observable in the form $\nabla(.)=(.)^{sat2}-(.)^{sat1}$ is given by:

$$\nabla\phi = \nabla\rho + c.\nabla dt + \nabla d_{ion} + \nabla d_{trop} \qquad \text{(A3-11)}$$

The corresponding model for the between-satellite single-difference carrier phase observable is:

$$\nabla\phi = \nabla\rho + c.\nabla dt + \lambda.\nabla N - \nabla d_{ion} + \nabla d_{trop} \qquad \text{(A3-12)}$$

This single-difference observable is free from receiver clock errors.

A3.2.2 Between-Receiver Double-Difference

If two satellites and two receivers at one epoch are considered, there are two possible differencing operations:

(a) two between-receiver single-differences, each involving a different satellite,

(b) two between-satellite single-differences, each involving a different receiver,

Regardless of the differencing option employed the two results, $\nabla\Delta(.)$ and $\Delta\nabla(.)$ are identical. The receiver-satellite double-differences for the carrier beat phase measurements is $\nabla\Delta\phi = \nabla\Delta\rho + \lambda.\nabla\Delta N - \nabla\Delta\delta d_{ion} + \nabla\Delta d_{trop}$.

The double-difference removes or greatly reduces the effects of:

(1) errors associated with the mis-alignment between the two receiver clocks,

(2) errors associated with mis-alignment between the two satellite clocks.

(3) satellite orbit errors,

(4) atmospheric delays of the observations.

Since the difference between epochs are not performed, whole cycle bias terms (composed of second-differences of the original whole-cycle biases) remain and are solved as part of the solution. A disadvantage of the double-differenced method is that another bias term would need to be added to the solution every time a loss of lock occurs in the data. For this reason, it is common to perform a pre-processing step to "repair" the loss of lock (or cycle slips) in the data. This is typically done by filtering the data, or some fitting the differenced data residuals with a polynomial.

A3.2.3 Receiver-Satellite-Time Triple-Differences

This triple-difference is the change in a receiver-satellite double-difference from one epoch to the next . The observation equation is:

$$\delta\nabla\Delta\phi = \delta\nabla\Delta\rho - \delta\nabla\Delta d_{ion} + \delta\nabla\Delta d_{trop} \qquad \text{(A3-13)}$$

The errors which are removed or reduced in double-differencing are also absent from triple-differences. In addition, for carrier beat phase measurements, the initial cycle ambiguity terms cancel. This property is the most attractive feature of this type of difference. It also allows much easier editing of cycle slips. With data which is known or suspected to contain cycle slips, by forming the triple-difference first, the cycle slips may be edited and then the repaired differences formed. The triple-difference eliminates biases in:

(1) receiver clocks,
(2) satellite clocks, and
(3) initial cycle ambiguity.

In addition, the number of observations has also been reduced.

Differencing takes advantage of the physical correlations which occur in the measurement by mathematically eliminating their equivalent functional correlations. When differencing occurs, the original carrier beat phase observations and their one-way phase mathematical model are modified. This means that both the functional and stochastical model will change. When differencing occurs the resulting covariance matrix will generally show stochastic correlation due to the law of propagation of variances. See Apendix 4.

Single-differencing between two receivers reduces both the number of observations and unknowns by the same amount, without introducing stochastic correlations. The variance of single-differences are twice as large as that of one-way data (MERMINOD, 1990). However, not all single-differences can be independent if more than two receivers are considered, since data from some receivers must be used several times to build a set of single-differences that includes all receivers in a session. Therefore, stochastic correlation is introduced into the single-difference mathematical model.

Double- and triple-differences exhibit stochastic correlation no matter what differencing operator is used. The covariance matrix of the double- and triple-difference observables can be made diagonal using a recursive algorithm (MERMINOD, 1990) but it is computationally cumbersome. It is simpler to take the correlations into account in the reduction process. Regardless of the level of differencing, differenced observables can be thought of as the result of applying a differencing operator **D** to the original undifferenced observables. For example (LINDLOHR & WELLS, 1985), consider the undifferenced equation:

$$\phi = \rho + c.(dt\text{-}dT) + \lambda.N - d_{ion} + d_{trop} \qquad \text{(A3-14)}$$

Any combination of the differencing operations examined (∇, Δ, δ) may be represented by an operator (matrix) **D** which is composed entirely of elements +1, -1, and 0 operating on a vector **l** of observed ranges ρ or ϕ. If the covariance matrix of the vector of ranges **l** is $\mathbf{C_l}$ then the covariance matrix of the differenced observations $\mathbf{D_l}$ is $\mathbf{C_{Dl}=DC_lD^t}$. This provides a straight-forward, if not computationally inconvenient, method of accounting for the mathematical correlations introduced by differencing (MERMINOD, 1990).

APPENDIX 4. GPS CORRELATIONS

A4.1 Correlations Between Observations (MacLEOD & RIZOS, 1988)

Correlations express the inter-dependence between variables. For two variables x, and y, in a linear relationship, the correlation between them is defined as:

$$\rho_{xy} = \sigma_{xy}/\sigma_x.\sigma_y \qquad \text{(A4-1)}$$

Where: σ_x = the standard deviation of x
σ_y = the standard deviation of y
σ_{xy} = the covariance of x and y.

When a correlation between two variables approaches the maximum of +1 or -1, the two variables are said to be highly correlated. Correlations in GPS can be either:

a. physical, and/or

b. mathematical (functional or statistical).

A4.1.1 Physical Correlation

Physical correlation refers to the correlations between the actual field observations, for example, an examination of the plots of the carrier phase data from different GPS satellites observed by the same receiver will vary in a similar manner showing physical correlation (LINDLOHR & WELLS, 1985). This occurs as the measurements are made using the same GPS receiver clock.

A4.1.2 Mathematical Correlation

This is related to the parameters in the mathematical model. It can therefore be divided into two types which correspond to those two components of the mathematical model:

a. **Functional Correlation.** Here the physical correlations can be taken into account by introducing adequate terms in the functional model of the observations. That is, functionally correlated quantities have the same parameter in the observation model. An example is a clock bias parameter which can be placed in the GPS observation equation to account for the physical correlation introduced into the measurements by the particular receiver or satellite clock involved with the measurement.

b. **Stochastic Correlation.** Stochastic correlation (also sometimes called statistical correlation) will occur between observations when off-diagonal elements are introduced into the covariance matrix of the observations. This correlation also appears when functions of the observations are considered, due to the law of the propagation of variances. Even if the covariance matrix of the observations is diagonal, (no stochastic correlation), the covariance matricies of the resultant least squares estimates of the parameters and residuals will generally be full covariance matricies, and therefore exhibit stochastic correlation.

A4.2 The Effects of GPS Correlations

A GPS campaign will generally set out to determine a set of coordinates for those stations within the network. During an observing session, several receivers simultaneously track carrier data over a number of epochs. The receivers are deployed over several sessions until all stations in the network have been occupied. The correlations which exist are easier to realise if the GPS campaign is considered in two distinct phases. Firstly, to determine positions within a session, and secondly, to combine the results of these sessions together in an optimal way.

For single receiver operation, the observations can be considered as a range (or ranges) between the satellite and the receiver. The accuracy of calculating the absolute position of a single point is a function of the accuracies of the observed quantities in the functional model. Due to the uncertainties in the satellite's position, satellite and receiver clock errors and propagation delays, the absolute point position will be accurate to no better than several dekametres.

When the satellite's signal is received at several receivers, the error sources in the GPS signals will exhibit some physical correlations. This means that for most survey applications it is necessary to use GPS in a **relative mode** in which two or more receivers observe the same satellites simultaneously. These physical correlations can be mathematically modelled as bias terms in the functional model. Using mathematical operations such as "differencing", these effects can be estimated or greatly reduced resulting in centimetre level accuracy.

The fundamental problem which is encountered when performing differencing operations is to accurately and completely determine to what extent various elements are correlated and how well they can be modelled. Physical correlations may be both spatial and temporal in nature, or a combination of both of these. An example of this is the propagation delay which a satellite's signal experiences as it is propagated through the ionosphere and the troposphere. This delay is a significant source of both spatial and temporal physical correlations. To completely account for the correlations from this one error source would require a realistic model of the atmospheric parameters. However, the viable alternatives for treating this delay are to estimate it, model or ignore it in the mathematical model.

A4.3 The GPS Functional Model

The observation model for GPS carrier beat phase measurements can be expressed as:

$$l = f(X_R, X^S, b) \qquad \text{(A4-2)}$$

Where: X_R = the position vector of the receiver,
X^S = the position vector of the satellite,
b = the position vector of the terms that are considered as biases, which are functions of the measuring process and the observations themselves.

The functional model described by MERMINOD (1990) defines the classic observation equation for the carrier phase from one satellite to one receiver as:

$$\phi_R{}^S(t) = \lambda^{-1}((X^S(t) - X_R(t))^2)^{0.5} + f.\varepsilon_R(t) - f.\varepsilon^s(t)\ n_R{}^S + (\phi_{atmos})_R{}^S$$

Where: f is the nominal frequency of the carrier wave
X is the position vector of the satellite or receiver
ε^s is the satellite oscillator (clock) error
ε_R is the receiver oscillator (clock) error
$n_R{}^s$ is the initial number of cycles of the carrier phase
ϕ_{atmos} is the phase delay due to the atmosphere

λ is the nominal wavelength of the GPS signal.

With the addition of a ϕ_{noise} term to the functional model (which considers the effects due to measurement noise and unmodelled influences, both of which are considered to be random) the above equation becomes the one-way phase observation equation. This equation has been discussed by authors such as GOAD (1985), LINDLOHR & WELLS (1985), and GRAFAREND et al (1985), however it is the basis for only a few GPS data reduction software programs (for example PHASER and USMASS - GRANT et al, 1990). Other functional models may be derived using linear combinations (examined earlier) of the carrier beat phase observable.

APPENDIX 5. METEOROLOGICAL INFLUENCES

A5.1 The Troposphere

The tropospheric (neutral atmosphere) refraction effects are minimised through the use of standard models which (in PoPS™) do not permit user interaction / modification. For surveys of less than a few tens of kilometres, the tropospheric delay will tend to be the same at both ends of the baseline and thus assuming similar tropospheric conditions prevail, the effects will largely cancel during the double-differencing operation. The error introduced in baseline coordinates under a wide range of conditions (including rain) will be less than 1 ppm.

Tropospheric delays are largely a function of the elevation of the satellite and the altitude of the receiver, and are dependent on the frequency of the carrier wave, atmospheric pressure, temperature, and water vapour pressure. There is significantly less tropospheric range delay error at higher altitude. For example, if two receivers are deployed, one at sea level and the other at 3,000m (ASL), assuming a satellite elevation angle of 20°, the tropospheric refraction error for the receiver at sea level will be approximately 7.8m. The receiver positioned at altitude will experience a tropospheric refraction error of the order of 4.8m. The refractive error results mostly in a degradation of the height component in the solution. There is only a minimal effect on latitude and longitude.

A5.1.1 Tropospheric Errors

Tropospheric errors can be divided into relative and absolute components. Relative tropospheric errors imply different prevailing tropospheric conditions existing at either end of a baseline during an observation session. For baselines over 100km, many GPS manufacturers (BEUTLER et al, 1989) suggest that meteorological measurements are made (temperature T, pressure P, Humidity e) at regular intervals throughout the observation session. This is not as successful for small networks due to influences of local meteorological conditions that may exist at individual stations (for example fog, temperature inversions, rain). This is further compounded by calibration errors of measuring instruments and observer error in the field. To investigate the magnitude of these relative tropospheric errors BAUERSIMA (1983) used a simplified version of Saastamoinen's formula, which expresses the tropospheric refraction correction for a given zenith distance:

$$\Delta r(z) = \Delta r(0)/\cos z \qquad \text{(A5-1)}$$

Where: $\Delta r(z)$ is the refraction correction at a zenith distance z
$\Delta r(0)$ is the corresponding correction at z=0

and

$$\Delta r(0) = 2.277\,[\,P + (\{1255/(273 + T)\}+0.05)\,e] \qquad \text{(A5-2)}$$

Where: $\Delta r(0)$ is the zenith correction in mm,
P is the atmospheric pressure in millibars,
T is the temperature in °C,

e is the water vapour pressure approximated by:

$$e = [H/100].\exp(-37.2465 + 0.213166\{T + 273\} - 0.000256908[T+273]^2) \quad \text{(A5-3)}$$

Where H is the humidity expressed in %.

In Table A5.1 below, the partial derivatives of $\Delta r(0)$ with respect to T, P, and H are given. These partials define the bias introduced into the observable ΔP or ΔP^*, if one of the stations records an incorrect temperature of 1^oC, a wrong pressure of 1 mbar, or an incorrect humidity of 1% respectively.

Table A5.1 The Tropospheric Refraction Correction's Meteorological Dependence $\Delta r(z=0)$ (BEUTLER et al, 1989)

T ^oC	P m bar	H %	$\delta\Delta r/\delta T$ mm/^oC	$\delta\Delta r/\delta P$ mm/mbar	$\delta\Delta r/\delta H$ mm/(1%)
0^o	1000	100	5	2	0.6
30^o	1000	100	27	2	4.0
0^o	1000	50	3	2	0.6
30^o	1000	50	14	2	4.0

From an examination the above table we can conclude that the partials are all of the order of millimetres and that in the worst observation case, a combination of high humidity and high temperatures, causes adverse effects to results. The consequences of using incorrect meteorological data in a small network are significant in terms of the large part per million (ppm) error that is introduced.

Absolute tropospheric errors will affect the baseline if identical meteorological conditions prevail at both ends of the baseline throughout the session. Under such conditions an assumption could be made that as a result, no bias will be introduced. This is not the case as the zenith distance of a satellite is different at each end of the baseline. The geometry describing tropospheric refraction is shown below in Figure A5.1 (BEUTLER et al, 1989).

The bias introduced into the observable by assuming uniform meteorological conditions existing at both ends of the baseline may be expressed as:

$$\varepsilon(a,z) = \Delta r(o)\{(\cos z_1)^{-1} - (\cos z_2)^{-1}\} \quad \text{(A5-4)}$$

Where $\Delta r(o)$ is the tropospheric zenith correction defined in equation (A5-2).

In PoPS™, there are four tropospheric models to choose from. The choice of model may be made from either Saastamoinen, Hopfield I, Hopfield II or no model at all. For small surveys of less than 10km, WILD recommends that modelled values of T, P, and e are used instead of observed meteorological values, as small systematic errors in meteorological values can often introduce larger biases than the assumption of a standard atmosphere. All of the various models were used in a test to determine the effect of each model on the coordinates determined for SPM8293. The differences that resulted in the coordinates of SPM8293 for the various models were insignificant. Saastamoinen's model was then used for all subsequent data processing.

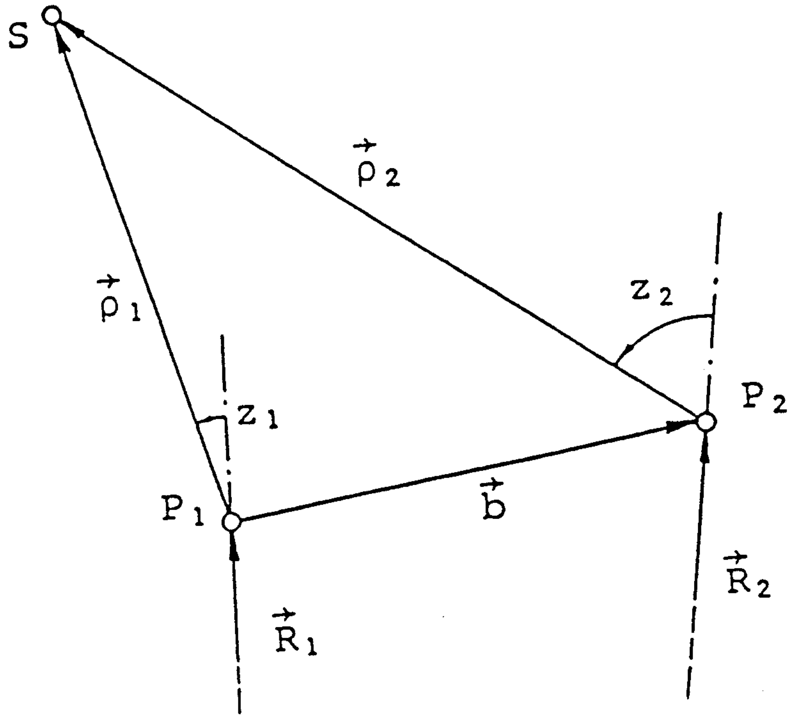


Figure A5.1 The Geometry Governing Tropospheric Refraction

Where: S is the observed satellite
P_1, P_2 are the receiver locations
R_1, R_2 are the geocentric locations of the satellites
ρ_1, ρ_2 are the topocentric positions of the satellites
b is the observed baseline
z_1, z_2 are the zenith distances of the satellite from P_1, P_2

A5.2 The Ionospheric Model

The ionosphere is that part of the atmosphere lying between 70 and 1000km above the earth's surface. In this layer, the sun's ultra-violet radiation ionises gas molecules which then lose an electron. These free electrons in the ionosphere influence the propagation of microwave signals as they pass through the layer. At night the ionospheric delay is approximately five times less than for day time observations. The refractive index of microwaves is a function of frequency and may be expressed:

$$n = 1 +/- (A \, . \, N_e) / f^2 \qquad \text{(A5-5)}$$

Where: A is a constant
N_e is the free electron density, and
f is the frequency

The nature of the sign will depend on whether the group (+) or the phase (-) refractive index is being calculated. The ionospheric zenith delay τ_o for a microwave propagating between a satellite and a receiver is given as (KING et al, 1987):

$$\tau_{ion} = \phi_{ion} / f = -(1.35 \times 10^{-7}) \cdot N_e / f^2 \qquad \text{(A5-6)}$$

Where: τ_{ion} is in seconds
ϕ_{ion} is in cycles
f is the signal frequency (Hz), and
Ne is the total electron content (TEC). TEC is expressed as the number of free electrons (at the zenith) per square metre

A5.2.1 The Total Electron Content (TEC)

There are a number of factors which influence the magnitude of TEC. These are the latitude of the receiver, the time of day the observation is made, and the level of solar activity at the time of observation. TEC is a maximum at mid to low latitudes and is a minimum at the poles. A diurnal cycle for TEC occurs with a maximum occuring two hours after solar noon (+/- 20^o latitude) and is a minimum before dawn. Ionospheric disturbances which can occur suddenly (and are very severe) also affect the value of TEC. Under such conditions the ionosphere is so perturbed that single frequency operations are virtually impossible.

Within PoPS™ there are two options available to deal with ionospheric refraction. Either a user may employ a single layer model or no model at all. The single layer model is the simplest way of accounting for ionospheric refraction. When using this model a number of important assumptions are made:

a. that all free electrons in the ionosphere are located in an infinitesimally thin spherical layer of height H above the Earth's surface (see Figure A5.2 below), and
b. TEC is of uniform density in the layer.

When observing L1 data only, a scale factor will be introduced which will shorten baselines if the ionosphere is neglected. This effect expressed in ppm is (BEUTLER et al, 1989):

$$\Delta L / L = -0.7 \cdot 10^{-17} \, TEC \qquad \text{(A5-7)}$$

The effect of this scale error can range from 0.35ppm to 3.5ppm for baselines determined from L1 observations.

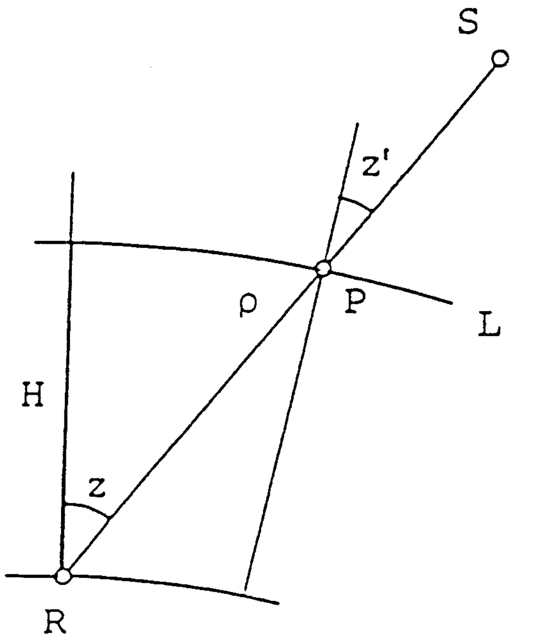


Figure A5.2. Single Layer Ionospheric Model
(adapted from BEUTLER et al, 1989)

Where:
- S is the satellite
- R is the receiver
- L is the ionospheric delay of infintesimal thickness
- H is the height of the ionospheric layer
- z zenith distance of the satellite as seen from the receiver
- z' the zenith distance of the satellite as seen from intersection point P of the satellite signal path with the ionospheric layer

Publications from

THE SCHOOL OF SURVEYING, THE UNIVERSITY OF NEW SOUTH WALES.

All prices include postage by surface mail. Air mail rates on application. (Effective Sept. 1990)

To order, write to Publications Officer, School of Surveying, The University of New South Wales, P.O. Box 1, Kensington N.S.W., 2033 AUSTRALIA

NOTE: ALL ORDERS MUST BE PREPAID

UNISURV REPORTS - G SERIES

Price (including postage): $6.00

G14. A. Stolz, "The computation of three dimensional Cartesian coordinates of terrestrial networks by the use of local astronomic vector systems", Unisurv Rep. 18, 47 pp, 1970.

G16. R.S. Mather et al, "Communications from Australia to Section V, International Association of Geodesy, XV General Assembly, International Union of Geodesy and Geophysics, Moscow 1971", Unisurv Rep. 22, 72 pp, 1971.

G17. Papers by R.S. Mather, H.L. Mitchell & A. Stolz on the following topics:- Four-dimensional geodesy, Network adjustment and Sea surface topography, Unisurv G17, 73 pp, 1972.

G18. Papers by L. Berlin, G.J.F. Holden, P.V. Angus-Leppan, H.L. Mitchell & A.H. Campbell on the following topics:- Photogrammetry co-ordinate systems for surveying integration, Geopotential networks and Linear measurement, Unisurv G18, 80 pp, 1972.

G19. R.S. Mather, P.V. Angus-Leppan, A. Stolz & I. Lloyd, "Aspects of four-dimensional geodesy", Unisurv G19, 100 pp, 1973.

G20. Papers by J.S. Allman, R.C. Lister, J.C. Trinder & R.S. Mather on the following topics:- Network adjustments, Photogrammetry, and 4-Dimensional geodesy, Unisurv G20, 133 pp, 1974.

G21. Papers by E. Grafarend, R.S. Mather & P.V. Angus-Leppan on the following topics:- Mathematical geodesy, Coastal geodesy and Refraction, Unisurv G21, 100 pp, 1974.

G22. Papers by R.S. Mather, J.R. Gilliland, F.K. Brunner, J.C. Trinder, K. Bretreger & G. Halsey on the following topics:- Gravity, Levelling, Refraction, ERTS imagery, Tidal effects on satellite orbits and Photogrammetry, Unisurv G22, 96 pp, 1975.

G23. Papers by R.S. Mather, E.G. Anderson, C. Rizos, K. Bretreger, K. Leppert, B.V. Hamon & P.V. Angus-Leppan on the following topics:- Earth tides, Sea surface topography, Atmospheric effects in physical geodesy, Mean sea level and Systematic errors in levelling, Unisurv G23, 96 pp, 1975.

G24. Papers by R.C. Patterson, R.S. Mather, R. Coleman, O.L. Colombo, J.C. Trinder, S.U. Nasca, T.L. Duyet & K. Bretreger on the following topics:- Adjustment theory, Sea surface topography determinations, Applications of LANDSAT imagery, Ocean loading of Earth tides, Physical geodesy, Photogrammetry and Oceanographic applications of satellites, Unisurv G24, 151 pp, 1976.

G25. Papers by S.M. Nakiboglu, B. Ducarme, P. Melchior, R.S. Mather, B.C. Barlow, C. Rizos, B. Hirsch, K. Bretreger, F.K. Brunner & P.V. Angus-Leppan on the following topics:- Hydrostatic equilibrium figures of the Earth, Earth tides, Gravity anomaly data banks for Australia, Recovery of tidal signals from satellite altimetry, Meteorological parameters for modelling terrestrial refraction and Crustal motion studies in Australia, Unisurv G25, 124 pp, 1976.

G26. Papers by R.S. Mather, E.G. Masters, R. Coleman, C. Rizos, B. Hirsch, C.S. Fraser, F.K. Brunner, P.V. Angus-Leppan, A.J. McCarthy & C. Wardrop on the following topics:- Four-dimensional geodesy, GEOS-3 altimetry data analysis, analysis of meteorological measurements for microwave EDM and Meteorological data logging system for geodetic refraction research, Unisurv G26, 113 pp, 1977.

G27. Papers by F.K. Brunner, C.S. Fraser, S.U. Nasca, J.C. Trinder, L. Berlin, R.S. Mather, O.L. Colombo & P.V. Angus-Leppan on the following topics:- Micrometeorology in geodetic refraction, LANDSAT imagery in topographic mapping, adjustment of large systems, GEOS-3 data analysis, Kernel functions and EDM reductions over sea, Unisurv G27, 101 pp, 1977.

G29. Papers by F.L. Clarke, R.S. Mather, D.R. Larden & J.R. Gilliland on the following topics:- Three dimensional network adjustment incorporating ξ , η and N, Geoid determinations with satellite altimetry, Geodynamic information from secular gravity changes and Height and free-air anomaly correlation, Unisurv G29, 87 pp, 1978.

From June 1979 Unisurv G's name was changed to Australian Journal of Geodesy, Photogrammetry and Surveying. These can be ordered from The Managing Editor, Australian Journal of Geodesy, Photogrammetry and Surveying, Institution of Surveyors - Australia , Nos 27 - 29 Napier Close, Deakin, ACT 2600, AUSTRALIA.

UNISURV REPORTS - S SERIES

S8 - S20	Price (including postage):		$10.00
S27 onwards	Price (including postage):	Individuals	$20.00
		Institutions	$28.00

S8 A. Stolz, "Three-D Cartesian co-ordinates of part of the Australian geodetic network by the use of local astronomic vector systems", Unisurv Rep. S 8, 182 pp, 1972.

S9 H.L. Mitchell, "Relations between MSL & geodetic levelling in Australia", Unisurv Rep. S 9, 264 pp, 1973.

S10 A.J. Robinson, "Study of zero error & ground swing of the model MRA101 tellurometer", Unisurv Rep. S 10, 200 pp, 1973.

S12. G.J.F. Holden, "An evaluation of orthophotography in an integrated mapping system", Unisurv Rep. S 12, 232 pp, 1974.

S14. Edward G. Anderson, "The Effect of Topography on Solutions of Stokes` Problem", Unisurv Rep. S 14, 252 pp, 1976.

S15. A.H.W. Kearsley, "The Computation of Deflections of the Vertical from Gravity Anomalies", Unisurv Rep. S 15, 181 pp, 1976.

S16. K. Bretreger, "Earth Tide Effects on Geodetic Observations", Unisurv S 16, 173 pp, 1978.

S17. C. Rizos, "The role of the gravity field in sea surface topography studies", Unisurv S 17, 299 pp, 1980.

S18. B.C. Forster, "Some measures of urban residual quality from LANDSAT multi-spectral data", Unisurv S 18, 223 pp, 1981.

S19. Richard Coleman, "A Geodetic Basis for recovering Ocean Dynamic Information from Satellite Altimetry", Unisurv S 19,332 pp, 1981.

S20. Douglas R. Larden, "Monitoring the Earth's Rotation by Lunar Laser Ranging", Unisurv Report S 20, 280 pp, 1982.

S27. Bruce R. Harvey, "The Combination of VLBI and Ground Data for Geodesy and Geophysics", Unisurv Report S27, 239 pp, 1985.

S28. Rod Eckels, "Surveying with GPS in Australia", Unisurv S28, 220 pp, 1987.

S29. Gary S. Chisholm, "Integration of GPS into hydrographic survey operations", Unisurv S29, 190 pp, 1987.

S30. Gary Alan Jeffress, "An investigation of Doppler satellite positioning multi-station software", Unisurv S30, 118 pp, 1987.

S31. Jahja Soetandi, "A model for a cadastral land information system for Indonesia", Unisurv S31, 168 pp, 1988.

S32. D. B. Grant, "Combination of terrestrial and GPS data for earth deformation studies" Unisurv S32, 285 pp, 1990.

S33. R. D. Holloway, "The integration of GPS heights into the Australian Height Datum", Unisurv S33, 151 pp.,1988.

S34. Robin C. Mullin, "Data update in a Land Information Network", Unisurv S34, 168 pp. 1988.

S35. Bertrand Merminod, "The use of Kalman filters in GPS Navigation", Unisurv S35, 203 pp., 1989.

S36. Andrew R. Marshall, "Network design and optimisation in close range Photogrammetry", Unisurv S36, 249 pp., 1989.

S37. Wattana Jaroondhampinij, "A model of Computerised parcel-based Land Information System for the Department of Lands, Thailand," Unisurv S37, 281 pp., 1989.

S38. C. Rizos (Ed.), D.B. Grant, A. Stolz, B. Merminod, C.C. Mazur "Contributions to GPS Studies", Unisurv S38, 204 pp., 1990.

PROCEEDINGS

Prices include postage by surface mail

P1. P.V. Angus-Leppan (Editor), "Proceedings of conference on refraction effects in geodesy & electronic distance measurement", 264 pp., 1968. Price: $10.00

P2. R.S. Mather & P.V. Angus-Leppan (Eds), "Australian Academy of Science/International Association of Geodesy Symposium on Earth's Gravitational Field & Secular Variations in Position", 740 pp., 1973. Price $15.00

MONOGRAPHS

Prices include postage by surface mail

M1. R.S. Mather, "The theory and geodetic use of some common projections", (2nd edition), 125 pp., 1978. Price $12.00

M2. R.S. Mather, "The analysis of the earth's gravity field", 172 pp., 1971. Price $8.00

M3. G.G. Bennett, "Tables for prediction of daylight stars", 24 pp.,1974. Price $5.00

M4. G.G. Bennett, J.G. Freislich & M. Maughan, "Star prediction tables for the fixing of position", 200 pp., 1974. Price $8.00

M5. M. Maughan, "Survey computations", 98 pp., 1975. Price $11.00

M7. J.M. Rueger, "Introduction to Electronic Distance Measurement", (2nd Edition), 140 pp., 1988. Price $17.00

M8. A.H.W. Kearsley, "Geodetic Surveying". 77pp., 1988. Price $11.00

M10. W. Faig, "Aerial Triangulation and Digital Mapping", 102. pp., 1986. Price $16.00

M11. W.F. Caspary, "Concepts of Network and Deformation Analysis", 183 pp., 1988. Price $25.00

M12. F.K. Brunner, "Atmospheric Effects on Geodetic Space Measurements", 110 pp., 1988. Price $16.00