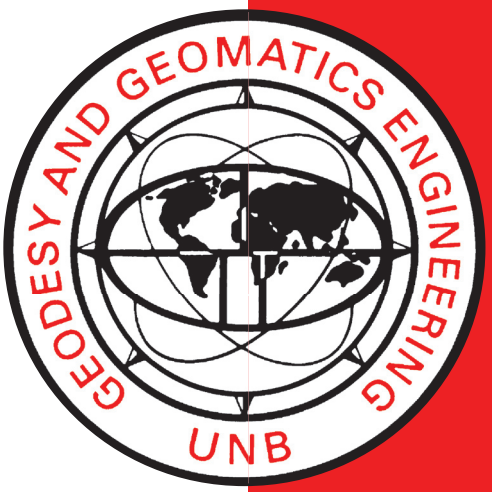
APPLICATION OF NAVSTAR/GPS TO GEODESY IN CANADA: PILOT STUDY

DAVID WELLS
PETR VANICEK
DEMITRIS DELIKARAOGLOU

March 1981



TECHNICAL REPORT
NO. 76

PREFACE

In order to make our extensive series of technical reports more readily available, we have scanned the old master copies and produced electronic versions in Portable Document Format. The quality of the images varies depending on the quality of the originals. The images have not been converted to searchable text.

THE APPLICATION OF NAVSTAR/GPS

TO GEODESY IN CANADA.

PILOT STUDY.

by

David E. Wells
Petr Vaníček
Demitris Delikaraoglou

Department of Surveying Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B., Canada
E3B 5A3

Technical Report No. 76
March 1981
Reprinted April 1985

PREFACE

The analyses contained in this report were performed under a contract "Pilot study of the application of NAVSTAR/GPS to geodesy in Canada" funded by the Geodetic Survey of Canada, awarded 17 February 1981, and delivered March 1981. The Scientific Authority for this contract was David Boal.

Part of the work contained herein was funded by two research grants from the Natural Sciences and Engineering Research Council of Canada. One of these, held by D.E. Wells, is an operating grant entitled "Arctic Marine Navigation Aids". The other, held by P. Vanicek, is a strategic grant entitled "Marine Geodesy".

This technical report is a revised version of the final contract report prepared for this contract.

We wish to thank the Geodetic Survey of Canada and the Natural Sciences and Engineering Research Council of Canada for providing funding for this research. Our understanding of GPS was considerably helped by discussions with John Bossler and Clyde Goad of the U.S. National Geodetic Service; Peter Bender and Douglas Larden of the Joint Institute for Laboratory Astrophysics; Mike Dyment and Patrick Hui of Canadian Marconi Company; Alex Hittel and Gerard Lachapelle of Sheltech Canada; and our colleague at U.N.B., Philip Parker.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. BASIC OPERATING PRINCIPLES OF NAVSTAR/GPS

1.1 Point Positioning and Navigation

The basic concept of point positioning through satellites is very simple. Given $\vec{\rho}$, the satellite position, and observing the range vector $\vec{r}$, the 3D point position $\vec{R}$ can be evaluated from (Figure 1.1):

$$\vec{R} = \vec{\rho} - \vec{r} \quad . \tag{1.1}$$

Clearly, the more accurate are $\vec{\rho}$ and $\vec{r}$, the more accurate will be the derived $\vec{R}$.

Since in geodetic positioning the point position $\vec{R}$ can be ordinarily considered fixed for a long period of time, it is possible to repeat the determination of $\vec{R}$ in time by getting $\vec{r}(t_i)$, i = 1,n. If the satellite positions are known at the same instants of time, we get observation equations for $\vec{R}$:

$$\vec{R} = \vec{\rho}(t_i) - \vec{r}(t_i), \; i=1,n \quad , \tag{1.2}$$

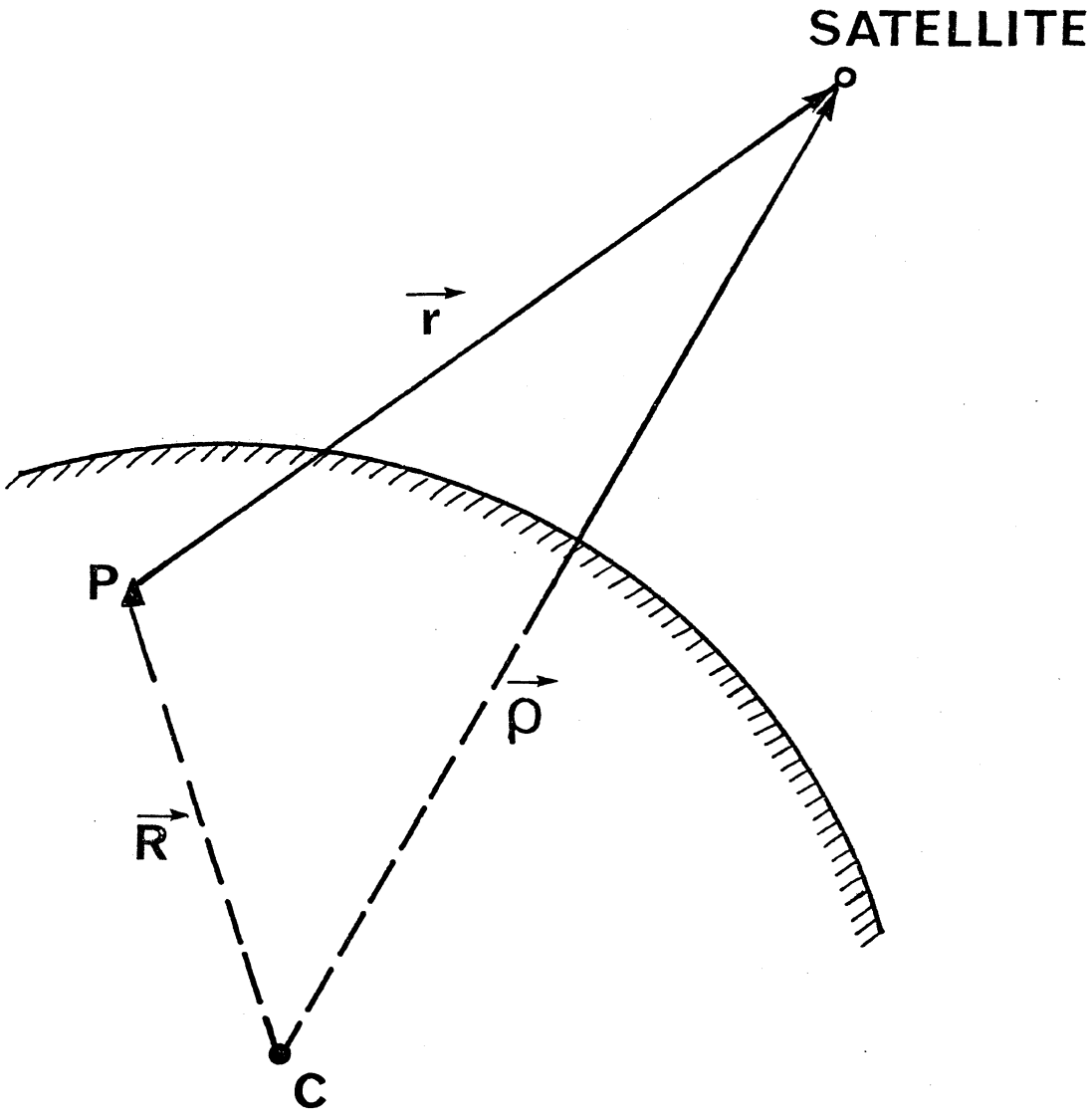
of which n-1 are redundant, thus serving to improve the accuracy of $\vec{R}$. This is clearly one distinct feature of geodetic positioning compared with navigation, since for navigation the position $\vec{R}$ of the vessel has to be considered a function of time as well. There are also other differences between geodetic point positioning and navigation and all these together are responsible for the different algorithms being used in the two tasks. In this report we shall concentrate on geodetic positioning and leave navigation alone.

1.2 GPS Operational Modes

To obtain the range vector $\vec{r}$, both length (range) and direction would have to be observed. Direction observations are generally much affected by refraction which is very difficult to account for. Also, since visible light is usually used for direction observations (for example optical observations), they are weather dependent and the devices needed are bulky. Therefore directions are not very popular observables to use.

Ranges alone can still determine the position of a point on the earth, but at least three satellite positions non-coplanar with the observer have to be used [Blaha, 1971]. However, the observation equation is different:

$$|\vec{\rho}(t_i) - \vec{R}| = |\vec{r}(t_i)| = r(t_i), \; i=1,n, \quad n \geq 3 \quad . \tag{1.3}$$



BASIC CONCEPT OF SATELLITE POINT POSITIONING

FIGURE 1.1

where $r(t_i) = r_i$ are the observed ranges (see Figure 1.2).

With three observed ranges, the sought position lies at the intersection of three spheres centred on satellite positions S_1, S_2, S_3 and having radii r_1, r_2, r_3. Clearly, since a satellite pass (the overhead part of the orbit) is approximately planar, the three positions S_1, S_2, S_3 should not be on one pass whose plane contains the observer. The angles subtended by pairs of satellite positions at the observer should ideally be selected to be close to 90^o. Actually, with four satellites, the best geometry appears to be one satellite at zenith and three near the horizon, separated 120^o in azimuth [Bogon, 1974].

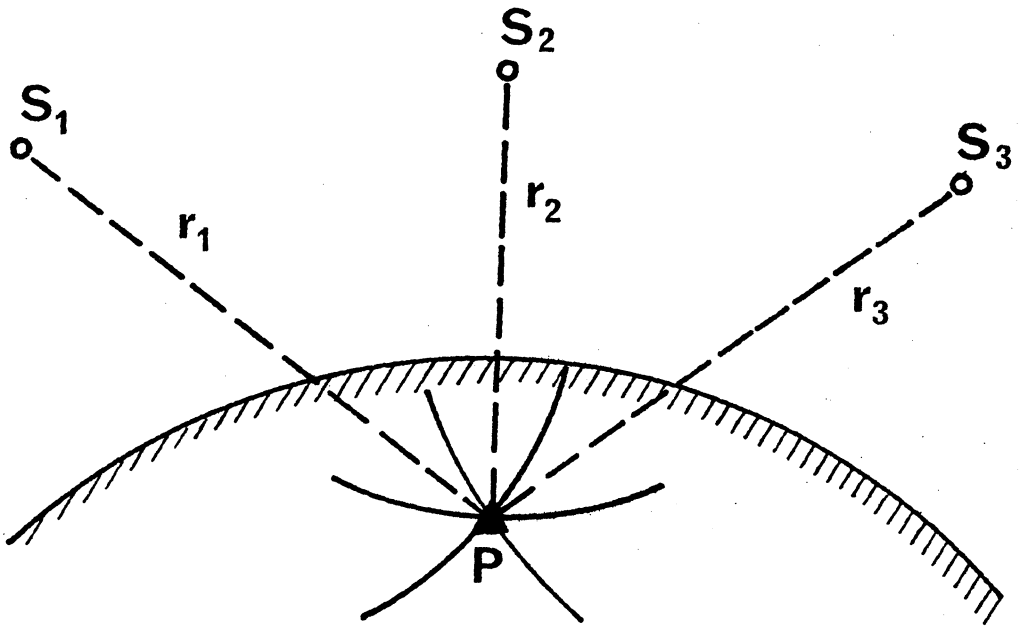
Naturally, with more ranges one gets redundancy and with it the possibility to improve the accuracy. But there is another complication involved here originating in the fact that the satellite and ground time scales are not perfectly synchronized. The synchronization can be achieved by observing redundant ranges, as shall be seen later. The solution to this problem varies with the way satellites are tracked: either the same satellite is tracked during different passes or different satellites are tracked simultaneously.

GPS provides for another possibility: to measure range differences Δr to two satellite positions $S(t_i)$, $S(t_i + \Delta t)$, Δt apart on the orbit [Anderle, 1980]. Each range difference gives one hyperbolic surface (see Figure 1.3) as the locus of the possible observer's positions. Two range differences give the intersecting curve of the two hyperbolic surfaces for the locus and three range differences indicate a point (if not all three range differences are taken along the same pass). The position $\vec{R}$ can then be obtained from the following system of equations:

$$|\vec{\rho}(t_i + \Delta t) - \vec{R}| - |\vec{\rho}(t_i) - \vec{R}| = \Delta r(t_i, \Delta t), \; i=1,2,3 \; . \qquad (1.4)$$

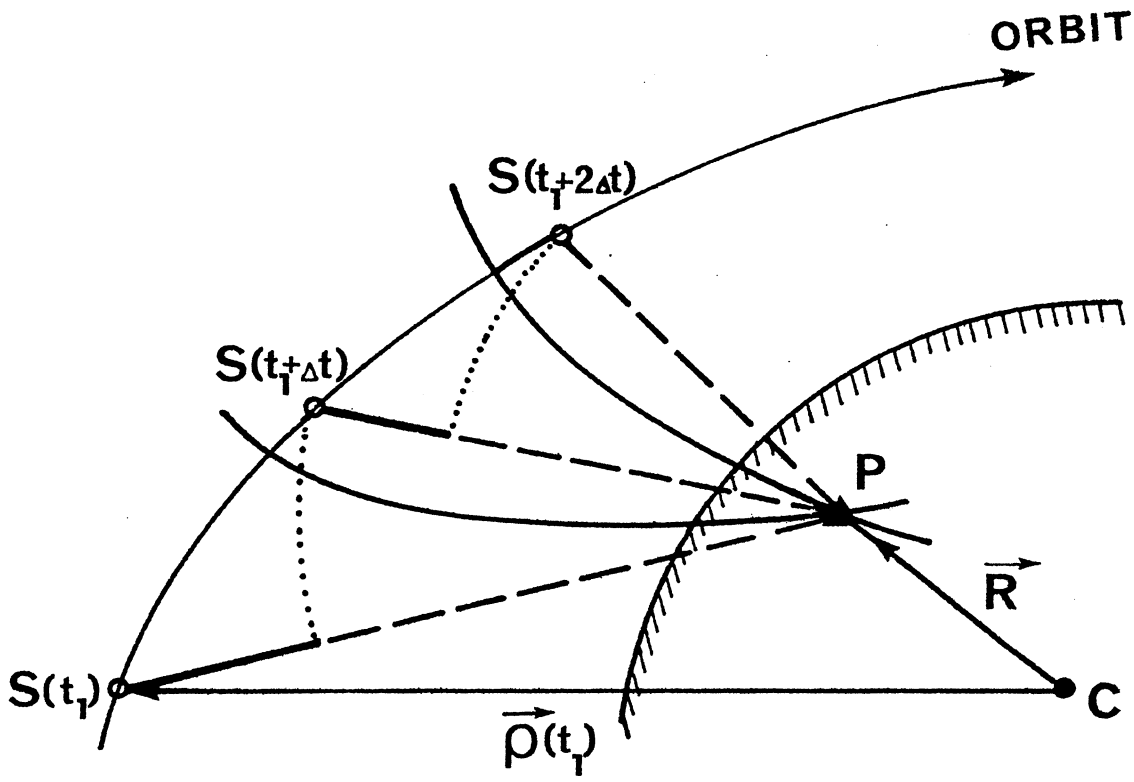
1.3 Sources of Errors

Before we enquire any further into the GPS capabilities, it is essential to understand what the main sources of errors are. The most conspicuous source is obviously the satellite positions involved. Whether the positions are predicted (from the past observed orbital motion) or computed after the mission, there are sizeable biases as well as random errors present [van Dierendonck et al., 1978]. Just what values these may attain in the GPS is not very well known yet.



SATELLITE MULTI-RANGING

FIGURE 1.2



SATELLITE RANGE-DIFFERENCE (HYPERBOLIC) POSITIONING

FIGURE 1.3

The next source of errors is the uncertainty in modelling the electromagnetic wave propagation through the ionosphere and troposphere (ionospheric and tropospheric refraction). Since ionospheric refraction (acceleration) is frequency-dependent, it can be, to a large extent, accounted for from the different behaviour of the two emitted carrier frequencies. On the other hand, the tropospheric retardation has to be modelled from the rather scanty knowledge of the meteorological structure of the troposphere. Relatively good models exist for the "dry" component of tropospheric refraction that use observed values of atmospheric pressure and temperature on the earth's surface [Hopfield, 1980]. The "wet" component is much smaller than the dry. Nevertheless, even the wet component becomes important when the ranging error falls below 10 cm. Yet it cannot be as well modelled as the "dry" component and may have to be measured--see Chapter 10. It is known from experience that, fortunately, those two effects are relatively constant over extended regions (to about 100 km) of the atmosphere [Anderle, 1980].

The geometrical configuration of the satellite positions used can affect the accuracy in an adverse manner. This should be quite obvious from Figures 1.2 and 1.3. This aspect will be extensively dealt with in Chapters 5 and 6.

As will become clear in the following, in either of the two GPS modes (ranging or range differencing), timing is very important. Thus the performance of the onboard satellite oscillator, as well as the ground oscillator, used for the timing, is rather critical. Some applications require long term stability of the oscillators, others need short term stability, as discussed in Chapter 9.

Also, the various available ways of timing the signals have different kinds of errors associated with them. To minimize these errors is one of the major goals in the GPS research going on today. There may also be systematic errors in the point determination from time delays experienced within the satellite receiver. These would be functions of the receiver make.

The specific error model used for the studies in this report is described in Chapter 3.

1.4 Relative Positioning

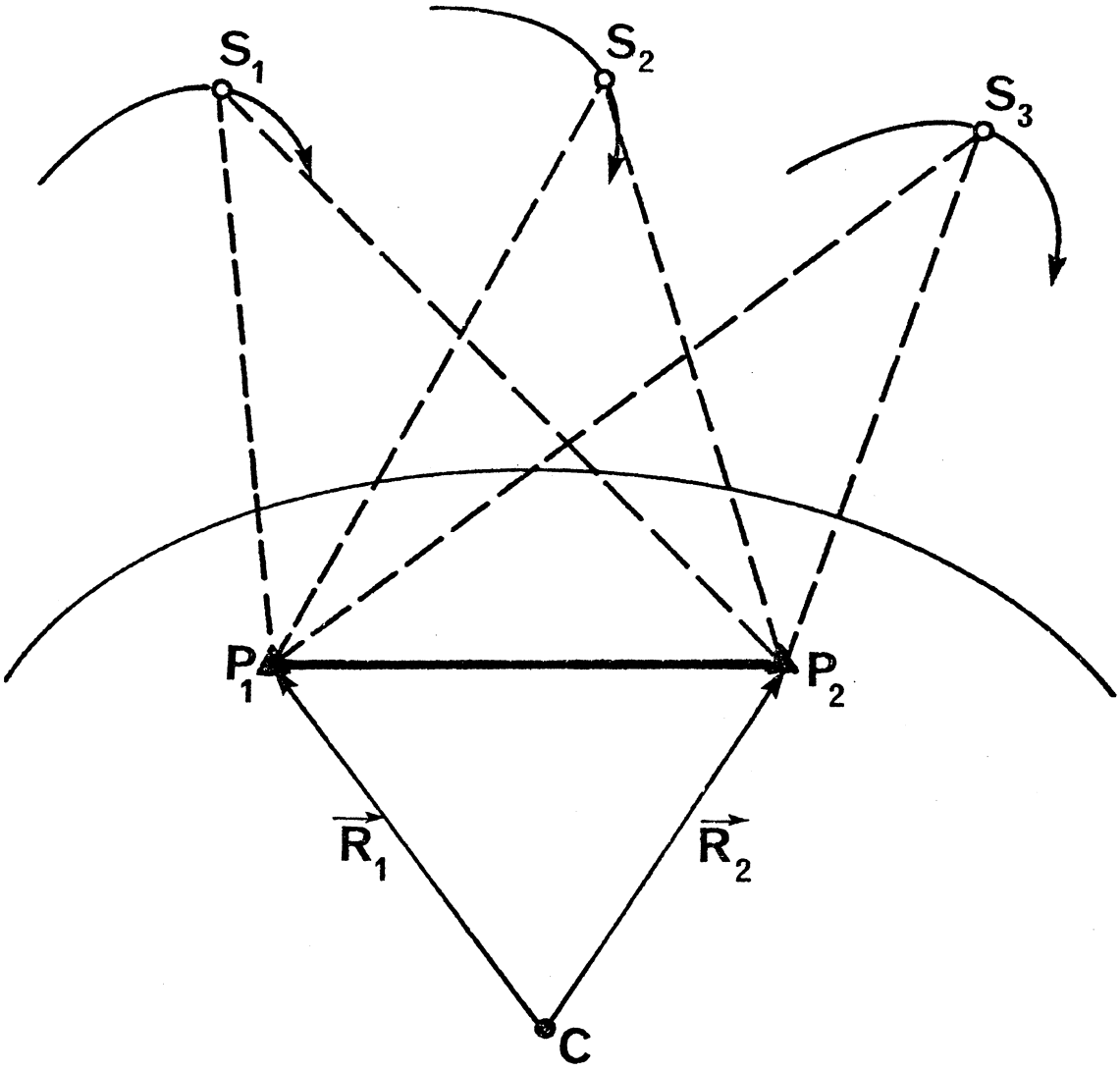
To counter the adverse effect of the biases in the used satellite positions as well as the effects of the incorrectly modelled atmospheric refraction and satellite oscillator performance, another mode of operation is employed--relative positioning. In this mode, two or more ground stations would observe the satellites simultaneously (see Figure 1.4). Either or both ranges and range differences can be used in this mode. In addition, other techniques are being contemplated, as shall be seen in Chapter 7. By using the same satellite positions for both ground stations, the station position differences are much less affected by the satellite position biases (also called orbital biases). Also, if the two ground stations are close enough, the radio signals from the satellite pass through essentially similar regions of the atmosphere and are hence retarded or advanced in essentially the same manner. It is thus foreseen that relative positions of one ground station with respect to the other could be determined to a significantly higher order of accuracy than the point positions [Bossler et al., 1981].

1.5 Satellites

It is envisaged that by the end of 1987 there will be 18 (and possibly even 24) NAVSTAR satellites available for use (Figure 1.5). These will have near circular orbits, inclinations to the equator of 55°, altitudes of 20,000 km, and orbital periods of about 12 hours [Milliken and Zoller, 1978]. With the full configuration, at least 4 satellites will practically always be seen at the same time at any point on Earth. Details of this full configuration, and of the configuration available before 1987, are given in Chapter 4.

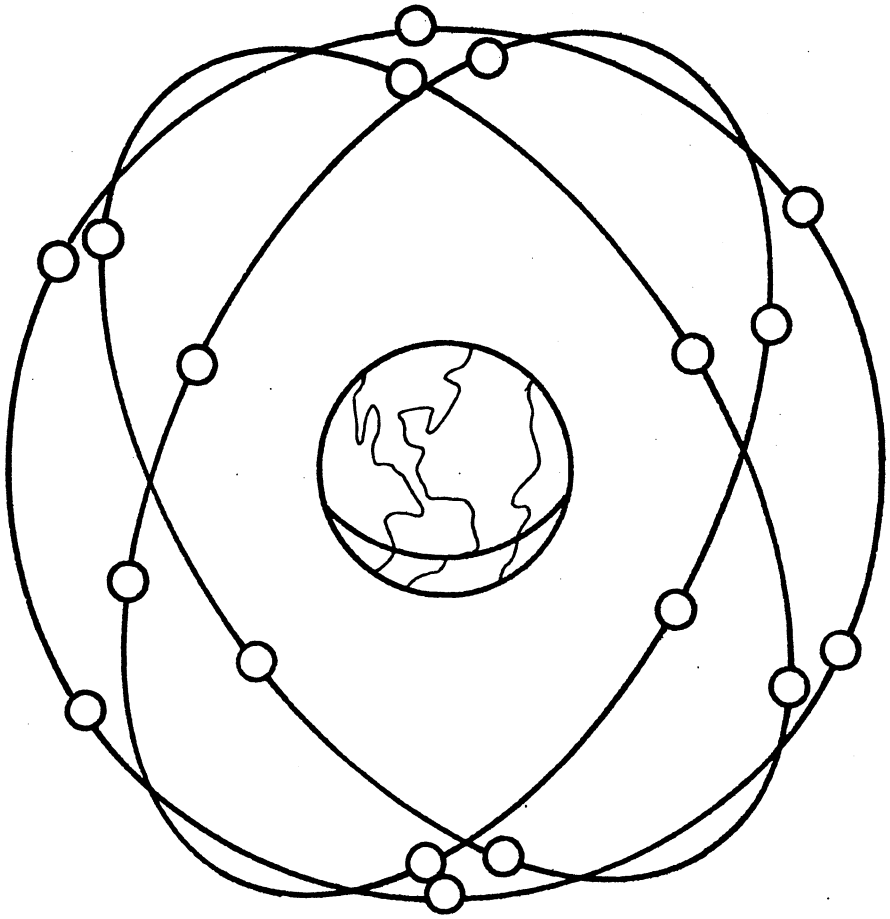
The physical appearance of the satellites, weighing approximately 430 kg, is shown in Figure 1.6 [Bossler et al., 1981]. The main parts of each satellite are the radio transmitter, receiver, antenna, five oscillators (one cesium, three rubidium, and one crystal) [Putkovich, 1980], and microprocessors which control the satellite's functions.

The satellites emit signals on two tightly controlled, coherent (generated by the same oscillator) frequencies. Since the signal structure is quite complicated and yet its understanding vital, we shall devote the whole next section to its description.



SATELLITE RELATIVE POSITIONING

FIGURE 1.4



GPS 18 - SATELLITE CONSTELLATION

FIGURE 1.5

A NAVSTAR/GPS SATELLITE

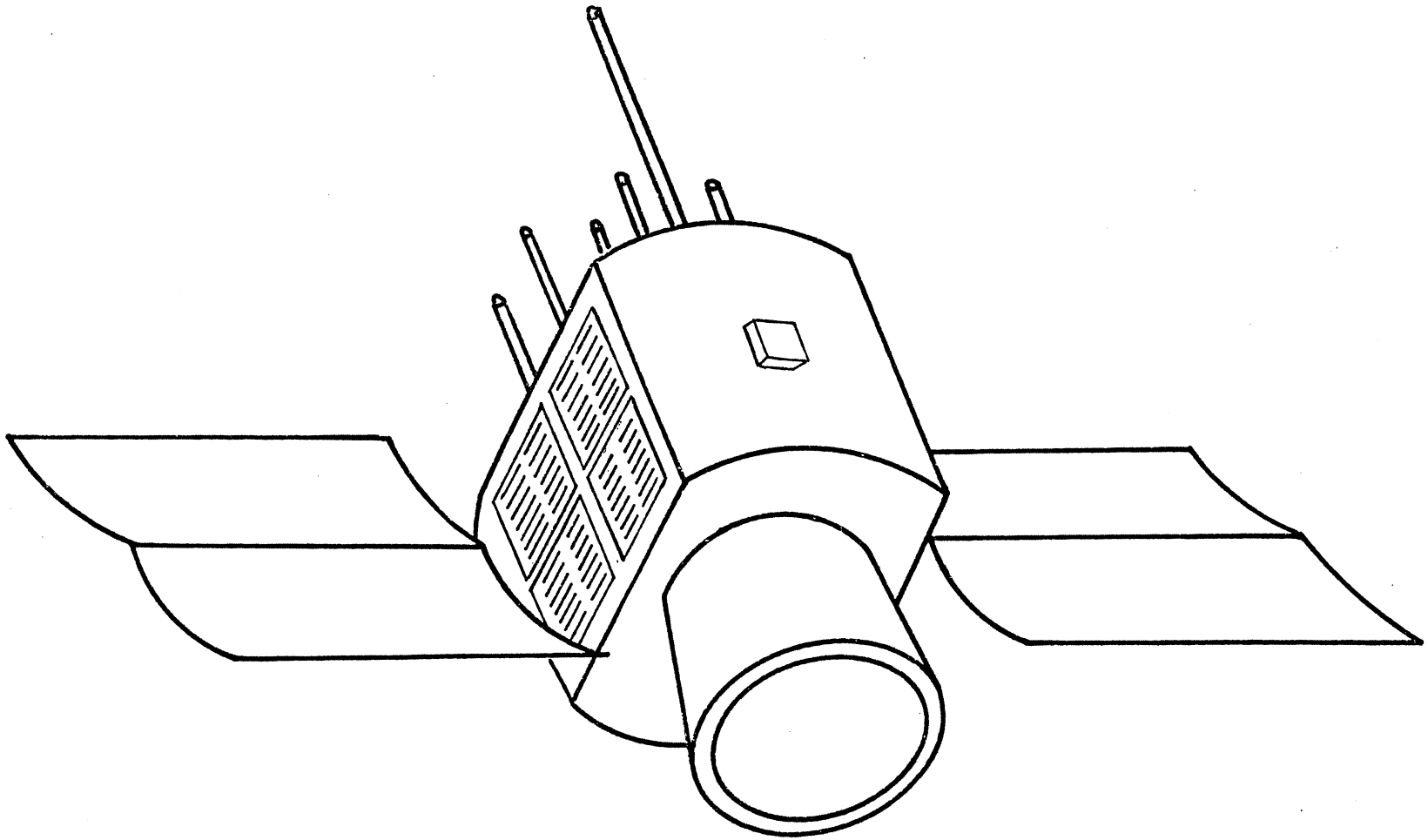


FIGURE 1.6

1.6 Signal Structure

GPS is a passive system in the sense that all the information the user needs to compute his position is automatically and continuously sent to him by the satellite-emitted radio signals. This information includes time marks embedded in the signals, the corresponding times of transmission from the satellite, and the position of the satellite at the time of transmission.

These satellite signals are controlled by 10.23 MHz satellite atomic clocks. The stability of these oscillators is better than a few parts in 10^{-13} during one satellite pass (several hours), which is good enough that we must worry about relativistic effects. If we take an oscillator whose frequency is exactly 10.23 MHz when at rest on the surface of the earth, and we launch it into a GPS orbit, then the time dilation property of special relativity (moving clocks run slower than stationary clocks) will cause this oscillator frequency to decrease by a factor of about 80×10^{-12}. At the same time, the gravitational blue shift property of general relativity (clocks in a weak gravity field run faster than clocks in a strong gravity field) will cause the oscillator frequency to increase by a factor of about 530×10^{-12}. Smaller relativistic frequency increases of a few parts in 10^{12} are caused by the earth's rotation and flattening [Krause, 1963]. The net effect is that the oscillator frequency increases by a factor of about 450×10^{-12}, so that the satellite clocks run fast by about 40 microseconds per day. To offset this, the oscillators are actually set to a rest frequency (before launch) which is a factor of 445×10^{-12} below the nominal value of 10.23 MHz (that is to 10.229 999 995 45 MHz), so that after launch the relativistically-shifted frequency will be very close to the nominal value. Hereafter, for simplicity, we will regard the frequency to be the nominal value of 10.23 MHz.

To permit measurement of the ionospheric refraction effect on the signals (which is frequency dependent), the satellites emit signals with two carrier frequencies:

L_1 at $\omega_1 = 154 \times 10.23$ MHz = 1575.42 MHz

L_2 at $\omega_2 = 120 \times 10.23$ MHz = 1227.60 MHz.

These carriers have wavelengths of 19.05 and 24.45 cm respectively.

Each of these two carriers has several kinds of modulation superimposed upon it.

The cosine wave of the carrier is modulated by a pseudo-random sequence $P_i(t)$ of step functions known as the P-code (P for "precise" or "protected"). It is a sequence generated by an algorithm which repeats itself every 267 days [Milliken and Zoller, 1978]. Each satellite is assigned a different 7-day segment of this 267-day period (see Table 4.1), and generates the P-code sequence associated with its assigned segment. At each Saturday/Sunday midnight of Universal Time, these sequences are reset to the start of the assigned 7-day segments. The frequency of the P-code is 10.23 MHz, that is each 154^{th} carrier wave of ω_1 sees a new modulating pulse. The wavelength of the P-code is 29.31 m.

The sine wave of the carrier is modulated by a unique sequence $G_i(t)$ of step functions called the C/A-code (C/A for "coarse" or "clear access"). $G_i(t)$ is generated in each satellite by Gold's algorithm [Gold, 1967] that repeats itself every millisecond. During 1 ms a string of 1023 step functions is generated at a frequency of 1.023 MHz. Thus every 1540^{th} carrier wave of ω_1 sees a new modulating pulse giving the C/A-code a wavelength of 293.1 m. Clearly a failure to identify the proper millisecond of the carrier arrival creates an error in the range determination of $k \times 300$ km, where k is an integer.

The reason for modulating the carrier wave with either P- or C/A-codes is timing. If the P- or C/A-sequences issued by the satellite and generated by the receiver are synchronous, the time shift between these two can be determined and this shift is the time taken by the signal to cover the distance between the satellite and receiver. Both codes consist of either +1 or −1 step functions: +1 does not change the carrier wave at all, −1 changes the phase by 180^{o} (see Figure 1.7).

The P-code, because of its higher frequency, gives a higher accuracy in timing and thus a higher accuracy in the range determination. On the other hand, because of its very long period, it is practically impossible for a receiver to locate the proper place in the 267 day long sequence without

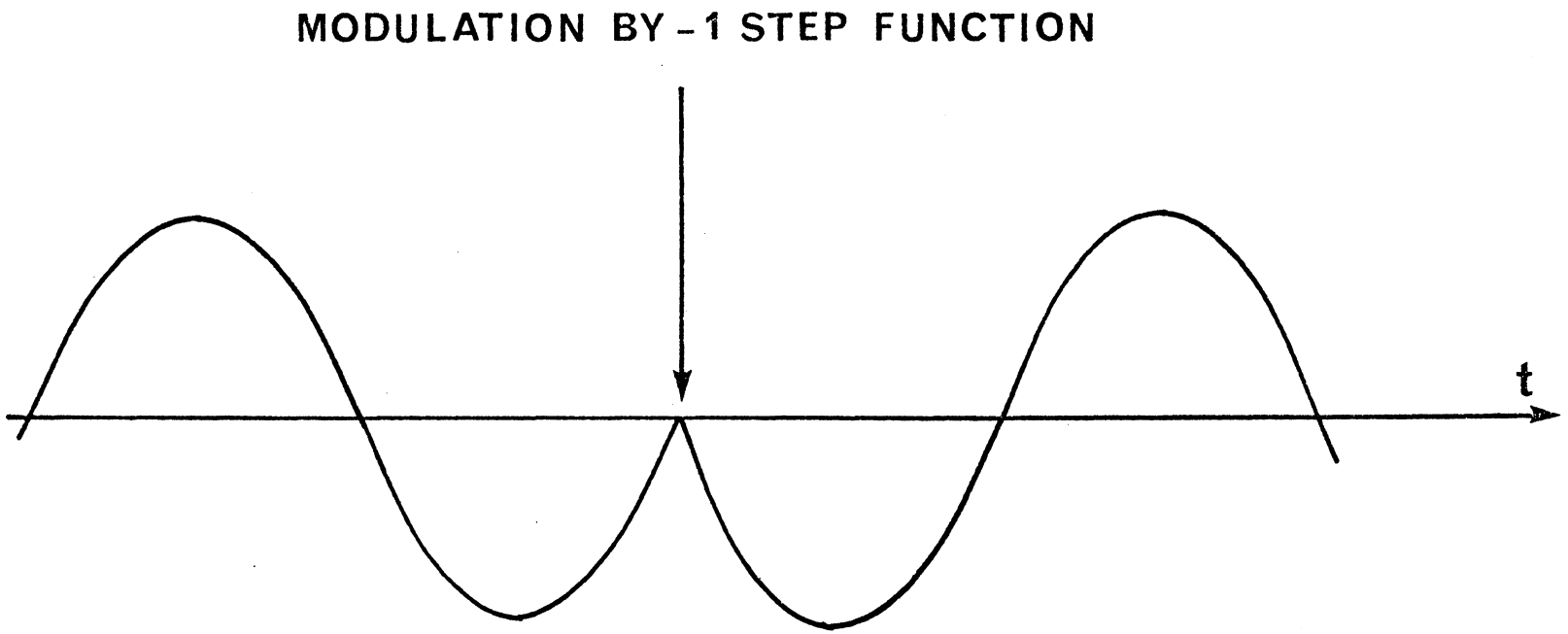


FIGURE 1.7

some assistance. This assistance may be, and probably will be, denied to users not connected in some capacity with the U.S. military, the proprietors of the system. We thus feel that it would be imprudent to rely on the future availability of the P-code for Canadian geodetic operations.

Both the cosine and sine waves are, in addition to the two codes, modulated by a low frequency stream of data $D_i(t)$ containing information on the satellite position (ephemeris), the satellite clock correction, its state of health, information about the other satellites of the system, and the "hand-over word" (HOW) designed to assist the initiated user to use the P-code. Details on how all this information is coded can be found in van Dierendonck et al. [1978].

$D_i(t)$ is a sequence of +1 and -1 step functions with a frequency of 50 bits per second. The duration of the sequence is 30 s divided into five subframes 6 s long. There are 300 bits of data in each subframe. The first word of each subframe multiplied by 4 gives the Universal Time (UT) of the beginning of the next frame expressed in terms of the so-called "Z-count", which is the integral number of 1.5 seconds of time elapsed since the beginning of the week (Saturday/Sunday midnight). The Z-count runs from 0 to 403 199.

The sine and cosine waves of the carrier have different amplitudes A_G, A_P. The sine wave, carrying the C/A-code, is 3 to 6 dB stronger than the cosine wave. Thus the C/A-code should be received with a significantly greater signal-to-noise ratio.

The complete structure of the first signal L_1 is [Spilker, 1978]:

$$L_1(t) = A_P\, P_i(t)\, D_i(t) \cos(\omega_1 t + \phi(t)) + A_G\, G_i(t)\, D_i(t) \sin(\omega_1 t + \phi(t)). \qquad (1.5)$$

Here $\phi(t)$ is noise, modelled as a sum of phase noise $\delta\phi(t)$ and oscillator drift $\delta\omega_1(t) \times t$.

The second signal L_2 so far consists of only the cosine (P-code) or the sine (C/A-code) part. It normally carries only the P-code [Spilker, 1978]. Eventually L_2 may carry both codes.

1.7 Satellite Ephemeris

Satellite ephemerides are determined by the system ground control using observations to satellites from a network of tracking stations of known terrestrial positions. The observed orbits are extrapolated into the

future, coded and uploaded into the satellite memories every 26 hours [Russell and Schaibly, 1978]. Each satellite then broadcasts its own ephemeris message every 30 s (see Section 1.6). This message can be decoded after reception and converted into a unique description of the satellite position for any desired instant of time.

The average accuracy of this broadcast ephemeris (combined with the satellite clock offset) is expected to amount to about 1.5 m in the range direction [Milliken and Zoller, 1978]. More precise ephemerides determined directly from tracking rather than by extrapolation may be available for post-mission use if the pattern of the TRANSIT operation is followed. In any case, the uncertainty in satellite ephemeris is one of the largest source of uncertainty in the determined positions.

1.8 GPS Receivers

The satellite signals are received on an omnidirectional antenna connected to a satellite receiver. The receiver is composed of the following parts: receiver proper (with amplifier), oscillator, digital processor that controls the functions of the receiver, and a memory. The function of the receiver is to decompose the received L_1, L_2 signals into their constituents; the P- and C/A-codes; the carriers; and the 50 Hz data stream. To do this circuits called tracking loops are used. Different types of loops are often used for the various constituents.

To track the P- or C/A-codes, the receiver must be able to generate its own "copy" of the code. The code is tracked by comparing the received code (generated by the satellite clock) with the internal code (generated by the receiver clock), and shifting the internal code in time until the two codes match (more precisely, until the correlation between the two codes is a maximum). This kind of tracking loop is called a delay lock loop. This time shift required to match the codes, when converted to an equivalent distance from the satellite to the receiver antenna by multiplying by the speed of light, is the measured pseudo range.

It is necessary to realize that matching the receiver generated C/A-code with the received C/A-code is relatively simple, since the C/A sequence repeats itself every millisecond. The situation with the P-code is much more involved. One obviously cannot wait for 267 days and even a substantially increased speed of generation of the P-code sequence within

the receiver would not bring down sufficiently the time needed for the search. It is thus necessary to have a key (time within the sequence coded in the HOW) to the P-code sequence to be able to use it at all.

To track the L_1 or L_2 carrier, a loop containing a Voltage Controlled Oscillator (VCO) is used. The carrier is tracked in frequency and phase by comparing the received carrier with the VCO output, and adjusting the VCO to keep the two matched. The VCO output is called the reconstructed carrier. This kind of tracking loop is called a phase lock loop (PLL). Note that the received carrier is usually first down converted to a lower, more convenient frequency before phase locking. However the phase information is kept intact.

To track the 50 Hz data stream, the carrier tracking PLL is of a special design called a carrier synchronizer, which additionally extracts the data bit stream [Gardner, 1979].

Although we have considered them separately here for simplicity, all of these tracking processes (code, carrrier, and data) are intimately interrelated, and for the receiver to operate properly, all the tracking loops must remain simultaneously locked. To first get to this state of simultaneous lock, the receiver must go through a complicated sequence called acquisition during which one loop after the other is brought closer to its locked state.

Three kinds of observations can be derived from the reconstructed carrier: the PLL output frequency will contain a Doppler shifted frequency component (as well as a constant centre frequency) which is a measure of the range rate to the satellite; this Doppler frequency can be integrated over some period, to provide a measure of the range difference to the satellite over that period; and the phase of the reconstructed carrier can be recorded at intervals determined by the receiver clock to provide information about the range to the satellite.

The integrated Doppler count is given by [Anderle, 1980]

$$N_{1,2} = \frac{c}{f_G} \int_{t_1}^{t_2} (f_G - f_R)\, dt \quad , \tag{1.6}$$

where c is the velocity of light in a vacuum, f_G is a reference frequency in the receiver, f_R is the received carrier frequency, and t_1, t_2 are selected time instants.

The reconstructed carrier phase measurement contains potentially the

most precise information about the satellite ranges one can obtain. The problem with utilizing this potential is, however, one of ambiguity: it is very difficult to locate accurately the cycle of the carrier whose phase is being measured (see more in Chapter 7).

From range measurements r_1 and r_2 made at the frequencies ω_1, ω_2 of the two signals L_1, L_2, we can evaluate the frequency dependent ionospheric refraction and correct the pseudo ranges for it. We obtain the range r using the following equation [Martin, 1978]:

$$r = \frac{r_1 - r_2(\omega_2/\omega_1)^2}{1 - (\omega_2/\omega_1)^2} \quad . \tag{1.7}$$

The accuracy of the correction $r - r_1$ (or $r - r_2$) determined this way is better than 1%.

As has already been mentioned, with any of the observing modes it is necessary to track several satellites from each point whose position one wants to determine. This can be done either simultaneously, when a (multichannel) receiver tracks several satellites on several channels operating independently, or sequentially when a (single channel) receiver switches from one "visible" satellite to another at a prescribed rate. The two important characteristics of a sequential receiver are the switching rate and the dwell, that is, the duration of the actual uninterrupted tracking of one satellite.

The main difference between geodetic and navigation receivers lies in their operation mode: navigation receivers must be either multichannel or have a fast switching rate and short dwell. Geodetic receivers on the other hand may be allowed to dwell on one satellite for hours without harming its deployment (although that may not be the best observing strategy).

Various receiver prototypes are now being tested with commercial production estimated to begin next year.

2. APPLICATIONS OF GPS TO GEODESY

2.1 Kinds of Applications

Where can GPS find its uses in geodesy? The answer to this question depends very much on the accuracy the GPS positions are going to have. If the accuracy is in the one metre range then we might expect the applications to be very different from those when the accuracy is in the centimetre range.

If the accuracy is of the order of a metre or submetre, then the GPS uses will be much the same as the uses of TRANSIT are today. These are: mapping control, ice tracking, staking out exploration claims, and geoidal control. GPS, even if it does not achieve its highest potential accuracy, will almost certainly replace the TRANSIT system in the 1990's, if the latter is then phased out as planned [Stansell, 1978].

If the accuracy of GPS shows to be in the 10 cm range, which most researchers believe is a conservative forecast, at least for the relative positioning, then additional vistas will open for its applications. These will certainly include boundary and lay-out surveys, detection of faster tectonic movements, and GPS may possibly be considered to replace levelling in hinterland areas.

In case the accuracy reaches the 1 cm level, GPS could then be used for practically all geodetic tasks. In addition to the above, GPS would be able to provide the framework for both horizontal as well as levelling networks. Also, remaining geodynamical tasks dealing with slower movements could be performed with the system.

Under these circumstances, even positioning for most engineering projects could be carried out with GPS. With these local uses, however, there will always be the decision involved as to where to stop using GPS and start using the standard geodetic techniques. This decision will largely depend on how much time and expense would be needed to locate the point by means of GPS.

2.2 Absolute Versus Relative Positioning

Absolute positions determined by GPS are in the "GPS coordinate system". This coordinate system is realized through the ephemerides of the operating

satellites, where it is assumed that the orbits are essentially planar with the centre of mass of the earth being in that plane. This is not quite true: the actual orbits deviate from being planar--the lower the satellites, the more significant this deviation.

For the GPS satellites, whose altitude is relatively high, the main deviation is due to the ellipticity of the earth's gravity field. The equipotential surfaces at the GPS satellite altitude display maximum departure from sphericity of about 1 km. Maximum departures due to the third order field amount to about 1.5 m and fourth order field to some 20 cm. The attenuation of higher orders at these altitudes is quite rapid [Vaníček and Krakiwsky, 1982]: eighth order does not contribute more than about 1 mm. In spite of this favourable behaviour, the lack of rigorous enforcement of geocentricity and of proper orientation of the GPS may result in the coordinate system having an offset from the geocentre as well as a misalignment vis-à-vis the natural geocentric coordinate system. A thorough testing of location and orientation of the GPS coordinate system with respect to other used systems will be necessary. This should be done by collocating some GPS determined points with points whose coordinates are known in other system(s). From this comparison, the necessary three translations and three rotations can be evaluated [Vaníček and Krakiwsky, 1982].

GPS will not be fully operational in time to affect the redefinition of the North American horizontal geodetic networks. These networks will thus be kept in place through the "supercontrol points" determined from the TRANSIT satellite system. It will be the TRANSIT's coordinate system that will ensure the geocentricity of our future geodetic coordinates. It will be therefore imperative to determine primarily the relative position of the GPS's coordinate system vis-à-vis the TRANSIT's.

It appears certain that relative positioning will yield significantly higher accuracy than point positioning. Hence, in geodesy, we would envisage the relative positioning mode to be used almost exclusively. This would imply that two approaches to positioning in particular will become applicable: "network-like" approach, of the kind used by the Geodetic Survey of Canada (GSC) for establishing the Canadian "Doppler" network [Kouba, 1978], and the "roving station" approach. In the latter approach, one station (preferably one belonging to the geodetic network) would be kept

fixed and permanently manned with a receiver while another receiver would be roving around the points to be determined.

2.3 Vertical Control

For GPS-determined positions to be of any use in vertical control, the geoid has to be known to an adequate degree of accuracy, that is, to about 10 cm. This is because the orthometric height is given as a difference between the geodetic height (above the reference ellipsoid) obtainable from a 3D-position and the geoidal height:

$$H = h - N \quad . \tag{2.1}$$

It seems clear that there is no hope of having the geoid known to this accuracy before the redefinition of the levelling network takes place in the mid-eighties, thus we do not see any possibility of using GPS to improve the vertical network within the context of redefinition. For the same reasons, we see little hope for GPS to be of help in learning more about sea surface topography. Other approaches will have to be used for this purpose prior to the redefinition.

After the heights have been redefined, the GPS determined positions can be used to significantly strengthen the solutions for the geoid. Conversely, towards the end of this century, the satellite dynamics and gravity data should be good and dense enough to give the geoid to a 10 cm accuracy even in the Canadian hinterland. Then, GPS can be used for regional height control in lieu of levelling. We do not foresee, however, that GPS will ever replace standard levelling for local height control because of economical reasons.

Last but not least, somewhere along the way, the GPS and the improved geoid should allow us to settle the argument about the appropriateness of different orthometric height systems [Vaníček, 1980]. Clearly, the whole problem of vertical control is an iterative one with all three components (see Equation (2.1)) needing an improvement.

2.4 Geodynamics

It is our conviction that the main applications of geodesy will be connected with solving geodynamical tasks. Repeated positions with subdecimetre accuracy are of interest to the geophysical/geodynamical community. In these new applications, however, there are both theoretical

and practical problems to be solved.

The outstanding theoretical problem is the treatment of repeated 3D positions in regional studies. The 3D approach has been advocated by Nyland [1977], Reilly [1980], and the U.S. National Committee on Geodesy [1980], but never really used in practice. The determination of optimal discretization in both space and time will have to be worked out for all the 3D, the horizontal as well as the vertical approaches.

The locations of regions of interest in Canada are: West Coast with the active tectonic plate margin phenomena, Prairies subsidence (of unexplained origin) if it exists, Hudson Bay postglacial uplift including the Great Lakes and the Maritime Provinces, St. Lawrence river valley dynamics including Anticosti Island, relative motion between Canada and Greenland as well as between the island of Newfoundland and the mainland. Generally, movements of the order of up to a few centimetres per year can be expected. Thus 10 years worth of repeated GPS position determinations (with subdecimetre accuracy) should suffice to give us the needed quantitive information about the movements.

Finally, a few points within the existing horizontal and levelling networks should be proclaimed basic points. The movements of these points in a geocentric coordinate system should be frequently monitored to keep track of the network position with respect to the earth. A scheme of this kind will be necessary to ensure the continuing usefulness of terrestrial geodetic networks for geodynamics.

2.5 Combinations of GPS with other Positioning Techniques

The main deployment of GPS in geodynamics will be in the regional investigations. Global needs will probably better be secured by Very Long Baseline Interferometry (VLBI) (for example, Ryan et al. [1978]), and local needs by terrestrial techniques, including the inertial positioning. The best ways of combining these techniques not only for geodynamics but also for position control will have to be worked out in detail.

Starting with the global aspect, the VLBI is inherently more sturdy and thus repeatability of results over a period of decades easier to ensure. On the other hand, GPS positions will be cheaper and quicker to obtain. The spacing between radio telescopes needed for ratio interferometry may be many thousands of kilometres, while the length of GPS baselines (in differential

mode) should probably be kept within a few hundred kilometres. All this appears to make GPS an ideal densification tool for VLBI.

Once the GPS points are established, they can be further densified using the inertial positioning system and the standard terrestrial techniques. It looks as if the inertial technique would be preferable when time is important, while the standard techniques can deliver positions of a higher accuracy. A combination of both these techniques could also be considered.

3. GPS ERROR MODELS

The accuracy of positions obtained from GPS is dependent on two general influences. In this chapter we consider error sources affecting the measurement of pseudo ranges and integrated Doppler counts. In the next three chapters we turn to the geometrical strength of the various measured ranges which contribute to the position computation.

3.1 Pseudo Ranging Error Models

Table 3.1 summarizes the effects of error sources which influence the range measurements, and is taken from Martin [1978] and Gilbert et al. [1979]. These error sources fall into three categories--those associated with the satellite, with signal propagation, and with the receiver. It is likely that errors from each of these sources will have complicated spectral properties, and that there will be correlations between some of these errors. However, at this early stage in GPS development, our error models are limited to the simple approach of predicting typical standard deviations of uncorrelated equivalent range errors from each error source.

Satellite errors consist of errors in the ephemeris (the satellite is not where the GPS data message tells us it is), and errors in the clock (the satellite clock is not perfectly synchronized to "GPS system time"). These satellite errors are uncorrelated between satellites, affect P-code and C/A-code navigation equally, and depend on the number and location of the tracking stations providing data for orbit determination, the orbit force model algorithm used, and the satellite constellation geometry [Fell, 1980a]. Detailed descriptions of GPS satellite errors can be found in Schaibly [1976], and Schaibly and Harkins [1979].

Propagation errors consist of ionospheric refraction errors, tropospheric refraction errors, and multipath errors. At GPS frequencies the full ionospheric range error may be more than 50 m (sunspot maximum, mid-day, satellite near horizon) or less than a metre (sunspot minimum, night, satellite at zenith). Since ionospheric refraction is frequency dependent, we can compare L_1 and L_2 pseudo-range measurements to estimate the effect. The standard deviation of this dual-frequency measurement is three times the standard deviation of the pseudo-range measurement noise

TABLE 3.1

GPS Pseudo-Range Error Budget* (m)

	P-CODE	C/A-CODE
Satellite		
ephemeris	1.5 (3.6) m	1.5 (3.6) m
clock	0.9 (2.7)	0.9 (2.7)
Propagation		
ionosphere-dual frequency	3.0	
ionosphere-models		0.5 - 15.0
troposphere	1.0	1.0
multipath	1.0	5.0
Receiver		
measurement noise	1.0	10.0
measurement truncation	0.3	3.0
computation	1.0	1.0
Combined effect (root sum squared)	4.0 (5.8)	12 - 20

*Standard deviations of uncorrelated equivalent range errors. Values in brackets refer to the prototype satellites.

[Martin, 1978], which from Table 3.1 is 1 m and 10 m for P-code and C/A-code respectively. According to this model, therefore, C/A pseudo-range measurements are not precise enough to determine ionospheric refraction range errors using the dual frequency technique. Also, at present only the P-code is transmitted on the L_2 frequency, so that dual frequency C/A measurements are not possible. Predictions of ionospheric refraction range errors from present day ionospheric models are unlikely to be more than 75% effective, so that ionospheric effects of 0.5 m and 50 m would be estimated leaving residual errors of up to 15 m.

Tropospheric refraction range errors vary from about 2 m with the satellite at zenith, to about 25 m with the satellite at 5^{o} elevation. This error must be modelled from surface weather measurements and vertical profiles of refractivity must be assumed to be known. A scale bias error in this tropospheric refraction model, which might have a typical standard deviation of 4% of the total tropospheric range error, results in a contribution to the GPS range error with 1 m standard deviation.

The P-code and C/A-code modulations on the GPS signal serve to reject multipath signals whose propagation path is more than one code wavelength longer than the direct signal. Within one code wavelength the effect of multipath will depend on the geometry of satellite, antenna and potentially reflecting surfaces, and will likely be a very irregular function of time. All that can be done at this stage is to indicate some typical rms values. Due to the wavelength dependence, the C/A-code value is larger than the P-code value.

The receiver errors consist of pseudo-range measurement noise, measurement truncation, and computation errors. The measurement noise standard deviation is directly proportional to the code wavelength. The measurement truncation values shown in Table 3.1 result from assuming that the receiver code can be matched to the satellite code to no better than $1/64^{th}$ of the code wavelength, and assuming that the resulting truncation error has a uniform distribution. The computation error is due to finite computer bit resolution, mathematical approximations, algorithm uncertainties, and execution and timing delays in the computations [Martin, 1978].

The combined effect of these error sources is range measurement error standard deviations of 4 m for P-code ranging, and of 12 m to 20 m for

C/A-code ranging, depending on the state of the ionosphere.

3.2 Doppler Range Difference Error Models

As with pseudo-range measurements, Doppler range difference or delta-range measurements to GPS satellites are affected by errors associated with the satellite, the signal propagation, and with the receiver.

Since Doppler measurements are fundamentally measurements of the change in the pseudo range toward a particular satellite over a finite time interval, errors essentially depend on the integration interval being employed. Table 3.2, taken from Canadian Marconi [1981], gives a summary of these errors for a typical single channel GPS receiver and an assumed integration interval of 30 s. The combined effect of these error sources is range difference measurement error standard deviations of 8 cm for P-code operations and about 19 cm for C/A-code operations.

TABLE 3.2

GPS Doppler Range Difference Error Budget* (cm)

	P-CODE	C/A-CODE
Satellite		
ephemeris	3.6	3.6
Propagation		
ionosphere-dual frequency	3.5	
ionosphere-models		5.2
troposphere	3.1	3.1
multipath	4.9	17.2
Receiver		
measurement noise	1.6	1.6
clock error	1.0	1.0
Combined effect (root sum squared)	8.0	18.7

*Standard deviations of uncorrelated equivalent range difference errors, for a single receiver, and assumed 30 s integrated interval.

4. GPS SATELLITE CONSTELLATIONS

4.1 Operational Satellite Constellation

For a multiple satellite navigation system like GPS, the geometric relationships between all the satellites in the system must ensure that continuous coverage from several satellites is equally available at every point on Earth. The higher the satellites, the fewer of them are needed, since each will have a wider coverage area. A simple solution, if it were possible, would be to have a network of geostationary satellites. Unfortunately, geostationary satellites must lie in the equatorial plane. One way to provide equal global coverage would be to add two polar orbit planes to such an equatorial orbit plane, to form a set of three mutually orthogonal orbit planes. Such a system would lack symmetry in having some satellites in equatorial orbit and some in polar orbit. However, we can establish symmetry simply by rotating this system of three orbit planes until each plane intersects the equator with the same inclination angle (which is 55°) but is separated from the other planes in right ascension by 120°. This symmetry is maintained whether the satellites have geosynchronous periods (24 hours) or not. If the satellites in each plane are equally separated around the orbit, then this is referred to as a three-plane uniform constellation.

A three parameter categorization of such uniform satellite constellations has been developed by Walker [1977]. The Walker constellation index has the form T/P/F, where T is the total number of satellites in all planes, P is the number of orbit planes, and F describes the relative phase between satellites in adjacent planes. This phase is expressed in terms of "pattern units", where one pattern unit is $360^{\circ}/T$, so that F can take on values between zero and P-1. For example, the Walker 24/3/2 constellation contains 24 satellites, divided equally among three planes, with relative phasing of two pattern units. Thus there are eight satellites uniformly spaced in each plane, separated along the orbit by 45°. Since in this case one pattern unit is $360^{\circ}/24 = 15^{\circ}$, this means that as a satellite in one plane crosses the equator from south to north, another satellite in the next plane to the east will be 30° along track above the equator. Note that the orbit plane inclination angle and orbit period are not specified by the Walker index, so that to each Walker index there

corresponds a set of constellations differing in period and/or inclination.

The satellite constellation originally planned for GPS was the Walker 24/3/2 constellation described above, with an inclination of 55° and an orbit period of 12 hours, corresponding to an altitude of 20 000 km. At this altitude, satellites are visible from the earth at spherical distances of up to 76°. Hence this constellation ensures that at least six satellites are always visible everywhere on earth. Because only four satellites need be continuously visible, and to reduce the cost of GPS, it is now likely that the fully implemented GPS system will use only 18 satellites (Figure 1.5). Several 18-satellite constellations have been proposed. One leading contender is a non-uniform constellation obtained by merely removing two adjacent satellites from each plane in the 24/3/2 constellation [Book et al., 1980]. Others are the uniform Walker 18/3/0 and 18/6/2 constellations.

The operational GPS satellites are scheduled to be placed in orbit by the Space Shuttle between 1985 and 1987. Twelve satellites should be in operation by the third quarter of 1986, providing global two-dimensional GPS fixing. The full set of 18 satellites are planned to be in orbit by the fourth quarter of 1987 [USDOD/DOT, 1980].

4.2 GPS Prototype Satellite Constellation

Until the operational GPS constellation is established, a constellation of prototype GPS satellites will be maintained. Originally, this was to have consisted of six satellites [Brady and Jorgensen, 1981]. To reduce costs this is now more likely to be the minimum number required to support the development of GPS (for which four satellites are necessary), and to support testing during the U.S. Navy Trident missile accuracy improvement program (for which five satellites are required). Thus there may be four, five, or six GPS satellites available at various times over the next four years.

Table 4.1 shows the current status of these prototype satellites. Six satellites have been launched. Four are at present fully operational, one is operating with a degraded clock, and one has failed. At least one replacement is scheduled for launch later this year.

This prototype six-satellite constellation is derived from a Walker 24/3/0 constellation, where satellite positions 1, 2, and 3 are spaced roughly 45° apart in one orbit plane, and positions 4, 5, and 6 are roughly

LAUNCH SEQUENCE NUMBER	ORBITAL POSITION NUMBER	ASSIGNED VEHICLE PRN CODE	INTERNATIONAL DESIGNATION	NASA CATALOGUE NUMBER	LAUNCH DATE (YYMMDD)	OPERATIONAL STATUS
1	2	4	1978-020A	10684	780222	crystal clock
2	4	7	1978-047A	10893	780513	not operating
3	6	6	1978-093A	11054	781007	operating
4	3	8	1978-112A	11141	781211	operating
5	1	5	1980-011A	11690	800209	operating
6	5	9	1980-032A	11783	800426	operating
7					8112	failed on launch
8					8211	to be launched
9					8309	to be launched
10					8403	to be launched
11					8503	to be launched

TABLE 4.1 GPS prototype satellites. A variety of numbers are used to identify each satellite. The orbital position number is used in this report. The assigned vehicle pseudo-random noise (PRN) code is transmitted as part of the GPS message, and is a number from 1 to 32 that indicates which of the 38 possible weeks within the P-code period has been assigned to this satellite. All three rubidium atomic clocks in each of the first two satellites which were launched have failed. Crystal clocks on these satellites still work, well enough on the first satellite to keep it in operation, but not so on the second satellite. The next four satellites are operating on either rubidium or cesium atomic clocks, and each has one or more as yet untested backup atomic clocks. Replenishment satellites are scheduled for future launch [Tennant, 1980].

45° apart on a second orbit plane, which is 120° in right ascension to the west of the first plane. The satellite positions cross the equator from south to north within five hours of each other. Position 6 crosses first, followed by 5 (90 minutes later), 3 (100 minutes later), 4 (180 minutes later), 2 (188 minutes later), and finally position 1 (270 minutes later than position 6). As seen from Table 4.1, positions 2 and 4 are not presently occupied by fully operational satellites.

These prototype orbit planes are inclined to the equator by 63° rather than the 55° planned for the operational constellation. The reason for this difference is that while the operational satellites will be launched from the Space Shuttle, the prototype satellites have been launched from the ground, which places restrictions on the inclination angles that can be used [Brady and Jorgensen, 1981].

The 12-hour orbit period produces ground tracks for the satellites which are reproducible from day to day. The orbit periods of all satellites are kept within a few seconds of each other, but are actually 122 seconds less than 12 hours (that is, 12 hours of solar time--the orbit periods are within a few seconds of 12 hours of sidereal time), so that each day the satellites reappear at a fixed site 243 seconds earlier than the day before. The orbit planes are not absolutely stationary in space, precessing westward in right ascension by about 12 degrees per year.

Ground tracks for the six-satellite constellation are shown on a polar stereographic projection in Figure 4.1, and on a Mercator projection in Figure 4.2. The combination of 12-hour period and 63° inclination results in ground tracks which follow meridian lines very closely between latitudes 40°S and 40°N, so that on a Mercator projection the ground tracks resemble square waves. For the operational constellation inclination of 55°, this effect will be much less pronounced.

While the operational constellation (of 18 satellites) is designed to provide at least four visible satellites at all times everywhere on Earth, the coverage available from the prototype constellation (of four, five or six satellites) is understandably less complete, both spatially and temporally. The spatial variation in this coverage in the vicinity of Canada is shown in Figure 4.3, expressed in terms of the number of hours per day for which three or more (of the four, five or six) GPS satellites are visible. The local trend is improved coverage the further north the

GPS PHASE II SIX SATELLITE CONSTELLATION

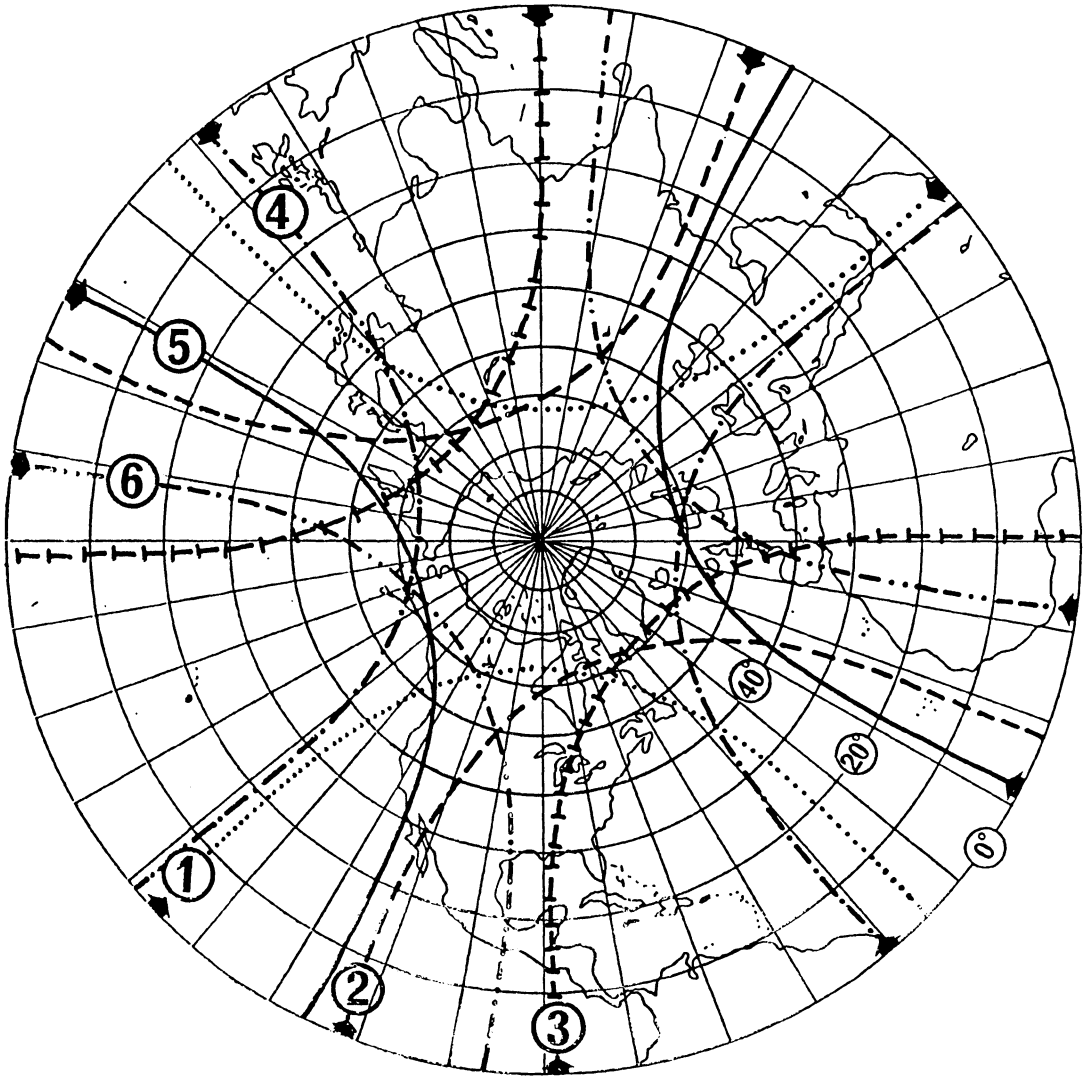


FIGURE 4.1

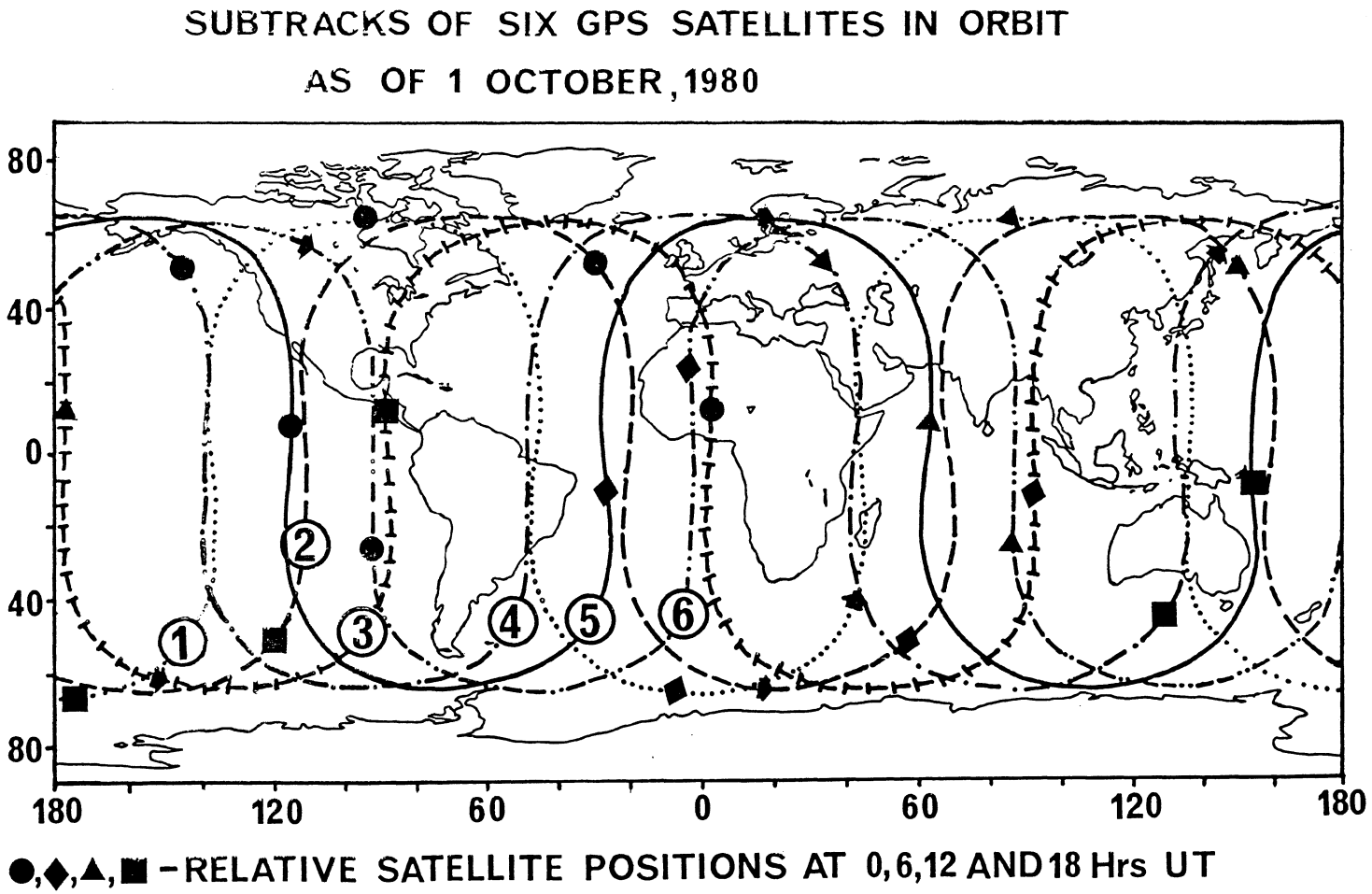

SUBTRACKS OF SIX GPS SATELLITES IN ORBIT
AS OF 1 OCTOBER, 1980
80
40
0
40
80
180
120
60
0
60
120
180
1
2
3
4
5
6
●,◆,▲,■ - RELATIVE SATELLITE POSITIONS AT 0,6,12 AND 18 Hrs UT

FIGURE 4.2

SPATIAL VARIATION OF PROTOTYPE GPS SATELLITE COVERAGE FOR CANADA.

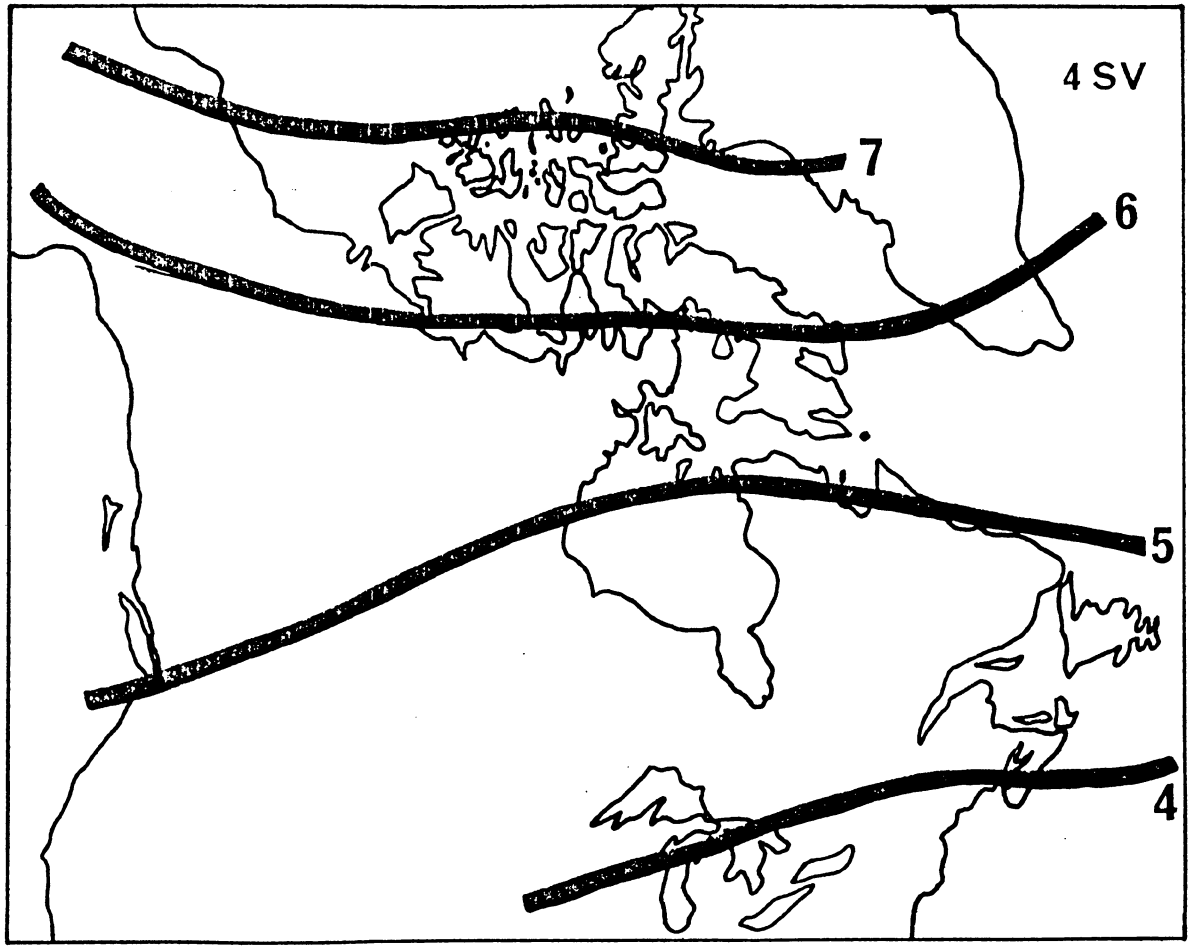


FIGURE 4.3(a)

SPATIAL VARIATION OF PROTOTYPE GPS SATELLITE COVERAGE FOR CANADA.

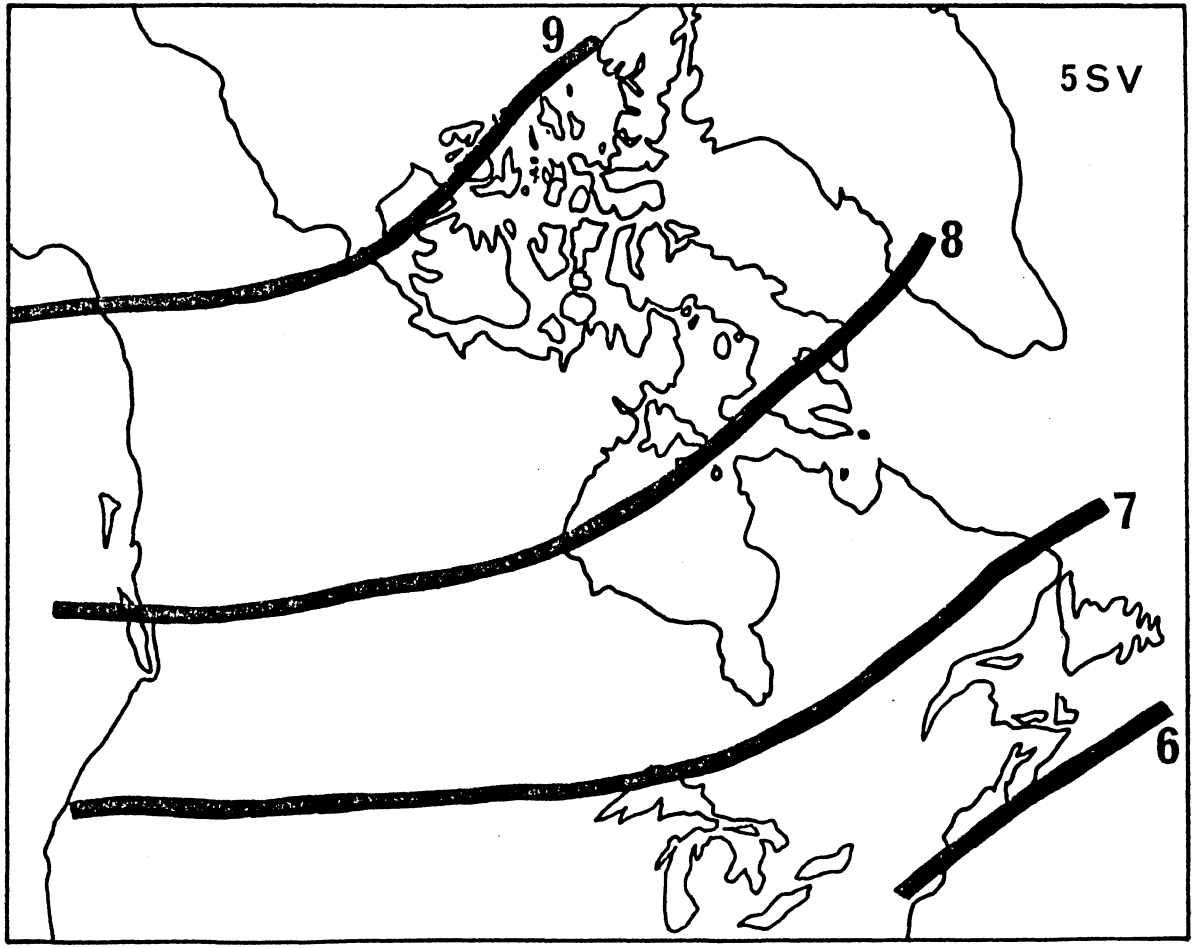


FIGURE 4.3(b)

SPATIAL VARIATION OF PROTOTYPE GPS SATELLITE COVERAGE FOR CANADA.

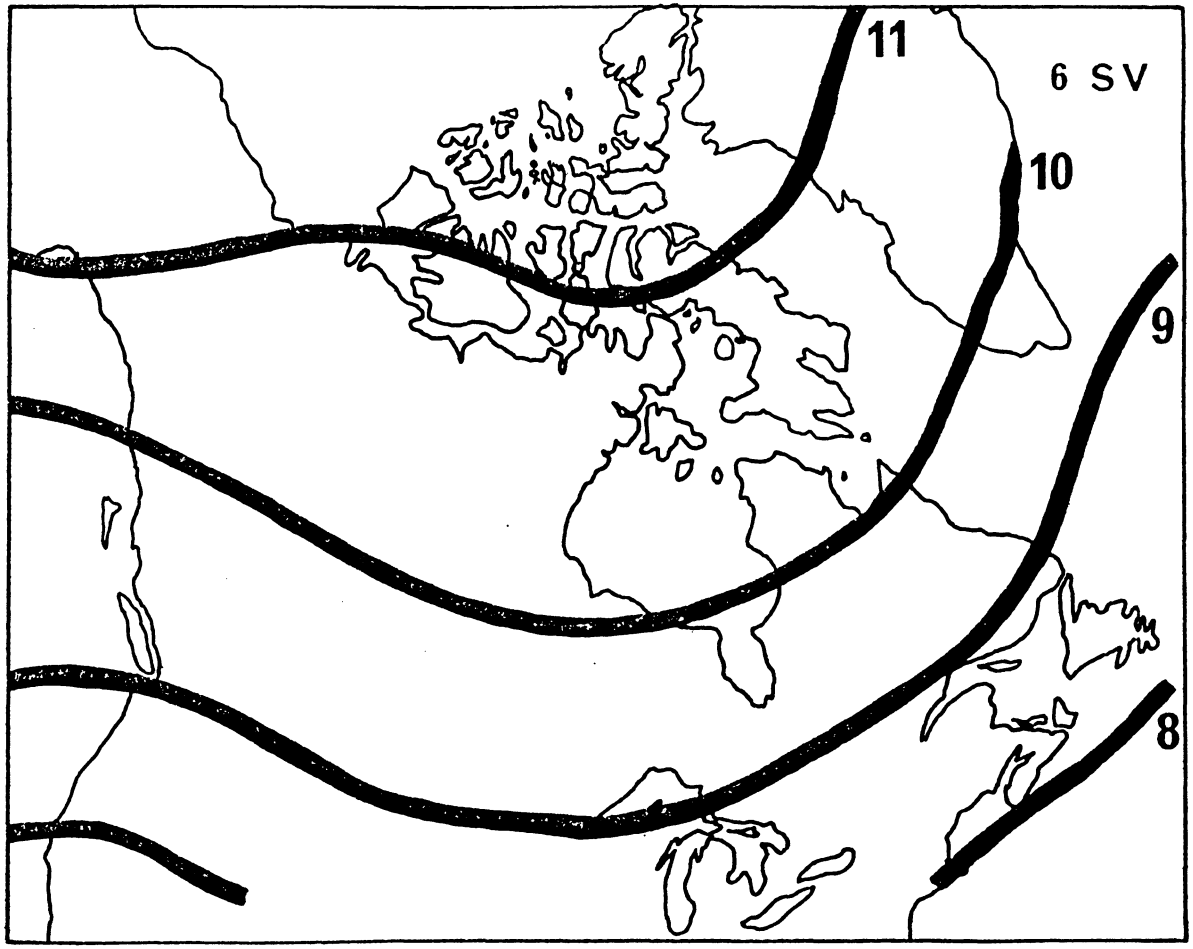


FIGURE 4.3(c)

observer. On a global basis, Brady and Jorgensen [1981] have shown that the present four-satellite constellation provides four-satellite coverage for a few hours per day in four areas of the world; North America, the South Atlantic, continental Asia, and southeast of Australia. On the other hand, these four satellites are never simultaneously visible in Peru, western Africa, and Indonesia. Figure 4.4 depicts the temporal limitations of this coverage, showing the satellite availability in the centre of Baffin Bay (75^{o} N, 68^{o} W) for 1 October 1980, with four, five and six satellite constellations. For the present four satellites, all four are visible for one hour per day, and at least two for 12 hours per day. If six satellites were in orbit, at least four would be visible for 8 hours per day, and at least two for 15 hours per day.

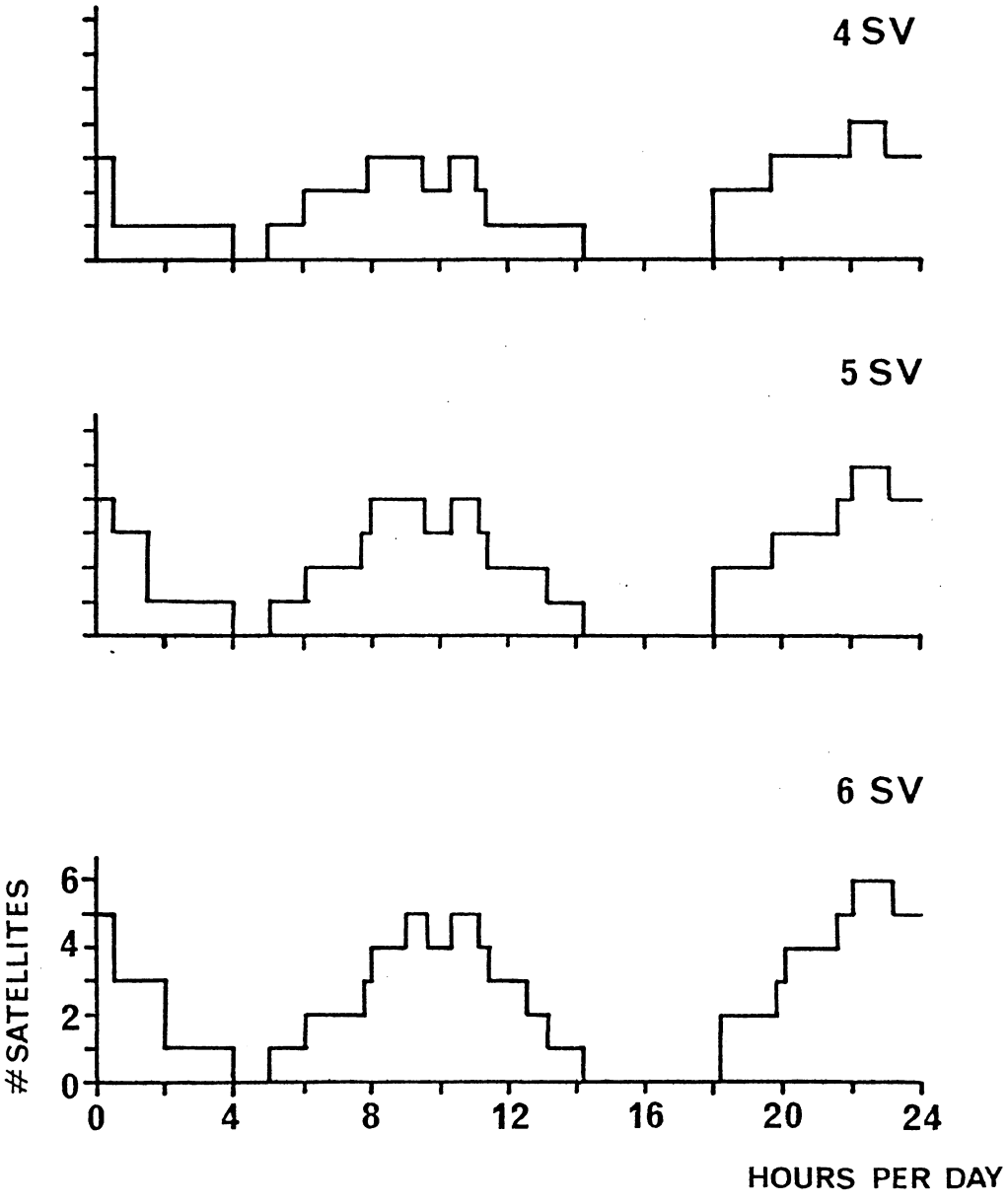


FIGURE 4.4

TEMPORAL VARIATION OF PROTOTYPE GPS
SATELLITE COVERAGE FOR BAFFIN BAY.

5. DESIGN OF THE SIMULATION STUDY

5.1 Pseudo-range Measurements

As already stated in Chapter 1, one of the GPS positioning modes involves passive ranging observations, known as pseudo ranges. These are defined as the time delays between the broadcast and the arrival time of the ranging signal, scaled by the speed of light c. Each pseudo range represents the geometric range between the receiver and the satellite, plus the effect of the synchronization error between the receiver and the satellite clocks. In addition, the observation is subject to other error sources such as satellite ephemeris error, errors due to tropospheric and ionospheric refraction, instability of satellite and receiver clocks, and other receiver timing errors. Ignoring temporarily these error sources, the observation equation for a pseudo range is given by

$$r(t_T) = c(t_R - t_T) = |\vec{\rho}(t_T) - \vec{R}| + c\,\Delta t \qquad (5.1)$$

$$= [(x(t_T) - X)^2 + (y(t_T) - Y)^2 + (z(t_T) - Z)^2]^{1/2} + c\,\Delta t,$$

where t_R and t_T are the GPS signal reception and transmission times respectively; $x(t_T)$, $y(t_T)$, $z(t_T)$ are the satellite coordinates at t_T in an adopted, earth-fixed coordinate system; X, Y, Z represent the receiver coordinates in the same coordinate system; Δt represents the time offset between the receiver and satellite clocks (GPS system time).

5.2 Doppler Range Difference Measurements

Although the pseudo-range measurement is expected to be the primary navigational mode of operation with GPS, Doppler observations will also be available (see section 1.2). For range measurements, the GPS carrier signal is not used and only a pseudorandom noise code sequence (either P or C/A), is exploited. However, the carrier can be recovered after the receiver has locked to the code sequence so that Doppler measurements can be made between the carrier and a receiver generated signal in a manner similar to the Doppler technique conventionally applied to the TRANSIT satellites. That is, a Doppler measurement can be obtained by differencing the reconstructed received carrier frequency f_R with the corresponding frequency f_G generated by the receiver equipment and counting the zero crossings of the resulting signal over some integration time interval. Usually referred to as

integrated Doppler, this measurement can be expressed as the range difference over the integration interval (t_j, t_k) as in equation 1.4:

$$\Delta r(t_j\ t_k) = |\vec{\rho}(t_k) - \vec{R}| - |\vec{\rho}(t_j) - \vec{R}|$$

$$= c/f_G[N_{jk} - (f_G - f_T)(t_k - t_j)] \qquad (5.2)$$

where N_{jk} is the measured integrated Doppler count (see equation 1.6) over the interval $(t_k - t_j)$, and f_s is the satellite reference frequency. As with ranges, this measurement is subject to ephemeris errors, atmospheric refraction, oscillator frequency variations, and other timing errors.

5.3 Methodology

Given a set of observation equations (such as 5.1 or 5.2), the GPS fixed point positioning solution is obtained by using least squares estimation. These observation equations contain three kinds of quantities: observables (to which we assign some standard deviation), "unknown" parameters (equivalent to assigning an infinite a priori standard deviation), and "known" parameters (equivalent to assigning an a priori standard deviation of zero). In practice, "unknown" parameters are usually not totally unknown, and "known" parameters often have some uncertainty in the value assigned them. It is common therefore to assign a finite standard deviation to the a priori values we specify for both "unknown" and "known" parameters, blurring the distinction between them and the observables. Both unknown and known parameters are then referred to as "quasi-observables". Let us denote the vector of observables by $\underline{\ell}$, the vector of "unknown" quasi-observables by $\underline{x}$, the vector of "known" quasi-observables by $\underline{y}$, and their corresponding a priori covariance matrices by $\underline{C}_\ell$, $\underline{C}_x$, and $\underline{C}_y$. Then a set of observation equations can be linearized by taking partial derivatives evaluated at the a priori values with respect to $\underline{\ell}$, $\underline{x}$, and $\underline{y}$ (which we denote by the matrices $\underline{B}$, $\underline{A}$, and $\underline{D}$ respectively) and evaluating the misclosure vector $\underline{w}$ (departure of each of the observation equations from equivalence when the a priori values for $\underline{\ell}$, $\underline{x}$, and $\underline{y}$ are used). We can write the resulting linearized equation as

$$\underline{w} + \underline{A}\,\hat{\underline{\delta}}_x + \underline{D}\,\hat{\underline{\delta}}_y + \underline{B}\,\hat{\underline{v}} = 0 \qquad (5.3)$$

where $\hat{\underline{\delta}}_x$, $\hat{\underline{\delta}}_y$, and $\hat{\underline{v}}$ are the least squares estimates of the corrections to be added to the a priori (or observed) values of $\underline{x}$, $\underline{y}$, and $\underline{\ell}$. Then the least

squares solution is given by

$$\hat{\delta}_x = - \hat{C}_x \, A^T \, (B \, C_\ell \, B^T + D \, C_y \, D^T)^{-1} \, w \tag{5.4}$$

where

$$\hat{C}_x = (C_x^{-1} + A^T \, (B \, C_\ell \, B^T + D \, C_y \, D^T)^{-1} \, A)^{-1} \tag{5.5}$$

is the least squares estimate of the relative covariance matrix of $\hat{\delta}_x$.

Specifically in our case we have $B = - I$; ℓ is a vector either of observed pseudo ranges (equation 5.1) or observed Doppler range differences (equation 5.2); x is a vector containing the coordinates of the ground station or stations (the elements of $\vec{R}$); and y is a vector containing other quantities we may wish to consider imperfectly known (such as orbit biases and residual refraction biases). Two of these biases are often included as elements of x: for pseudo ranges the clock bias Δt; and for Doppler range differences the frequency offset $\Delta f = f_G - f_S$. This is the approach taken here. Other possible biases are not considered here (that is $D = 0$). Then equations 5.4 and 5.5 simplify to

$$\hat{\delta}_x = - \hat{C}_x \, A^T \, C_\ell^{-1} \, w \tag{5.6}$$

$$\hat{C}_x = (C_x^{-1} + A^T \, C_\ell^{-1} \, A)^{-1} \quad . \tag{5.7}$$

The problem addressed here is, given A, C_ℓ, C_x, to compute $\hat{C}_x$ from equation 5.7.

Each observation equation results in one row of the partial derivative matrix A. Denoting a typical row of A by a_i, pseudo-range observations from equation 5.1 will result in

$$a_i = [\frac{\partial r}{\partial X} , \frac{\partial r}{\partial Y} , \frac{\partial r}{\partial Z} , \frac{\partial r}{\partial \Delta t}]$$

$$= \frac{1}{r^o(t_T)} [x(t_T) - X^o, \, y(t_T) - Y^o, \, z(t_T) - Z^o, \, C \, r^o(t_T)] \tag{5.8}$$

and from Doppler range differences from equation 5.2

$$a_i = [\frac{\partial \Delta r}{\partial X} , \frac{\partial \Delta r}{\partial Y} , \frac{\partial \Delta r}{\partial Z} , \frac{\partial \Delta r}{\partial \Delta f}]$$

$$= \frac{f_G}{c} [\frac{x(t_j) - X^o}{r^o(t_j)} - \frac{x(t_k) - X^o}{r^o(t_k)} , \frac{y(t_j) - Y^o}{r^o(t_j)} - \frac{y(t_k) - Y^o}{r^o(t_k)} ,$$

$$\frac{z(t_j) - Z^o}{r^o(t_j)} - \frac{z(t_k) - Z^o}{r^o(t_k)} , t_k - t_j] \quad , \tag{5.9}$$

where the superscript 'o' indicates a priori values.

The covariance matrix of the observations $\underline{C}_{\ell}$ will be diagonal if the observations are statistically independent (uncorrelated). If all observations have the same variance σ_{ℓ}^2, then

$$\underline{C}_{\ell} = \sigma_{\ell}^2 \underline{I} \quad . \tag{5.10}$$

Some sources of error will produce observations from the same set which do not have equal variances. One example is the effect of unmodelled tropospheric refraction, which is much greater when the satellite is near the horizon than when it is at higher elevations.

Some sources of error will produce observations which are serially correlated. One example is the jitter in the timing of Doppler observation integration epochs. If one integration interval ends sooner than the model predicts due to timing jitter, then the next interval will also start earlier. The two intervals are 100% negatively correlated, and

$$\underline{C}_{\ell} = \sigma_{\ell}^2 \begin{bmatrix} 1 & -0.5 & 0 & 0 \\ -0.5 & 1 & -0.5 & \cdot \\ 0 & -0.5 & 1 & \cdot \\ 0 & \cdot & \cdot & 1 \end{bmatrix} \quad . \tag{5.11}$$

In this report, $\underline{C}_{\ell}$ is always assumed as given by equation (5.10).

5.4 Simulation Outline

The simulation consists of changing $\underline{A}$, σ_{ℓ}^2, and $\underline{C}_x$ according to a certain plan and evaluating $\hat{\underline{C}}_x$ from equation 5.7. In standard geodetic terminology, we thus carry out a "preanalysis" based on some assumed geometrical configurations and accuracies. In the literature, this task is sometimes called "formal estimation of accuracy". The effect of user location on the variability of satellite geometry, and hence on GPS performance is evaluated using the set of grid points shown in Figure 5.1.

To summarize, the problem addressed here is to evaluate the scalar function

$$A(\tau;\ X,\ G,\ S,\ C,\ M,\ N,\ D,\ I,\ \ldots) \tag{5.12}$$

where A is the predicted GPS point positioning accuracy achieved after a total observing time τ.

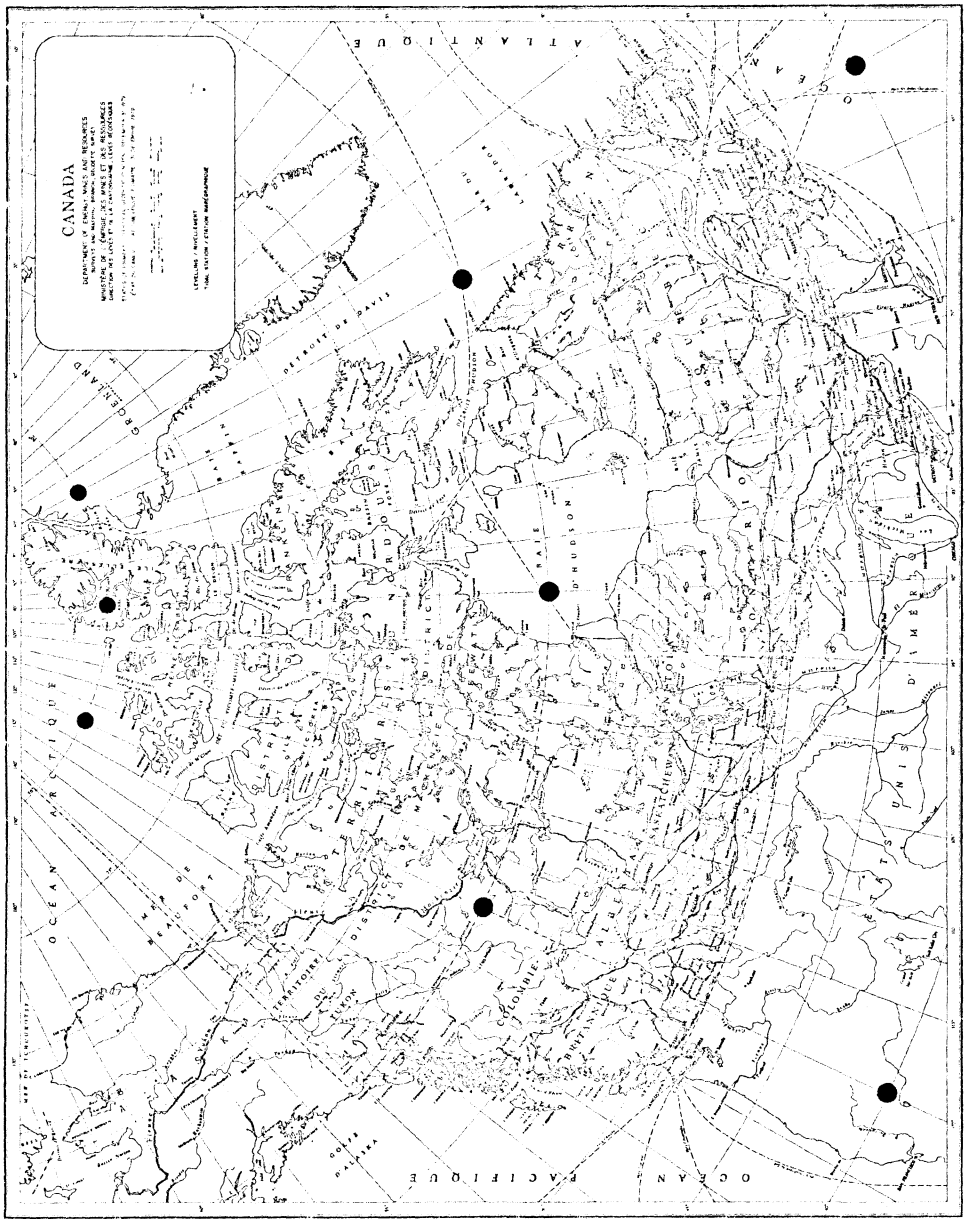


FIGURE 5.1 Simulation Grid Points

$A(\tau)$ depends on the following parameters:

(a) Coordinate-related parameters

X = coordinate selected for study (ϕ, λ, h, etc.),

G = observer's geographical position, to show variations in GPS performance over a set of grid points.

(b) Satellite-related parameters

S = satellite constellation (present 6 satellites, several proposed 18 and 24 satellite constellations),

C = GPS code used (P or C/A).

(c) Receiver-related parameters

M = observing mode (range or range difference),

N = number of receiver channels (1, 2,...,n),

D = satellite dwell time (2.5 ms to 3 hours),

I = switching rate (10 seconds to 30 minutes).

Results are presented in two forms--tabulated final values of $A(\tau)$, and plots of $A(\tau)$ against τ.

6. RESULTS OF SIMULATION STUDY

This chapter presents the results of several simulations carried out to determine the relative geometric strength of point positioning using the range and range-difference modes. For the purpose of this report, only the problem of point positioning has been addressed. The results are based on preanalysis using either only range or only range-difference observations. No solutions based on both observables were considered although the possibility of such a combination exists [Jorgensen, 1980b]. There are three basic variables considered in this analysis: different GPS satellite constellations, receiver design, and the geographic location of the user.

6.1 Pseudo Range Results

In general, the results of this section were computed for cases where observations were taken over five minute intervals. Range measurements were assumed statistically independent with unit variance or standard deviation of one metre. Only observations above 5 degrees of satellite elevation were utilized. A priori knowledge of the station coordinates was assumed to be 1000 m in all three dimensions (X, Y, Z).

Since several options of tracking geometry are possible with multiple satellites in view, a criterion for satellite selection was adopted to ensure that numerous samplings of satellite geometry were utilized. This selection was based on the minimization of the rms position error

$$\sigma_P = \sqrt{(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)} \quad , \tag{6.1}$$

which relates to the Position Dilution of Precision (PDOP) by

$$\text{PDOP} = \frac{\sigma_P}{\sigma_{obs}} \quad , \tag{6.2}$$

where σ_{obs} is the measurement error. PDOP can be easily determined geometrically by relating it to the volume V of a special tetrahedron, as shown in Figure 6.1 [Bogon, 1974; Spilker, 1978]. That is, it can be shown that PDOP ~ 1/V so that the PDOP becomes smaller as the volume V becomes larger.

For 8 hours of total observation time, assuming the use of a single channel receiver switching from one satellite to the next every five minutes, simulation runs for each of the selected grid points were carried

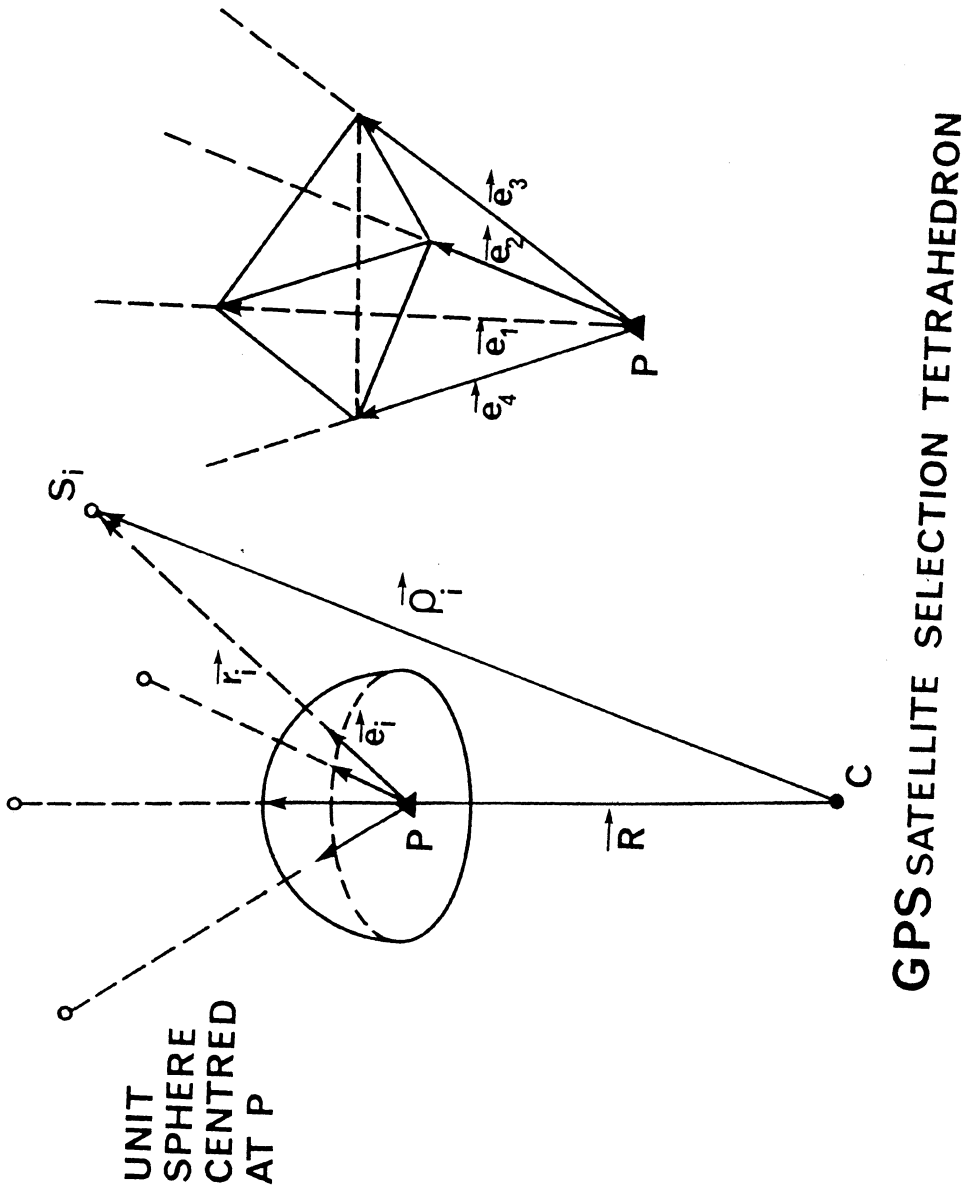

GPS SATELLITE SELECTION TETRAHEDRON
UNIT SPHERE CENTRED AT P
S_i
$\vec{Q}_i$
$\vec{r}_i$
$\vec{e}_i$
P
C
$\vec{R}$
$\vec{e}_1$
$\vec{e}_2$
$\vec{e}_3$
$\vec{e}_4$
P

FIGURE 6.1

out. The results of the performance of the primary contender non-uniform 18/3/2 constellation and the present six satellite constellation are shown in Tables 6.1 and 6.2 respectively.

From the results of Table 6.1, there are some discernible trends due to the location of the user. In general, there is an increase in the position uncertainty as the latitude of the tracking station increases. Clearly stations toward the pole show an acute longitude uncertainty, which in turn forces the overall position uncertainty to remain large. This can be attributed to the orbital characteristics of the GPS satellites. Because of their high altitude (20 000 km), their twelve hour period and the small precession of their orbit planes, GPS ground tracks repeat from day to day nearly exactly displaying a pattern almost like a square wave on a Mercator projection plot. Since the majority of satellite passes are predominantly north to south, the estimated latitude has a smaller standard error than longitude and height. This is also illustrated in Figures 6.2, 6.2(a) to 6.2(c) displaying typical plots of the uncertainty on the position components of a particular station as a function of time.

An examination of the results in Table 6.2 indicates that the present set of six satellites can provide results comparable to the full constellation provided that tracking is confined to periods of multisatellite visibility. From Figures 6.3(a) to 6.3(c) it is noted that the estimated errors in all three components receive immediate improvement when three or more satellites are visible, then the uncertainty decreases in an exponential fashion until about two satellites remain visible and finally it levels off until the last visible satellite sets. The trends in the six satellite constellation performance as a function of the location of the user again appear to be the same.

To investigate the effect of varying lengths of tracking each observed satellite, a number of different tests were carried out for one particular station arbitrarily selected at $\phi = 60^{\circ}$ N, $\lambda = 240^{\circ}$ E. The results of these tests, shown as part of Table 6.3, are based on 8 or 24 hours of continuous observation where the time alloted for tracking each satellite was fixed at one, two, or three hours. For each observing schedule, again, the satellite selected for observation was the one whose observations, when included with prior data, minimized the trace of the parameter covariance matrix.

For the same total observation time of 8 hours, the results of these

TABLE 6.1 Point Positioning Performance of the Non-uniform 18/3/2 Constellation as a Function of Geographic Location. Standard Error of Station Parameters Based on Range Observations (8 hours total tracking).

LATITUDE		40°			60°			80°		
LONGITUDE		240°	270°	300°	240°	270°	300°	240°	270°	300°
POSITION ACCURACY (m)	σ_ϕ	0.19	0.20	0.19	0.17	0.18	0.18	0.17	0.17	0.17
	σ_λ	0.23	0.22	0.24	0.36	0.34	0.36	0.93	1.00	1.00
	σ_h	0.36	0.34	0.35	0.38	0.35	0.34	0.46	0.47	0.45
	$\sqrt{Tr(\Sigma_X)}$	0.46	0.45	0.47	0.55	0.52	0.52	1.05	1.12	1.11

Comments: Results are based on range observations every five minutes with measurement uncertainty of 1 (m).

TABLE 6.2 Point Positioning Performance of the Present Six Satellite Constellation as a Function of Geographic Location. Standard Error of Station Parameters Based on Range Observations (~ 8 hours tracking).

LATITUDE		40°			60°			80°		
LONGITUDE		240°	270°	300°	240°	270°	300°	240°	270°	300°
POSITION ACCURACY (m)	σ_ϕ	0.21	0.18	0.14	0.25	0.22	0.22	0.29	0.27	0.19
	σ_λ	0.29	0.36	0.29	0.41	0.51	0.54	0.84	1.09	1.55
	σ_h	0.43	0.42	0.30	0.41	0.44	0.46	0.50	0.48	0.43
	$\sqrt{Tr\ (\Sigma_x)}$	0.56	0.58	0.44	0.63	0.71	0.74	1.02	1.22	1.62

Comments: Results are based on range observations every five minutes with measurement uncertainty of 1 (m).

EFFECT OF VARIATIONS IN RECEIVER SWITCHING RATE (I) AND TOTAL OBSERVING TIME (τ) ON RANGE ACCURACY

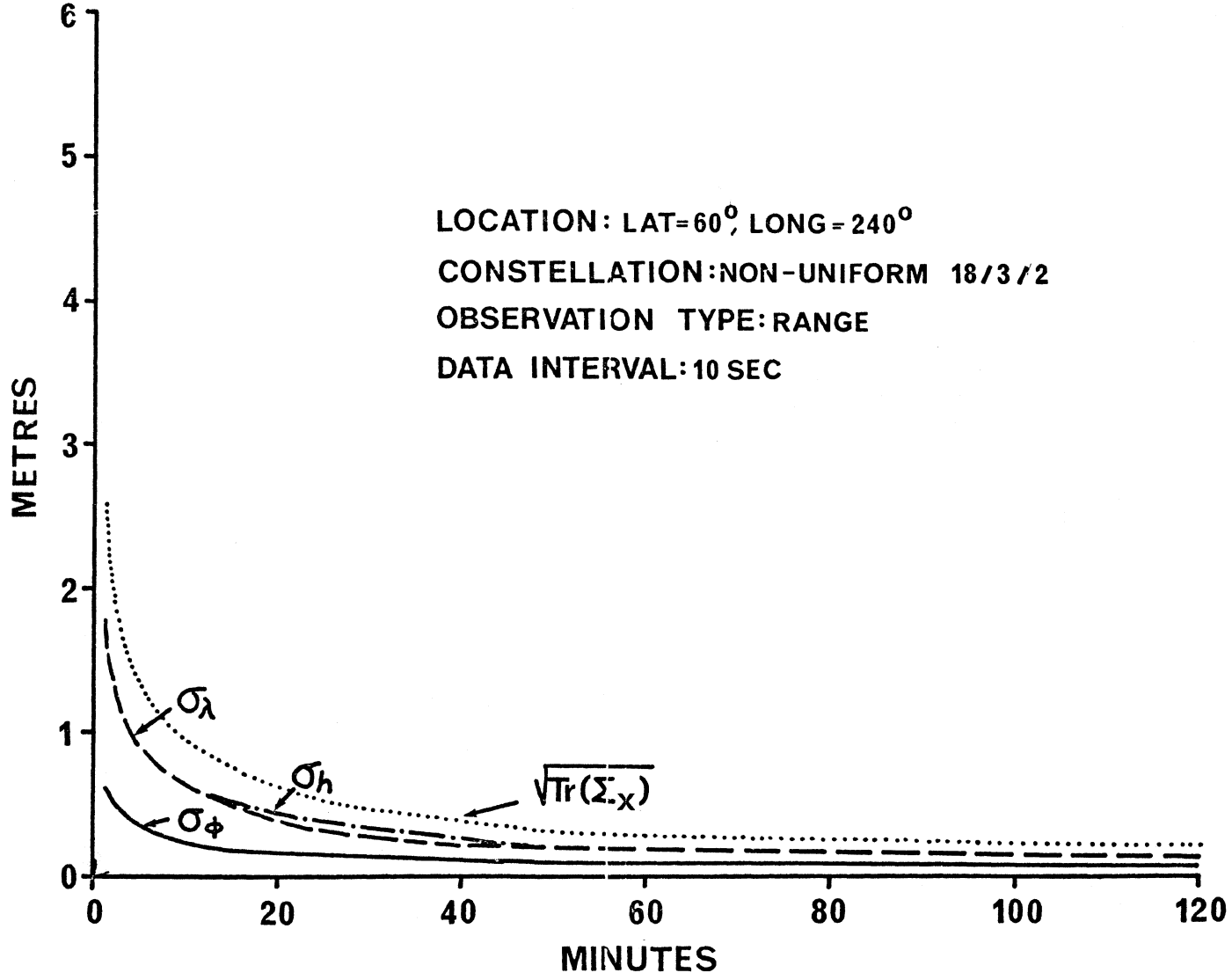


FIGURE 6.2(a)

EFFECT OF VARIATIONS IN RECEIVER SWITCHING RATE (I) AND TOTAL OBSERVING TIME (τ) ON RANGE ACCURACY.

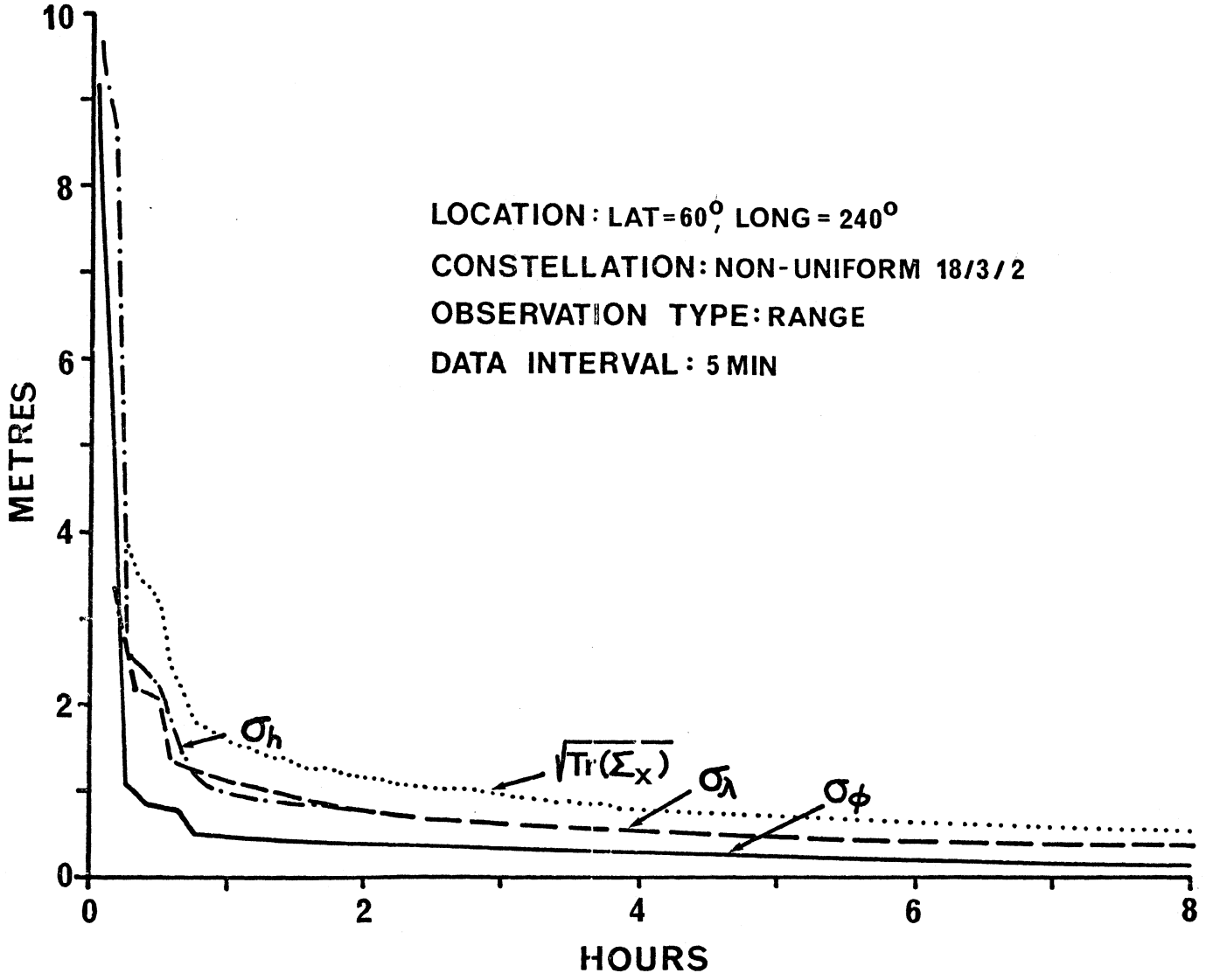


FIGURE 6.2(b)

EFFECT OF VARIATIONS IN RECEIVER SWITCHING RATE (I) AND TOTAL OBSERVING TIME (τ) ON RANGE ACCURACY.

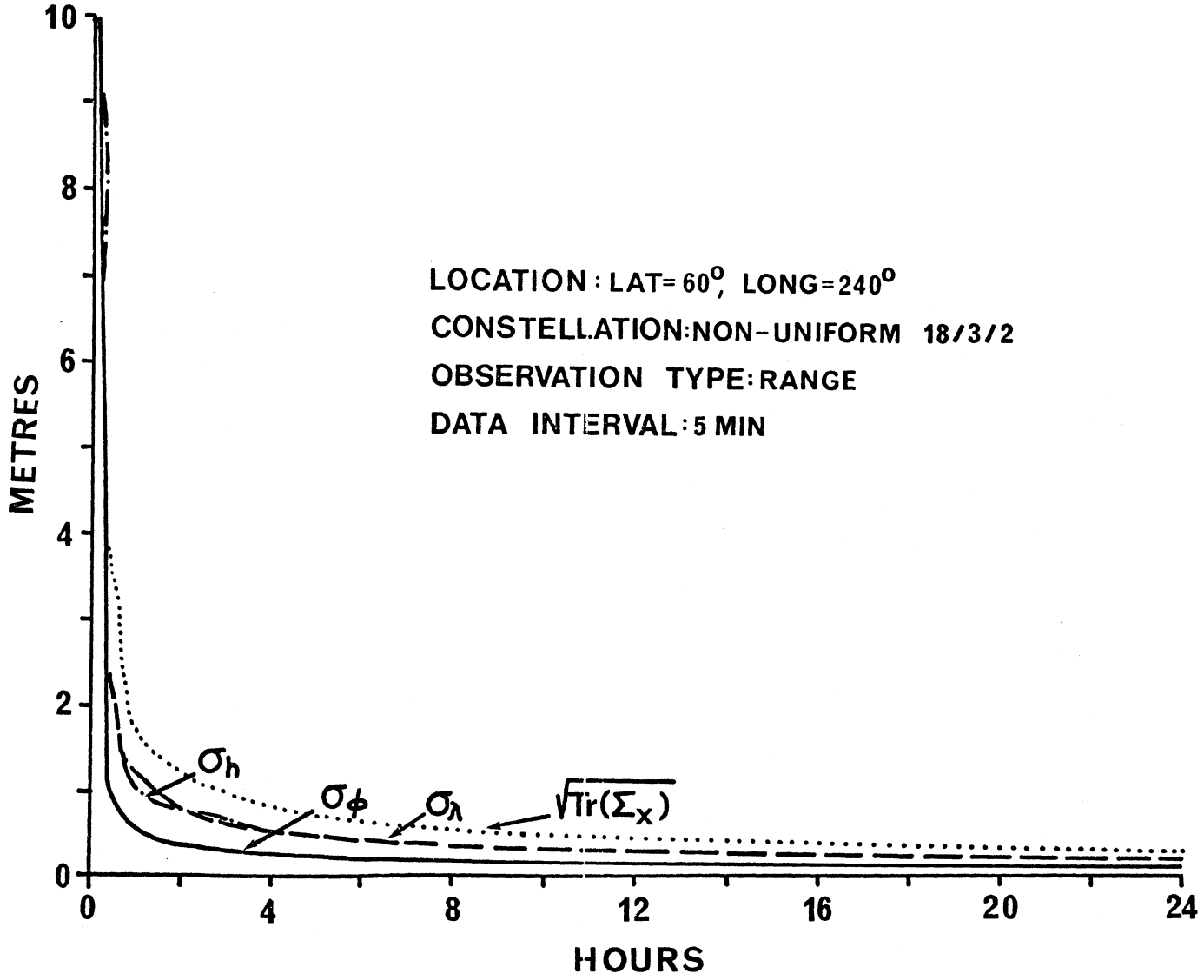
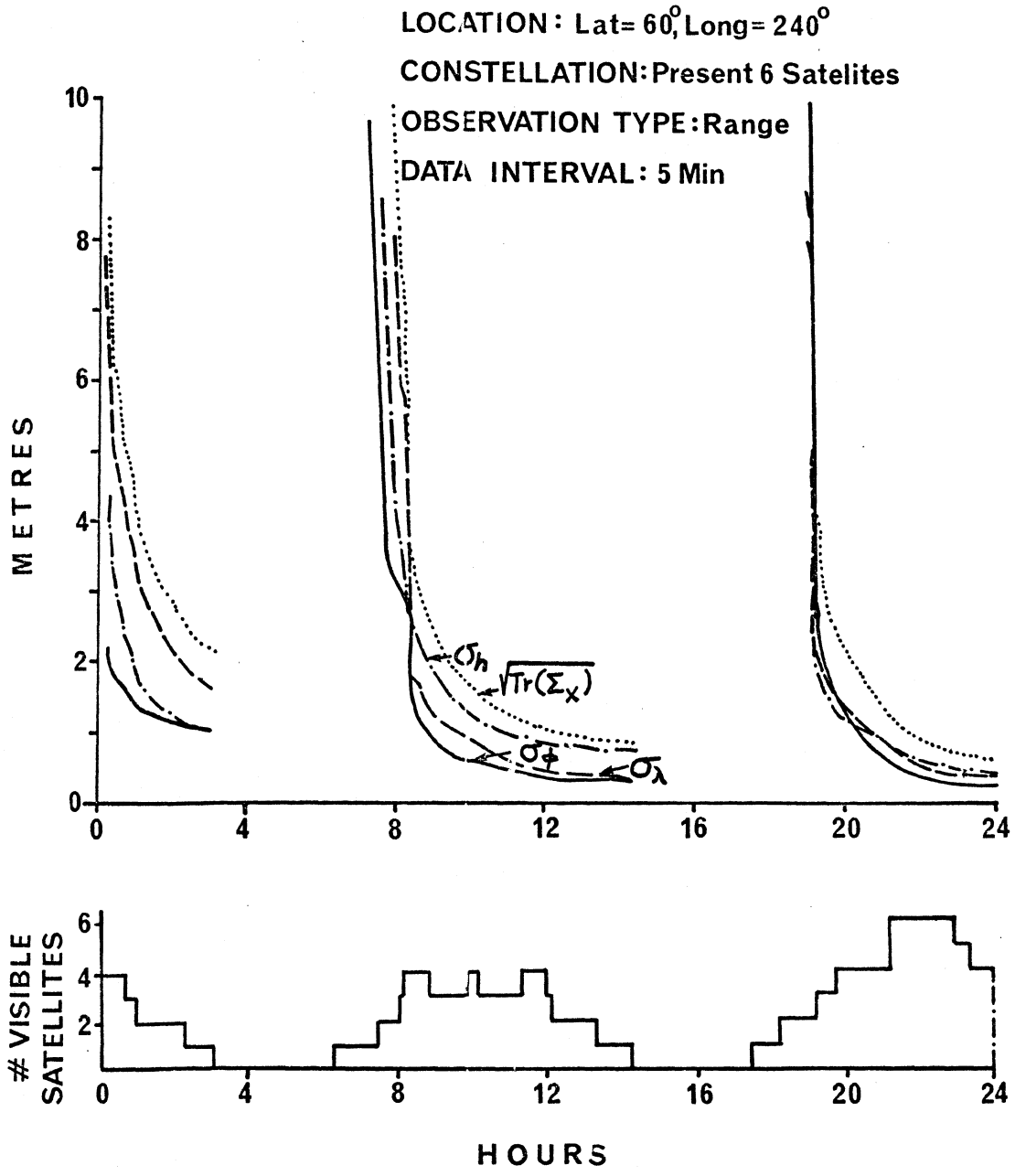


FIGURE 6.2 (c)

FIGURE 6.3(a)

EFFECT OF VARIATIONS IN THE OBSERVER'S GEOGRAPHIC LOCATION (G) ON RANGE ACCURACY.



EFFECT OF VARIATIONS IN THE OBSERVER'S GEOGRAPHIC LOCATION (G) ON RANGE ACCURACY.

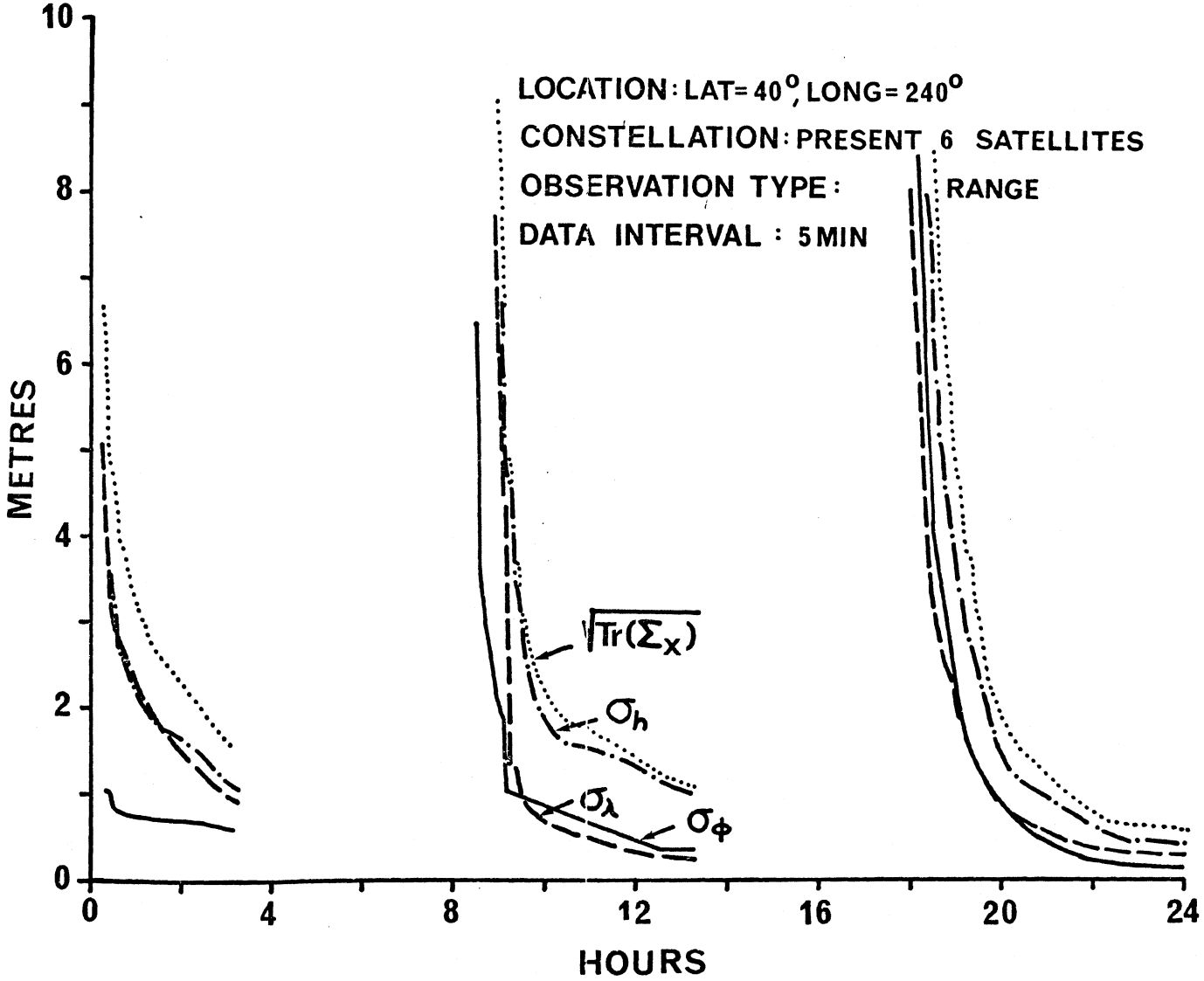


FIGURE 6.3(b)

EFFECT OF VARIATIONS IN THE OBSERVER'S GEOGRAPHIC LOCATION (G) ON RANGE ACCURACY.

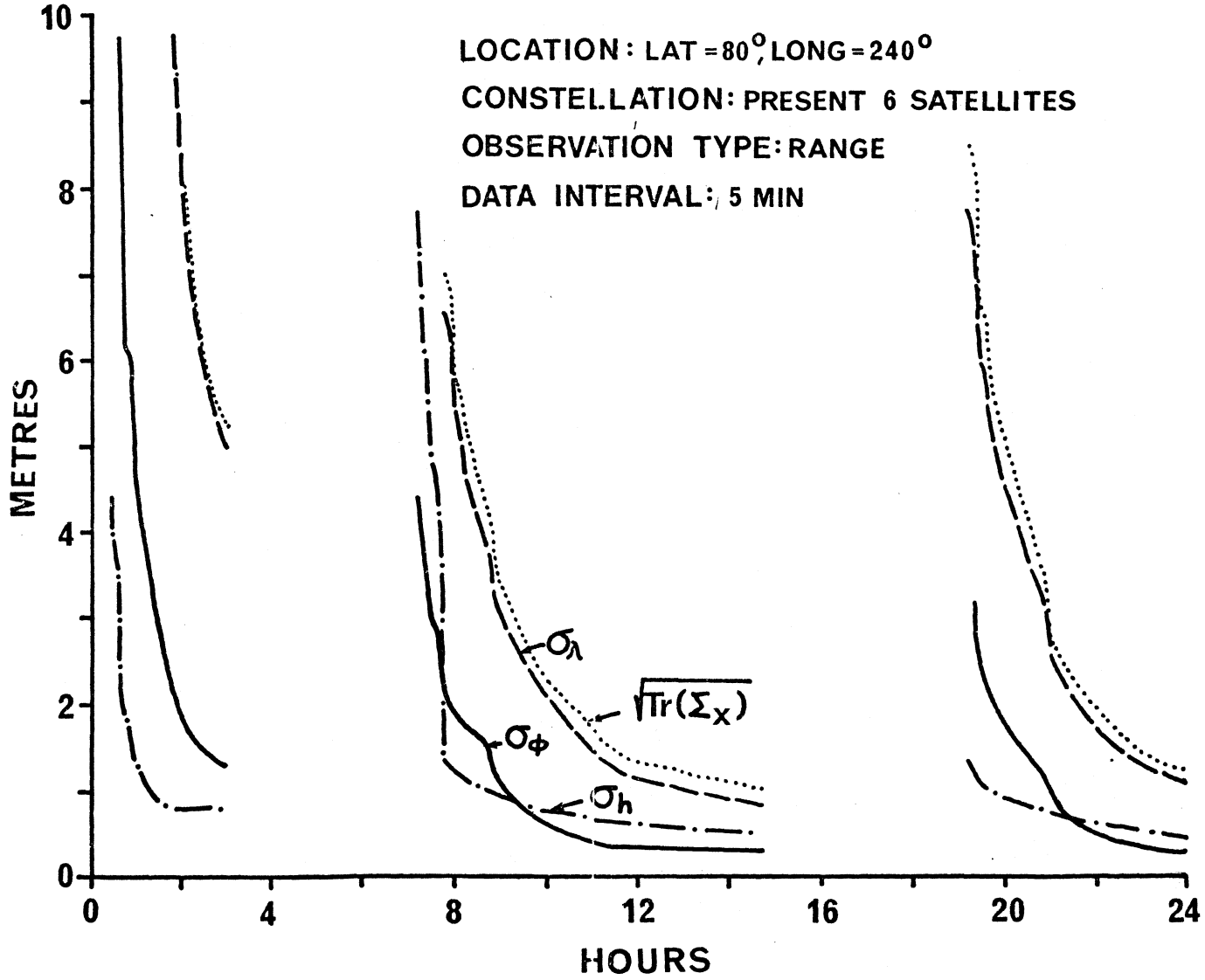


FIGURE 6.3(c)

TABLE 6.3 Point Positioning Performance of the Non-uniform 18/3/2 Constellation as a Function of Tracking Interval (ϕ = 60° N, λ = 240° E). Standard Error of Station Parameters Based on Range and Doppler Observations Every Five Minutes with Measurement Uncertainties of 1 and 1.4 m respectively.

TOTAL OBSERVATION TIME		8 Hours						24 Hours					
TRACKING INTERVAL		1 Hour		2 Hours		3 Hours		1 Hour		2 Hours		3 Hours	
OBSERVATION TYPE		R	D	R	D	R	D	R	D	R	D	R	D
POSITION ACCURACY (m)	σ_ϕ	0.37	1.20	0.35	0.63	0.40	0.65	0.11	0.97	0.12	0.57	0.12	0.59
	σ_λ	0.52	2.94	0.55	1.63	0.54	1.33	0.20	2.01	0.21	1.39	0.22	1.20
	σ_h	0.52	1.31	0.73	0.77	0.76	0.98	0.23	0.91	0.66	0.23	0.22	0.78
	$\sqrt{Tr(\Sigma_x)}$	0.82	3.43	0.98	1.92	1.03	1.78	0.32	2.42	0.70	1.25	0.33	1.55

tests indicate that if a longer tracking interval (i.e., longer duration of actual uninterrupted tracking of one satellite) is utilized, uncertainties in the estimated position are to be expected to be larger than in the case whereby satellite switching between satellites in short time intervals is allowed. An increase in the tracking interval from one to two, then to three hours, shows a substantial increase in the overall station position uncertainty, with the uncertainty in the height component displaying the most significant increase. This is likely due to the sampling of the selected passes reflecting somewhat the increased frequency of high elevation observations noted in Figure 6.4. As is to be anticipated, if a longer site occupation (e.g., 24 hours) is utilized, the results can be reduced further to approximately the 30 cm level.

6.2 Doppler Range Difference Results

Doppler observations were simulated for one tracking station (at $\phi = 60^{\circ}$ N, $\lambda = 240^{\circ}$ E) over tracking intervals of one, two, and three hours and total observation time from 8 to 24 hours. Ranging observations taken every five minutes were treated as correlated range differences with standard deviation of 1.4 m corresponding to equivalent range measurement uncertainty of 1 m (note the defining equation (3.2) for range differences and equation (3.12)). Satellites were again selected on the basis of minimizing the trace of the coordinate covariance matrix. The results of these tests are summarized in Table 6.3 where the results of range observation under the same conditions are also given for comparison.

For correlated Doppler observations based on one-hour tracking intervals there is a two- to five-fold increase in the uncertainty of the coordinate components, and a four-fold increase in the overall position uncertainty compared with the corresponding range results. As the tracking interval increases to two and three hours, the Doppler results improve significantly, remaining however inferior to the corresponding range results. Increasing the total observation time to 24 hours shows a decrease in the uncertainty of the Doppler observations. Figures 6.5(a) to 6.5(c) show results for tracking intervals of 1, 2, and 3 hours for a station at 60° latitude. Figures 6.6(a) to 6.6(c) show equivalent results for 40° latitude. However, the rate of decrease is smaller than the decrease produced by the range results over the same period.

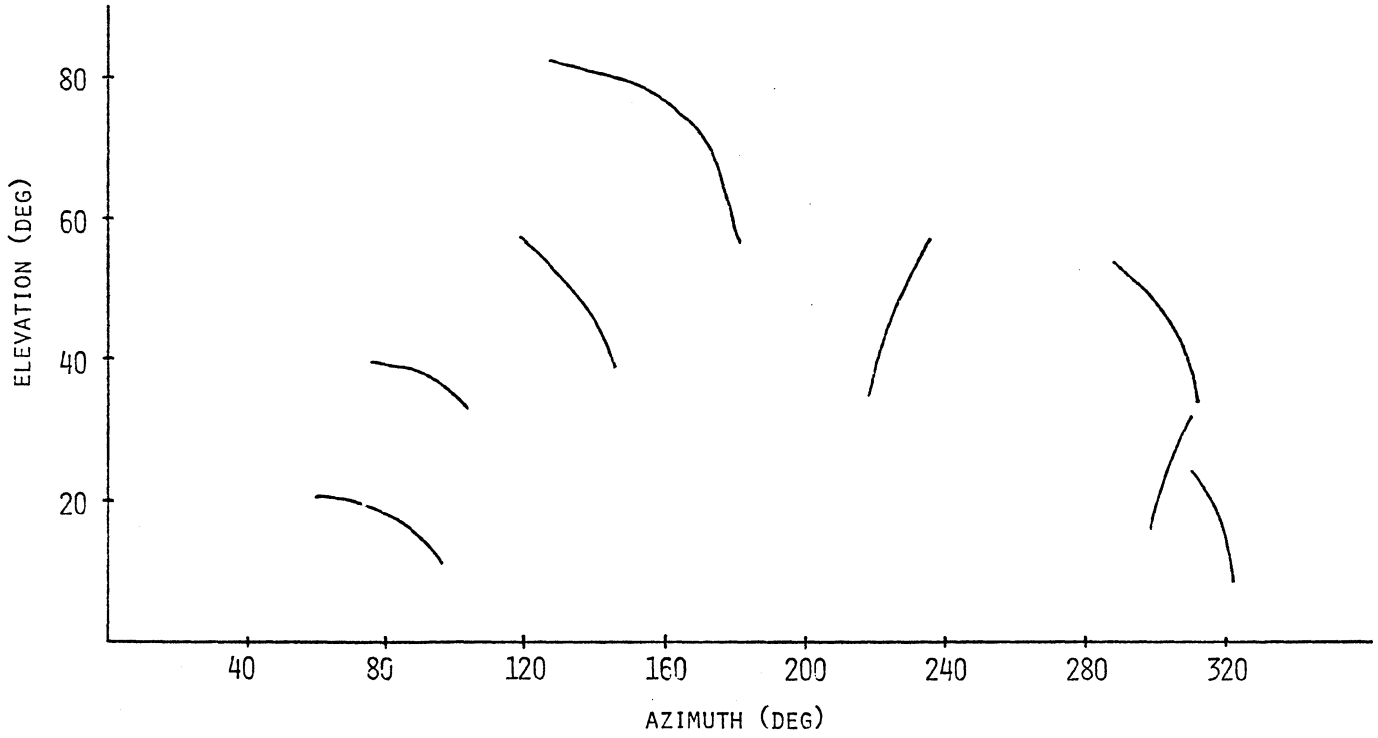


FIGURE 6.4 Satellites selected for Doppler Simulation

EFFECT OF VARIATIONS IN SATELLITE DWELL TIME (D) (ϕ = 60°) ON DOPPLER ACCURACY.

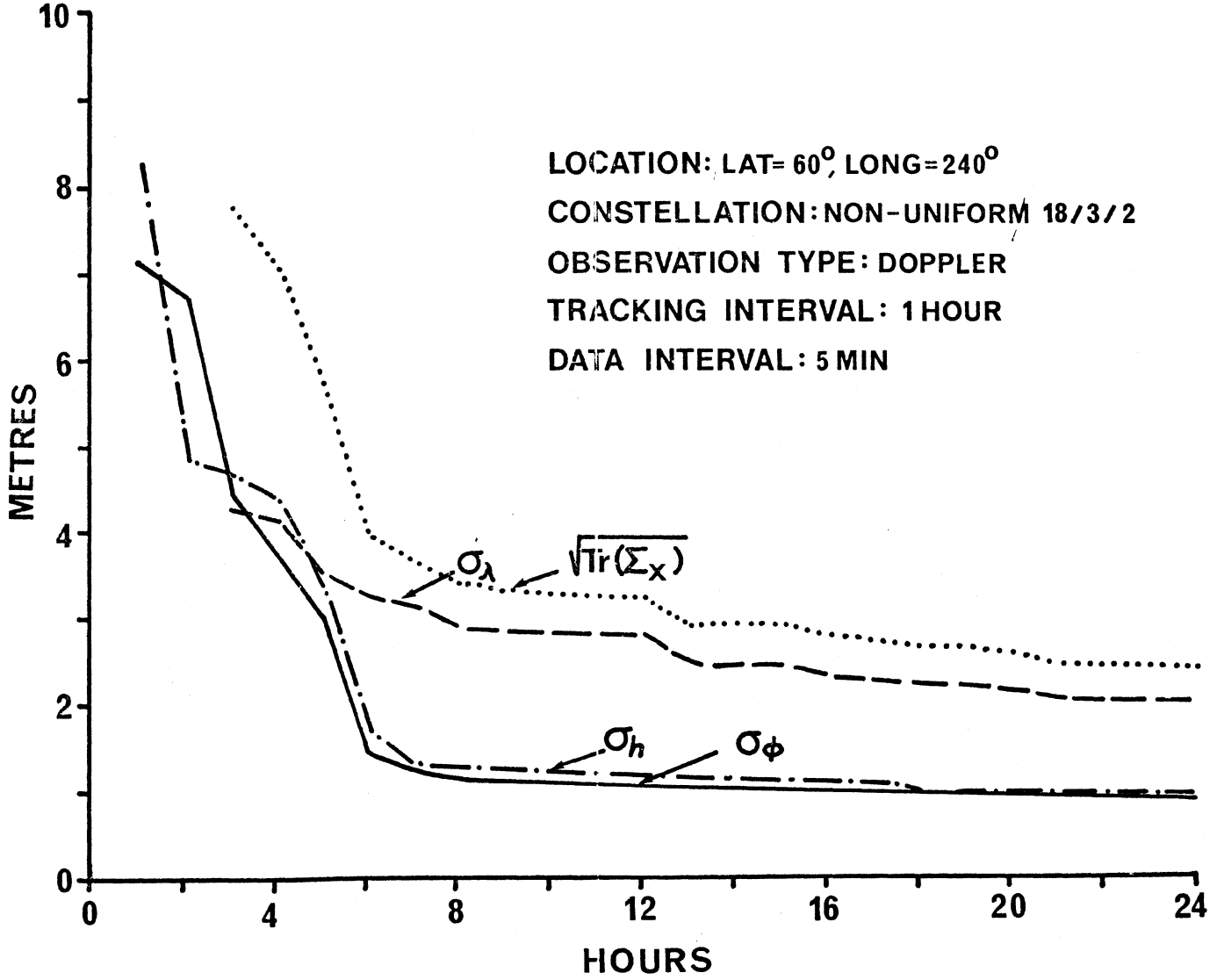


FIGURE 6.5(a)

EFFECT OF VARIATIONS IN SATELLITE DWELL TIME (D) (ϕ = 60°) ON DOPPLER ACCURACY.

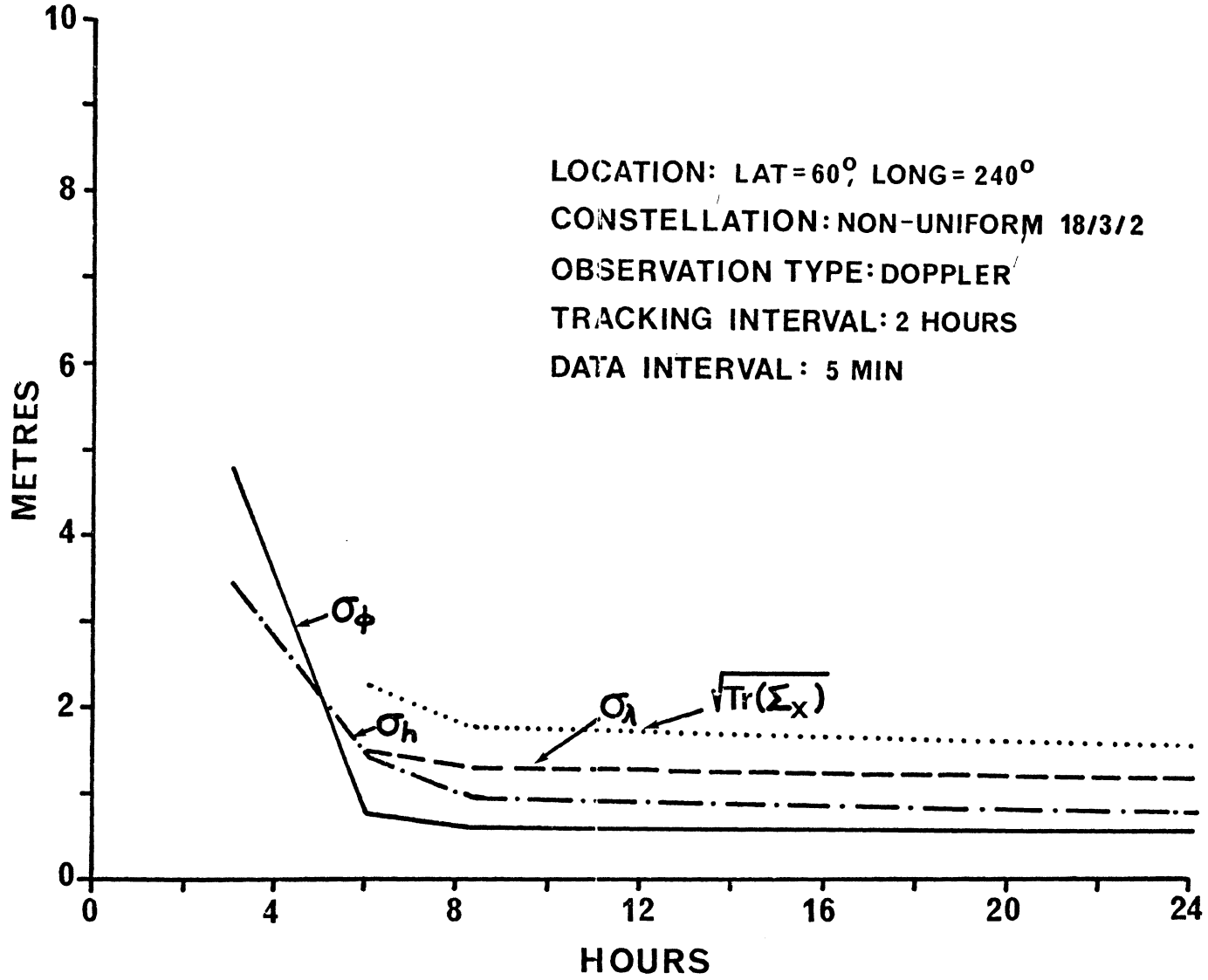


FIGURE 6.5(b)

EFFECT OF VARIATIONS IN SATELLITE DWELL TIME (D) (ϕ = 60°) ON DOPPLER ACCURACY.

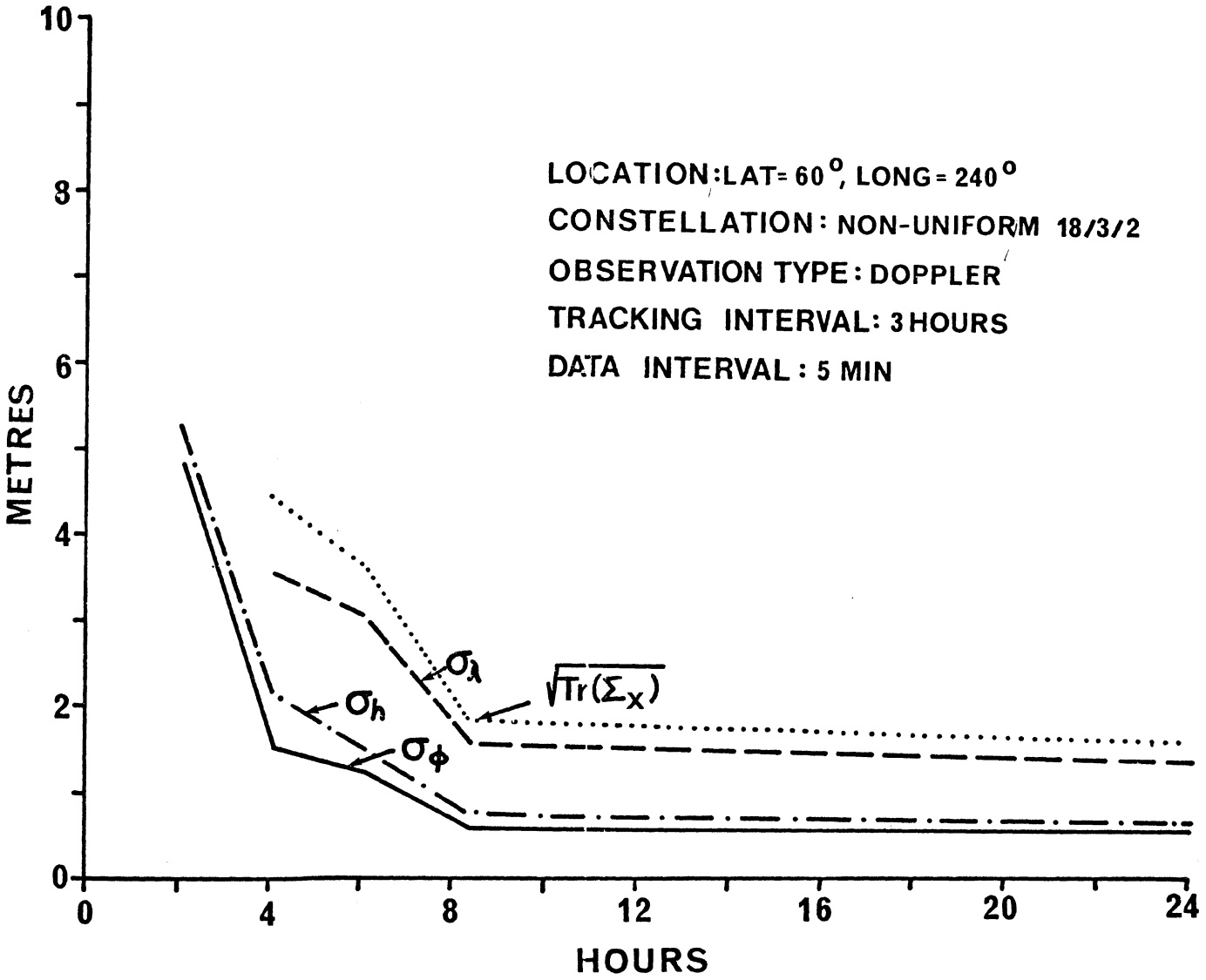


FIGURE 6.5(c)

EFFECT OF VARIATIONS IN SATELLITE DWELL TIME (D) (ϕ = 40°) ON DOPPLER ACCURACY.

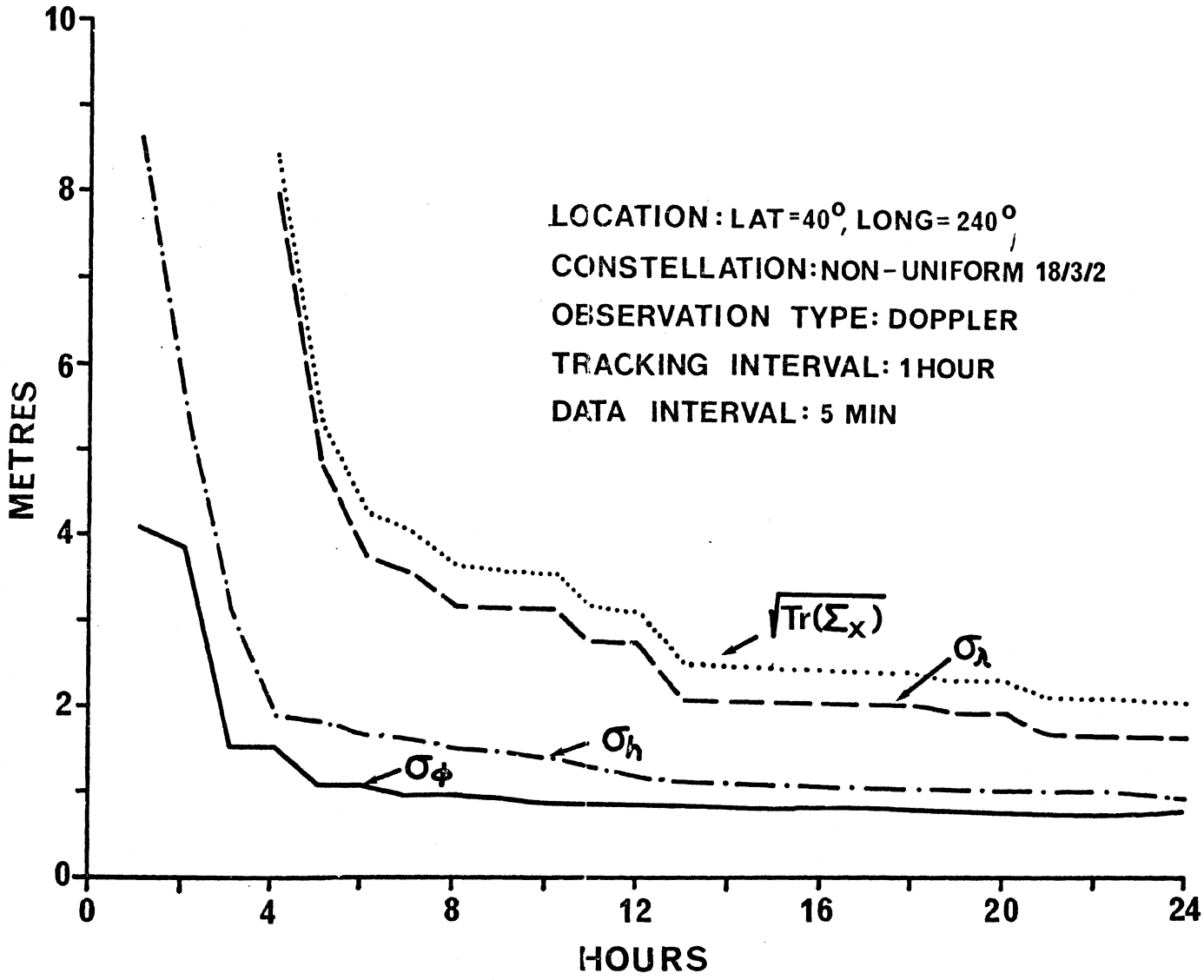


FIGURE 6.6(a)

EFFECT OF VARIATIONS IN SATELLITE DWELL TIME (D) (ϕ = 40°) ON DOPPLER ACCURACY.

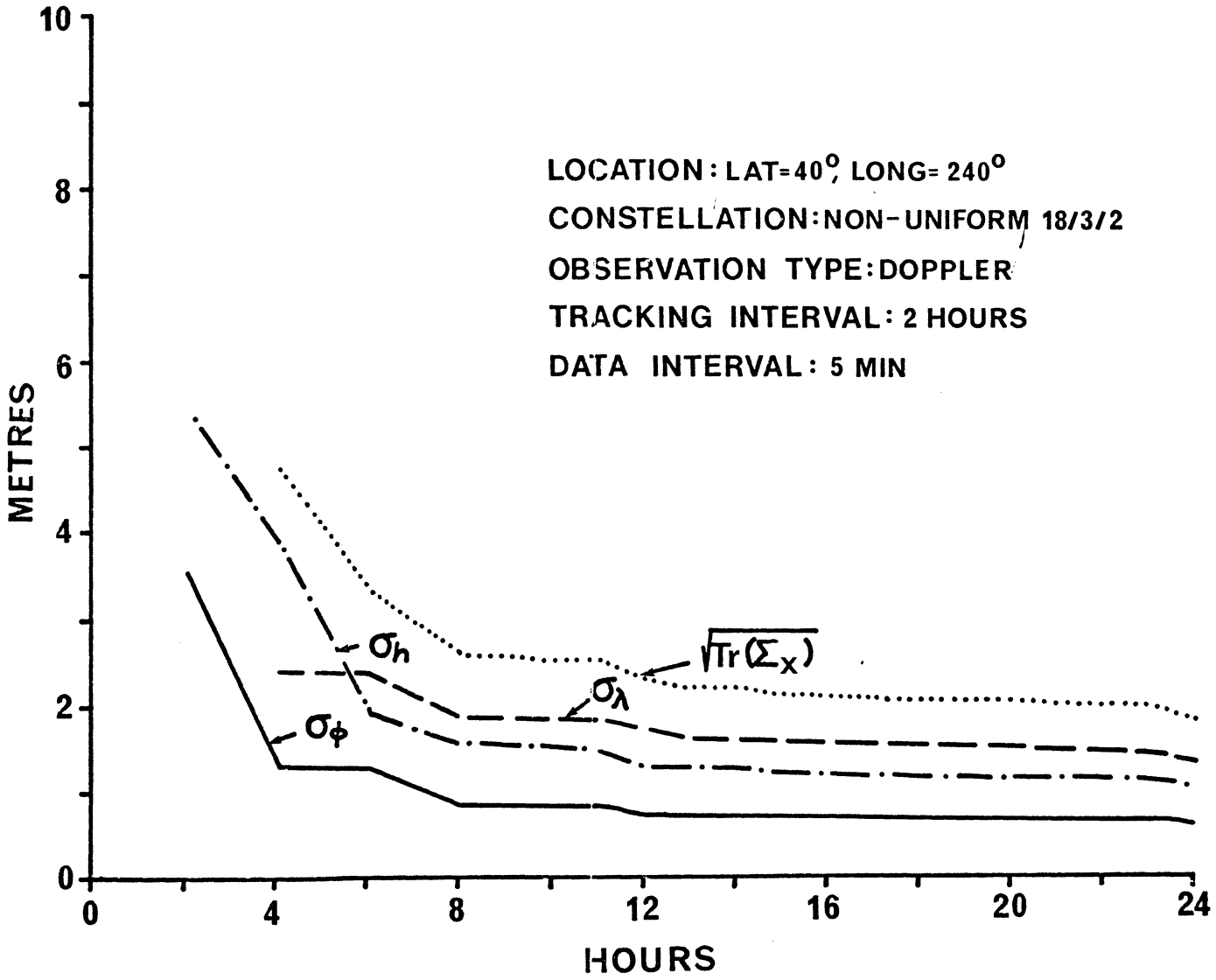


FIGURE 6.6(b)

EFFECT OF VARIATIONS IN SATELLITE DWELL TIME (D) (ϕ = 40°) ON DOPPLER ACCURACY.

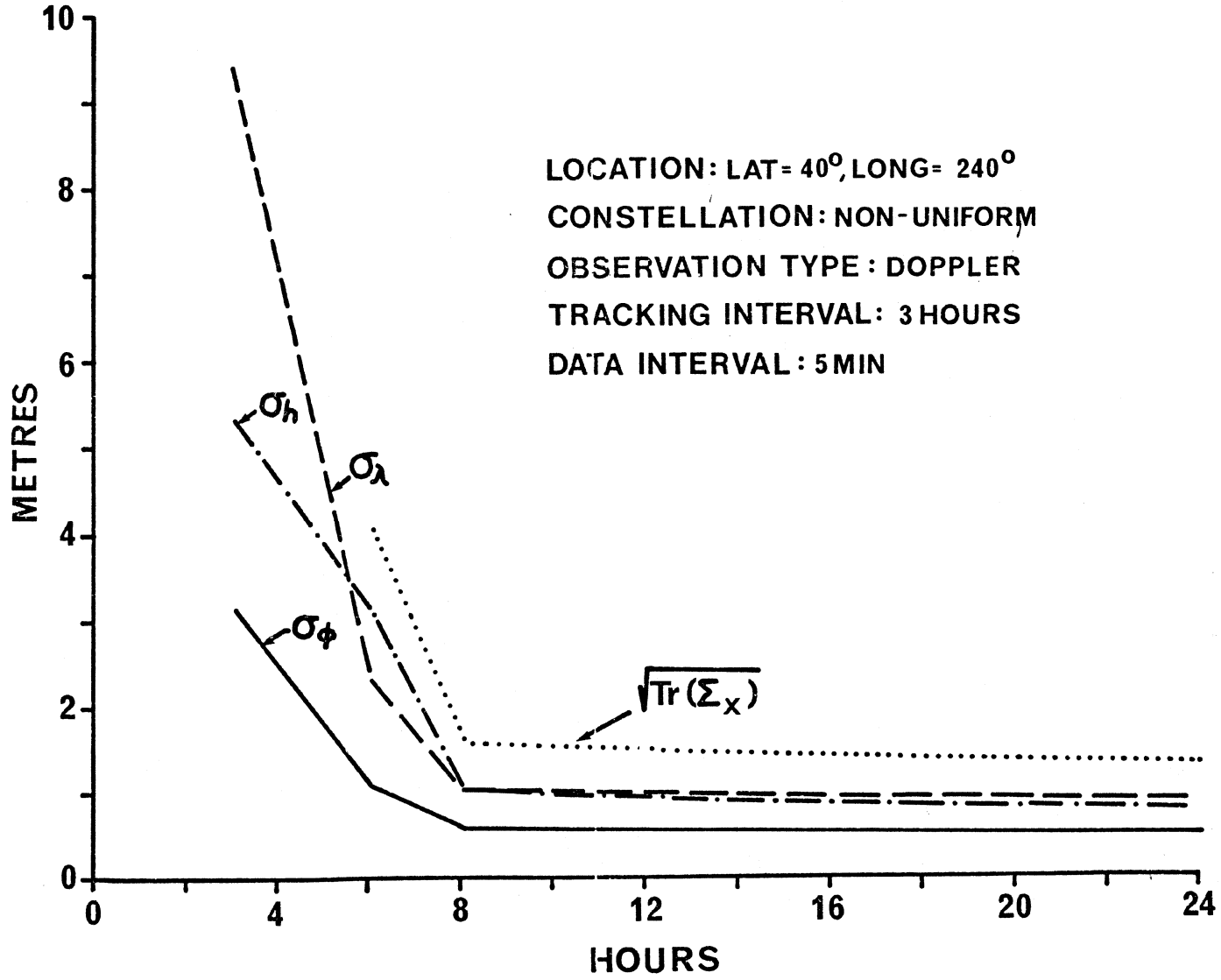


FIGURE 6.6(c)

To obtain a measure of the variation of the Doppler results as a function of the user's location, the covariance computations for the two-hour tracking interval and total observation time of 8 hours were repeated for the test grid points of Figure 5.1. From Table 6.4, some variations in the results are evident. As with range, polar stations have a larger standard deviation for longitude than the latitude and height components, compared with stations at lower latitudes. Again, this is likely due to the distribution of observing elevations and azimuths at higher latitudes.

6.3 Additional Results

So far in the tests already described, there were three major assumptions involved:

(a) the use of the non-uniform 18/3/2 satellite constellation,

(b) the use of a single channel sequential receiver,

(c) a postulated range measurement uncertainty of one metre.

Alternatives would be to assume:

(a) That in its final deployment a different full constellation will be utilized. To investigate the effect of such an event, some additional tests considering the non-uniform 18/3/0 and 18/6/2 constellations, as well as the original 24/3/2 constellation, were carried out.

(b) The use of a multiple channel receiver. Tests assuming the use of such a receiver, simultaneously tracking four satellites, were also carried out.

(c) A measurement error close to the total range error budget value which reflects the contributions of a number of effects such as those shown in Tables 3.1 and 3.2. The same procedure can be used then to estimate the influence of these effects without modelling a higher dimension parameter vector which includes these quantities as unknowns.

For these tests, only range observations were considered. The results are summarized in Table 6.5. Case 1 has been denoted as the standard test case with cases 2 through 13 indicating the results of tests deviating from the standard case by some particular aspect (e.g., constellation, measurement error, or tracking operation).

Test cases 3 to 5 indicate that there is no significant difference in the results using any of the suggested final 18 or 24 satellite constellations.

TABLE 6.4 Point Positioning Performance of the Non-uniform 18/3/2 Constellation as a Function of Geographic Location. Standard Error of Station Parameters Based on Doppler Observations (tracking interval 2 hours, total observation time 8 hours)

LATITUDE		40°			60°			80°		
LONGITUDE		240°	270°	300°	240°	270°	300°	240°	270°	300°
POSITION ACCURACY (m)	σ_ϕ	0.60	0.53	0.74	0.62	0.63	0.61	0.60	0.93	0.82
	σ_λ	1.09	0.87	0.97	1.62	1.36	1.33	4.86	3.84	2.96
	σ_h	0.79	0.95	1.05	0.77	0.87	0.93	1.10	0.66	1.09
	$\sqrt{Tr\ (\Sigma_x)}$	1.47	1.40	1.61	1.91	1.74	1.74	5.02	4.01	3.26

Comments: Results are based on Doppler observations every five minutes with measurement uncertainties of 1.4 m.

TABLE 6.5

Effect of Variations in S, C, N, and I

STANDARD TEST CASE		DEVIATIONS FROM STANDARD TEST CASE															
TEST CASE		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NON-UNIFORM 18/3/2 CONSTELLATION		✓	6 SVs	18/3/0*	18/6/2*	24/3/2	✓	✓	✓	✓	✓	✓	✓	✓			
MEASUREMENT ERROR = 1 m		✓	✓	✓	✓	✓	4 m	4 m	4 m	16 m	16 m	16 m	✓	✓			
TOTAL OBSERVATION TIME = 8 hrs		✓	✓	✓	✓	✓	✓	24 hrs	✓	✓	✓	✓	2 hrs	2 hrs			
DATA INTERVAL = 5 MIN		✓	✓	✓	✓	✓	✓	✓	10 s	✓	10 s	10 s**	10 s	5 min**			
POSITION ACCURACY (m)	σ_ϕ	0.17	0.25	0.18	0.18	0.17	2.66	1.60	0.52	25.55	8.35	3.98	0.07	0.19			
	σ_λ	0.36	0.41	0.35	0.35	0.35	5.30	3.24	0.01	48.94	15.80	7.88	0.14	0.38			
	σ_h	0.38	0.41	0.33	0.36	0.33	4.95	3.48	1.52	45.57	17.35	8.91	0.14	0.39			
	$\sqrt{Tr(\Sigma_x)}$	0.55	0.62	0.51	0.53	0.51	7.72	5.02	1.90	71.59	24.91	12.54	0.21	0.58			

* Non-uniform constellations.
** Simultaneous 4 satellite tracking (multichannel receiver).

Test cases 6 to 8 and 9 to 11 show the effect of code availability, i.e., the P-code and C/A-code respectively. For an increased total error budget value, the performance of the system can be improved by either increasing the total observation time (i.e., case 7 versus case 6), or utilizing shorter satellite switching intervals (e.g., 10 s instead of 5 min; cf. case 8 versus case 6 or case 10 versus case 9). In the event of the latter, less time of site occupation is required (e.g., case 12). Utilizing a multiple channel receiver, simultaneously tracking four satellites at a prescribed rate reduces the estimation uncertainty by a factor of 2 compared to the sequential, single channel receiver results (e.g., case 11 versus case 10; also case 13). However, this cannot be looked at outside the context of receiver designs and the different receiver requirements relevant to each case [Collins 1981].

7. DIFFERENTIAL GPS TECHNIQUES

7.1 Advantages of a Differential Approach

As has been demonstrated for the TRANSIT Doppler positioning system, and for other radio positioning systems such as Omega, the interstation vector between two or more stations which simultaneously track the same radio signals can be more accurately determined than the absolute position vector of either station [Wells, 1976].

GPS error sources can be divided into three groups: errors originating at the satellite (ephemeris, satellite timing, oscillator stability); errors due to propagation (ionospheric and tropospheric); and errors originating in the receiver (receiver delays, noise, oscillator stability). The effects of the first two groups of errors are correlated between stations within a few hundred kilometres of each other, simultaneously tracking GPS. The process of taking advantage of this correlation to improve the relative position accuracy has been called "translocation" for the TRANSIT system. We will apply this label to differential GPS as well. The improvement is usually of one order of magnitude or more [Anderle, 1980].

Several levels of GPS translocation may be considered. At the lowest level the receivers are required to track simultaneously, but not necessarily the same satellites. In this case the effect of the system timing biases, and propagation errors would be correlated. One level higher, the receivers are forced to track the same set of satellites, although not necessarily in the same sequence. In this case the effect of ephemeris errors and individual satellite timing biases would be correlated. At the highest level, the receivers are forced to track the same wavefront from the same satellite simultaneously, in which case the correlation between the effects of satellite and propagation error sources is the maximum. The choice of which level of translocation to use depends on tradeoffs between desired accuracy and sophistication of the equipment and processing.

As the distance between the two "differential" receivers increases, the correlation between error effects decreases. The effective range of GPS translocation should not be significantly different from that for TRANSIT, despite the higher GPS orbits, due to the limiting correlation distance of atmospheric conditions affecting refraction. This is typically about

200 km. According to Fell [1980a], the effect of orbital errors and satellite clock errors on relative positions up to a few hundred kilometres apart is less than 20 cm.

In the remainder of this chapter, we discuss (from the differential point of view) the three measurement techniques which we have not dealt with in detail so far; Doppler range differences, reconstructed carrier phase differential range measurements, and interferometric differential range and range rate measurements. While pseudo range measurements can be used for differential positioning as well, the precision of the pseudo range measurement (as compared with the magnitude of the satellite and propagation errors which can be reduced by differential techniques) is not as advantageous as for the other three techniques.

7.2 Doppler Range Difference Measurements

In this mode, the receiver locks onto the received signal by means of a phase-lock loop which reconstructs the carrier (see Section 1.8). The beat frequency between the reconstructed carrier and the ground oscillator generated signal, that is, the Doppler frequency $f_G - f_R$ is the main source of information here. Depending on the way the Doppler count

$$N(t_1, t_2) = \int_{t_1}^{t_2} (f_G - f_R)\, dt \qquad (7.1)$$

is integrated, we can treat a sequence of integrated Doppler counts as being uncorrelated, or as correlated (for continuously counted integrated Doppler (CCID)). In either case, the integrated Doppler counts can be converted to range differences by scaling them by the velocity of light divided by the transmitted frequency.

Clearly, if a single channel receiver is to accumulate CCID data from several satellites, it must not lose counts while switched to another satellite. This requires that the receiver maintain a predicted count until re-acquiring the original satellite, and that this count have a prediction error of less than one cycle. This imposes constraints on the switching rate (see Chapter 8). Various techniques for the count prediction are being investigated. So far it appears that the Doppler count cannot be measured without the knowledge of the P-code. Lack of access to the P-code thus renders this technique inoperative.

Oscillator instability has a profound effect on Doppler data collection

techniques when used in the point positioning mode (see Chapter 9). This effect can be, however, considerably reduced if two satellites are tracked nearly simultaneously by the two receivers in the differential mode.

The first tests conducted by Anderle [1980] with experimental geodetic GPS receivers showed promising results. Subsequent simulations showed that errors can be expected to be of the order of metres for uncorrelated Doppler counts and decimetres for the CCID approach (with 10 cm being the minimum for 8 hours of tracking). The mathematical model used for the differential positioning is the same as that used with the TRANSIT in the translocation mode [Kouba, 1978] and we shall not dwell on it here. It is interesting to note, however, that the ionospheric propagation correction to the range differences evaluated from the two frequencies L_1 and L_2 is good to about 5 mm [Anderle, 1980].

7.3 Reconstructed Carrier Phase Differential Range Measurements

The reconstructed carrier phase (variation) measurements are based on a rather different principle. It is useful to regard the phase of the incoming (received) signal as the indicator of the origin of the signal. If the transmitter stays at the same fixed distance from the receiver, this phase also remains constant. Viewed from this perspective, it is rather clear that the phase of the received signal changes with the change of range between the satellite and receiver. The phase changes by 360° when the range changes by one wavelength λ of the carrier.

The variation of the incoming signal carrier's phase is identical with that of the reconstructed carrier's because they are phase locked. Thus the phase variation can be measured on the reconstructed carrier with respect to the ground oscillator generated signal, since the phase of the ground oscillator remains constant. Now, there is a well-known relationship between the phase ϕ and frequency f of any sine wave:

$$df = \dot{\phi} = d\phi/dt \quad . \tag{7.2}$$

Hence the phase change $\delta\phi(t)$ of the reconstructed carrier can be written simply as

$$\delta\phi(t) = \phi(t) - \phi(t_o) = \int_{t_o}^{t} \delta f_R \, dt \quad . \tag{7.3}$$

Realizing that δf_R is simply the beat frequency $(f_G - f_R)$ because the frequency f_G of the ground oscillator is constant, we can see that the phase

changes can be easily obtained from the continuously integrated Doppler counts (equation (7.1)).

The phase change can be measured to an accuracy of about 8^{o} [Counselman et al., 1980] which for L_1 corresponds to approximately 4 mm and for L_2 to about 5 mm. The problem is, however, that the measurable $\delta\phi(t)$ changes only from 0^{o} to 360^{o}. If the change is larger than 360^{o}, then the number of full cycles n--corresponding to the change of the range by $n\lambda$--is lost. Thus the phase (or range) change can be measured quite accurately but ambiguously, with the ambiguity being equal to an integral number of cycles (wavelengths λ). The success of this approach then hinges on the capability to resolve this ambiguity.

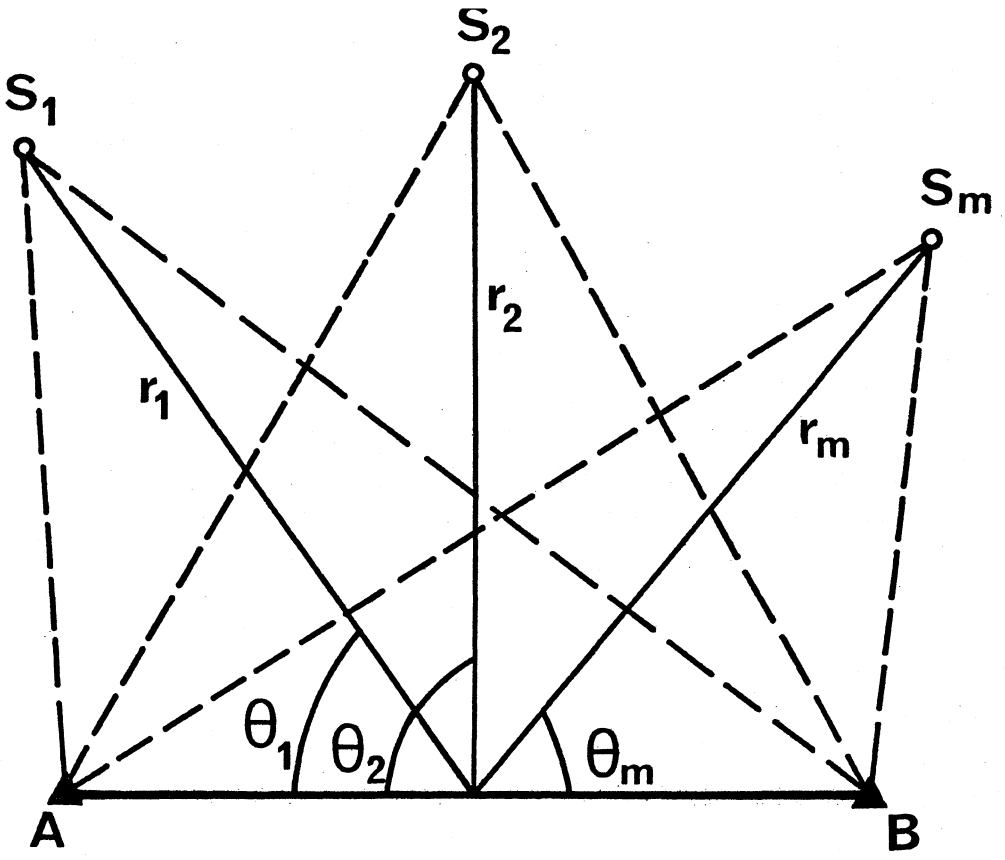
The mathematical model for m satellites reads [Bossler et al., 1981]:

$$D \cos\theta_j(1 + (D^2/8r_j^2)\sin^2\theta_j) \simeq (n_j + \phi_j)\lambda + P_j, \quad j=1,m, \tag{7.4}$$

where D is the length of the baseline AB (to be determined), r_j is the range from the mid-point of the baseline to the j^{th} satellite, θ_j is the angle between the j^{th} range and the baseline (see Figure 7.1), $P_j = P_j^B - P_j^A$ is the difference between the tropospheric propagation corrections to ranges for the two end-points and the j^{th} satellite, $\phi_j = \phi_j^B - \phi_j^A$ is the difference between the two measured phases of the reconstructed carrier by the two receivers at A and B, and n_j is an integer denoting the difference of lost cycles. All these values refer to a specified instant of time t. Realizing that D cos θ_j is a projection of the baseline onto the j^{th} range, the components of the baseline vector $\vec{D}$ can be resolved when several satellites are used.

Larden and Bender [1980] claim the effective ambiguity to be an integral multiple of $\lambda/2$ while Counselman et al. [1980] argue for an integral multiple of λ. This ambiguity can be resolved in several ways. The rather obvious one is to use the pseudorranges to determine the approximate relative positions of A and B, from which n_j can be determined. Another approach, using the idea of solving for biases in each baseline projection D cos θ_j, is discussed in detail by Larden and Bender [1981]. The latter approach requires the water vapour radiometer's measurements for baselines longer than about 5 to 10 km.

It seems, from the work done so far, that two potential problems exist



GPS PHASE MEASUREMENT GEOMETRY

FIGURE 7.1

with the reconstructed carrier phase mode of differential positioning. The first relates to the multipath effect [Spilker, 1978]; the other is the uncertainty in the evaluation of tropospheric propagation correction due to the water vapour content in the air. More about the latter will be said in Chapter 10. Also, knowledge of the P-code is a prerequisite for being able to measure phase because of the necessity to reconstruct the carrier.

The simulations carried out so far show that on baselines 10 to 100 km long with 6 hours of tracking, an accuracy of 2 to 5 cm can be attained without using water vapour radiometers and 1 to 2 cm with radiometers [Anderle, 1980]. Similar accuracies were obtained by Bossler et al. [1981] from merely 2 hours of simulated tracking. The latter authors present a comprehensive list of items for further research into this technique.

7.4 Interferometric Differential Range and Range Rate Measurements

The differential modes that we have discussed so far require that the receiver generate a signal (the reconstructed carrier) that is phase-locked to the incoming GPS signal. To implement such a phase-lock loop requires that we know the coded modulations on the GPS carrier (the C/A- and P-codes). In these modes, the observations consist of either the phase or frequency difference between the reconstructed carrier and a stable local oscillator.

It is also possible to time the arrival of the incoming GPS signal by a stable local oscillator, using an open loop rather than a phase-lock loop, and hence without requiring knowledge of the coded modulations. In this case the incoming signal is treated as random noise with known frequency content [e.g., MacDoran, 1979]. The effect of the P- and C/A-codes is to spread the GPS signal spectrum over 20 MHz and 2 MHz bandwidths respectively. By sampling the signal through several narrow frequency "windows" near the GPS carrier (centre) frequency, the usual VLBI "observables" of differential range (or delay) and differential range rate (or fringe frequency) between pairs of GPS stations can be obtained using conventional VLBI correlation techniques. Repeating this for both GPS carriers provides for ionospheric refraction correction in the usual way. A tracking strategy has been suggested of switching all receivers simultaneously through two frequency windows each on L_1 and L_2, and repeating this on each of four satellites, allowing 250 ms per window, 1 s

per satellite, and 4 s per sequence [MacDoran, 1979].

If the tracking is done in the pure VLBI mode, satellite selection and identification can no longer be made by means of Doppler frequency and C/A-, P-code methods. Therefore, directional antennae must be used. Crude (to 100 km) knowledge of GPS and ground station positions are required in order to control the antenna pointing. Conventional VLBI processing techniques assume that plane waves from distant sources have been received. In order to apply corrections to the actually spherical wavefronts from GPS satellites (so that these same techniques can still be used), the satellite position must be known to about 1 km. These positions could be obtained either by means of a conventional GPS receiver, or, if complete independence from the GPS data stream is desired, by a VLBI-type monitor network.

Assuming water vapour radiometers are used (see Chapter 10), accuracies in determining the baseline vector between stations separated by between 2 and 200 km have been predicted to be 0.5 to 3 cm respectively [MacDoran, 1979]. A data span of 2 hours would be required to achieve this. Receiver costs are estimated to be competitive with those for the other differential GPS modes, but the VLBI data processing may represent an appreciable additional cost.

To prove the feasibility of this concept, it must be demonstrated that

(a) meaningful delays and fringe frequency information can be extracted from the GPS signals, treated as noise;

(b) inexpensive processing techniques can be devised which will handle this data;

(c) the cost and accuracy predictions for this technique are realistic.

Two kinds of prototype receivers, one with the acronym SERIES (satellite emission radio-interferometric earth surveying) [MacDoran, 1979], and the other with the acronym MITES (miniature interferometric terminals for earth surveying) [Counselman and Shapiro, 1979] are under construction. If their performance validates the above points, then this technique has a considerable advantage over the other differential GPS techniques--it is available to all users no matter what decisions may eventually be made restricting the availability of the full GPS signal structure.

8. RECEIVER CONSIDERATIONS

A number of proposals for geodetic GPS receiver designs have been made. These vary considerably in the observing technique to be used, as described in Chapter 7. In this section we restrict our attention to the more basic decision of how the space vehicles are to be selected and tracked. At one end of the spectrum of possibilities is a multichannel receiver tracking several satellites simultaneously. At the other end of the spectrum is the single channel receiver tracking one satellite at a time, and switching satellites perhaps every few hours. From the geodetic point of view, the multichannel receiver has two advantages--the demands on oscillator stability are less (as described in the next section), and the satellite/receiver geometries sampled lead to a stronger solution in less time. However, the single channel receiver has advantages of lower size, weight, power, cost, and component count (hence lower failure rate). A third option which combines these two sets of advantages is the time-division-multiplexed (TDM) or "fast switching" single channel receiver [Ward, 1980; Collins, 1981]. This is a single channel receiver that looks like a multichannel receiver, by switching between satellites so quickly that it can keep track of all the signal phases simultaneously, and never really "lose" track. This receiver design has one final advantage that neither of the other two has. It is capable of making effectively simultaneous, accurate differential measurements to two different GPS satellites. Multichannel receivers will have channel-to-channel delay variations that would limit the accuracy of such differential measurements. A slow switching single channel receiver would have oscillator stability variations that would limit the accuracy of such differential measurements.

9. OSCILLATOR STABILITY

The GPS usage is based on precise time synchronization and time measurements. The timekeeping ability of the receiver is limited by instabilities in the reference frequency provided by the receiver oscillator. So critical is this aspect of user equipment performance, that much effort has been expended to study the receiver oscillator stability requirements in detail. We first introduce some concepts and provide an overview. Since the various GPS observing modes each have particular oscillator stability requirements, we then consider and compare these (in a simplified way).

9.1 Characterization of Oscillator Stability

Stability is a general, and hence vague, term when referring to the performance of an oscillator. We may mean long term stability, in the sense of a slow secular drift of the oscillator frequency, or we may be concerned with short term stability in the sense of random variations in phase. We may want to consider stability from the spectral purity point of view, or from the fractional frequency deviation over various sampling intervals.

Accordingly, two precise definitions for the measure of frequency stability have been established [Barnes et al., 1971]. One of these, $\sigma_y(t)$, is the square root of the Allan variance, which measures frequency stability in the time domain. The other, $S_y(f)$, is the spectral density of the instantaneous fractional frequency fluctuations, which measures frequency stability in the frequency domain. In general, there are two kinds of frequency variations--random and non-random (secular, periodic, sudden, etc.). Both $\sigma_y(t)$ and $S_y(f)$ respond to both random and non-random variations. We will now define these two measures.

If the instantaneous oscillator output voltage is

$$V(t) = [V_o + \varepsilon(t)] \sin [2 \nu_o t + \phi(t)] \tag{9.1}$$

where V_o and ν_o are the nominal amplitude and frequency, then

$$y(t) = \dot{\phi}(t)/(2\pi\nu_o) \tag{9.2}$$

is the instantaneous fractional frequency deviation from ν_o, where we assume both $|\varepsilon(t)/V_o|$ and $|y(t)|$ are always very small. If we define the average

value of y(t) over the averaging period τ by

$$y_k \equiv \frac{1}{\tau} \int_{t_k}^{t_k + \tau} y(t)\, dt \quad , \tag{9.3}$$

and if the averaging measurement is repeated N times at a repetition interval T (i.e., $t_{k+1} = t_k + T$, k=0,1,...,N), then

$$\langle s_y^2(N, T, \tau) \rangle \equiv \left\langle \frac{1}{N-1} \sum_{n=1}^{N} \left(\bar{y}_n - \frac{1}{N} \sum_{k=1}^{N} \bar{y}_k \right)^2 \right\rangle \tag{9.4}$$

is analogous to the sample variance of y_k, where < > denotes an infinite time average. The Allan variance then is the two-sample variance (N = 2), with no dead time between measurements (T = τ), that is

$$\sigma_y^2(\tau) \equiv \langle s_y^2(2, \tau, \tau) \rangle \quad , \tag{9.5}$$

which can also be written

$$\sigma_y^2(\tau) \equiv \langle (\bar{y}_{k+1} - \bar{y}_k)^2/2 \rangle \quad . \tag{9.6}$$

The square root of the Allan variance is the preferred measure of frequency stability in the time domain, and it is basically the normalized rms, i.e., relative error, of the differences between adjacent pairs of frequency measurements.

The spectral density $S_y(f)$ of y(t) is the preferred measure of frequency stability in the frequency domain. However, often $S_\phi(f)$, the spectral density of the phase fluctuations, and L(f), the single-sideband phase-noise-to-signal-ratio are used. These three measures are related by

$$S_y(f) = (1/\nu_o)^2 f^2 S_\phi(f) \quad , \tag{9.7}$$

and when the fractional frequency deviation (or phase deviation) of the modulation is small

$$L(f) \simeq S_\phi(f)/2 \simeq \nu_o S_y(f)/(2f^2) \quad . \tag{9.8}$$

Spectral purity is the degree to which the frequency spectrum of an oscillator output consists of one narrow line centred on ν_o. The choice of whether to use time or frequency domain characterizations of oscillator stability depends on which one most easily identifies and separates the variations under consideration. Random amplitude noise $\varepsilon(t)$ produces variations in the signal envelope in the time domain, and broadening of the spectrum in the frequency domain. Random phase noise $\dot{\phi}(t)$ produces variations in zero crossing times in the time domain, and broadening of the spectrum in the frequency domain. This spectral broadening may be only a

few hertz, or may be in megahertz. "Close-in phase noise" refers to spectral offsets from ν_o of less than 5 kHz. "Far-out phase noise" refers to offsets from ν_o of more than 5 kHz. Low close-in phase noise is a critical oscillator performance criterion for applications such as GPS.

Let us now apply these characterizations of oscillator stability to some typical commercially available oscillators. Types of oscillators which are candidates for use in GPS user equipment are crystal, rubidium, and cesium oscillators. We have chosen three crystal, one rubidium, and two cesium oscillators for comparison. The data were obtained from recent specification sheets from Hewlett-Packard, Austron, and Frequency and Time Systems. Table 9.1 and Figure 9.1 give the time domain stability measure $\sigma_y(\tau)$. Table 9.2 and Figure 9.2 give the frequency domain stability measure L(f). In general, we see that the crystal oscillators exhibit higher spectral purities above offsets of 1 Hz than the atomic oscillators, but that their Allan variances are worse for averaging times longer than about 1 second, i.e., they display a worse long-term stability.

9.2 Effect of Oscillator Stability on GPS Receiver Performance

The rigorous approach in evaluating the influence of oscillator stability on GPS receiver performance is to assume a model for either the time domain or frequency domain representation of stability, and then to propagate its effect through to the observations, and thence to the positions. For example, using a model for the Allan variance of cesium oscillators, Fell [1980a, 1980b, 1980c] found the effect on point positioning coordinate errors to be 5 cm and 7 cm for pseudo range and range-difference measurements respectively, tracking for 24 hours, switching satellites once per hour. A similar analysis for reconstructed carrier phase measurements was performed by Collins [1981], who found phase prediction errors of 1 cm due to oscillator noise for "slow" switching between satellites.

Here we look at the problem from a more simplistic viewpoint to see if we can gain some "insight" into the nature of the influence of oscillator stability on GPS results. First let us look at pseudo range, range difference, and phase measurements to see how oscillator stability enters into the measurements. For simplicity, we restrict our attention to a single channel receiver capable of tracking only one satellite at a time.

AVERAGING TIME τ (seconds)	SQUARE ROOT OF ALLAN VARIANCE $\sigma_y(\tau)$					
	CRYSTAL			RUBIDIUM	CESIUM	
	HP 10811	AUSTRON 1120S	FTS B-5400	HP 5065A	HP 5061A/004	FTS 4050/004
10^{-3}	1.5×10^{-10}	3×10^{-12}				
10^{-2}	1.5×10^{-11}		2.5×10^{-11}	1.5×10^{-10}	1.5×10^{-10}	2×10^{-10}
10^{-1}	5×10^{-12}		2.5×10^{-12}	5×10^{-12}	5×10^{-12}	5×10^{-12}
10^{0}	5×10^{-12}	2×10^{-11}	7×10^{-13}	1.6×10^{-12}	2.7×10^{-12}	3×10^{-12}
10^{1}	5×10^{-12}		7×10^{-13}	5×10^{-13}	8.5×10^{-13}	10×10^{-13}
10^{2}	1×10^{-11}		7×10^{-13}			3×10^{-13}
10^{3}						1×10^{-13}
10^{4}						1×10^{-13}

TABLE 9.1

OSCILLATOR STABILITY IN TIME DOMAIN.

OSCILLATOR STABILITY IN TIME DOMAIN.

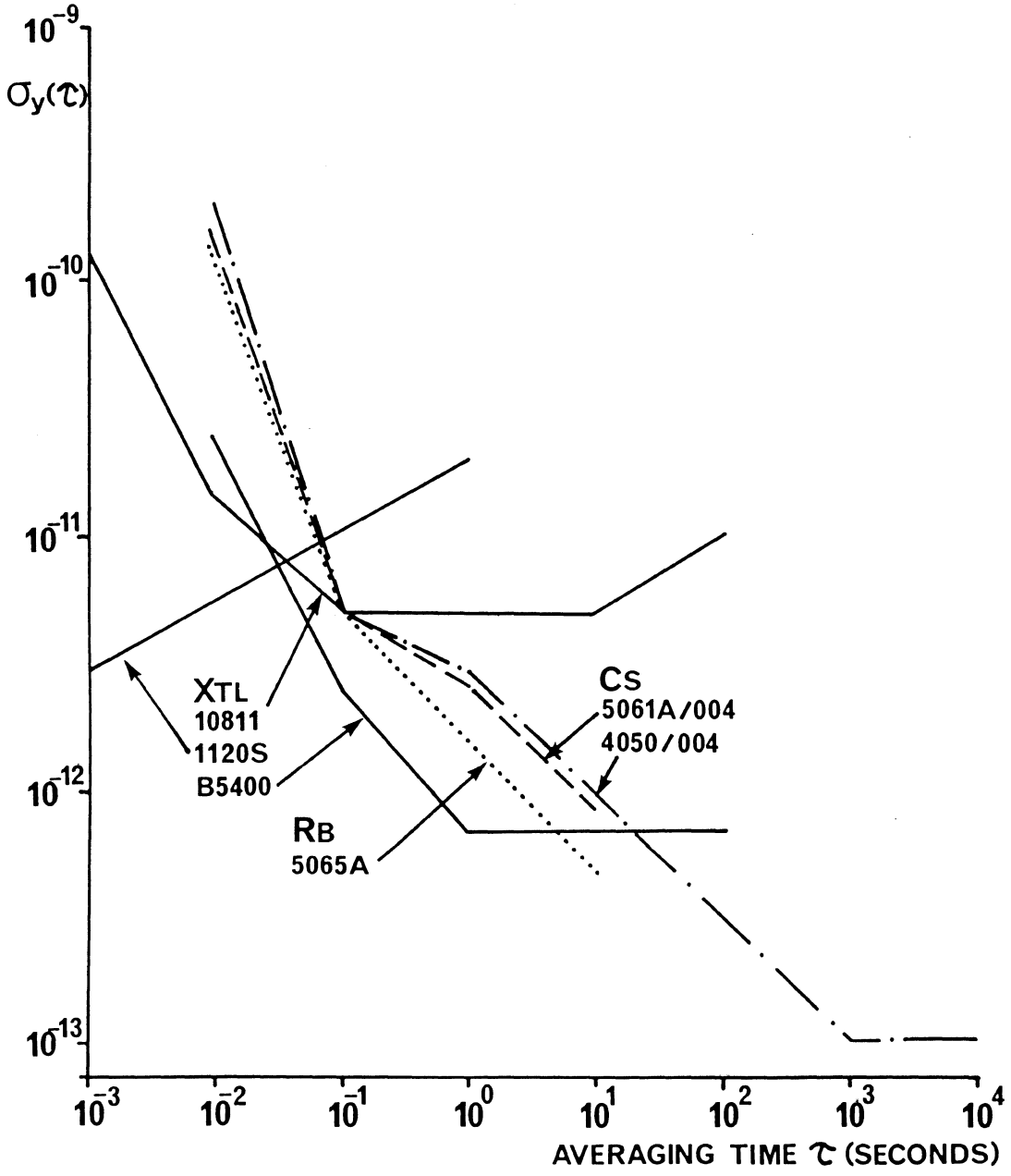


FIGURE 9.1

OFFSET FROM CARRIER f (Hz)	SINGLE SIDEBAND PHASE NOISE-TO-SIGNAL RATIO (1 Hz BANDWIDTH) $\mathcal{L}(f)$ in dB					
	CRYSTAL			RUBIDIUM	CESIUM	
	HP 10811	AUSTRON 1120S	FTS B-5400	HP 5065A	HP 5061A/004	FTS 4050/004
10^{-3}				- 25	- 28	- 28
10^{-2}				- 52	- 48	- 48
10^{-1}				- 72	- 68	- 68
10^{0}	- 90	- 90	-117	- 93	- 96	-100
10^{1}	-120		-140	-120	-120	-130
10^{2}	-140	-165	-144	-126	-125	-140
10^{3}	-157		-144	-140	-140	-140
10^{4}	-160	-180				

TABLE 9.2

OSCILLATOR STABILITY IN FREQUENCY DOMAIN.

OSCILLATOR STABILITY IN FREQUENCY DOMAIN.

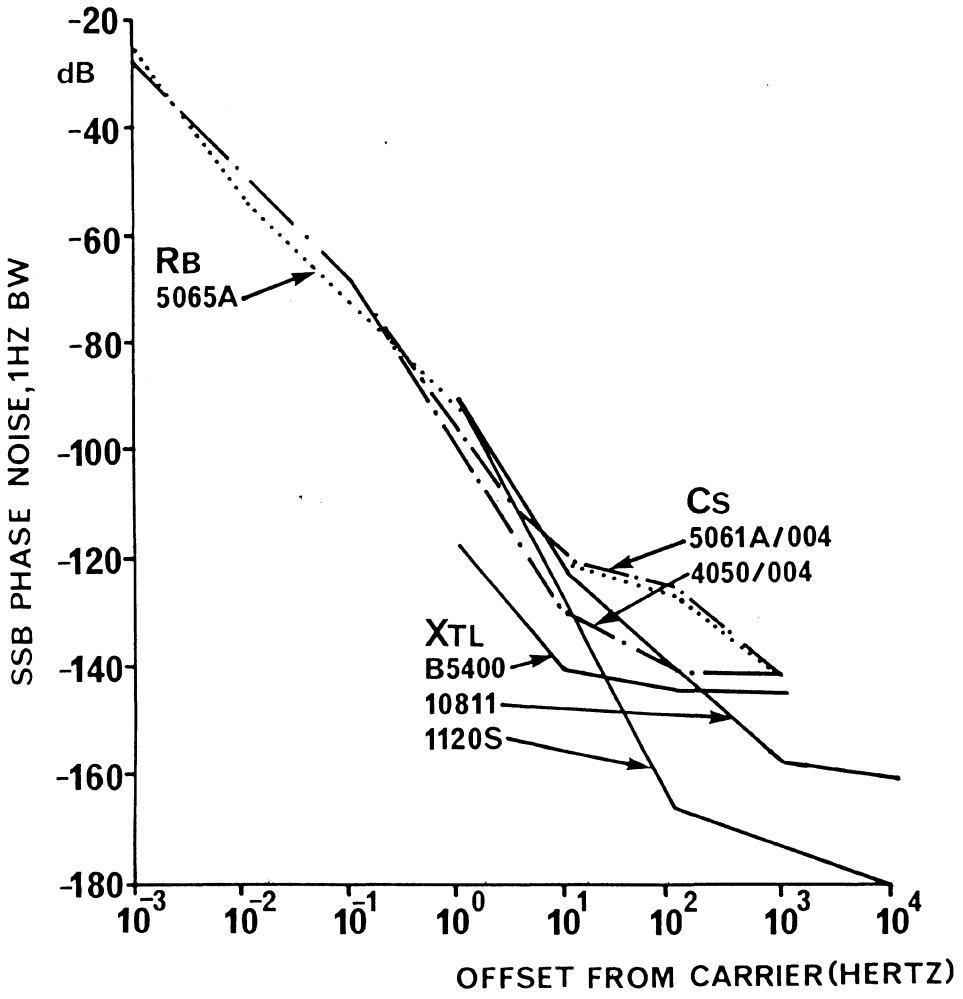


FIGURE 9.2

If we were to use a multichannel pseudo ranging receiver (say 4 channels), then we would obtain enough simultaneous observations to solve for three coordinates of position and to synchronize our clock to the GPS time scale. Other than for purely hardware tracking loop stability (which we will not consider in detail here), there are no particular demands on oscillator stability. However, once we drop back to a single channel receiver, we must collect a time sequence of data from one or more satellites, in order to obtain our clock synchronization. This means that, at least for this data collection period, we must assume either that our receiver oscillator frequency was constant, or perhaps that we can model its systematic behaviour. If we require time τ to collect enough data to establish clock synchronization, and if our unmodelled oscillator fractional frequency offset during this time is $\Delta f/f$, then the error in our clock synchronization, expressed as an equivalent range error, will be

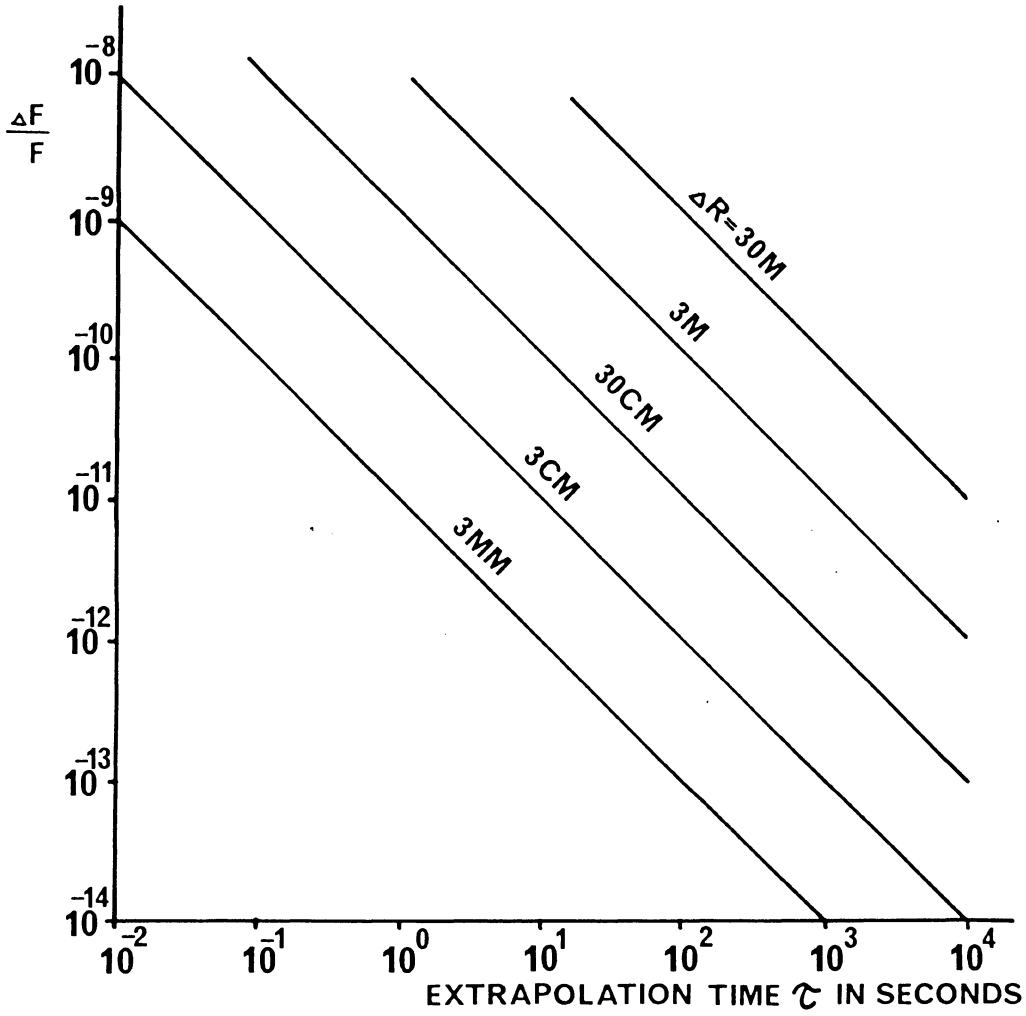
$$\Delta r = c\,\tau\,\Delta f/f \quad , \tag{9.9}$$

where c is the velocity of light. Figure 9.3 is a plot of lines of equal Δr, given τ in the range 10^{-2} to 10^{4} seconds, and $\Delta f/f$ in the range 10^{-9} to 10^{-14}. For example, we note that if we require $\tau = 10^{2}$ seconds, and our $\Delta f/f$ is 10^{-11}, we introduce an oscillator-related error of 30 cm.

Turning to Doppler range-difference measurements, again the Doppler observation equation assumes that the oscillator frequency is either stable or that we can model its behaviour. In this case an unmodelled fractional frequency offset $\Delta f/f$ will result in an error which accumulates over the entire tracking period on one satellite. Hence we again have equation (9.9), only in this case τ is the satellite dwell time and Δr is the equivalent maximum range difference error due to $\Delta f/f$. For dwell times of 10^{3} seconds, if $\Delta f/f$ is 10^{-11}, we experience an error Δr of 3 m.

Finally, with a single channel phase tracking receiver, as we re-acquire each satellite signal, we must be able to predict the phase of that satellite's signal carrier accurately enough to position ourselves in the correct carrier cycle. Since the GPS L_1 signal carrier has a wavelength of 19 cm, we probably need to predict the phase to within a few centimetres. An unmodelled $\Delta f/f$ again satisfies equation (9.9), where Δr is our phase prediction error, and τ is the time taken to cycle through all satellites. If $\Delta f/f = 10^{-11}$, to obtain $\Delta r = 3$ cm requires $\tau \leq 10$ seconds.

Using a similar approach to the above, MacDoran [1979] has shown that



RANGE ERROR AS FUNCTION OF FRACTIONAL FREQUENCY OFFSET AND EXTRAPOLATION TIME

FIGURE 9.3

for the SERIES (VLBI) approach, there are four specific oscillator stability requirements for different dwell time intervals:

τ	$\Delta f/f$
0.25 s	8×10^{-10}
4 s	2×10^{-12}
16 s	6×10^{-11}
100 s	4×10^{-9} .

What we can assume from this simple analysis is that oscillator stability is a critical element in range-difference measurements; that it demands fast satellite switching rates in phase measurements; and that it is less critical in pseudo range and VLBI-mode measurements. We have blindly substituted a systematic bias $\Delta f/f$ for the random characteristic $\sigma_y(\tau)$; have ignored the dependence of oscillator stability on sampling time; and have considered time synchronization prediction, rather than propagation of covariances. Our numbers are not likely reliable, however the conclusions we have been able to draw have in fact been confirmed by the more rigorous analyses cited in the text.

10. TROPOSPHERIC REFRACTION AND MICROWAVE RADIOMETRY

10.1 Tropospheric Refraction

Electromagnetic waves travel through the troposphere at a velocity v which is less than their velocity c through a vacuum. Consequently, electromagnetic measurements yield apparent distances which are slightly greater than the equivalent geometrical (vacuum) distances. To obtain the geometrical distances we must know the index of refraction $n = c/v$ (or equivalently the refractivity $N = (n-1) \times 10^6$). However, the electromagnetic wave velocity v, and hence the index of refraction n, is not a constant. It depends on the composition of the troposphere, its temperature and pressure, and the frequency of the electromagnetic wave.

The troposphere contains over a dozen gasses in measurable quantities. It has been found, however, that the refractive behaviour of most of these are similar, and can be modelled as if there was only one constituent. The two exceptions are carbon dioxide and water vapour. For expediency it is usual to assume that carbon dioxide makes up 0.03% of the troposphere, and to include its effect with the remaining gasses. Thus the usual models for refractive index include only two components, the "dry" component (everything except water vapour), and the "wet" component (water vapour). The effect of each component depends on its abundance (partial pressure), and on temperature.

As we shall see below, above about 20 GHz refraction depends on frequency as well. Considering, for the moment, only frequencies much lower than 20 GHz, the standard procedure for correcting electromagnetic distances for tropospheric refraction is to obtain point measurements or estimates of total atmospheric pressure, the partial pressure of water vapour, and atmospheric temperature, and to obtain n from an empirical formula (e.g., Essen and Froome [1951] or Smith and Weintraub [1953]). For infrared and visible frequencies, the dependence of refractive index on frequency must be accounted for [Edlen, 1966; Barrell and Sears, 1939].

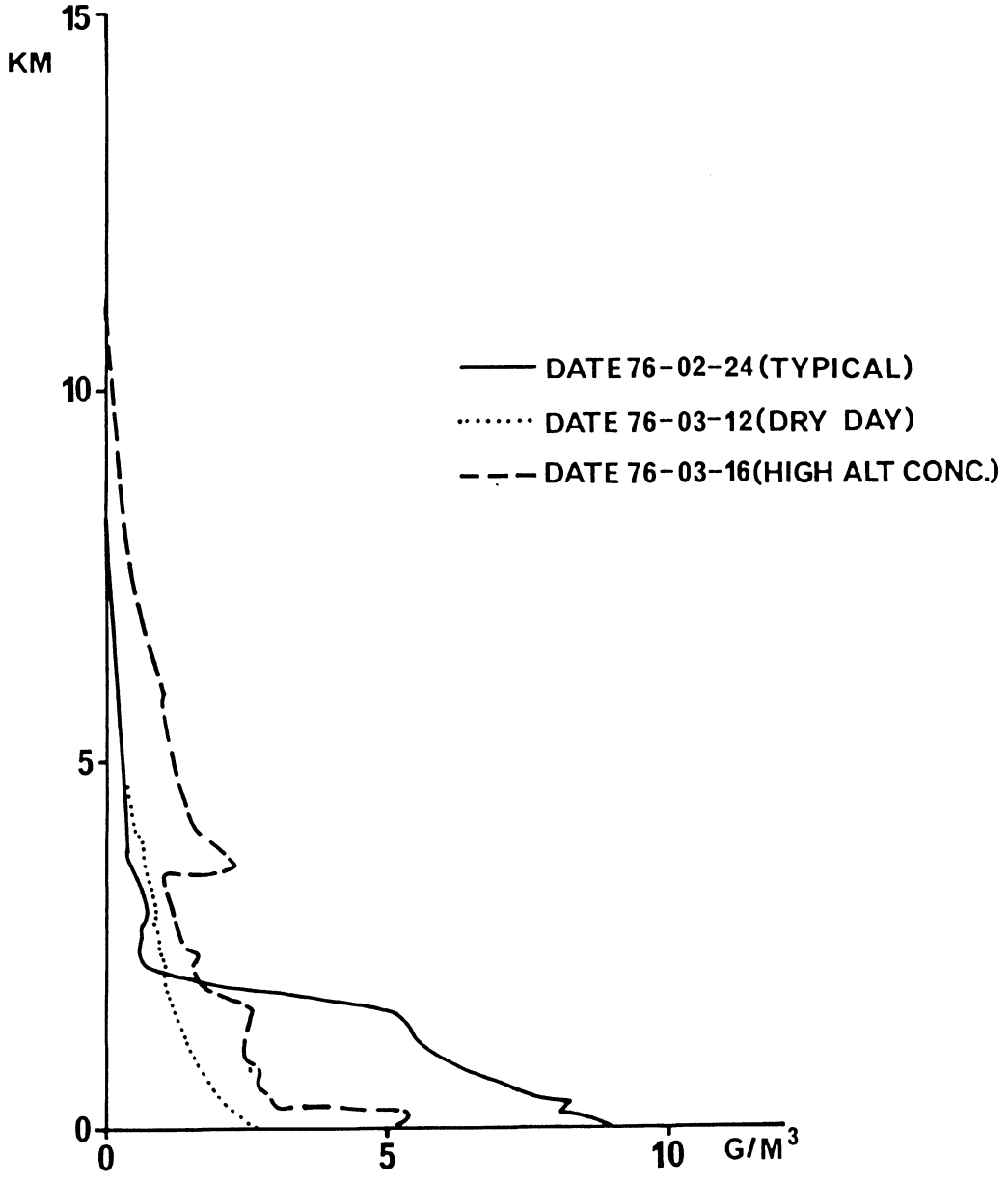
10.2 Variations in Refractive Index

What we have considered so far are techniques for estimating the refractive index at a point (for which we know pressure and temperature). However, the troposphere is not homogeneous in temperature, pressure, and

water vapour content. The three predominant features of the nonhomogeneity are relatively stable vertical profiles of pressure and temperature; variable horizontal profiles in pressure and temperature; and the variability in time, horizontal profiles, and vertical profiles of water vapour content. Figure 10.1 illustrates some time variation of the vertical profile of water vapour content (from Wu [1979]). This figure typifies what we know most about, concerning variations in water vapour content. Radiosonde data, shown here, give us information about vertical profile variations between samples separated by several hours. Not much is known, however, about the small scale high frequency (periods of minutes) variations in vertical profile. Little also is known about horizontal profile variations, since the only relevant data are radiosonde vertical profiles from different weather stations (typically a few hundred kilometres or more apart).

When we make distance measurements through the troposphere, we are interested in the profile of refractive index (or more simply its integral) along the wave path. For terrestrial measurements, point values of the refractive index can be obtained along the length of this line, or at least at its end points. For satellite measurements, this line traverses the entire tropospheric layer. Measurements made when the satellite is at zenith require the vertical profile of refractive index. Measurements made off zenith require horizontal profiles as well. Because these are not easily measured, the approach has been to use models for the vertical profile, based on archived profiles measured principally with radiosondes, and to treat the horizontal profiles as homogeneous. Because of the stable vertical profiles of pressure and temperature, such models, anchored by surface measurements of pressure and temperature, have proven to be remarkably good for the dry component of refraction [Hopfield, 1980]. However, the variability of water vapour has limited the performance of such models. For lack of a better approach, it has been usual to use the same shape for the vertical profile of the "wet" component of refraction as for the "dry" component, changing only the scale, and the effective thickness of the troposphere [Hopfield, 1980]. Since water vapour is a relatively minor constituent of the troposphere, its total effect on refraction is about 10% of that of the dry component. For extraterrestrial range measurement precisions of worse than about 10 cm, the error due to water vapour

FIGURE 10.1



VERTICAL PROFILES OF TROPOSPHERIC WATER VAPOUR CONTENT AT POINT MUGU, CAL.

refraction using the above approach is not significant. However, at the 1 cm level, it is significant.

10.3 Tropospheric Refraction Frequency Dependence

To consider the effect on refraction of the electromagnetic frequency, we must consider the process from the molecular viewpoint. Any propagation medium other than a vacuum will be made up of molecules with which the electromagnetic waves will interact, in one of two ways. In the first case (non-resonant interaction), the effect is simply to dissipate and retard the electromagnetic wave, resulting in attenuation of the signal, and the slower effective propagation velocity v. In considering the second case (resonant interaction), we must take into account the quantum nature of the energy associated with each type of molecule. We take the water vapour molecule as an example.

Part of the energy associated with the water vapour molecule is explained by describing this molecule as an asymmetric rotor (like a complicated gyroscope) with no two of its three principal moments of inertia equal. This rotor is restricted by quantum theory to rotate with only a few discrete values of rotational angular momentum. Each of these values has a unique energy level associated with it. A transition (or jump) between different rotation states thus is exhibited by a jump in energy level. According to Planck's law, $E = h\nu$, each change in energy E is related to an associated frequency ν through Planck's constant h. This energy change must be supplied from outside the molecule (if the molecule jumps to a higher energy state), or must be somehow disposed of by the molecule (if the jump is to a lower energy state). One mechanism for these energy transfers is by means of electromagnetic energy. Therefore if an electromagnetic wave has a frequency which matches that of a specific energy level transition in a molecule, part of the wave energy may be absorbed by the molecule, resulting in a jump to a higher molecular energy level. Similarly, a jump to a lower energy level may result in the emission of electromagnetic waves whose frequency matches that of the molecular transition. This is the resonant interaction between electromagnetic waves and the molecules of the propagation medium. Near one of these resonant frequencies, the attenuation and delay of the electromagnetic waves varies strongly with frequency, and dispersion (the variation of refractive index with frequency) occurs.

In the troposphere, the lowest resonant frequency is due to one of the water vapour molecule rotational energy state transitions mentioned above. It occurs at a frequency of 22.235 15 GHz. The next lowest frequency resonance is due to several closely-spaced transitions in the oxygen molecule, centred at about 60 GHz. Figure 10.2 (from Straiton [1975]) is a plot of oxygen and water vapour attenuation of electromagnetic signals, as a function of frequency, in the range 10 to 400 GHz. A plot of refractive index variation over this frequency range would have a similar structure. The radio frequencies used for distance measurement are deliberately kept well below these resonances (GPS for example uses 1.2 and 1.6 GHz). These resonances are deliberately used, however, to measure water vapour content by the microwave radiometer technique, as we shall see below.

10.4 Thermal Radiation

Radiometers are instruments for measuring radiation. The microwave radiometers we are concerned with measure a particular kind of radiation called thermal radiation, emitted in the microwave region of the electromagnetic spectrum. Electromagnetic radiation may be produced by a substance in a number of ways, such as γ-ray radioactivity, flourescence, emission of radio waves by an AC conductor, and the thermal emission of electromagnetic radiation by all substances whose temperature is above absolute zero (^{o}K) [Jakob, 1949]. Thermal emission is isotropic. All substances also absorb thermal radiation. An idealized substance which absorbes all thermal radiation incident upon it (a perfect absorber) is called a "black body". Two important laws governing the thermal emission properties of these perfect absorbers are that

(a) the total amount of radiant energy emitted by a black body is proportional to the fourth power of its absolute temperature (^{o}K) (Stefan-Boltzmann law);

(b) the frequency spectrum of the radiant energy emitted by a black body has a temperature dependent distribution function given by Planck's equation.

No physical substance is a perfect black body. However, the thermal characteristics of any substance can be described in terms of the temperature of a black body which would give the same total intensity of thermal emission as the substance. This is called the "brightness

temperature" of the substance. Any substance above absolute zero will emit thermal radiation, and will have a brightness temperature.

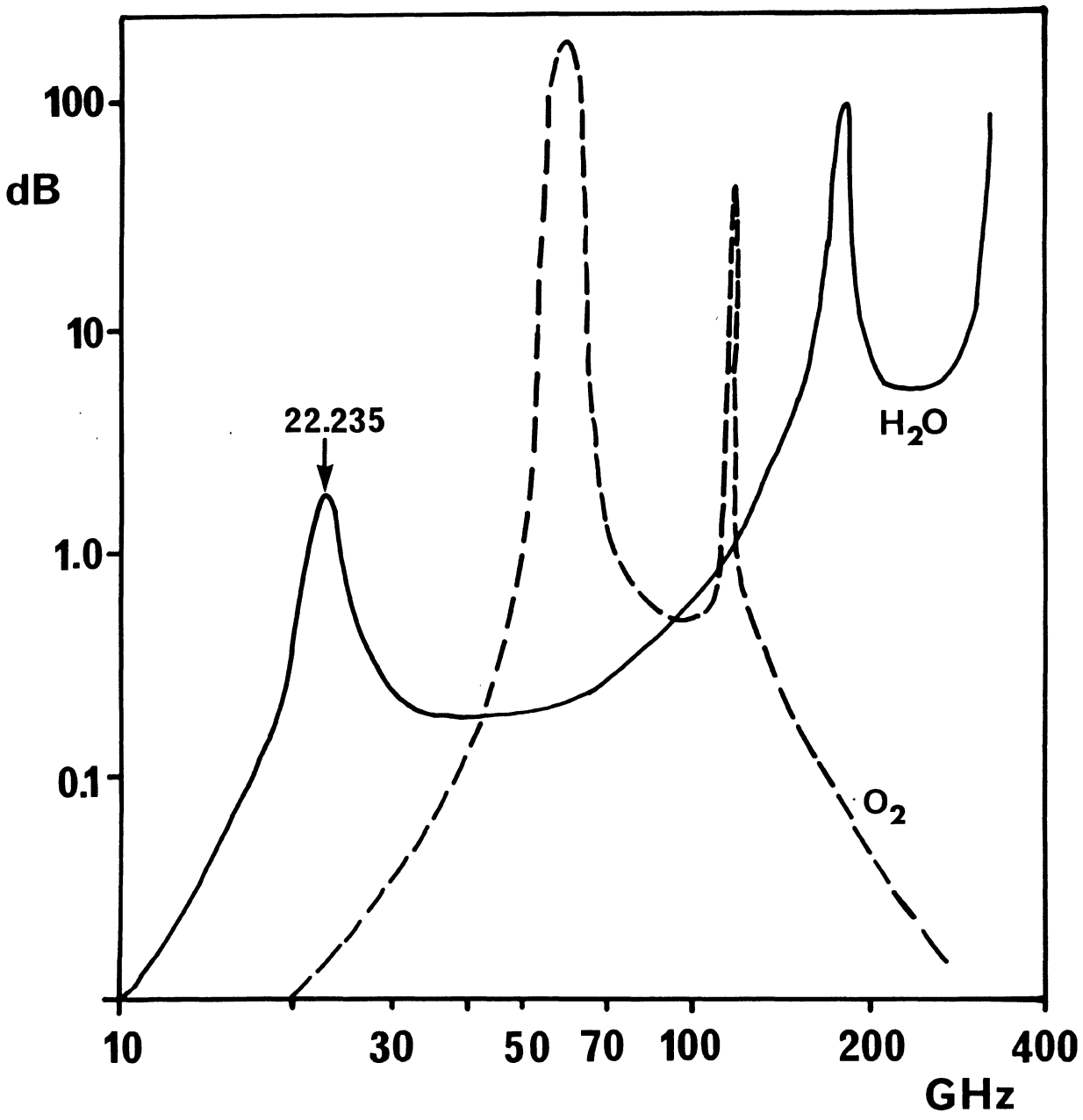
At any point in the troposphere, there are several thermal radiation processes underway. Each tropospheric point will have an associated brightness temperature and can be considered as a thermal absorber/emitter. At each point then, radiation will be emitted and incoming radiation will be absorbed from cosmic radiators (echo of the "big bang"), and from other radiating points in the troposphere. Each of these processes has a characteristic frequency spectrum.

10.5 Microwave Radiometers

A microwave radiometer consists of a directional antenna pointed at the sky, connected to a microwave receiver which measures the thermal radiation intercepted by the antenna. Since the antenna itself is absorbing (from the sky and elsewhere) and emitting (to the receiver) thermal radiation, its thermal characteristics at any instant can be characterized by a brightness temperature, called the antenna temperature. The thermal radiation intercepted by the antenna will be from all layers of radiators in the atmosphere, along the line of sight of the antenna pointing. By tuning the microwave receiver to different frequencies (or by using several fixed-frequency receivers) various points on the frequency spectrum associated with the antenna temperature can be mapped. This spectrum has been related theoretically to the combination of the spectra of the continuum of thermal radiators in the troposphere which influence the antenna temperature [Dicke et al., 1946].

When the radiometer is tuned to frequencies that lie within the attenuation spectra of Figure 10.2, there will be two different processes at work--the attenuation due to the molecular processes of Figure 10.2, and thermal absorption and reradiation. Since the antenna temperature due to the thermal processes can be predicted, the actual antenna temperature becomes a measure of the attenuation due to the molecular processes, which in turn depend on the density profile of the molecules involved along the line of sight of the antenna pointing. Thus by sampling in the vicinity of the 22 GHz water vapour resonance, the integral of the water vapour content of the troposphere is measured. This is usually expressed in terms of equivalent depth of liquid water (total precipitable water vapour content),

FIGURE 10.2



MICROWAVE ABSORPTION SPECTRA FOR WATER VAPOUR AND OXYGEN

or, more appropriately for our purposes, in terms of the radio range correction--which is a factor of about six times the equivalent water depth.

10.6 Radiometer Calibration

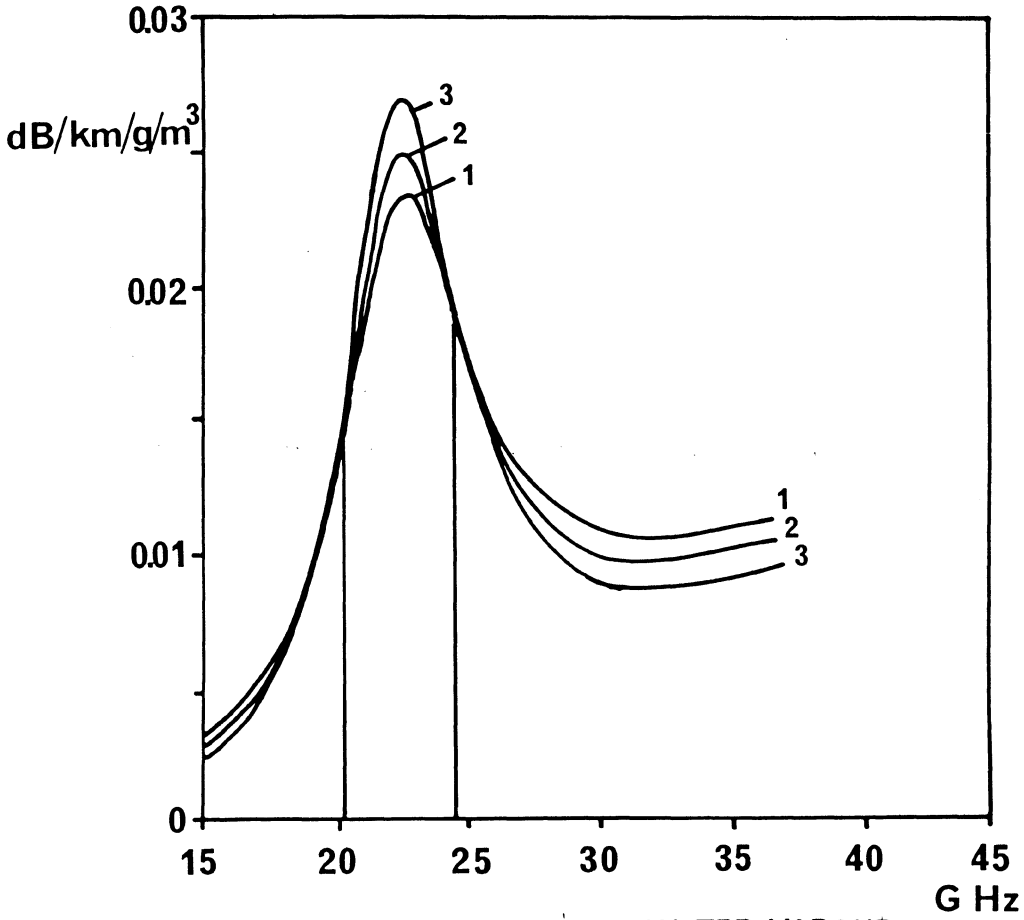
Several calibration procedures are used with microwave radiometers. First, the receiver is switched between sampling the antenna and sampling one or more waveguide terminations which are heated to a constant known temperature by means of a heating coil, and are called "hot loads" or "cold loads" depending on their temperature.

Second, the variation in the antenna temperature with zenith distance is used as a measure of the fraction of the radiation absorbed by the atmosphere in the zenith direction. However, this calibration technique (called the "tipping curve" technique) is based on the assumptions that the brightness temperature of the atmosphere is homogeneous and isotropic, and that the absorption properties of the atmosphere are isotropic in the horizontal direction. These assumptions are not consistent with the use of the tipping curve calibration technique, when we want to study horizontal variabilities.

Selection of the precise microwave frequencies to sample, in order to optimize the water vapour content measuring performance of the radiometer, has been extensively studied [Westwater, 1978; Guiraud et al., 1979; Wu, 1979]. The height and shape of the attenuation peaks shown in Figure 10.2 are functions of pressure, temperature, and atmospheric constituents present. This means that at various heights the absorption peaks will have different shapes. Figure 10.3 (taken from Westwater [1978]) shows the expected variation in the 22 GHz peak shape, indicating that at 20.6 GHz, the attenuation is independent of variations of the peak shape.

Liquid water has a different molecular structure than water vapour, and therefore a different absorption spectrum than that shown in Figure 10.2. Liquid water also does not influence the "wet" component of tropospheric refraction. However, liquid water present in the form of clouds and rainfall will contaminate the thermal radiation part of the measurement, leading to errors in the determination of water vapour content. To calibrate the effect of liquid water alone, a second frequency well off the 22 GHz peak is sampled. The valley between the 22 GHz water vapour peak and the 60 GHz oxygen peak, at 31.4 GHz, is used.

FIGURE 10.3



VARIATIONS IN 22 GHz WATER VAPOUR
SPECTRAL PEAK

Absolute calibration of the water vapour content measured by microwave radiometers against external standards is a difficult problem. Radiosonde measurements are at zenith only, are not very accurate, and do not sample small scale and high frequency variations. Refractometers give point values only, not profile integrals. Laser radar (LIDAR) does not go high enough. The best prospect is multifrequency ranging to aircraft at about 5 km height. As Figure 10.1 shows, most of the variations in water vapour content occur below this height, and the amount of water vapour above that height can be reliably extrapolated from below.

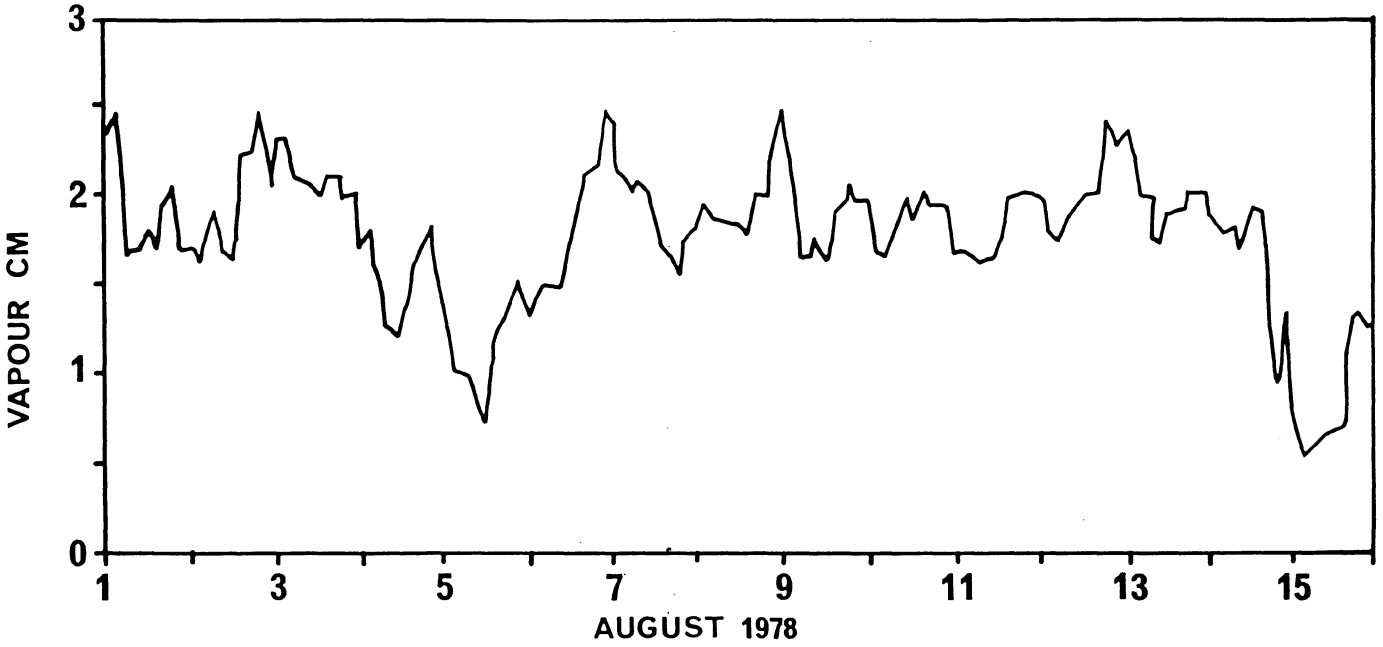
10.7 Status

Two groups have developed microwave radiometers capable of being applied to geodetic work. Radiometers developed at the Jet Propulsion Laboratory, California Institute of Technology, are scheduled for installation at several VLBI stations. These radiometers are estimated to measure water-vapour radio frequency delay effect with an accuracy of 2 cm for elevations above 17° [Resch and Claflin, 1980].

Radiometers developed at the Environmental Radiometry, Wave Propagation Laboratory, NOAA/Environmental Research Laboratories, Boulder, Colorado, have been in use for meteorological research for some years. Results from these instruments are shown in Figures 10.4 and 10.5 [Guiraud et al., 1979], which show typical short term (over 10 hours) and long term (over 15 days) variations in zenith water vapour content at Boulder, which is not a location noted for its humid climate. The parameter shown is equivalent water depth. Expressed in terms of zenith radio ranging error, these figures show a maximum of 15 cm, a minimum of 3 cm, and maximum rates of change of 1 cm in 10 minutes and 10 cm in 8 hours. At a zenith distance of 70° (elevation 20°), assuming horizontal isotropy, these ranging errors would be a factor of 3 larger.

Construction of a radiometer of the NOAA design would involve about $30 000 in microwave parts, and very careful construction, costing an additional $100 000 [Bender, private communication]. The instrument is in the "transportable" class, requiring a housing about the size of a travel trailer.

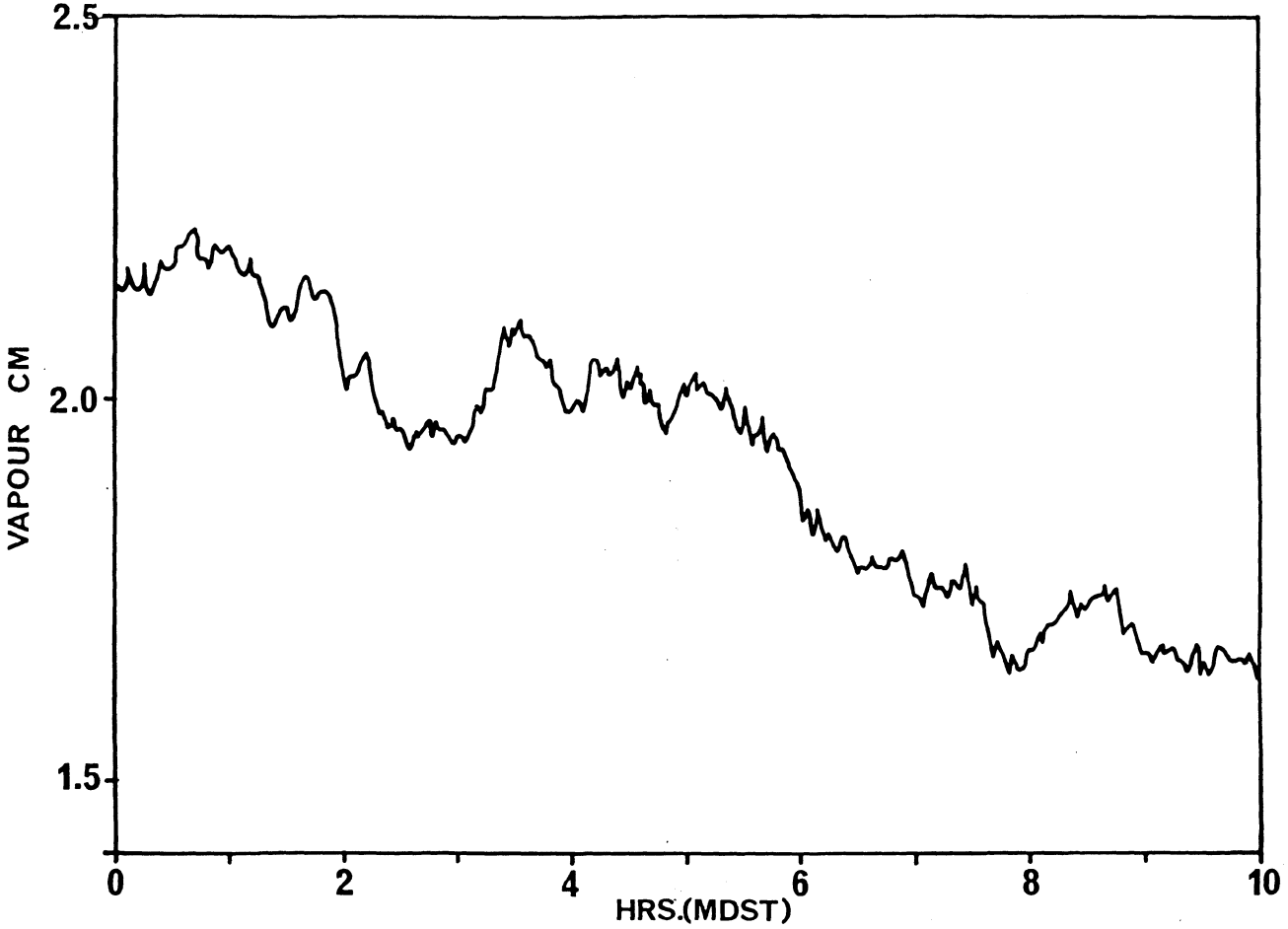
FIGURE 10.4



TOTAL ZENITH PRECIPITABLE WATER VAPOUR CONTENT AT BOULDER

FOR AUGUST 1-16,1978

FIGURE 10.5



TOTAL ZENITH PRECIPITABLE WATER VAPOUR CONTENT AT BOULDER FOR AUGUST 29, 1978

10.8 Recommendations

Microwave water vapour radiometers represent a potentially very significant breakthrough in solving the outstanding problem in modelling the tropospheric refraction effect on radio signals from space. However there are three areas of concern, in which it is recommended that further research be pursued:

(a) VLBI base line determinations over 4000 km distance have shown repeatabilities of 4 cm over 2 years without using microwave radiometers [Ryan et al., 1978]. It has also been demonstrated that local horizontal coordinate determinations are little affected by the errors neglected when radiometers are not used [Resch and Claflin, 1980]. Yet the limited water vapour content variability data available from radiometers indicates that significant effects should be evident from ignoring such variations. Does this inconsistency reflect our lack of understanding of what the radiometry is actually measuring; or does it indicate that the variabilities in water vapour between well separated VLBI sites are somehow averaged out in the VLBI measurement process leaving no appreciable bias?

(b) The radiometric technique is a very complex measurement, requiring expensive unique equipment, sophisticated calibration procedures, and operators who are well grounded in the theory behind the measurements. It is not clear how best to transfer this technology to the user community--whether regional permanent radiometric monitor sites would be sufficient (perhaps in conjunction with meteorological work), or whether a radiometer attachment to each GPS receiver is necessary (or even feasible). We need to know more about both the time and spatial variability of water vapour.

(c) The calibration of radiometers, both in terms of the tipping curve technique, and against external standards, needs further study.

11. CONCLUSIONS AND RECOMMENDATIONS

The prospect of GPS positioning in a differential mode being significantly better than the TRANSIT translocation results is very good. Thus, very probably, GPS will replace the TRANSIT system for geodetic positioning in about the year 1995. Unfortunately, GPS, including the receivers, will not be in place in time to have any effect on the redefinition of horizontal networks.

Before GPS can completely replace TRANSIT, a series of tests have to be conducted to

(a) determine if the TRANSIT and GPS coordinate systems are compatible, and

(b) establish possible biases that may exist in the GPS differential mode.

It is equally clear that GPS will not be able to provide useful information either about sea surface topography or for the levelling network in general. This is because of the unavailability of accurate enough geoidal heights.

There are only very incomplete data available on the onboard oscillators' stability. It would thus be premature to draw any final conclusions about the oscillators' performance. The only conclusion we can offer at this stage is that the selected switching rate and dwell shall have to reflect this performance to obtain the best possible results.

Our numerical simulations of GPS point positioning have shown that pseudo range measurements have inherently better geometrical strength than range difference (Doppler) measurements. A combination of both ranges and range differences can be used.

For pseudo range measurements, shorter switching intervals (e.g., 5 minutes) from satellite to satellite provide more accurate results than the longer intervals (e.g., 1 hour). The situation is completely reversed for range differences, whose geometry is weak for shorter switching intervals.

There is no appreciable difference in positioning accuracy achievable using any of the suggested constellations of either 18 or 24 satellites. Curiously, if all of the present six satellites were available, the results obtained for periods when several satellites are visible simultaneously would have an accuracy comparable to that obtained from the full set.

It seems clear to us that steps should be taken to ensure that GPS can

replace levelling on a regional scale in the hinterland areas toward the end of this century. The main obstacle to this deployment is again the lack of an accurate enough knowledge of the geoid, which the GPS can also help to get. More research into this topic is needed.

GPS should become the main tool for studying regional crustal dynamics. To do this, long term stability of GPS point position and relative position determinations should be tested first. That will, of course, require years to do and thus the tests should start as soon as possible, complemented by theoretical research on the optimization of station separations, and measurement simultanaity, interval and type.

The performance of GPS in conjunction with radio-astronomy, inertial positioning, and standard terrestrial techniques should be tested practically. There are indications that by properly combining these techniques the propagation of errors can be checked; optimum configurations for these should be determined.

We feel that the problem of the cycle ambiguity in the reconstructed carrier phase approach has not been unequivocally resolved. There may exist other techniques that should be thoroughly investigated. The optimal satellite configurations for differential positioning should be also studied.

More research is needed to establish if a water vapour microwave radiometer is likely to improve significantly the GPS performance in the differential mode. Radio-astronomy appears to have achieved very good repeatability without the radiometers. In addition, the radiometers at present are very complex and expensive.

Finally, we feel we must emphasize the point that the "VLBI mode" of GPS differential positioning appears at present to be the most promising one in light of the denial of the P-code and probable degradation of the system contemplated by the U.S. military authorities. Thus high priority should be placed on further research into this mode.

REFERENCES

Anderle, R.J. (1980). Application of NAVSTAR GPS geodetic receiver to geodesy and geodynamics. NSWC TR 80-282, Dahlgren, Va.

Barnes, J.A., A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot, and G.M.R. Winkler (1971). "Characterization of frequency stability". IEEE Transactions on Instrumentation and measurement, IM-20, pp. 105-120.

Barrell, H. and J.E. Sears (1939). "The refraction and display of air for the visible spectrum". Royal Society of London, Philosophical Transactions, A-238, pp. 1-64.

Bender, P. (1980). Private communication.

Blaha, G. (1971). Investigations of critical configurations for fundamental networks. Department of Geodetic Science Report 150, Ohio State University, Columbus, Ohio.

Bogon, A.H. (1974). Geometric performance of the global positioning system. Aerospace Corporation, AD-783210, June.

Book, S.A., W.F. Brady, and P.K. Mazaika (1980). "The non-uniform GPS constellation". IEEE Plans 80, pp. 1-9.

Bossler, J.D., C.C. Goad, and P.L. Bender (1981). "Using the GPS for geodetic positioning". Bulletin Geodesique, v. 54, n. 4, pp. 553-563.

Brady, W.F. and P.S. Jorgensen (1981). Worldwide coverage of the Phase II NAVSTAR satellite constellation. Presented at U.S. Institute of Navigation National Aerospace meeting, Philadelphia.

Canadian Marconi Company (1981). Final report on the study of satellite signal processing techniques applicable to GPS equipment. CMA Document No. 0680-1001, D017/D017A-81045, Montreal, June.

Collins Government Avionics Division (1981). Feasibility concept for a translocation survey GPS receiver design. Contract report NA-80-SAC-00670.

Counselman, C.C. and I.I. Shapiro (1979). "Miniature interferometric terminals for earth surveying". Proceedings of the Second International Geodetic Symposium on Doppler Satellite Positioning, Austin.

Counselman, C.C., I.I. Shapiro, R.L. Greenspan, and D.B. Cox (1980). Backpack VLBI terminal with subcentimetre capability. NASA CP-2115, pp. 409-414.

Dicke, R.H., R. Beringer, R.L. Kyhl, and A.B. Vane (1946). "Atmospheric absorption measurements with a microwave radiometer". Physical Review, 70, pp. 340-348.

Edlen, B. (1966). "The refractive index of air". Metrologia, 2, p. 71.

Essen, L. and K.D. Froome (1951). "The refractive indices and dielectric constants of air and its principal constituents at 24,000 mc/s". Proceedings of the Physical Society of London, B-64, pp. 862-875.

Fell, P.J. (1980a). Geodetic positioning using a global positioning system of satellites. Department of Geodetic Science Report 299, Ohio State University, Columbus, Ohio.

Fell, P.J. (1980b). "An analytic technique for statistically modelling random atomic clock errors in estimation". Proceedings, Twelfth Annual Precise Time and Time Interval Applications and Planning Meeting, PTTI Conference, Goddard Space Flight Center, December.

Fell, P.J. (1980c). "Geodetic positioning using a global positioning system of satellites". Proceedings, IEEE Plans 80, pp. 42-47.

Gardner, C.S. (1976). "Effects of horizontal refractivity gradients on the accuracy of laser ranging to satellites". Radio Science, 11, pp. 1037-1044.

Gardner, C.S. (1977). "Correction of laser tracking data for the effects of horizontal refractivity gradients". Applied Optics, 16, p. 2427.

Gardner, F.M. (1979). Phaselock Techniques. Wiley, Toronto.

Gilbert, G.A., E.H. Martin, D. Symes and C. Mathews (1979). "Civil applications of NAVSTAR/GPS", In: Principles and Operational Aspects of Precision Positioning Systems, Ed., C.T. Leondes, NATA AGARDograph No. 245.

Gold, R. (1967). "Optimal binary sequences for spread spectrum multiplexing". IEEE Trans. on Info. Theory, pp. 615-621.

Guiraud, F.O., J. Howard, and D.C. Hogg (1979). "A dual-channel microwave radiometer for measurement of precipitable water vapor and liquid". IEEE Trans. on Geoscience Electronics, GE-17, pp. 129-136.

Harris, R.L. (1973). "Introduction to spread spectrum techniques", In: Spread Spectrum Communications, NATO AGARD Lecture Series No. 58.

Hopfield, H. (1980). "Improvements in the tropospheric refraction correction for range measurements". Proceedings of the Symposium on Satellite Doppler Tracking and its Geodetic Applications, The Royal Society of London, Oct. 10-11, 1979.

Jakob, M (1949). Heat transfer. Wiley, New York.

Jorgensen, P.S. (1980a). "NAVSTAR/global positioning system 18 - Satellite constellations". Navigation, v. 27, n. 2, pp. 89-100.

Jorgensen, P.S. (1980b). "Combined pseudorange and Doppler position for the stationary NAVSTAR user". IEEE Plans 80, pp. 450-458.

Kouba, J. (1978). Geodetic satellite Doppler positioning and application to Canadian test adjustment. Collected papers 1978, Geodetic Survey of Canada, Surveys and Mapping Branch, Ottawa.

Krause, H.G.L. (1963). "Relativistic perturbation theory of an artificial satellite in an arbitrary orbit about the rotating oblated earth spheroid and the time dilation effect for this satellite", In: The Use of Artificial Satellites for Geodesy, Ed. G. Veis, North Holland, pp. 69-107.

Larden, D.R. and P.L. Bender (1981). Expected accuracy of geodetic baseline determinations using the GPS reconstructed carrier phase method. (In preparation.)

MacDoran, P.F. (1979). "Satellite emission ratio-interferometric earth surveying SERIES-GPS geodetic system". Bulletin Geodesique, 53, pp. 117-138.

Martin, E.H. (1978). "GPS user equipment error models". Navigation, Vol. 25, pp. 201-210.

Milliken, R.J. and C.J. Zoller (1978). "Principle of operation of NAVSTAR and system characteristics". Navigation, v. 25, pp. 95-106.

Nyland, E. (1977). "Repeated geodetic surveys as experiments in geophysics". Canadian Surveyor, v. 31, n. 4, pp. 347-359.

Putkovich, K. (1980). "USNO GPS program". Proceedings, Twelfth Annual Precise Time and Time Interval Applications and Planning Meeting, PTTI Conference, Goddard Space Flight Center, December.

Reilly, W.I. (1980). Determination of geometric and gravimetric earth deformation parameters from geodetic observations. Geophysics Division report 140, Department of Scientific and Industrial Research, New Zealand.

Resch, G.M. and E.S. Claflin (1980). Microwave radiometry as a tool to calibrate tropospheric water-vapor delay. NASA CP-2115, pp. 377-384.

Russell, S.S. and J.H. Schaibly (1978). "Control segment and user performance". Navigation, v. 25, n. 2, pp. 74-80.

Ryan, Clark et al. (1978). "Precision surveying using radio interferometry". Journal of the Surveys and Mapping Division, American Society of Civil Engineers, v. 104, pp. 25-34.

Shaibly, J.H. (1976). Simulated and projected performance of the NAVSTAR/GPS control segment. Presented at the AIAA Guidance and Control Conference, San Diego.

Schaibly J.H. and M.D. Harkins (1979). "The NAVSTAR Global Positioning System control segment performance during 1978". Proceedings of the Second International Geodetic Symposium on Satelliite Doppler Positioning, Austin.

Smith, E.K. and S. Weintraub (1953). "The constants in the equation for atmospheric refractive index at radio frequencies". Proceedings of the Institute of Radio Engineers, 41, pp. 1035-1037.

Spilker, J.J. (1978). "GPS signal structure and performance characteristics". Navigation, v. 25, n. 2, pp. 29-54.

Stansell, T.A. (1978). The TRANSIT navigation satellite system. Magnavox, October.

Straiton, A.W. (1975). "The absorption and reradiation of radio waves by oxygen and water vapor in the atmosphere". IEEE Transactions on Antennas and Propagation, TAP-23, pp. 595-597.

Tennant, D.M. (1980). "NAVSTAR Global Positioning System (GPS) clock program: present and future". Proceedings of the Twelfth Annual Precise Time and Time Interval Application and Planning Meeting, NASA conference publication 2175, pp. 703-718.

U.S. DOD/DOT (1980). Federal radionavigation plan. Four volumes. U.S. Departments of Defense and Transportation. (NTIS accession numbers ADA 093774-7.)

U.S. National Committee on Geodesy (1980). Geodetic research and development in the ocean survey. National Academy Press, Washington, D.C.

van Dierendonck, A.J., S.S. Russell, E.R. Kopitzke, and M. Birnbaum (1978). "The GPS navigation message". Navigation, v. 25, n. 2, pp. 55-73.

Vaníček, P. (1980). "Heights based on observed gravity". Proceedings of the NAD Symposium, Ottawa, 26-30 May.

Vaníček, P. and E.J. Krakiwsky (1982). Geodesy: the concepts. North-Holland, Amsterdam.

Walker, J.G. (1977). Continuous whole-earth coverage by circular-orbit satellite patterns. TR77044, Royal Aircraft Establishment, Farnborough, U.K.

Ward, P. (1980). "An advanced NAVSTAR/GPS multiplex receiver". Proceedings, IEEE Plans 80, pp. 51-58.

Waters, J.W. (1976). "Absorption and emission by atmospheric gases". Chapter 2.3 in Vol. 12-B Astrophysics-Radio Telescopes, Ed. M.L. Meeks, in the series Methods of Experimental Physics, pp. 142-176.

Wells, D.E. (1975). Doppler satellite control. Department of Surveying Engineering Technical Report 29, University of New Brunswick, Fredericton, N.B.

Wells, D.E. (1976). "Concept of satellite Doppler positioning using translocation techniques". Proceedings of the First International Symposium on Satellite Doppler Positioning, Las Cruces, New Mexico, pp. 77-96.

Westwater, E.R. (1978). "The accuracy of water vapor and cloud liquid determination by dual-frequency ground-based microwave radiometry". Radio Science, v. 13, pp. 677-685.

Westwater, E.R. and F.O. Guiraud (1980). "Ground-based microwave radiometric retrieval of precipitable water vapor in the presence of clouds with high liquid content". Radio Science, v. 15, pp. 947-957.

Wu, S.C. (1979). "Optimum frequencies of a passive microwave radiometer for tropospheric path-length correction". IEEE Transactions on Antennas and Propagation, AP-27, pp. 233-239.