GPS TIME CORRELATION AND ITS IMPLICATION FOR PRECISE NAVIGATION

WILLIAM D S ROBERTS

BSc (Hons) Surveying Science

Thesis submitted for the Degree of

Doctor of Philosophy

Department of Surveying

University of Newcastle upon Tyne

June 1993

ABSTRACT

The Global Positioning System (GPS) is a satellite navigation system which, when fully operational, "will provide highly accurate position and velocity information in three dimensions, as well as precise time, to users around the globe 24 hours a day" (Anon, 1990). GPS can be operated under all weather conditions, with the only restriction being that the user must be able to receive radio signals from the satellites. Such a comprehensive positioning system has never been previously available, and thus GPS is currently being used for a diverse range of applications.

This thesis is focused at the application of GPS for the offshore oil industry which is requiring increasingly higher instantaneous positioning accuracies. The GPS system and real-time positioning techniques are described, along with the main error sources that limit the available accuracy. The suitability of using GPS observations in a standard set of mathematical algorithms, the Kalman filter, in order to obtain position and velocity information has been examined. This is carried out by analysing the observations in order to determine some statistical properties that are usually ignored during the processing and data spanning a two year period has been analysed. The effect that these properties have on the resultant position and its precision was ascertained, finding that position discrepancies were insignificant but their associated precisions were highly dependent on the statistical properties were highly dependent of the data sets. Along similar lines, the ability of the Kalman filter to detect blunders, or gross errors, within GPS-type observations was analysed showing that the relevant test statistics performed sub-optimally and, again, this was dependent on the properties of the data.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Paul Cross, for his continuous patience, enthusiasm and encouragement throughout my research period. I am also indebted to all those within the Department's geodetic group, with special thanks going to Moira Tighe, Paul Denys, Peter Rands, Duncan Hawksbee and Simon Corbett. I would also like to acknowledge Helen Cunningham for making my time in the Department most enjoyable. I also appreciate the support from Shell UK, who were the industrial sponsors of this project.

Final thanks must go to my mother and father and the rest of my family who have always been supportive of my studies in Newcastle, and I dedicate this thesis to Alison, Holly and Ian.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

OVERVIEW

GPS is operated by the US Department of Defence through the Joint Program Office. It is a military system that was designed with military objectives in mind, and, due to security reasons, has not been made fully available to the civilian community. Presently, a positional accuracy of 100 metres 95% of the time is guaranteed to all users. This level of accuracy is not high enough for many purposes and techniques have been developed and enhanced to improve this in real-time. Differential GPS is now well established, providing accuracies in the region of 2 - 10 m, and is used extensively in the offshore industry for operations such as general navigation, survey, seismic acquisition, rig moves and dredging.

The offshore seismic industry has an estimated world-wide annual market of US$950 million, $30 million of which is for the precise positioning of seismic vessels, and $46 million is for the positioning of oil rigs (Jensen, 1992). Great emphasis is therefore placed on the reliability of positioning systems which are being employed. Traditionally, vessels have had two navigation systems onboard; a primary and a secondary system. The primary system is used the vast majority of the time, and the secondary system is used when the primary system is either no longer available (due to coverage limitations), or is malfunctioning. Having two systems allows individual performance to be checked. Differential GPS techniques are now being widely used as the primary (and occasionally the only) navigation system. These allow positioning at distances of up to 1000 km from a land-based reference station with two-dimensional accuracies of 10 m, and better. There are no other navigation systems capable of these accuracies at such distances, for example

Syledis operates at ranges of up to 80 km, and no reliable comparisons of the GPS positions can therefore be made. Due to this, considerable attention is being paid to the real-time quality control and the quality assurance of GPS positioning. An example of this is through the workshops, recommending standards and guidelines for GPS offshore positioning, set up by the UK Offshore Operators' Association (UKOOA) which represents the combined interests of operators within the North Sea region. The workshops, which have been held since 1991, comprise of representatives from oil companies, seismic exploration companies, GPS receiver manufactures, navigation system operators, and academics.

Part of a positioning service includes the processing of the observation types by various mathematical techniques in order to estimate the position, velocity and acceleration of a vessel. These techniques have been adapted to incorporate different navigation systems, such as Syledis, Pulse-8, Hyperfix, and now GPS. One such method is the Kalman filter which incorporates information on the dynamics of the vessel along with the observations from the positioning system.

THESIS OBJECTIVES

This thesis examines the suitability of using GPS observations within the Kalman filter algorithms applied to offshore positioning. A mathematical assumption of the Kalman filter is that the observations are not correlated in time; in other words, there is no mathematical correlation between subsequent observations. GPS, being a satellite based system, is prone to many error sources; some unintentional, such as atmospheric delays, and some intentional, such as selective availability, which have been imposed by the US military. These error sources introduce temporal correlation into the observations, a fact that is ignored within the Kalman filter. This could have implications when using such a method for the processing of GPS,

or other similar, observables. This may include the final positions being incorrect and their precisions, or quality assessments, being over, or under, optimistic. The ability for a mathematical procedure to detect large unmodelled errors, blunders, in observations is also very important since these errors will propagate into position which, in turn, could have economic and safety implications. There are well accepted algorithms within the Kalman filter equations that allow for the detection of blunders but, again, these ignore the fact that observations may be temporally correlated.

The objectives of this thesis are therefore two fold. Firstly to examine the GPS observables and measure the actual extent of the temporal correlation and other statistics, and give possible reasons for changes that may be observed. This is carried out by modelling the observables using a time series model followed by mathematically estimating their characteristics. A thorough analysis of GPS observations recorded with several receiver types and under differing conditions has been carried out. The second objective is to determine the effect of different levels of temporal correlation within the Kalman filter. Differences in position and precision can be measured by using two separate approaches; one that can take the correlation into account, and a second that ignores it. Deterioration in any blunder detection (reliability) routines can be assessed by processing many data sets with different correlation statistics and with blunders of varying sizes. The detection routines should detect an error of a predefined size a certain number of times, for instance they will detect a 10 m error 95% of the time. Tabulating the actual performance when processing correlated data sets measures the extent of which the correlation affects these routines.

THESIS OUTLINE

Chapter One introduces GPS and how it is used for real-time positioning. This includes describing the GPS system, its observables and the operations that are carried out within a receiver in order to measure these observables. The main error sources are then described with particular attention being paid to intentional errors that are part of the accuracy denial policies of the US military. The technique of differential GPS is also described.

Chapters Two and Three are concerned with the algorithms that are used for positioning with GPS. It order to compute a position, it is first necessary to determine the position of the satellites and this is described in Chapter Two. Chapter Three describes the processes that are then needed to combine the GPS observables with the satellite positions in order to determine a receiver's coordinates. This is carried out by describing the operations of a computer program, RINEXPOS, that has been developed to process GPS observations in either stand-alone or differential modes. Further algorithms that may be introduced at a later date, or in other similar programs, are also documented.

Chapters Four and Five concentrate on the analysis of the GPS observables and the methods involved in determining certain characteristics. Chapter Four introduces time series analysis, certain models that can be used and their associated correlation and variance properties. The chapter also introduces the concept of Fourier analysis which allows the breakdown of a data set into differing cyclic patterns of varying amplitudes and frequencies. This technique allows the characteristics of GPS error types to be described, in particularly the intentional errors. Chapter Five analyses different GPS data sets that have been recorded over the period March 1991 to March 1993 using several different receiver types and processing strategies. All the

data sets are described in terms of their error magnitude (their variance), their temporal correlation and the trends detected from their Fourier breakdown.

Chapters Six and Seven describe the Kalman filter and the effect that temporal correlation has on the results. Chapter Six introduces the Kalman filter algorithms for determining position and detecting blunders. The statistical assumptions and terms that are often associated with the Kalman filter are also summarised. Chapter Seven then describes the effect that temporal correlation has on the results from the standard Kalman filter equations. This is carried out in terms of positional differences and precision difference for a simplified positioning system. The performance of the reliability routines is also analysed by processing differently correlated GPS observations through the Kalman filter.

Finally, the research conclusions and suggestions for further analysis are given.

CHAPTER ONE
GPS FOR REAL-TIME NAVIGATION

1.1 INTRODUCTION

The NAVSTAR Global Positioning System (GPS) was designed as a replacement to the existing TRANSIT satellite navigation system which was developed in the early 1960's. TRANSIT is capable of providing a two-dimensional navigational accuracy of approximately 250 metres at the 95% level, yet a position fix is only obtainable, on average, every 1.5 hours due to the poor satellite coverage (Ackroyd and Lorimer, 1990). GPS was designed to overcome these problems and to provide an instantaneous position with an accuracy of 100 metres (95%) almost anywhere on the globe at any time, and in any weather. With the system it is also possible to maintain centimetre relative accuracies between two points at distances up to 1000 km apart, which satisfies the most stringent of requirements within the fields of surveying and geodesy. These, along with the fact that GPS only requires visibility to the satellites, has revolutionised today's survey practices in terms of achievable accuracies, speed, and cost.

This chapter describes the structure of the GPS system, the signals that sent from the satellite, and the measurements made within the receiver. The chapter then concentrates on the differential GPS (DGPS) technique which is now the common procedure used for real-time positioning by users requiring accuracies substantially better than 100m.

1.2 THE GPS SEGMENTS

The policy making body for GPS is the Joint Program Office which represents the US armed forces, NATO, and the civilian Department of Transport (DoT) and the Defence and National Mapping Agencies (Ackroyd and Lorimer, 1990). Originally designed by the US Department of Defence (DoD), GPS comprises of three main components: the control segment, the space segment, and the user segment (Fig 1.1).

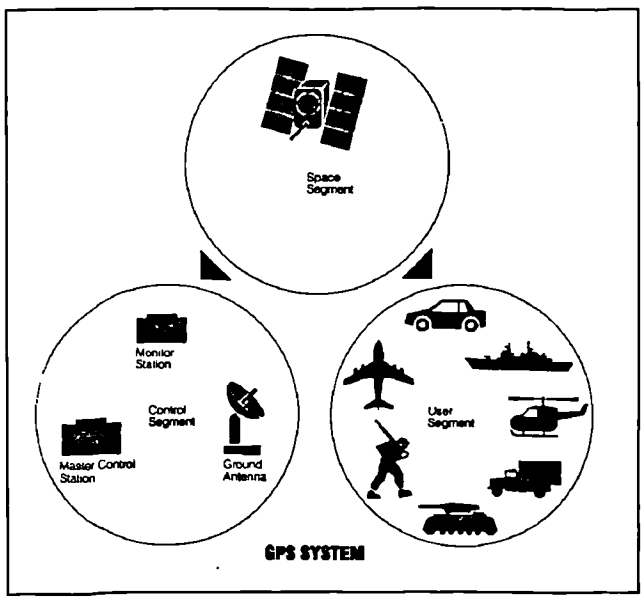


Fig 1.1 The GPS control, space and user segments (Anon)

1.2.1 The Control Segment

The control segment comprises of five stations located at Colorado Springs, Hawaii, Ascension Island, Diego Garcia and Kwajalein. All of these stations monitor the GPS satellites as they pass overhead by measuring distances to them every 1.5 seconds (Hofmann-Wellenhof et al, 1992). This data is then smoothed using ionospheric and meteorological information before 15 minute normal points are generated and sent to the Master Control Station at the US Air Force's Space

Command facilities at Colorado Springs. It is here that parameters describing the satellites' orbit and clock performance are estimated, as well as assessing the health status of the satellites and determining if any re-positioning may be required. This information is then returned to three uplink stations (collocated at the Ascension Island, Diego Garcia and Kwajalein monitor stations) which transmit the information to the satellites.

1.2.2 The Space Segment

The space segment comprises of a network of satellites in near circular orbits at a nominal height of 20,183 km above the earth and with a period of 12 sidereal hours. The original constellation was for 24 satellites, in 3 orbital planes and inclined to the equator by 63° (Spilker, 1980), but these plans have since been changed and the satellites are currently placed into six different planes and have a 55° inclination (see Fig 1.2).

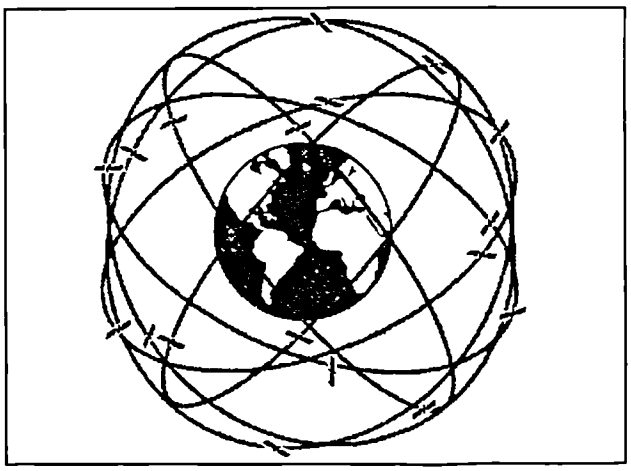


Fig 1.2 The GPS constellation

The satellites are approximately 800 kg in weight and are powered by batteries supplemented with solar panels. Manoeuvring within the orbit is carried out by the means of three thrusters. The various signals that are transmitted lie in two

frequency bands: the L-Band which is used for positioning purposes, and the S-Band which is used for the management of navigation data from the control station (Anon, 1989). There are four categories of GPS satellite, of which two types have been currently launched (as of April 1993).

The first prototype Block I satellite was launched in February 1978 from the Vandenberg Air Force Base in California. Since then a further 10 satellites have been launched up until 1985 and there are still 4 operational. These satellites have a design life of 4.5 years (Hofmann-Wellenhof et al, 1992) and were placed in the original 63° inclination. The main difference between these and the later generation satellites is that there was no ability to degrade the transmitted signals thus providing the civilian users of GPS with a reduction in the achievable positional accuracy (see 1.4.1).

The second category of GPS satellite is the operational Block II which were first launched in 1985. These have the capability to degrade the signal, have a design life of 7.5 years, and are in an inclination of 55°. Originally they were to be put into orbit from the Space Shuttle, but due to the 1986 disaster, they were re-enforced and are now launched using the Delta II rocket; a process that delayed the whole GPS program. The Block IIA satellites are a slight modification of the original design.

The Block IIR satellites (R for replenishment) are designed to have a longer life, to be capable of satellite to satellite communications, will be launched from 1996 (Graviss, 1992) and be used to maintain the full constellation. A further follow-on category Block IIF has also been planned.

The space segment will be declared initially operational when there are 24 working Block I, II, and IIA satellites in orbit (expected June 1993). Full operational

capability will be when there are 24 Block II and IIA in orbit and the testing of this could last up to one year.

1.2.3 The User Segment

The user segment comprises of the receivers that have been designed to decode the signals transmitted from the satellites for the purposes of determining position, velocity or time. To decipher the GPS signals, the receiver must perform the following tasks: (Anon, 1989)

- selecting one or more satellites in view
- acquiring GPS signals
- measuring and tracking
- recovering navigational data

The DoD has its own policies for GPS receivers for use within the armed forces. This is essentially the testing and purchasing of different receiver types - for example precision lightweight receivers, standard lightweight receivers, miniaturised airborne receivers, and timing receivers.

1.3 THE GPS POSITIONING SIGNALS

The GPS satellites transmit two L-Band signals which can be used for positioning purposes. The signals, which are generated from a standard frequency of 10.23 MHz, are L1 at 1557.42 MHz and L2 at 1227.60 MHz and are often called the carriers. Since the carriers are pure sinusoids, they cannot be used easily for instantaneous positioning purposes and therefore two binary codes are modulated onto them: the C/A (coarse acquisition) code and P (precise) code. Also it is

necessary to know the coordinates of the satellites and this information is sent within the navigation data message which is also modulated onto the carriers.

1.3.1 C/A Code

The C/A code is a pseudo random (PN) binary code (states of 0 and 1) consisting of 1,023 elements, or chips, that repeats itself every millisecond. The term pseudo random is used since the code is apparently random although it has been generated by means of a known process, hence the repeatability. Due to the chipping rate (the rate at which each chip is modulated onto the carrier) of 1.023 Mbps, the chip length corresponds to approximately 300 m in length and due to the code length, the ambiguity is approximately 300 km - ie the complete C/A code pattern repeats itself every 300 km between the receiver and the satellite.

The code is generated by means of a linear feedback register which is a hardware device representing a mathematical PN algorithm, as shown in Fig 1.3.

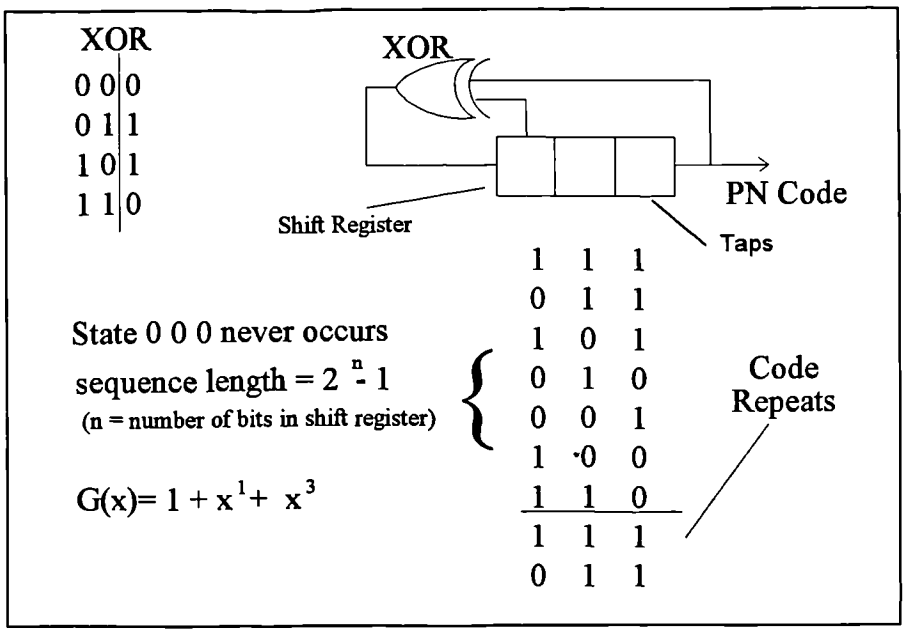


Fig 1.3 Example of a three stage linear feedback register

The sequences that are used are known as Gold codes which have particularly good autocorrelation and cross correlation properties. If two identical codes were perfectly aligned, then the correlation (the sum of the products of all aligned chips) will be at its greatest. As the two codes are the same, this is known as the autocorrelation. Codes said to have good autocorrelation properties mean that the autocorrelation value will be greatly reduced if the two identical codes are misaligned. Fig 1.4 represents the autocorrelation function for a PN code, were n is the number of taps in that code. When the two codes are perfectly aligned, the value of the function is at its maximum (1,023 for the C/A code) and has a value of -1 when the codes are separated by one chip or more.

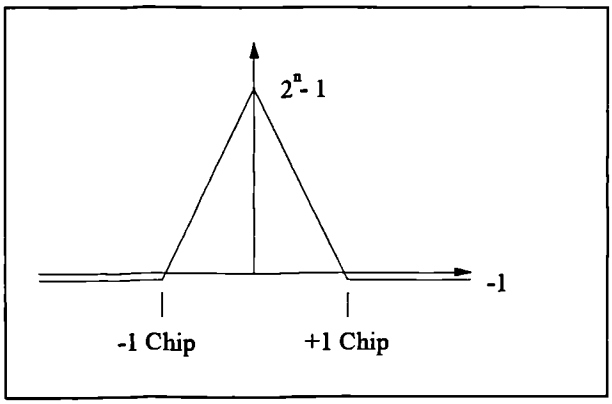


Fig 1.4 PN autocorrelation function

The cross correlation properties of the Gold codes are such that the correlation function between two different sequences is low. Both these properties are used within the GPS receiver for the identification of the code for the required satellite, and then for the alignment between the transmitted code and one that has been internally generated.

The C/A code is a product of two PN codes, G1 and G2, resulting in the Gold sequence, G. G2 is dependent on the actual satellite that is generating the codes and is therefore used as a means of distinguishing the satellite that has transmitted the

signal. These are represented by:

$$G1(x) = 1 + x^3 + x^{10}$$
$$G2(x) = 1 + x^2 + x^3 + x^6 + x^8 + x^9 + x^{10}$$
$$G(x) = G1(x) \times G2_i(x) \qquad (1.1)$$

where i is the satellite dependent integer delay

The x^2, x^3, etc are the tap positions used within the linear feedback register. From (1.1), it can be seen that ten taps are used, producing a code length and peak autocorrelation value of 2^{10} -1 (or 1,023) for the C/A code.

1.3.2 P code

The P code, or precise code, is a much longer binary code that would repeat only every 38 weeks (Pratt, 1992). Despite the code being shortened to a one week repeatability because each satellite transmits a different weekly section of the code, there is still no ambiguity between the satellite and receiver. Rapid access to the relevant part of the code for a particular satellite is carried out by means of a hand-over-word obtained from the broadcast data message. The chipping rate is at 10.23 MHz resulting in a chip length of approximately 30 m.

1.3.3 Broadcast Data Message

The data message includes information describing the positions of the satellites, their health status, and the aforementioned hand-over-word. Each satellite sends a full description of its own orbit and clock data (within the ephemeris information) and an approximate guide to the orbits of the other satellites (contained within the almanac information). The data is modulated at a much slower rate of 50 bps and

thus it takes 12.5 minutes to transmit all of the information. To reduce the time it takes to obtain an initial position, the ephemeris and clock data is repeated every 30 seconds (Langley, 1990). Parameters representing the delay caused by signal propagation through the ionosphere are also included within the data message.

1.3.4 The Modulation Techniques

For purposes of imposing the binary data onto the carriers, all of the codes are transferred from the 0 and 1 states to the -1 and 1 factors respectively. The broadcast data message is then modulo-2 added to both the C/A code and the P code. This inverts the code (see Fig 1.5) and has the effect of also inverting the autocorrelation function, ie the maximum value of the function between a code and an identical (but inverse) code is $-(2^n-1)$.

Binary biphase modulation (also known as binary phase shift keying [BPSK]) is the technique that is used to modulate the codes onto the initial carrier waves (see Fig 1.5). The codes are now directly multiplied with the carrier, which results in a 180° phase shift of the carrier every time the state of the code changes.

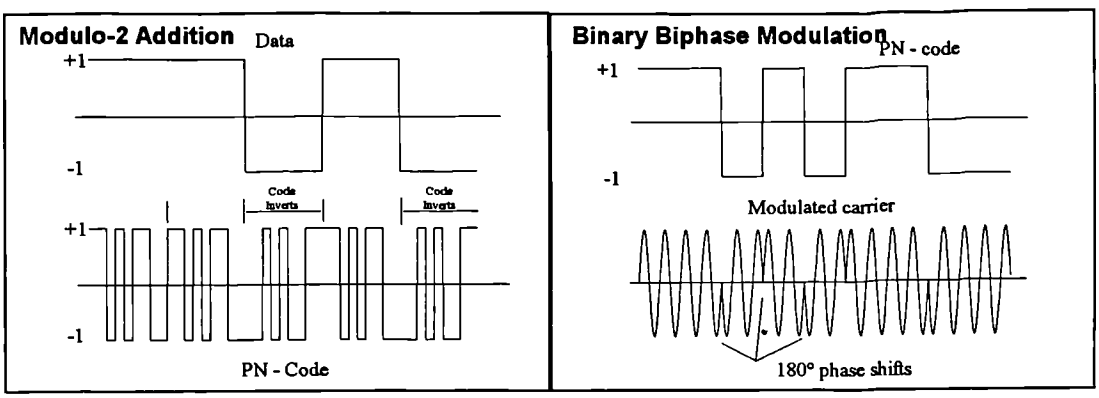


Fig 1.5 GPS modulation techniques

The modulation techniques also have the properties of widening the transmitted signal over a much wider frequency band than the minimum bandwidth required to

transmit the information which is being sent (Pratt, 1992). This is known as spread spectrum modulation and has the benefits of developing processing gain in the de-spreading operation within the receiver, and it helps prevent possible signal jamming.

1.3.5 The L1 and L2 Signals

As mentioned in 1.3, the L1 signal is centred at 1575.42 MHz and the L2 at 1227.60 MHz both generated from the fundamental frequency, f_0 = 10.23 MHz. The L1 frequency is $154f_0$ and has a wavelength of almost 19 cm, and L2 is $120f_0$ with a wavelength equating to approximately 24 cm. The purpose for using two different frequencies is so that errors introduced by ionospheric refraction can be eliminated.

The L1 signal is modulated by both the C/A code and the P code, in such a way that the two codes do not interfere with each other. This is done by modulating one code in phase and the other in quadrature (ie they are at 90° to each other) and is shown in Fig 1.6. The C/A code is also amplified so that it is between 3 and 6 dB stronger than the P code (Spilker, 1980). A diagrammatic representation of the L1 signal can be seen in Fig 1.7.

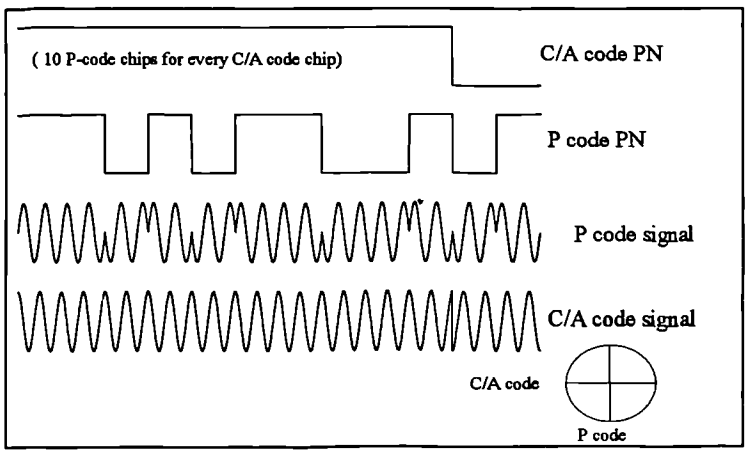


Fig 1.6 L1 P code and C/A code combination

For L2, it is stated that the signal is modulated by P code or the C/A code (Spilker, 1980) although normal operation has seen the P code being used. It should be noted that the precision obtained from P code measurements is thought not to be in the interests of US national security and therefore will be restricted for civilian users. This is the same for both the L1 and L2 frequencies, and for further details refer to section 1.4 on anti-spoofing.

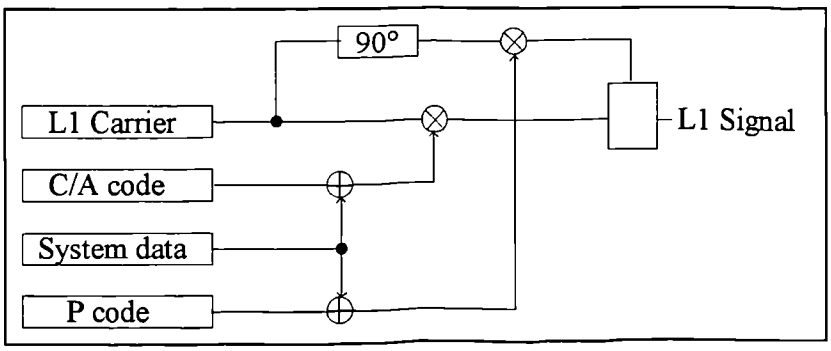


Fig 1.7 The L1 signal structure

1.3.6 The Role of the GPS Receiver

A GPS receiver has to detect and convert the signals transmitted from all of the satellites into useful measurements. During the propagation of the signals through the atmosphere, a loss in signal strength occurs and it is because of this that the spread spectrum and correlation properties of the signals are required. The processing steps for the L1 signal will be briefly described.

The antenna is usually designed to be omni-directional with a gain of 3 dB, meaning that 50% of all surrounding signal is ignored (those coming from below the horizon, or antenna ground plane). The antenna is connected to the receiver by a coaxial cable through which a voltage is sent from the receiver to a pre-amplifier at the antenna end. This pre-amplifier increases the power of the detected signal so that it

can be sent along the cable into the receiver.

The signal is immediately passed through a high-pass filter that rejects all parts of the signal which are not within the L1 bandwidths (typically a filter with a central frequency of 1575.42 Mhz and bandwidth of 20 MHz) and this results in the radio frequency (RF) signal. The signal is then modulated with a sinusoid generated by the local oscillator, resulting in the generation of a signal with two different frequency components. The frequency produced by the local oscillator is selected so that a signal at a low frequency (approximately 40 KHz) will be created, from the rule cos(f1)*cos(f2) = 0.5(cos(f1 - f2) + cos(f1 + f2)). This lower frequency is then separated by passing the signal through a low-pass filter which will also eliminate some further noise. The bandwidth of this filter is dependent on the type of measurements that are required. If the P code is required, the bandwidth will be set to approximately 20 MHz (the distance between the first two nulls [Fig 1.8]) and will be around 2 MHz if only the C/A code is tracked. The remaining analogue signal (known as the IF, or intermediate frequency signal) is then sampled as fast as possible to convert it to a digital form. An in-phase (I) and quadrature (Q) sample of the signal is generated by further sampling with an in-phase and a delayed (by 90°) frequency from the local oscillator.

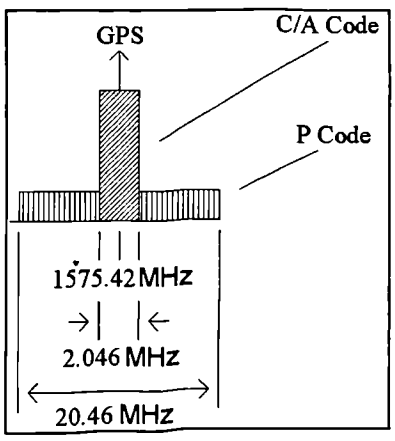


Fig 1.8 L1 power spectrum

In order to perform the de-spreading operation the IF signal is mixed down to zero frequency and copies of the signal are sent into separate channels, each of which extract the code and carrier information for a particular satellite. A replica of the C/A or P code is generated by the numerically controlled oscillator (NCO) and this is correlated with the noisy IF signal. The correlation process also de-spreads the signal, moving it to above the noise floor. The pseudorange is measured as the time shift required to align the internally generated signal with the IF signal, scaled by the speed of light. Three replica codes are in fact used for the correlation purposes - one is directly aligned with the IF signal (punctual), one is delayed (late) and one is advanced (early). The early and late codes lie on the slope of the autocorrelation function (Fig 1.4) either side of the peak and are used to aid the continuous tracking of the code, and to reduce the tracking error. New advances in receiver design have reduced the spacing between the early and late correlators which results in the cancellation of much of the noise inherent in the measurement process.

The signal is then processed for the data modulation and the carrier phase measurements. A locally generated carrier is generated by the carrier NCO and is correlated with the IF signal within a unit called the Costas loop. Once the two signals have been correlated, the data message bits can be decoded by detecting the shift in phase. The continuous phase observable is obtained by counting the elapsed cycles and by measuring the fractional part of the phase of the correlated locally generated signal. Cycle slips occur within this measurement when the elapsed cycles are not correctly counted, and loss of lock when the two signals are no longer continuously correlated.

1.3.7 Positioning with GPS

There are essentially two broad categories of GPS positioning which can be described as real-time navigation and high precision carrier phase positioning. Navigation uses a minimum of four pseudorange measurements to four satellites which are used to solve for the three-dimensional coordinates of the receiver and the clock offset between the receiver oscillator and GPS system time. An extension to this mode is differential GPS (DGPS) which again uses the pseudorange observable for positioning, but also incorporates real-time corrections for the errors inherent in the measurements.

The second category uses the much more precise carrier phase observations to compute baselines between two locations. Since the two carriers have short wavelengths (19 and 24 cm for L1 and L2 respectively), they cannot be used in the same manner as the pseudorange. The whole number of complete wavelengths (integer ambiguities) between the satellite and receiver must first be determined and this is usually carried out by post processing using linear combinations of the two frequencies and differencing techniques (see Talbot [1992] or Chen [1992] for further details).

Differences between these two modes are becoming less distinguishable. Combining the pseudorange with the phase data reduces the noise error within the pseudorange measurement resulting in a much higher positioning accuracy. New techniques are also being developed to solve for the integer ambiguities in a single epoch leading to very high baseline positioning in real-time. These are known as on-the-fly or fast ambiguity resolution techniques have already proved to provide accuracies of less than 1 cm on moving platforms over short baselines.

1.3.8 The WGS84 Coordinate System

WGS84 (World Geodetic System 1984) is the coordinate system designed for the use of GPS since this is the system in which the data describing the orbit of the GPS satellites is given. It is an earth-fixed system (the coordinates of locations do not change, see 2.1.1) which is related to the earth by means of a best fit ellipsoid. The system, which replaced WGS72 in 1985, is based primarily on coordinates derived from TRANSIT measurements from a total of 1591 locations (Cross, 1989). This initially provided coordinates in the NSWC9Z-2 TRANSIT system, which were then scaled, shifted, and oriented using collocated Satellite Laser Ranging (SLR) and Very Long Baseline Interferometry (VLBI) measurements.

1.4 GPS Error Sources

DGPS techniques were always necessary for users requiring positional accuracies better than approximately 16 m (95%) simply due to the effect of hardware and atmospheric error sources. However, due to US national security concerns, much larger intentional errors have been placed on the GPS system thus limiting accuracies obtainable from civilian users. This has reduced the GPS accuracy to 100 m (95%) but has had the effect of rapidly advancing DGPS techniques. The two aspects of this degradation are known as selective availability (SA) and anti-spoofing (AS).

1.4.1 Accuracy Denial

1.4.1.1 The evolution of accuracy denial

In order to appreciate fully the nature and impact of AS and SA it is necessary to

provide some technical details of GPS and to discuss the evolution of the civilian use of the system. Also, it is worth remarking at the outset that GPS, although conceived (in around 1973) entirely as a military system, was certainly partially sold to the US Congress on the basis of its civilian utility and it is thought to come with the enormous price tag of some thirteen billion US dollars.

The initial system concept was that all absolute point positioning, military and civilian alike, would essentially be carried out using the P-code. The C/A code was simply conceived as a means of accessing the P-code. Hence its name: Coarse/Acquisition code. The problem is that the P-code has such a long repeat period that users need an approximate position (from the C/A-code pseudoranges) to help determine which part of it to generate in order to carry out the code correlation process needed for pseudorange measurement.

It is almost certain that accuracy denial was not part of the original GPS design since the Block I satellites, launched from 1978 through to 1985, did not even have the capability for signal degradation. Nevertheless the US Department of Defence (DoD) must have considered the matter of restricting usage rather early in the system development. For instance in Lasiter and Parkinson (1977) it is stated that the full navigation capabilities after Phase II (full scale development of GPS) would only be available to a limited group of users. There is no assessment of this group nor any indication of the accuracies that others might expect. The earliest indication of such figures is in Stansell (1980) which quotes a letter dated (1978) from the DoD which indicates that navigation information at the 100-200 metre level would be available for civil use under all conditions.

Despite these indications most of the technical literature on GPS continued to give an incredibly optimistic view of its performance and its civil applications. For example as late as 1982, (Payne, 1982) in a "military" paper at a civilian

symposium, gives results from P-code absolute point positioning tests carried out in 1979 in which real-time rms position errors of as low as seven metres were achieved with the P-code. No mention of C/A-code positioning was made and there was only a brief reference to possibilities of accuracy denial. Most civilians were probably under the impression that accuracy denial or any other kind of restriction on the use of GPS simply would not occur. Nevertheless the DoD were clearly becoming concerned over the national security implications of GPS being adopted by a potential enemy.

The US Department of Transport (DoT) was not part of the Phase I GPS programme, but was encouraged by the US Congress to consider GPS in the context of the total navigation and positioning needs of the nation. The DoT compared the accuracy requirements of civil users with the national security interests laid down by the DoD. As a result the DoT/DoD formed a joint Federal Radionavigation Plan (FRP) to consider all common navigation systems. The first FRP, published in 1980, stated, according to Beser and Parkinson (1984), that the C/A signal would be available for civil use "at the highest level of accuracy consistent with national security". This accuracy is quoted as 200 metres CEP for the first full year of GPS operation with an improvement as time passes. The FRP included three aspects: degraded C/A-code results, restriction on the use of the P-code and user charges, although detailed policies for all three were not identified.

In March 1982 a second edition of the FRP was published (Cook, 1983) with slightly different figures. It was the first formal mention of the SPS (Standard Positioning Service) and PPS (Precise Positioning Service) as titles for the services available (at that time) to civil and military users. A further revision took place in 1983, (Scull, 1984) when the SPS accuracy was increased to 100 metres 2drms (which is equivalent to about 44 metres CEP). This change was brought about by pressure from civil users, in particular from the Federal Aviation Administration

after the KAL007 disaster when a Korean airliner drifted into Soviet air space. At that time President Reagan reiterated that GPS would continue to be available for international civil use.

Another change occurred in August 1984 when a policy for civil access to the PPS was approved. These policies were stated again in 1988, see (Scull, 1989), in the 1988 FRP and represent the current state of affairs. Note that in all FRPs since 1984 no further mention has been made of user charges.

The 1992 FRP mentioned differential GPS for the first time, stating that the only restriction is as for all communication links, including DGPS correction broadcasts, which are subject to control of the nation's National Command Authority. The policy also confirmed that the system will consist of 24 operational satellites, not 21 plus 3 active spares. The military Full Operational Capability (FOC) is planned for 1995 (Montgomery, 1993).

1.4.1.2 Measures of accuracy

There have been many different measures used for describing the accuracies obtainable from GPS. The most common two terms are CEP and 2drms and their exact meaning will be described in the following. Full statistical details can be found in Kalafus and Chin (1986).

CEP is an acronym for Circular Error Probability and refers to the radius of a circle in which 50% of the values occur, i.e. if a CEP of 100 metres is quoted then 50% of absolute horizontal point positions should be within 100 metres of the true position. For most offshore positioning applications 50% is rather too small a probability to be useful and a higher percentage is more valuable, typically 95% is often quoted and the term R95 used (with this notation CEP is R50). R95 (or CEP)

could be determined as follows. Simply plot the results from a large number of GPS fixes and draw a circle, centred on the mean (or known position) that contained 95% (or 50%) of the results. The radius of the circle would be the R95 (or CEP) value. Of course the disadvantage of using this measure is that it says nothing about the remaining 5% (or 50%) of the data (unless the probability density function of the errors is known or assumed) and may hence hide the possibility of there sometimes being some very large outages.

2drms is now the more commonly used term, and the one used by the most recent FRPs. It refers to twice the d_{rms} (distance root mean square error) and not, as many people seem to believe, to the two-dimensional rms. In order to compute the d_{rms} from a set of data it is simply necessary to compute the rms of the radial errors, i.e. the linear distances between the measure and known (or mean) positions. It can be predicted using covariance analysis by multiplying the HDOP, a measure of the satellite geometry, by the standard deviation of the observed pseudoranges and it is largely this predictability that makes it a much more convenient measure in practice.

A disadvantage of the 2drms measure is that is does not have a constant probability attached to it. This point is rather complex and has been analysed in detail by Chin (1987). Essentially the associated probability is a function of the ellipticity of the relevant error ellipse resulting from a particular satellite geometry. On the assumption that the pseudorange errors are normally distributed this probability will be in the range 95.4% to 98.2%. A representation of R95 and 2drms is shown in Fig 1.9.

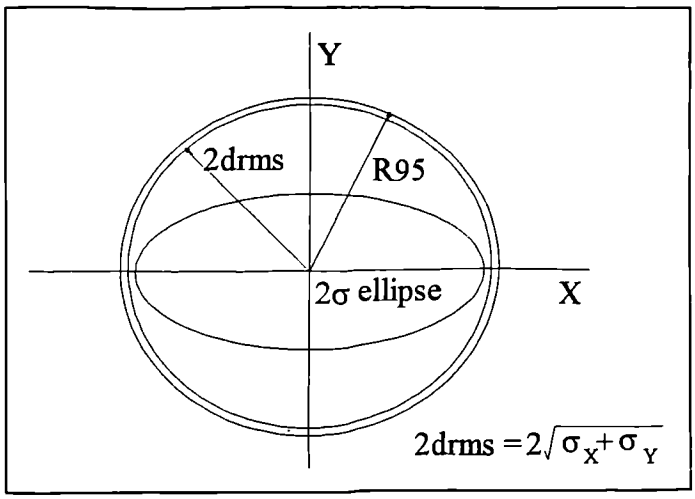


Fig 1.9 R95 and 2drms measures of accuracy

1.4.1.3 Selective availability

SA essentially consists of two different components, known as dither and epsilon. Dither is an intentional manipulation of the satellite clock frequency resulting in the generation of the carrier waves and the codes with varying wavelengths. In other words, under SA, the distance between each C/A code chip will be variable, and no longer the designed 300 m (see Fig 1.10). The replica code generated within the receiver will still assume the chip length to be 300 m and pseudorange measurements are based on this.

The epsilon component of SA refers to errors imposed within the description of the satellite orbit in the ephemeris data sent in the broadcast message (Fig 1.10). For positioning purposes, the coordinates of the satellites are derived using this incorrect information, and errors in these coordinates propagate into the position of the receiver.

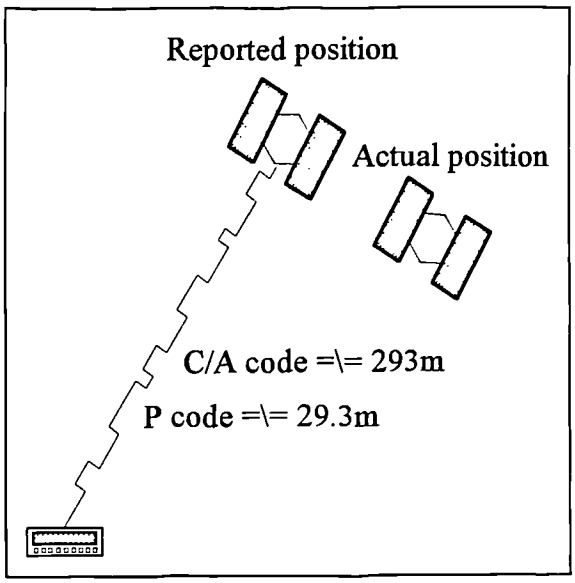


Fig 1.10 SA dither and epsilon

1.4.1.4 Anti-spoofing

AS further alters the GPS signal by changing the characteristics of the P code by mixing it with a so-called W code resulting in the Y code. It is the latter that is modulated onto the carriers and is thus designed to prevent the ability of the receiver to make P code measurements. Many receiver manufacturers have already developed techniques to still make P code measurements with only a small addition in added noise, see for instance Talbot (1992) or Ashjaee and Lorenz (1992).

1.4.2 Satellite Errors

Further unintentional errors within the space system still exist and will propagate into a position solution. These include errors in the modelling of the satellite clock offset and drift using a second order polynomial, and also errors that exist within the Keplerian representation of the satellite ephemeris information. The cause of these errors is primarily due to the manner in which the satellite ephemeris and clock is monitored. Tracking data for all observed satellites recorded at the GPS Monitor Stations is sent to the Master Control Station which then uses this data to

predict the parameters for the future. These predictions are then returned to the uplink stations where they are transmitted to the satellites. The latency of the tracking data and the prediction routines used at the Air Force Base therefore directly effect the satellite system errors. Although similar to the SA errors, these are much smaller in magnitude.

1.4.3 Atmospheric Propagation Errors

The satellite signals propagate through atmospheric layers as they travel from the satellite to the receiver. Two layers are generally considered when dealing with GPS: the ionosphere which extends from a height of 70 to 1000 km above the earth, and the troposphere which from the ground level to 70 km (Gu et al, 1993).

As the signal propagates through the ionosphere, the carrier experiences a phase advance and the codes experience a group delay. In other words, the GPS code information is delayed resulting in the pseudoranges being measured too long as compared to the geometric distance to the satellite (Hofmann-Wellenhof et al, 1992). The extent to which the measurements are delayed depends on the Total Electron Count (TEC) along the signal path which is a measure of the electron density. This is dependent on three further factors: the geomagnetic latitude of the receiver, the time of day and the elevation of the satellite. Significantly larger delays occur for signals emitted from low elevation satellites (since they travel through a greater section of the ionosphere), peaking during the daytime and subsiding during the night (due to solar radiation). In regions near the geomagnetic equator or near the poles, the delays are also larger (Spilker, 1980).

The ionospheric delay is frequency dependent and can therefore be eliminated using dual frequency GPS observations, hence the two carrier frequencies in the GPS design. Single frequency users, however, can partially model the effect of the

ionoshere using the Klobuchar model (see Hofmann-Wellenhof et al 1992). Eight parameters for this model are transmitted with the broadcast data for the satellites, and are used as the coefficients for two third order polynomial expansions which are also dependent on the time of day and the geomagnetic latitude of the receiver. These polynomials result in an estimate of the vertical ionospheric delay, which is then combined with an obliquity factor, dependent on satellite elevation, producing a delay for the receiver-satellite line of sight. The final value provides an estimate within 50% of the true delay (Cohen et al, 1992) and produces delays ranging from 5m (night) to 30m (day) for low elevation satellites and 3-5m for high elevation satellites at mid latitudes (Gu et al,1993).

The troposphere causes a delay in both the code and carrier observations. Since it is not frequency dependent (within the GPS L band range) it cannot be cancelled out by using dual frequency measurements but it can, however, be successfully modelled. The troposphere is split into two parts: the dry component which constitutes about 90% of the total refraction, and the wet part which constitutes the remaining 10%. Values for temperature, pressure and relative humidity are required to model the vertical delay due to the wet and dry part, along with the satellite elevation angle which is used with an obliquity/mapping function. Models put forward by Hopfield, Black and Saastamoninen are all successful in predicting the dry part delay to approximately 1 cm and the wet part to 5 cm.

1.4.4 Multipath

Mutipath is the phenomena by which the GPS signal is reflected by some object or surface before being detected by the antenna. The signal can be reflected off a part of the satellite (for instance the solar panels) although this is usually ignored as there is nothing that can be done by the user to prevent this. Mutipath is more commonly considered to be the reflections due to surfaces surrounding the antenna and can

cause range errors as high as 15 cm for the L1 carrier and of the order of 15-20 m for the pseudoranges (Rodgers, 1992). The surface most prone to multipath is water, whilst sandy soil is the least (abid.).

1.4.5 Receiver Noise

Errors which are due to the measurement processes used within the receiver are typically grouped together as receiver noise. These are dependent on the design of the antenna, the method used for the analogue to digital conversion, the correlation processes, and the tracking loops and bandwidths (Pratt, 1992). Noise within the pseudorange measurements can be reduced by a factor of 50% by combining with the more precise carrier phase observations (Goad, 1990).

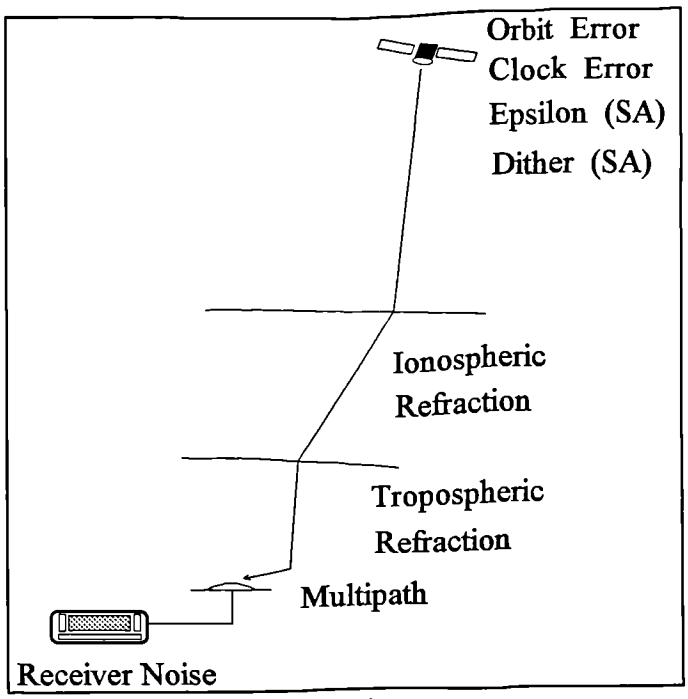


Fig 1.11 GPS error sources

1.5 DIFFERENTIAL GPS

Differential GPS (DGPS) relies on the concept that the errors in the position at one location are similar to those for all locations within a given area. By recording GPS measurements at a point with known coordinates, these errors can be quantified and corrections can be applied to the other locations. By applying these corrections in real-time, the accuracy of GPS for instantaneous positioning is reduced from 100 m (95%) to typically 5m (95%). DGPS is now a well practised technique for areas such as navigation, offshore surveying and seismic surveying.

1.5.1 DGPS Reference Station

Under DGPS operations a station is established at a known location with its role being to generate corrections which will later be applied at the unknown stations. The reference station consists of a GPS receiver and antenna and a data link (modulator and antenna).

The corrections that are generated could be of two different types: 3D positional errors or individual pseudorange errors. Both of these approaches give identical results if, and only if, the same satellites are observed at the reference station and the unknown mobile station(s) (Drewett, 1989). There are several reasons why this may not always be the case (Anon, 1990):

- The receiver criterion for selecting satellites may be different (eg 12-channel vrs 5-channel receivers).
- The satellite geometry at the two stations may be different due to the distance between them, resulting in differing rise and set times for the individual satellites.
- Terrain obstructions may exist at either station
- The receivers at both stations may not necessarily use all observed satellites to

determine its position.

Because of these reasons, the pseudorange corrections are generated at the reference station which usually adopts an all-in-view policy. By transmitting individual corrections for all satellites, the mobile station uses the corrections for the satellites observed at that station thus avoiding any errors that may be introduced. As well as a pseudorange correction for each satellite, its rate of change is also computed and transmitted to the mobile station and is used to model the time varying characteristics of the corrections over the period in which the corrections are generated at the reference station and applied at the mobile station (the age of correction).

A central processor (a dedicated PC linked to the GPS receiver) is often used to generate the corrections, although many of todays receivers can generate the corrections in the "box". The processor also converts the corrections into a standard binary format (RTCM SC-104) which is then sent to the data link equipment and modulated onto a carrier frequency which is subsequently transmitted. Details of how the corrections are generated can be found in 3.3.7.

1.5.2 DGPS Mobile Station

The mobile station consists of a GPS receiver and antenna, a data link receiver and demodulator, and usually a PC acting as a central processor. The pseudorange and range rate corrections are received via the data link equipment, demodulated, and the binary message is then sent to the PC. The corrections are then translated and applied to the individual pseudorange corrections observed by the mobile station's GPS receiver which is also connected to the PC. The final correction is added to the observed pseudoranges and is derived from (Anon, 1990):

$$PRC(t) = PRC(t_0) + RRC(t_1 - t_0) \qquad (1.2)$$

where PRC(t) is the correction to be applied to the appropriate mobile pseudorange,

$PRC(t_0)$ is the correction generated at the reference station,

RRC is the range rate correction,

t_0 is the time at which the correction was generated at the reference station, and

t_1 is the time at which the mobile pseudorange data was observed.

Occasionally a further correction (the delta correction) needs to be applied if the reference station has used different ephemeris data to generate the pseudorange corrections than is available at the mobile station. All satellites that have been "corrected" are then used within a position computation (see 3.3.4.2) in which the position and appropriate statistics for the mobile station are derived. Fig 1.12 represents the DGPS operation.

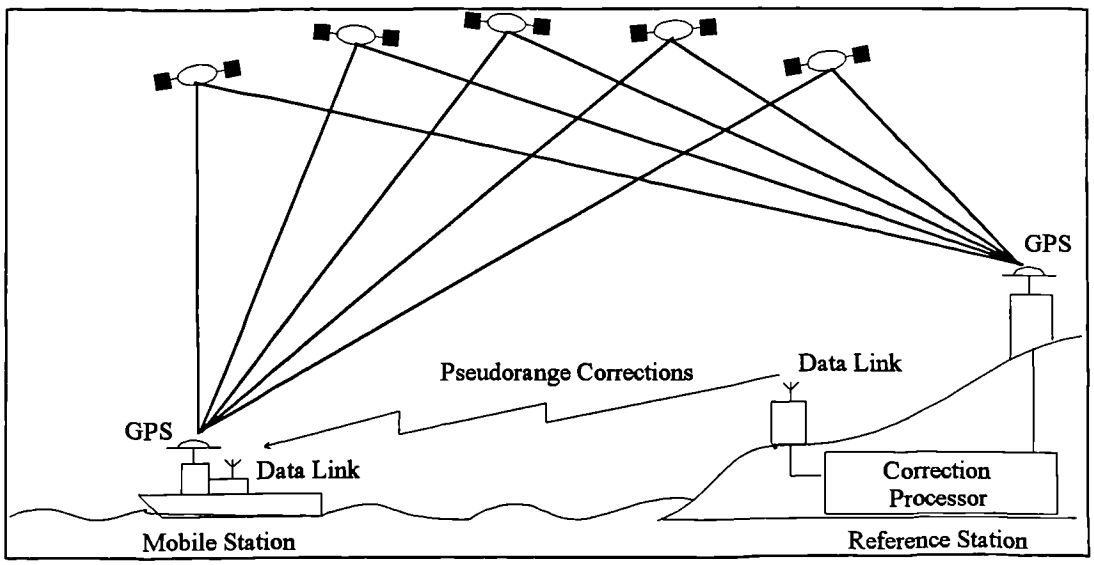


Fig 1.12 Differential GPS

1.5.3 Data Links

The data link provides the connection between the reference and mobile station, and the medium that is used must be such that it can transmit over the required reference-mobile separation, the binary data can be modulated fast enough to provide a certain age of correction, and it must be reliable in terms of probability of good reception (Barboux, 1993). The choices that are available can be divided into land based and satellite based links.

The land based data links are predominantly radio transmitters operating at various frequencies dependent on application. Low frequency (LF) links can be used which provide a wide coverage due to good surface wave propagation. However, these systems require substantial antenna and power, and provide a relatively slow age of correction (Barboux, 1993) due to their long wavelengths. Choosing higher frequencies generally reduces the power consumption, requires simpler hardware, reduces the age of correction, but also leads to a reduction in the distance over which the signal can be transmitted. Alternatives to the LF option that are in operation include the medium (MF), high (HF), and ultra-/very-high (UHF/VHF) frequencies and their performances are given in Table 1.1. Other options for land based data links use existing radio transmissions which avoids the need to establish a dedicated link and the necessity to obtain frequency allocation permission. These include the use of existing radiolocation systems (for example Pulse-8 and Hyper Fix), local FM radio broadcasts, cellular telephones, and marine radiobeacons. The latter uses a existing network of direction finding radiobeacons which, although no longer used for positioning purposes, are well maintained, have frequency allocation and a range of several hundred kilometres. The US Coast Guard (USCG) and the International Association of Lighthouse Authorities (IALA) are supporting the use of this medium, and the USCG have pledged to provide a DGPS radiobeacon network for continental US coastal waters, southern Alaska, Hawaii and Puerto

Rico by early 1996 (Hallmann, 1993). One final option for ground based transmissions is for the use of pseudolites which are designed to transmit the data at the GPS frequencies. The advantage of these is that the standard GPS antenna can be used to receive the corrections, and the signal has the possibility of being used as a additional pseudorange measurement. However, over land, the signal range is poor and this along with expense issues, has meant that the option is seldom viable.

	Frequency Band	Range (km)	Typical Age of Correction (s)
LF	30 - 300KHz	> 700	< 20
MF	300KHz - 3MHz	< 500	5 - 10
HF	3 - 25MHz	< 200	5 - 10
UHF/VHF	30 - 300MHz	< 100	< 5

Table 1.1 Ground based data links

The satellite based data links are now common place and provide a near global coverage with a fast age of correction (less than 5 seconds) and use the INMARSAT Standard-A telephone and telex link. INMARSAT is an internationally owned cooperative set up to meet the communication needs of the world shipping industry and operates geostationary satellites in four ocean region areas: the Atlantic Ocean Region West, the Atlantic Ocean Region East, the Indian Ocean Region and the Pacific Ocean Region (Al-Nakib, 1992). Information is uploaded to the satellites from Coast Earth Stations (CES) which is usually operated by the local signatory - in the UK this is British Telecom and the CES is located at Goonhilly in Cornwall. A Ship Earth Station (SES) is required to receive the signals from the geostationary satellites. For DGPS operations, the correction message is sent from the reference station to the CES via leased land lines, is uploaded to the satellites and subsequently transmitted for the use of the mobile.

1.5.4 RTCM SC-104 Message Format

It is necessary for DGPS operators to keep to standards concerning the format of the correction message and matters concerning the reference and mobile station operations, and the data link. These standards have been recommended by the Radio Technical Commission for Maritime Services Special Committee 104 (RTCM SC-104) which was originally established in November 1983 and were first published in November 1987.

The data format has been modelled on the GPS navigation message with the word size, word format and parity algorithms being the same. In version 2.0, there are a possible 64 different message types of which 21 have been defined. The Type 1 message contains the pseudorange and range rate corrections, the issue of data (IOD), and the user differential range error (UDRE). The IOD parameter allows the mobile station to identify the satellite navigation data used by the reference station - if this is different than the delta corrections (Type 2) should be included. The UDRE is an estimate of the standard deviation of the differential error as determined at the reference station and can be used to weight the solution at the mobile. A list of some of the message types is provided in Table 1.2 (Ackroyd and Lorimer, 1990).

Message Type	Message Title
1	DGPS corrections
2	Delta DGPS corrections
3	Reference station parameters
4	Carrier surveying information
5	Constellation health
6	Null frame
7	Marine radiobeacon almanacs
8	Pseudolite almanacs
9	High rate DGPS corrections
10	P code DGPS corrections
11	C/A code L1/L2 delta corrections
12	Pseudolite station parameters
13	Ground transmitter parameters
14	Surveying auxiliary message
15	Ionospheric/tropospheric message
16	Special message
17	Ephemeris almanac
18-59	Undefined
60-63	Differential Loran C messages

Table 1.2 RTCM SC-104 message types

1.5.5 DGPS Error Budget

The ability to obtain instantaneous positioning to within 5 m can be summed up by looking at an error budget that offsets stand-alone GPS and DGPS. The error sources described in section 1.4 can be divided into two groups: those which are independent of the user and those which are dependent. Table 1.3 shows the pseudorange error budget for DGPS under SA (adapted from Ackroyd and Lorimer [1990] and Gu et al, [1993]) with the baseline distance between the known and unknown stations of 500 km and with an age of correction of 5 seconds.

Error Source	**Stand-alone (m)**	**Differential (m)**
User Independent		
Satellite clock	15.0	0.1
Ephemeris	40.0	1.0
Orbit	5.0	0.13
Ionosphere	12.0	1.0
Troposphere	3.0	0.5
User Dependent		
Multipath	2.0	2.8
Receiver noise	0.5	0.7
Total root sum squared	44.8	3.3

Table 1.3 DGPS error budget

The table clearly shows which of the error sources cancel under the DGPS operation. The reduction of satellite clock errors depends on the age of correction of the pseudorange errors - the greater the latency, the less the clock error cancels. Remaining orbital parameters depend on the baseline length between the reference and mobile stations and the error imposed on the satellite. These are largely independent on the age of correction since the largest error is due to epsilon which can only be altered with every new ephemeris data for a particular satellite, which occurs every hour. The atmospheric error sources will cancel if the baseline length is short, and/or there are similar atmospheric conditions at both sites. It should be noted that the user dependent error sources, multipath and receiver noise, are increased under DGPS since similar errors at two stations are now combined.

1.5.6 GPS Integrity Monitoring

".. a simple definition of integrity has been adopted by the Radio Technical Commission for Aeronautics that states:

Integrity - the ability of a system to provide timely warnings to users when the system should not be used for navigation." (Al-Nakib, 1992).

Monitoring the GPS system and assessing its quality in real- (or near real-) time must therefore be carried out and action taken appropriately. Unhealthy satellites, incorrect pseudoranges, poor DGPS corrections, excessive atmospheric interference and problems at particular reference stations must be detected and the mobile stations must be informed. There are three main approaches that are being used to perform these tasks.

Monitor stations. A monitor station acts as a mobile station in that it receives the DGPS corrections from a reference station and computes its position. The difference is that the monitor is at a known location and can therefore assess the quality of the solution. Any abnormal occurrences are then sent back to the reference station and are then transmitted to the mobile users via the RTCM SC-104 ascii message.

RAIM. Receiver autonomous integrity monitoring (RAIM) is a technique whereby the performance of the GPS system is determined within the receiver itself. Four pseudoranges are needed to compute a position at any particular instance. If five are available then five different position solutions can be computed by using all combinations of four satellites (Al-Nakib, 1992). By examining these five solutions, ones which are using an erroneous pseudorange will be similar, and different from the remaining solution. In this way poor ranges can be isolated, but the technique obviously relies on a full geometry to be able to detect multiple errors.

GIC. The GPS integrity channel (GIC) uses satellites to transmit the warnings

which are detected by monitor stations. Geostationary satellites are used to transmit the integrity data to the mobiles, and two different signal formats have been considered. Wideband GIC transmits a signal very similar to the C/A code measurements at the same frequency as L1. This has the advantages that the GPS antenna can be used to receive the signal, the data can be demodulated within the receiver, and the signal from the geostationary satellite could be used as an additional pseudorange to augment the GPS coverage. The next generation of INMARSAT satellites (INMARSAT 3), which will be launched in late 1994, will have this capability. The second signal format is the narrowband GIC which uses the telephone and telex communication bands to transmit the message. This requires separate receiving antenna and is merely an extension of the space-based DGPS correction services.

CHAPTER TWO
COMPUTATION OF A GPS SATELLITE'S POSITION

2.1 COORDINATE REFERENCE SYSTEMS

In satellite geodesy there are two fundamental earth-centred coordinate systems. The first is the earth fixed terrestrial coordinate system which is required as a reference frame in all positioning on the earth's surface. The second is the inertial, or fixed celestial coordinate system, and is required since the equations of motion of satellites (both artificial and the planets) are often expressed in this system.

2.1.1 The Earth Fixed Terrestrial Coordinate System

Earth fixed systems are right handed systems with the origin at the geocentre, or the earth's centre of mass. The Z-axis is aligned with the Conventional International Origin (CIO) pole - the mean spin axis of the earth between 1900 and 1905. The X-axis points towards the zero of longitude (approximately Greenwich) and is in the equatorial plane. The Y-axis completes the right handed system. WGS84, the coordinate system designed for use with GPS, is such a system and a representation of this is shown in Fig 2.1.

2.1.2 The Inertial Coordinate System

The celestial coordinate system (Fig 2.1) is a right handed system which is either fixed in space, or has uniform motion (ie constant in time). The origin is at the centre of mass of the earth, the Z-axis coincides with the earth's axis of rotation, and the XY-plane is the equatorial plane. The X-axis points towards the vernal

equinox, which is the intersection of the equatorial plane and the plane of the earth's orbit around the sun. Theories of the earth's precession and nutation state that the equator and ecliptic, and hence the vernal equinox, are in a constant state of motion (Sharif, 1989). Therefore to correctly define an inertial coordinate system, the effect of precession and nutation must be accounted for.

2.1.3 Orbital Coordinates

A third type of coordinates are used for satellite coordinate computations. These orbital coordinates represent the position of the satellite within its orbit. The system is again right handed and is centred at the geocentre. For purposes that will become clear later, the system will be defined as follows. The X-axis points towards the Ascending node and the XY-plane is in the orbital plane with the Y-axis pointing towards an argument of latitude of 90°. The Z-axis is perpendicular to the orbital plane, completing the right handed system.

2.1.4 Coordinate Representation

The following notation is used to represent the 3D cartesian coordinates of a point P in the three different coordinate systems

Inertial $$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^{P}_{I} = \mathbf{X}^{P}_{I}$$

Earth fixed $$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^{P}_{E} = \mathbf{X}^{P}_{E}$$

Orbital $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_O^P = \mathbf{X}_O^P$

The relationship between the three systems is shown in Fig 2.1.

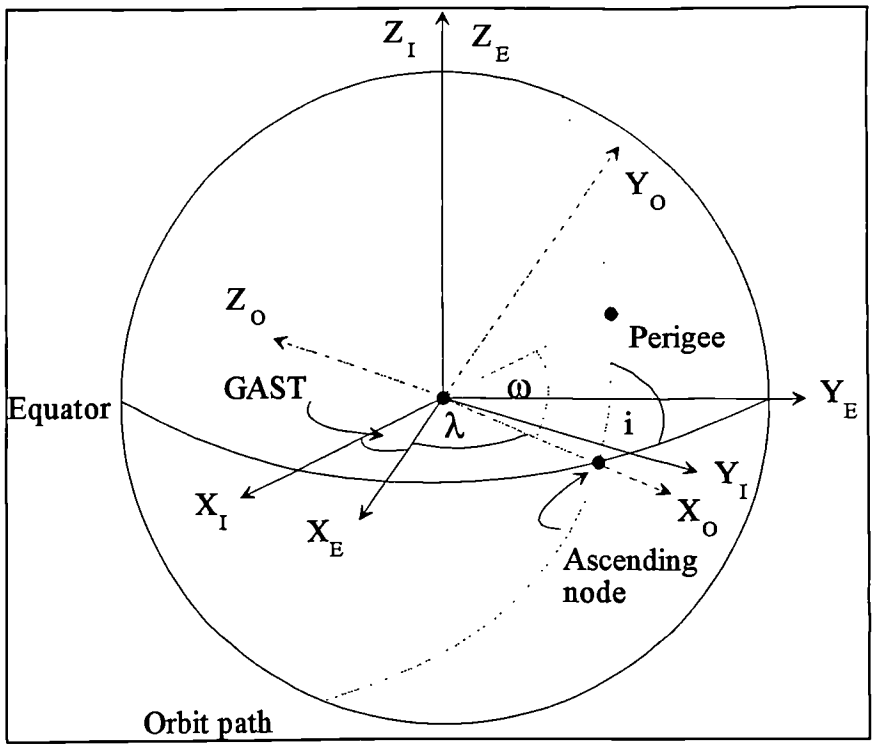


Fig 2.1 Inertial, earth-fixed and orbital coordinates.

2.2 REPRESENTATION OF A SATELLITE'S ORBIT

2.2.1 Kepler's Laws

Johannes Kepler (1571 - 1630) was a German astronomer and mathematician who developed three laws, based on the observations of Tycho Brahe (1546 - 1601), that described the motion of the planets around the sun. His first two laws were published in Astronomia Nova in 1609 and his third law was published in Harmonices Mundi in 1619.

This planetary motion, or Keplerian motion, is now used to describe the path of an orbiting satellite around the earth with the only force acting on it being the gravitational attraction of the earth. Kepler's three laws are

1. The orbit is an ellipse in a plane with the centre of mass of the attracting body at one of its foci. This indicates that a satellite orbiting the earth will not be the same distance from the earth at all times, unless the orbit is circular. The point where it is closest to the earth is known as the perigee and the farthest point is known as the apogee.

2. The satellite's radius vector sweeps out equal area in equal time. This means that the speed of the satellite is not constant and will be at its fastest at the apogee and its slowest at the perigee.

3. The ratio between the square of the orbital period (the time taken for a satellite to complete one orbit) and the cube of the semi major axis of the elliptical orbit is the same for all satellites. This indicates that two satellites with the same orbital semi major axis but with different eccentricities (the flattening of the ellipse), will take the same time to complete one revolution of their orbits.

2.2.2 The Keplerian Elements

Just six parameters (the Keplerian elements) can be used to describe a satellite following Keplerian motion with respect to the earth. Their definitions are

a	the semi major axis of the elliptical orbit
e	the eccentricity of the elliptical orbit
i	the inclination of the orbital plane to the equatorial plane
Ω	the right ascension of the ascending node
ω	the argument of the perigee
t_p	the time at which the satellite passes the perigee

The size and shape of the orbital ellipse is defined by a and e, the orientation by i and Ω, the rotation of the ellipse within the orbital plane by ω, and the instant when the satellite is at a known location by t_p.

These parameters are illustrated in Fig 2.2 in which the coordinate system is an XYZ inertial right ascension coordinate system. The satellite's orbit and the earth's equator have been projected onto a unit sphere.

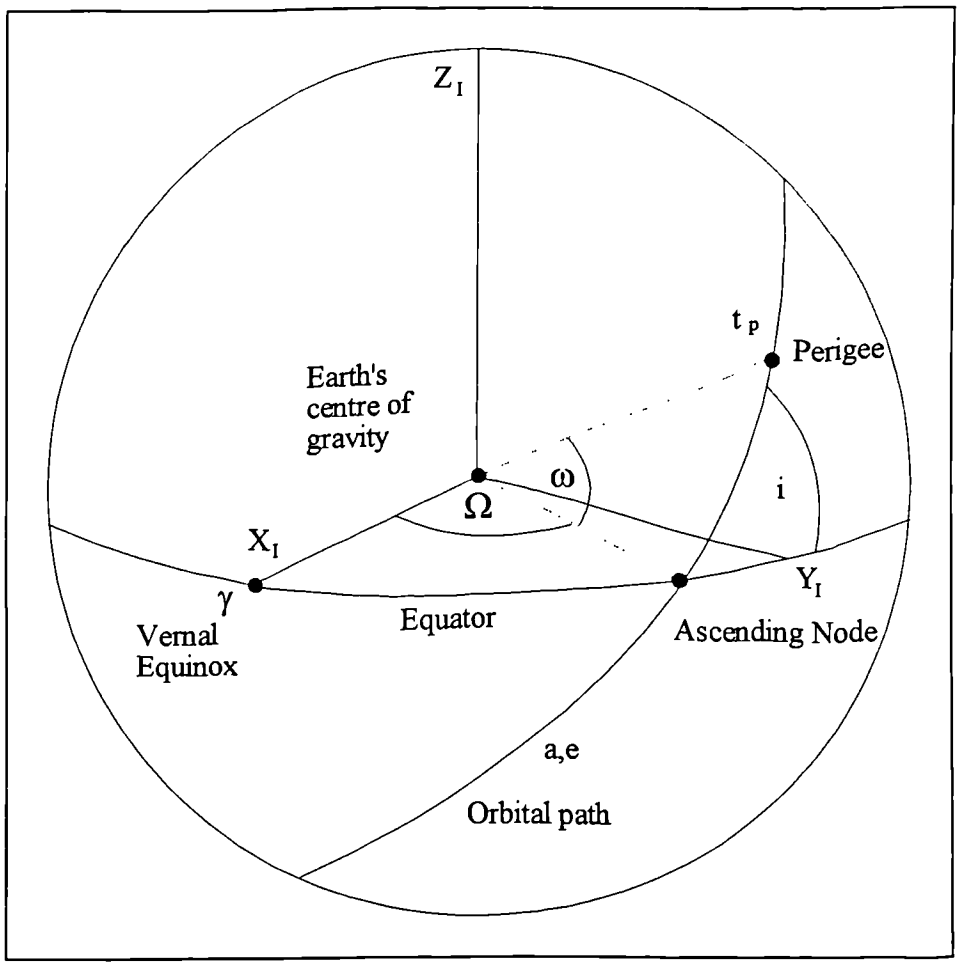


Fig 2.2 The Keplerian elements depicting a normal satellite orbit

2.2.3 Alternatives to the Keplerian elements

There are various ways of representing a satellite's orbit around the attracting body. Polynomial or harmonic functions, as well as 3D cartesian coordinates and velocities at a specific time all could be used to model the orbit (Van Dierendonck et al, 1980). The Keplerian elements are often used for the broadcast ephemeris of satellite positioning systems (eg Transit and GPS) instead of these alternatives. They have the benefits that they are efficient to store (there are only six basic parameters), they have an obvious meaning and they degrade gracefully with time (Ashkenazi and Moore, 1986). They do have one disadvantage in the computational time to compute a satellite position from the given elements - however, today's processing speed has made this totally negligible. It should be noted that Keplerian

elements are not always used to represent an orbit. The precise ephemeris (post-mission) for Transit and GPS, and the broadcast ephemeris for Glonass all use a 3D cartesian coordinate system providing the XYZ position and velocity components at a specific moment in time.

2.2.4 Perturbing forces on a GPS satellite

Kepler's laws are for an idealised satellite orbit where the only attracting force is a spherical gravity field. For any satellite orbiting the earth this is not the case and its Keplerian position will be affected by the following perturbing forces

- The earth is not a perfect sphere and has an uneven density distribution. The effect that this has on the earth's gravitational field is represented by spherical harmonic coefficients which are used to compute the disturbing potential (the difference of potential on the geoid and on a reference ellipsoid) at a particular location. The largest coefficient in the expansion is known as J_2 (or C_{20}) and represents the effect of the flattening of the earth (the earth's equatorial bulge) on the gravitational field. This coefficient is about 1000 times larger than the other coefficients, although those of up to degree and order 36 are significant for satellite orbit computations (Leick, 1990). This is dependent on the actual satellite, and for GPS a harmonic function of degree and order 8 are adequate.

- Other planets (in particular the moon and the sun) have their own gravity fields and exert an attraction on the satellite. This is known as the third body effects. The third body gravitational attractions have an additional effect on the satellite orbit since they cause earth and ocean tides. The change in the earth's mass distribution and shape resulting from these tides alters the gravity field and thus the forces acting on any orbiting body. The magnitudes of these forces are extremely well modelled and therefore their effects can be greatly reduced.

- The satellite is not travelling in a perfect vacuum and will experience frictional atmospheric drag. This is a function of the atmospheric density at the orbital height and of the satellite's mass and surface area. It is negligible for GPS orbits as they are orbiting at approximately 21000 km above the earth.

- The satellite will experience the impact of light photons emitted by the sun both directly and indirectly (the albedo effect). This is known as solar radiation pressure and will be a function of the satellite's effective area (the surface area normal to the radiation), the surface reflectivity, the luminosity of the sun, and the distance to the sun (Leick, 1990). For GPS satellites, this effect cannot be ignored, it is difficult to model, and therefore represents the largest unknown error source.

Most of these effects can be modelled, but this is difficult for the solar radiation pressure and albedo perturbing forces (de Jong, 1991). All of the perturbing forces alter the gravitational pull on the satellite and are quantified in terms of their disturbing accelerations. If ignored, these disturbing accelerations will have an effect on the GPS satellite position which will grow with time. Table 2.1 lists the accelerations and their influence on GPS satellite orbits over 1, 2, and 3 hours.

Perturbing force	Acceleration (m/s^2)	Growth of orbit error (m)		
		1 hour	2 hours	3 hours
J_2	5×10^{-5}	400.00	1700.00	4000.00
Third body	7×10^{-6}	20.00	100.00	400.00
Other harmonics	3×10^{-7}	1.50	7.20	40.00
Solar radiation	6×10^{-8}	1.00	5.00	25.00
Earth tides	1×10^{-9}	0.01	0.02	0.16
Ocean tides	5×10^{-10}	0.00	0.01	0.07

Table 2.1 Growth of GPS orbit perturbations (Sharif, 1989) and (de Jong, 1991).

2.3 THE GPS BROADCAST EPHEMERIS PARAMETERS

2.3.1 The GPS Keplerian Elements

The broadcast ephemeris, sent down within the GPS navigation message, uses the Keplerian elements to represent the idealised GPS orbit and incorporates additional terms to account for the effects of the perturbing forces. The main Keplerian elements are

$\sqrt{a}$	square root of the semi-major axis
e	eccentricity
t_0	ephemeris reference time
M_0	mean anomaly at reference time
i_0	inclination angle at reference time
Ω_0	right ascension at reference time
ω	argument of the perigee

The elements t_0 and M_0 replace the usual time of perigee parameter, t_p, creating seven, instead of six, Keplerian elements. The reason for this is that the parameters are defined to a reference time which will be within a few minutes of the time of evaluation, decreasing the sensitivity of the time derivative parameters and making no increase in the number of words required within the navigation message (Van Dierendonck et al, 1980).

The term Ω_0 within the GPS ephemeris is not strictly the right ascension of the ascending node at t_0. More correctly, it is the angle between the Greenwich meridian at the beginning of the GPS week and ascending node at the ephemeris reference time. It should be more correctly described as the longitude of the ascending node from the start of the GPS week to the ephemeris reference time. For more details on this parameter refer to Wells et al (1986) or Drewett (1989).

2.3.2 The GPS Perturbation Parameters

In addition to the Keplerian elements, extra parameters have to be included to model the position of the satellite from its normal orbit and are provided as secular drift terms and harmonic coefficients. These parameters are (Ashkenazi and Moore, 1986)

Δn	correction to the computed mean motion
$\dot{\Omega}$	rate of change of right ascension
$\dot{i}$	rate of change of inclination
C_{uc}, C_{us}	amplitude of cos and sin correction terms to the argument of latitude
C_{rc}, C_{rs}	amplitude of cos and sin correction terms to the geocentric radius

C_{ic}, C_{is} amplitude of cos and sin correction terms to the inclination of the orbital plane

The GPS Keplerian elements and the perturbing parameters are shown in Fig 2.3.

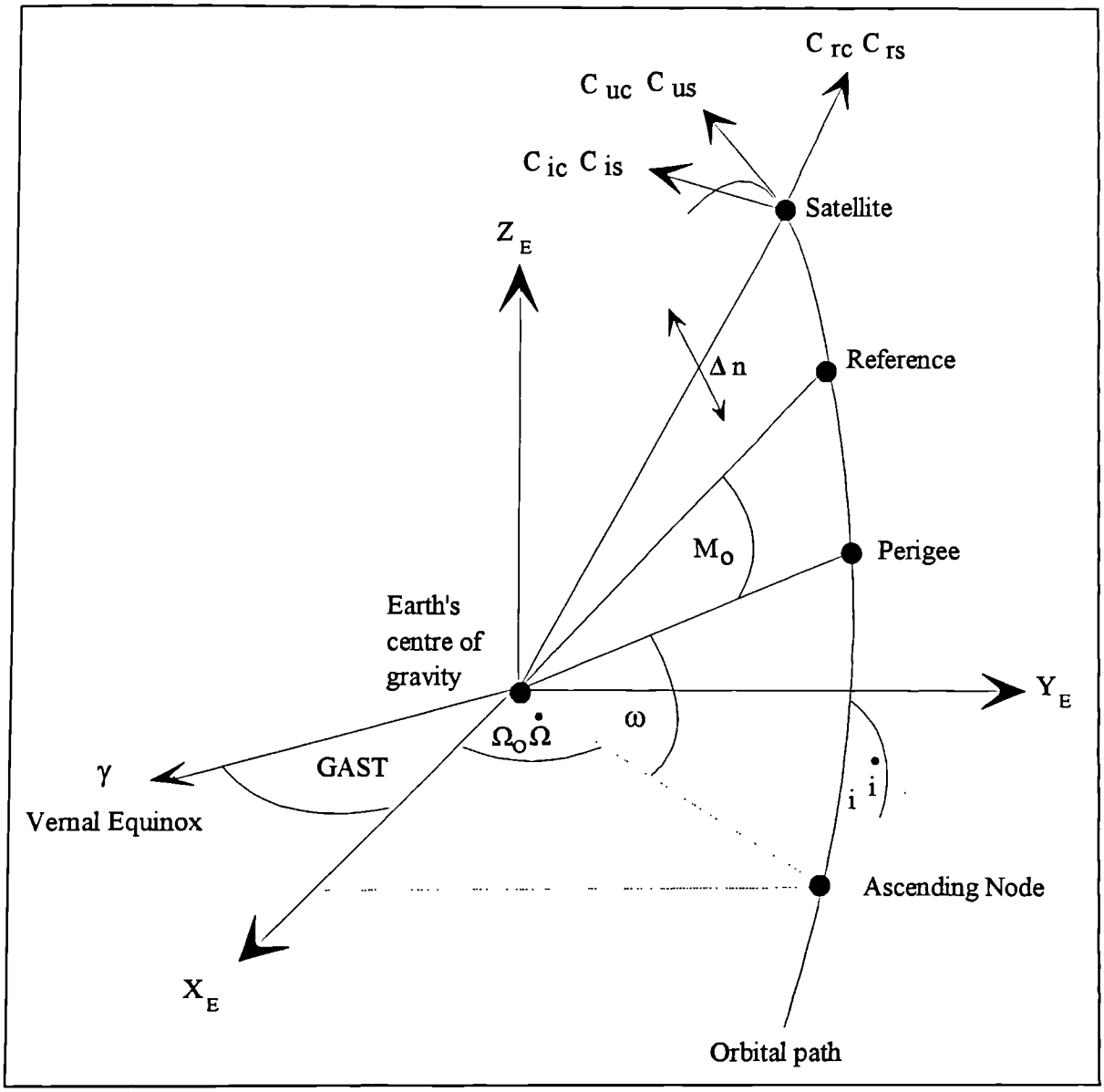


Fig 2.3 The GPS Keplerian elements

2.3.3 The Drift Terms

The correction to mean motion, Δn, is used to describe the combined rate of change of the ω and M_0 angles as shown in Fig 2.3. The parameter mainly describes the effect of J_2.

The rate of change of right ascension, $\dot{\Omega}$, describes the precession of the satellite resulting from J2. The earth's equatorial bulge exerts torques on the satellite moving within its orbit causing the orbital path to migrate along the equator. This effect is exactly analogous to the earth's precession. $\dot{\Omega}$ is also used to model the precession and nutation of the earth.

The rate of change of inclination, $\dot{i}$, describes short term periodic motion which is due to the J_2 effect. $\dot{i}$ is also used to model the precession and nutation of the earth.

The effect of the zonal harmonics on satellite orbits is given in many texts on physical geodesy - for instance Heiskanen and Moritz (1967) or Lambeck (1988). The effect is often split into short term periodic motions (periods of 1, 2, 3, ... revolutions) and long term periodic and secular motions. The secular terms are constant over one revolution and increase steadily with time. Long term periodic terms again are constant over one revolution but have a slow periodic motion. The semi-major axis of the orbit does not change secularly or long periodically. The eccentricity and the inclination undergo long-period, but not secular, variations, whereas the right ascension and the perigee argument change both secularly and long periodically (Heiskanen and Moritz, 1967).

Tesseral harmonics (describing longitudinal variations in the earth) have a much smaller effect than the zonal harmonics. These are the cause for the rate of change of inclination rapidily changing its sign.

The approximate magnitudes of the drift terms for a GPS satellite are given in Table 2.2. Polar motion (precession and nutation) has the same effect on $\dot{\Omega}$ and $\dot{i}$, but has a magnitude of approximately 1×10^{-13} radians / second and is therefore not seen in the table.

Parameter	Radians / sec	Seconds of arc / sec
Δn	5×10^{-9}	1.3×10^{-3}
$\dot{\Omega}$	-8×10^{-9}	-1.6×10^{-3}
$\dot{i}$	$\pm 1 \times 10^{-10}$	$\pm 2.1 \times 10^{-5}$

Table 2.2 Approximate magnitudes of GPS perturbing drift terms

2.4 COMPUTATION OF A GPS SATELLITE'S POSITION

The position of a GPS satellite can be found in an earth fixed coordinate system from the Keplerian elements and the parameters describing the effects of the perturbing forces. The following equations have been adapted from Van Dierendonck et al (1980) and Drewett (1989).

2.4.1 The Position of the Satellite within the Orbital Frame

The position of the satellite can be related to the transmission time, t, by using the true anomaly, f, or the eccentric anomaly, E, along with the time the satellite passed the perigee, t_p. The true anomaly represents the geocentric angle in the orbital plane between the perigee and the satellite. The eccentric anomaly is a more convenient way of locating the satellite within its orbit and is shown in Fig 2.4. A circumscribing circle, of radius a and centred at the centre of the orbital ellipse (C),

is drawn over the ellipse. The eccentric anomaly is the angle subtended from the centre of the earth (G) to the perigee and a sub-satellite point (S') which has been projected onto the circle. The distance r represents the radial distance to the satellite from the earth's centre. It is apparent from Fig 2.4 that the orbital coordinates can be found via trigonometry once r and $(f+\omega)$ have been determined.

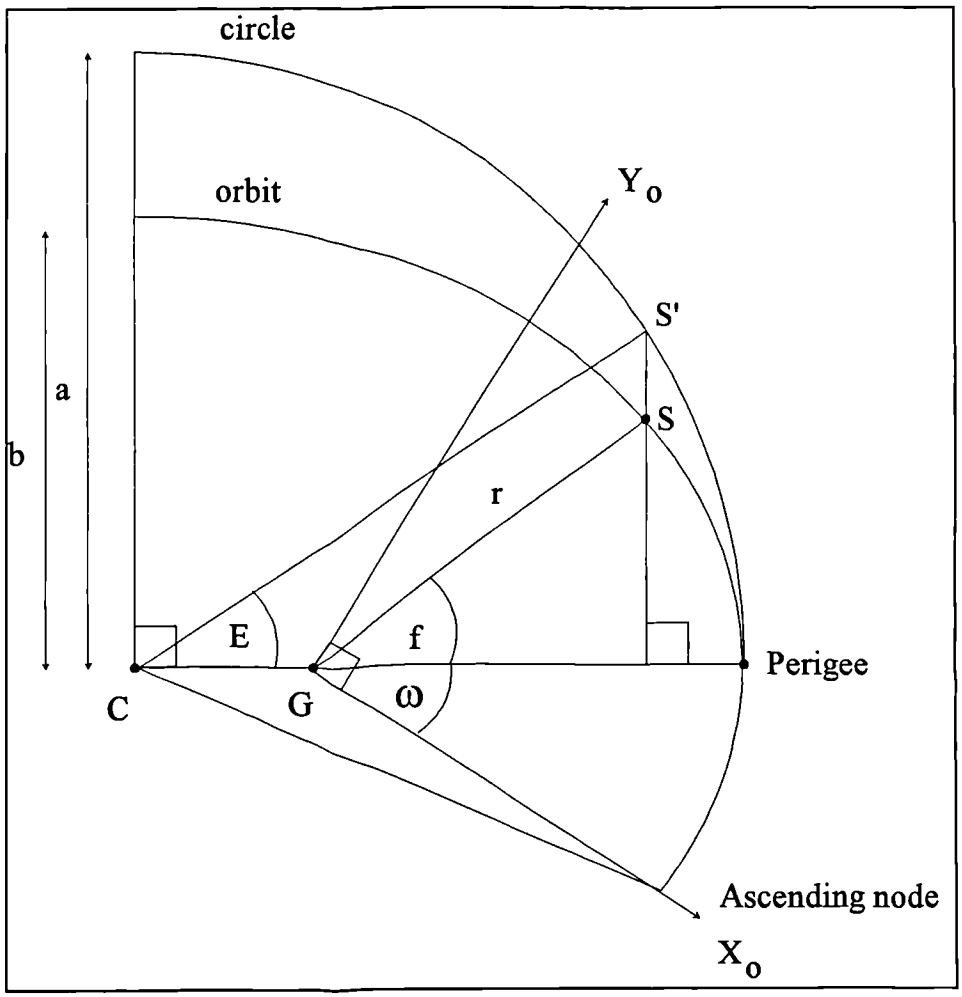


Fig 2.4 The eccentric and true anomalies of the satellite.

To find E, it is first necessary to calculate the mean anomaly, M, which is the true anomaly of an imaginary satellite moving with uniform angular velocity and with the same period of the GPS satellite. The mean anomaly is linked to the mean motion, n, of the satellite which is obtained from Kepler's third law

$$GM = a^3 n^2 \tag{2.1}$$

since it is related to the orbital period by,

$$n = 2\pi / T \tag{2.2}$$

where

GM	is the gravitational constant times the mass of the earth. (Defined as 3.986005×10^{14} $m^3 sec^{-2}$ for WGS84, Anon [1991]).
T	is the orbital period of the satellite

Due to the perturbed orbit, the mean motion must be corrected by

$$n' = n + \Delta n \tag{2.3}$$

It is convenient to introduce the time delay, t_k, between the time, t, at which the satellite's coordinates need to be determined (the transmission time) and the ephemeris reference time, t_0. Thus

$$t_k = t - t_0 \tag{2.4}$$

This time difference must account for beginning and end of week crossovers.

The mean anomaly at t_k can be found by

$$M_k = n' t_k \tag{2.5}$$

Kepler's relates the mean and eccentric anomalies by

$$M = E - e\sin E \tag{2.6}$$

which can be rearranged and solved iteratively for E_k. Hence

$$E_{k_i} = M_k - e\sin E_{k_{i-1}} \tag{2.7}$$

The true anomaly, f_k, is then found by

$$\cos f_k = (\cos E_k - e) / (1 - e\cos E_k) \tag{2.8}$$

The next step is to calculate the argument of latitude - that is the angle subtended at the geocentre between the satellite and the Ascending node. For a normal orbit, this is given by

$$\phi_k = f_k + \omega \tag{2.9}$$

Due to the orbit perturbations, a correction is necessary and is determined by

$$\delta u_k = C_{us}\sin 2\phi_k + C_{uc}\cos 2\phi_k \tag{2.10}$$

The corrected argument of latitude u_k is thus given by

$$u_k = \phi_k + \delta u_k \tag{2.11}$$

The radial distance, r_k, from the geocentre to the satellite is computed in a similar manner

$$r_k = A(1 - e\cos E_k) + \delta r_k \tag{2.12}$$

where

$$\delta r_k \approx C_{rc}\cos 2\phi_k + C_{rs}\sin 2\phi_k \tag{2.13}$$

The orbital coordinates of the satellite are then obtained

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_o^s = \begin{bmatrix} r_k \cos u_k \\ r_k \sin u_k \\ 0 \end{bmatrix} \tag{2.14}$$

2.4.2 The Satellite's Position within an Earth-Fixed Frame

In order to bring the system into the earth-fixed WGS84 coordinate system, the orbital coordinates must be rotated. The system must first be rotated about the X-axis until the XY-plane coincides with the equatorial plane. The system is then rotated about the Z-axis until the X-axis is aligned with the WGS84 X-axis (the zero meridian).

The corrected inclination angle between the orbital and equatorial planes and referenced to time t_k is given by

$$i_k = i + t_k\dot{i} + C_{is}\sin 2\phi_k + C_{ic}\cos 2\phi_k \tag{2.15}$$

The angle between the X-axis directions of the orbital and earth-fixed systems is known as the longitude of the ascending node. The traditional method of obtaining this angle is by subtracting the time oriented GAST angle from the right ascension of the ascending node. As mentioned in 2.3.1, the GPS Keplerian element Ω_0 is the longitude of the ascending node from the beginning of the GPS week. This

avoids the necessity to compute GAST, but the rotation of the earth must be accounted for. The corrected longitude of the ascending node, λ_k, is found by

$$\lambda_k = \Omega_0 + (\dot{\Omega} - \omega_e)t_k - \omega_e t_o \tag{2.16}$$

where

ω_e is the WGS84 rate of the earth rotation ($7.292115.10^{-5}$ radians / sec).

The earth-fixed WGS84 coordinates of the satellite are finally computed by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^S_E = \begin{bmatrix} \cos\lambda_k & -\cos i_k \sin\lambda_k & 0 \\ \sin\lambda_k & \cos i_k \cos\lambda_k & 0 \\ 0 & \sin i_k & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^S_O \tag{2.17}$$

During the time taken for the signal to travel from the satellite to the receiver, the earth fixed coordinates of the satellite will have changed due to earth rotation. For positioning purposes, therefore, it is usual to rotate the coordinates to the time of signal reception. Hence if t1 is the time of transmission from the satellite and t2 the time of reception at the receiver, then the coordinates will be rotated about the Z axis by $(t2 - t1)\omega_e$.

CHAPTER THREE
GPS POSITIONING ROUTINES AND THE RINEXPOS UTILITY PROGRAM

3.1 PREAMBLE

A utility program has been developed to perform a variety of GPS pseudorange positioning computations. This includes standalone point positioning and the traditional two station DGPS positioning. The program is called RINEXPOS (RINEX POSitioning program) and has been written in ProPascal Version 5.2[1].

The options available in RINEXPOS Version 92.003 along with the algorithms which have been incorporated are described in this chapter. Different algorithms that can be easily implemented are also given where appropriate, thus enabling future enhancement of the program. All the screen views of the program having been captured using a PCX grab utility and have then been converted to show black text on a white background. This improves the clarity when seen on paper, although all screen colours and menu highlighting are not shown.

3.2 THE RINEXPOS MENU STRUCTURE

The program is activated by simply typing RINEXPOS at the command line, which results with the main menu as shown in Fig 3.1. The three options within this menu are to input the data files, to change the default running parameters and to

[1]ProPascal is a copyright of Prospero Software, Inc.

carry out the positioning computations. The latter will be described later in this chapter since it depends on the selections made within the other menus.

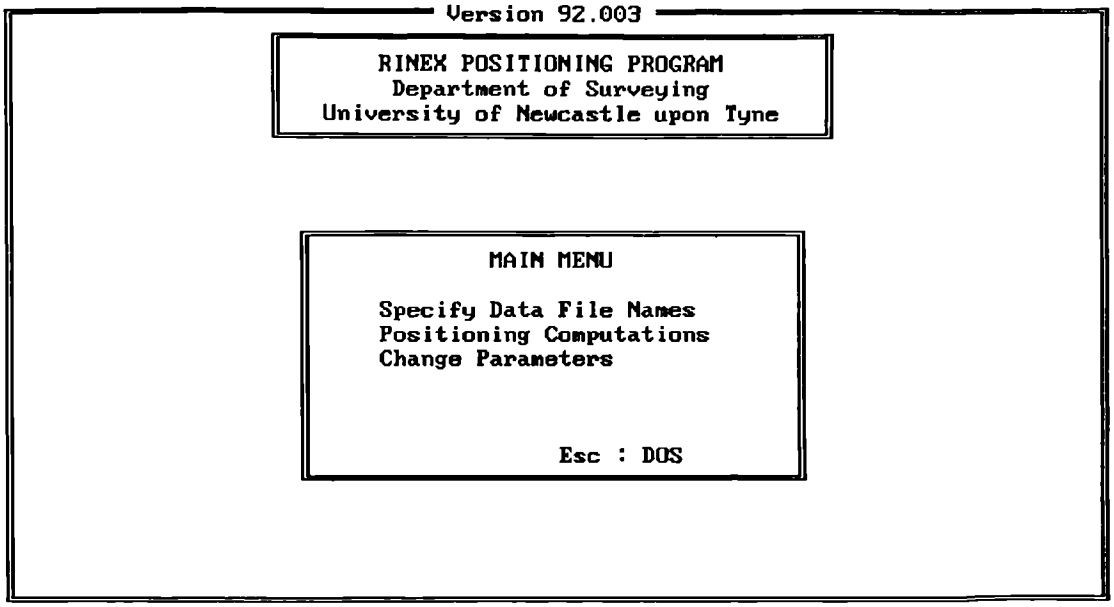


Fig 3.1 The RINEXPOS Main Menu

3.2.1 Specify Data File Names

The data files required for the input are:

- A navigation file containing the ephemeris data for all the satellites viewed during the observation period.
- An observation data file containing the pseudorange observations recorded at the reference station.
- A similar file for the observations recorded at the mobile station.
- The name for the results file into which the final positions will be placed.

Depending on the mode of operation (see 3.2.2) that has been selected, it is not necessary to input all the data files. For instance if the satellite positioning mode has already been selected, it is only necessary to input the ephemeris file and a results file, and for point positioning the mobile station data file is not required.

Since the default mode is for DGPS all the files are initially required if this option is selected first, although the unnecessary files are not opened if options have been changed at a later stage. Only files that exist are allowed to be input, except for the results file for which an overwrite warning is produced if a file of the same name is present in the working directory. All the data input files (ie except for the results file) must be in the RINEX format.

3.2.1.1 The RINEX data format

There are many different companies producing GPS receivers all of which hold the data recorded by the receiver in a different manner. The data is usually split into various binary files which need decoding particular to the manufacturer's format. Almost without exception, one of the files will contain the ephemeris information sent down with the GPS navigation message, and one will contain the observables recorded from the two frequencies transmitted from the satellites. Other files that might exist include one for site information, and one for the meteorological data.

The Receiver INdependent EXchange (RINEX) format has been developed so that data from different receiver types can be easily exchanged and can be processed using software written to accept one common format. Most software for different receivers contain a program which will convert from the original binary formats to a common RINEX ascii format. Apart from now being in a well documented standard format, the data files can be easily viewed since they are now ascii. This in turn enables an immediate understanding of the quantities involved.

The first proposal for the RINEX format was developed by the Astronomical Institute of the University of Berne (Gurtner and Mader, 1990). For details of the present formats (RINEX Version 2) and a more precise description of the data, the

reader should refer to ibid. and Gurtner (1993). The recommend naming convention for RINEX files is (Gurtner, 1993)

ssssdddf.yyt

where	ssss	is a four character station name	
	ddd	is the modified Julian day (the day of year)	
	f	the file sequence number	
	yy	the last two digits of the year	
	t	file type :	
		O	observation file
		N	navigation file
		M	meteorological data file

All the different types consist of lines with a maximum of 80 characters. There is general header information at the beginning of each file before the actual data is provided. Short descriptions of the two main RINEX file types (the navigation and the observation files) will suffice for this thesis and examples are given in Appendix A.

3.2.1.2 The navigation message file

The RINEX navigation message provides all the necessary elements for the determination of the satellites positions. The header information simply comprises of descriptions of the conversion programs and their executed dates with the RINEX version that has been generated. Optional records within the header include the eight Klobucher ionospheric model constants ($\alpha_{0..3}$ and $\beta_{0..3}$), and polynomial parameters representing the offset of GPS time and UTC. The main section of the

navigation message comprises of the following information for each satellite at all hourly updates.

- The ephemeris time tags (the time of ephemeris (t_{oe}) and the transmission time of message).
- The Keplerian elements (the seven [see 2.3.1] elements describing the normal GPS orbit).
- The elements modelling the perturbing forces acting on the satellite,
- The three coefficients of the second order polynomial describing the satellite clock correction.
- The satellite health and accuracy measures (the Health flag and the user equivalent range error (UERE)) as well as ephemeris confidence flags (the age of data ephemeris (AODE) and the age of data word (AODC)).
- Atmospheric refraction delay parameters in the form of the Time Group Delay (T_{GD}).

3.2.1.3 The observation data file

The header record within the observation record contains general information on the RINEX data version, receiver type used, the operators, and the conversion programs that generated the RINEX data. The header also contains some more information including

- The approximate position of the antenna at the start of the survey in WGS84 cartesian coordinates.
- The antenna height and offset from the survey marker used.
- The wavelength of the carrier phase observables that will follow within the main section. This provides the knowledge of whether or not a squaring procedure has been used to recover the carrier thus halving the relevant wavelengths.

- Details of the types of observations that are included for all epochs within the data file and their order within the file. There are eight different observables that can be included for GPS data files, which are
  - C/A code full pseudorange (ie the 300 km ambiguity has been resolved) on the L1 carrier. Measured in metres.
  - P code full pseudoranges on L1 and L2. Measured in metres.
  - A derived C/A code full pseudorange on L2 (the C/A code on L1 plus the difference between the measured P or Y code distances on L1 and L2). Measured in metres.
  - Carrier phase observations on L1 and L2. Measured in cycles.
  - Doppler measurements on L1 and L2. Measured in Hertz.
- Information on the observation times, the observation interval and the number of different observation types made to each satellite.

After the header, the main section consists of information recorded at each epoch during the session. There is capability for comment lines describing operations during a kinematic or pseudo kinematic survey, but in general each epoch will consist of

- The time tag to which all observations have been related. This is given within the receiver time frame.
- The satellite PRN numbers which have been observed at this particular epoch. There is provision for GPS, GLONASS and Transit satellites which may have been observed with a combined/integrated receiver. For the case of Transit, there is a further data type for the Integrated Doppler on the two Transit frequencies (150 and 400 MHz).
- The receiver clock offset as determined within the receiver relating the receiver time frame to GPS system time.

- For each satellite the values for all the observation types follow with signal strength and loss of lock indicators. Information for the different satellites is given on different lines.

3.2.2 Change Parameters

The main menu option for changing parameters leads to a further menu as shown in Fig 3.2. All the options available in this menu are briefly described.

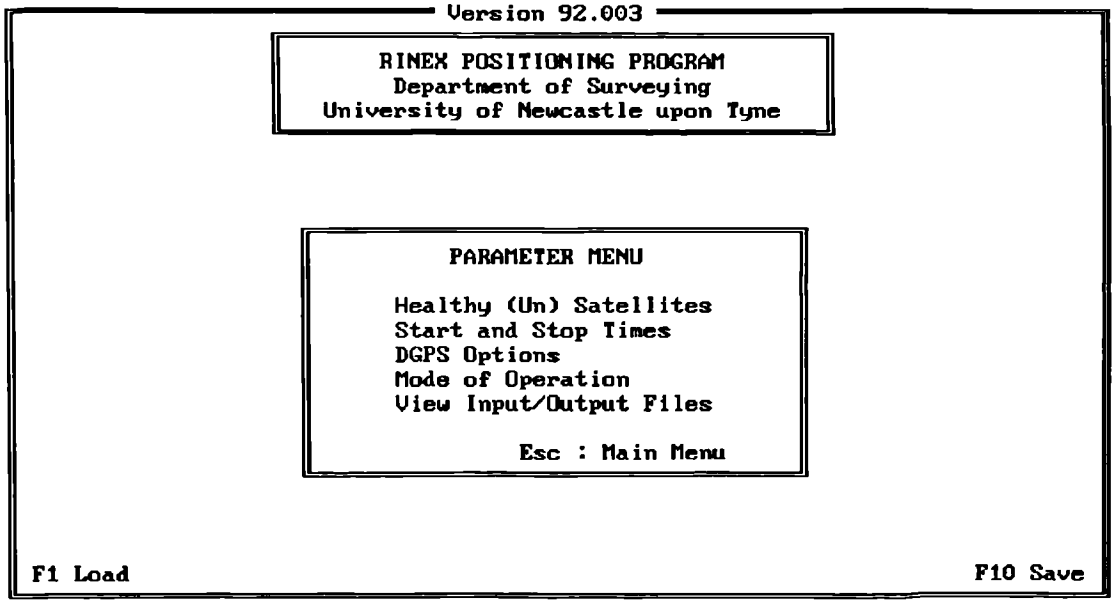


Fig 3.2 The RINEXPOS Parameter Menu

3.2.2.1 Healthy (un) satellites

This permits the user to turn on and off satellites which may have been flagged unhealthy. It also allows just the Block I or II satellites to be used. The default setting for this option is to use all satellites with PRN numbers 2, 3, 6, 9 and 11 through to 32 as shown in Fig 3.3. There is no provision within the program to automatically turn off unhealthy satellites as flagged within the navigation message.

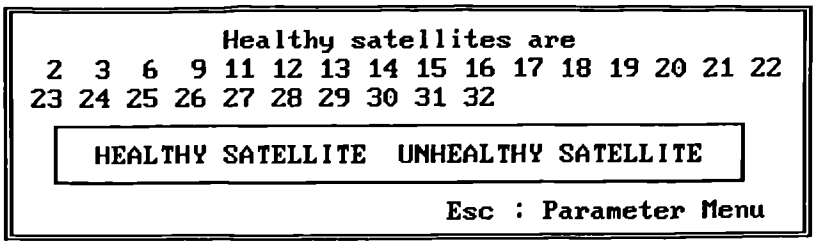


Fig 3.3 RINEXPOS healthy satellite selections

3.2.2.2 Start and stop times

This option permits the user to specify the amount of data that will be processed. The choices are either by epoch number (epoch one being the first observation within the reference station data file), by time of day, or by processing the whole file. All these choices relate to the reference station file except when carrying out satellite positioning when the time of day indicates the interval for which these parameters are required. An example from the program with a specified start/stop time is shown in Fig 3.4.

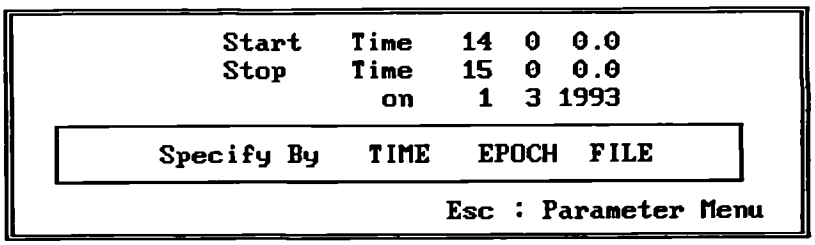


Fig 3.4 Start / stop times selection

3.2.2.3 Mode of operation

Three different modes of operation are allowed within RINEXPOS, namely DGPS positioning, point positioning and satellite positioning with the default being DGPS. The satellite positioning option deserves a further explanation since it is used for

comparison of satellite positions output from other programs developed at Newcastle.

The satellite positioning option is used to compute the WGS84 coordinates of all the satellites within the navigation file during a period of the day which is specified through the "Start and Stop Times" option. The coordinates are in their cartesian form (ie XYZ) and the times relate to the GPS time of transmission from the satellites and there have been no earth rotation effects applied. The algorithms used are those described in chapter two, and the time interval between successive epochs is set to five seconds.

3.2.2.4 View input/output files

This option allows a scrolled view of all the input and output files that have been specified or created. Since all the files are in an ascii format, this simply reads the appropriate file and scrolls the output to the screen. Progression down the file is either by each line, by a page, or continuous.

3.2.2.5 DGPS options

Choosing this option enters a third menu allowing further selections to be made.

3.2.3 The Differential GPS Menu

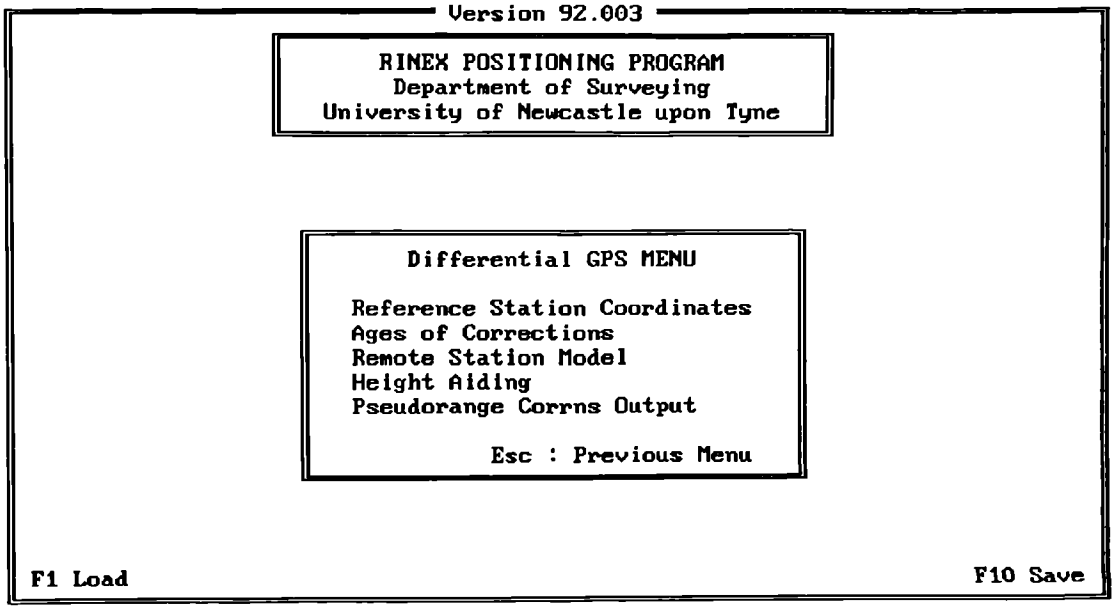


Fig 3.5 The Differential GPS Menu

3.2.3.1 Reference station coordinates

This option allows for the input of the reference station coordinates which are used to determine the differential pseudorange corrections. The algorithm for this is given later in this chapter in section. The coordinates are all WGS84 ellipsoidal coordinates and an example is given in Fig 3.6.

REFERENCE STATION COORDINATES		
Name		Newcastle
Latitude	N	54 58 40.2100
Longitude	W	1 36 56.7500
Height		134.0000

Fig 3.6 The Reference station coordinates

3.2.3.2 Ages of correction

This option attempts to imitate the latency in a true differential system (due to the data link required to transfer the pseudorange corrections from the reference station to the remote station) by introducing a delay within the two data files. This also allows for an investigation on the importance of the rate at which a data link can transfer the corrections which will be affected mainly by high frequency SA dither. The age of correction can be input so that the program will process the data set with any specified latency between the reference station and the remote. A tolerance level is also required which accounts for the fact that the observation times at the two stations will not be completely synchronous. Fig 3.7 is an example of the default settings for this option and the program will process the data sets in the following manner. The pseudorange corrections will be derived at the reference station and applied to the remote station data that has been recorded at exactly the same epoch (age of correction zero seconds) with the tolerance applied. The two data sets are then processed in this way. The data is then processed again but with an age of correction of five seconds - this means that the pseudorange corrections are applied to the remote station at the epoch that was observed five seconds after the one from which they were generated at the reference station. This is continued until the age of correction is greater that sixty seconds.

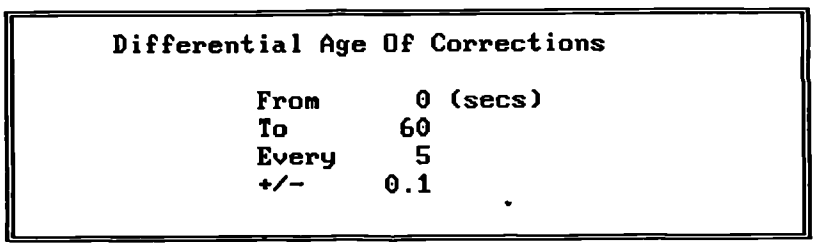


Fig 3.7 Default age of corrections

3.2.3.3 Remote station model

When processing in the differential mode, there is an option to allow the mobile station to be processed using a least squares solution or by using the Kalman filter prediction and filtering algorithms.

The purpose of the latter is to reduce the level of noise within the solution by filtering in the position domain and relies on the fact that the dynamic nature of the mobile station is known, ie for instance there is previous knowledge that the station is moving in a straight line. Three different models for these dynamics are catered within RINEXPOS and these are a stationary model, a straight line model and a model allowing the mobile to turn. The number of parameters that are solved for within the solution thus also change from five (three position coordinates plus a receiver clock offset and its drift) to eight (three additional velocity components acting along the coordinate axes) to ten (horizontal accelerations) respectively. The algorithms used within the Kalman filter solution are provided in chapter six, equations 6.18 through 6.23. An important consideration for the Kalman filter is the values within the stochastic models which are used to determine the amount of importance that is placed on the knowledge of the mobile's dynamics. These are input via a control file named "kalman.inp" which comprises of the following lines:

1. The number of parameters to be solved for (depending on the required model). The options are 10, 8, or 5 with 10 as the default.
2. The standard deviation of the latitude forcing function. For instance, for the 10 parameter model, the forcing function is the rate of change of acceleration in latitude and the units will be in radians/second^3. A default value of $1x10^{-11}$ is used which equates to approximately $6x10^{-5}m/s^3$ for the United Kingdom latitudes.

3. The standard deviation for the longitude forcing function. This is the same as above and uses the same default values (equating to approximately $5x10^{-5}m/s^3$).
4. The standard deviation for the forcing parameters in height. The 10 state model, this is height accelerations and the units will be in m/s^2, with the default value set to $1cm/s^2$.
5. The standard deviation of the forcing function for the receiver clock offset. For all models, this is its acceleration and a default of $1x10^{-8}s/s^2$ is set.
6. The standard deviation of the GPS pseudorange observations in metres. The best value for this parameter has been found to be three metres which is used as the default. By reducing this number (ie increasing the precision of the pseudorange observations) less emphasis is placed on the mobile's dynamics.
7. A switch (set to 1 for on, or 0 for off) that controls the amount of output to be written to file. The on setting enables all the stages of the Kalman filter processing to be output, whilst off surpresses all but the final positions.

The settings described above play an important role in the performance of the Kalman filter. If too much weight is placed on the dynamic model of the mobile, then the position information derived from the GPS observations may be ignored. This can produce an extremely smooth vessel track which may appear to be a very good solution, however any wandering of the vessel (eg due to tidal or wind changes, or simply turning a corner) may not be picked up within the solution. The default values have been selected by determining the most appropriate values for DGPS test data.

The least squares solution produces a position using only the GPS pseudorange observations with no input from the knowledge of the motion of the mobile. This has the advantage over the Kalman filter as the problems outlined above will not occur, but the solution will not appear as smooth. In order to create this "smooth" track via the least squares approach, the noise within the actual GPS observations

must be reduced - a procedure that is not carried out within this version of the software, but is often performed internally within the receiver. The least squares approach simply solves for four parameters (three positional components and the receiver clock offset) and this is the only type of solution allowed when in point positioning mode.

3.2.3.4 Height aiding

This option allows for an external height observation to be used within the solution (either traditional least squares or Kalman filter). This is a height above the ellipsoid which must be input along with its associated standard deviation. If a least squares solution is used, the pseudorange observations are weighted according to their standard deviations as calculated from their User Range Accuracy (URA) values within the GPS ephemeris file - see Drewett (1989) for further details. If a Kalman filter solution is used, the pseudorange observations are weighted according to the input file as described in 3.2.3.3. The algorithms used to implement height aiding are discussed later in section 3.3.6.

3.2.3.5 Pseudorange corrections output

This option enables the output of the pseudorange corrections as determined at the reference station to be output to file for later analysis.

3.3 ALGORITHMS FOR GPS PSEUDORANGE POSITIONING

The following section describes many of the algorithms which have been used for the positioning computations within RINEXPOS. The details on determining the satellite positions, the reader should refer to Chapter 2. Where appropriate, further

equations have been documented and can be easily incorporated into this program or into future GPS positioning programs.

3.3.1 Satellite and Receiver Time Frames

When considering GPS computations, it is typical to refer to three different time frames. The first two are local frames which are maintained by a local oscillator which will experience drift. These are the receiver time frame (all receivers will keep a slightly different time) and the satellite time frame (again, this will be different for all satellites). It is necessary to relate these time frames to a common frame, and GPS system time is most typically used.

GPS system time is a continuous frame that is kept by the atomic clocks at the GPS Master Control Station and was synchronised to UTC (Universal Time Coordinated) at midnight on 5/6 January 1980. UTC experiences leap second increments to keep in line with the secular drift of the Earth's rotation, and is therefore now offset from GPS system time by a number of integer seconds. A continuous offset of 19 seconds relates GPS system time with International Atomic Time which is the fundamental time scale for all time-keepers on the Earth and is kept at the Bureau International de l'Heure in Paris (Anon, 1989).

The relation between the local time frames kept by the receivers and satellites and GPS system time can be summarised as

$$T_{GPS} = T_R - dT_R \qquad (3.1)$$

where T_{GPS} is GPS system time

T_R is local receiver time

dT_R is the local receiver clock offset

$$T_{GPS} = T_S - dT_S \tag{3.2}$$

where T_{GPS} is GPS system time

T_S is local satellite time

dT_S is the local satellite clock offset

3.3.2 Pseudorange Transmit Time

In order to compute the satellite's earth fixed coordinates, it is first necessary to determine the time at which the particular observable was transmitted. This is needed in the GPS time frame and can be determined via the pseudorange measurement in the following manner. The superscripts are used to describe the time at which the range was transmitted, and the subscripts for the time at which the range was received. The pseudorange (PR) is defined for the RINEX format (Gurtner and Mader, 1990) as

$$PR/c = T_{GPS} + dT_R - (T^{GPS} + dT^S) \tag{3.3}$$

where T_{GPS} is the true received time in the GPS time frame

T^{GPS} is the true transmission time in the GPS time frame

c is the speed of light in a vacuum.

From (3.3) and (3.1), it is clear that the transmission time of a particular pseudorange is obtained from

$$T^{GPS} = PR/c - T_R - dT^S \tag{3.4}$$

where T_R is the received time in the receiver time frame.

T_R is the actual time tag to which all the observations are related and is thus known, and dT^S is determined from an algorithm which requires a few iterations (see 3.3.3). This technique assumes that errors within the pseudorange measurement would be small enough to have no significant effect on the determined satellite's position. A large pseudorange error could be typically 100 m (a summation of SA, ionospheric, multipath, and receiver noise errors), which will be a time error of approximately 333 ns. A GPS satellite travels at approximately 3000 m/s and therefore covers a distance of one mm over the time error interval.

3.3.3 Satellite Clock Offset

The satellite clock offset, dT_S, is the relationship between GPS system time and the time frame kept within each satellite including all relativistic effects. Its value is estimated by a second order polynomial whose parameters are transmitted within the navigation message. The polynomial equation is (Van Dierendonck et al, 1980)

$$dT_S = a_0 + a_1(t - t_{oe}) + a_2(t - t_{oe})^2 \qquad (3.5)$$

where a_0, a_1, a_2 are the transmitted polynomial coefficients
t is the time of signal transmission
t_{oe} is the time of ephemeris which is transmitted within the navigation message.

The a_2 term often has a value of zero, and therefore a first order (straight line representation) is suffice to model the satellite clock offset. In order to compute the time of transmission from equation (3.4), it is necessary to first calculate dT_S and this is carried out by iteration of (3.5) and (3.2) with the initial value of t being the received time. The sensitivity of the clock offset polynomial to the value of t is not

great enough to differentiate between the signal transmission time in the satellite frame or in GPS system time.

3.3.4 Receiver Point Positioning

The algorithms presented below are used within RINEXPOS for the positioning of a single station (point position mode), and for the positioning of the differentially corrected pseudoranges both using least squares or the Kalman filter. The algorithms have been derived from the basic mathematical models which are then solved for within the program using the least squares observation equation approach or the Kalman filter. No justification is given for using a least squares approach as this is well documented and further details can be obtained from Gelb (1974) or Cross (1983).

3.3.4.1 Basic mathematical model

Simultaneous pseudorange observations from different satellites are used as lines of position to compute the three dimensional coordinates of the receiver (or more correctly the position of the phase centre of the antenna) along with it's clock offset from GPS system time. If we define the true coordinates of the receiver in a cartesian system and the clock offset as $\left(\overline{X}_R \quad \overline{Y}_R \quad \overline{Z}_R \quad d\overline{T}_R\right)$, then a vector of unknowns (the parameters we wish to solve for) can be written as:

$$\overline{x} = \left[\overline{X}_R \quad \overline{Y}_R \quad \overline{Z}_R \quad d\overline{T}_R\right]^T \tag{3.6}$$

We can also define l as a vector of observed pseudoranges to different satellites which is related to their true values $\overline{l}$ by:

$$\overline{l} = l + v \tag{3.7}$$

where v is a vector of residuals.

The general relationship between the vector of observations and the vector of unknowns is given by: (Cross, 1983)

$$F(\bar{x}) = \bar{l} \tag{3.8}$$

For each satellite in view, (3.8) can be written in a more specific form as: (Drewett, 1989)

$$\sqrt{\left((\bar{X}_R - X_S)^2 + (\bar{Y}_R - Y_S)^2 + (\bar{Z}_R - Z_S)^2\right)} - d\bar{T}_R.c = \bar{l} \tag{3.9}$$

where c is the speed of light in a vacuum, and

$\mathbf{X_S}$ are the true coordinates of the satellite.

3.3.4.2 Solving the mathematical model

It is usual to solve for the vector of unknown parameters using the least squares observation equation approach (see Cross [1983] for further details). Since equation (3.9) is non-linear, it is necessary to first linearise the equation by using the Taylor's series expansion. For this, the unknowns need to be split into two parts - one being an approximation to their true values ($\tilde{x}$) and the other being a correction to this approximation (Δx). The general function (equation 3.8) can now be expressed as

$$F(\bar{x}) = F(\tilde{x} + \Delta x) = l + v \tag{3.10}$$

which when expanded via the Taylor's series to the first differentials produces

$$F(\tilde{x}) + (\partial F / \partial \tilde{x})\Delta x = l + v \tag{3.11}$$

This now represents a linear set of equations relating the new parameters, Δx, to the observables, l. Equation (3.11) can be rearranged so that the parameters are on the left hand side and the measured and known quantities are on the right hand side, thus forming the well known least squares equations, $\mathbf{Ax} = \mathbf{b} + \mathbf{v}$.

The pseudorange equation (3.9) is partially differentiated with respect to the provisional values in the following way.

$$\partial F / \partial \tilde{X}_R = \frac{\tilde{X}_R - X_S}{\sqrt{\left((\tilde{X}_R - X_S)^2 \quad (\tilde{Y}_R - Y_S)^2 \quad (\tilde{Z}_R - Z_S)^2\right)}} \tag{3.12}$$

By representing the geometric distance between the satellite and the receiver as ρ_{RS}, (3.12) can be written as

$$\partial F / \partial \tilde{X}_R = \frac{\tilde{X}_R - X_S}{\rho_{RS}} \tag{3.13}$$

Similarly, the remainder of the partial differentials are :

$$\partial F / \partial \tilde{Y}_R = \frac{\tilde{Y}_R - Y_S}{\rho_{RS}} \tag{3.14}$$

$$\partial F / \partial \tilde{Z}_R = \frac{\tilde{Z}_R - Z_S}{\rho_{RS}} \tag{3.15}$$

$$\partial F / \partial d\tilde{T}_R = -c \tag{3.16}$$

Suppose now there are n satellites for which pseudorange observations have been observed. Rearranging equations (3.11) to (3.16) produces a system of n equations, such :

$$
\begin{aligned}
&(\partial F_1 / \partial \tilde{X}_R)\Delta X + (\partial F_1 / \partial \tilde{Y}_R)\Delta Y + (\partial F_1 / \partial \tilde{Z}_R)\Delta Z - c.\Delta dT = l_1 - F_1(\tilde{x}) + v_1 \\
&(\partial F_2 / \partial \tilde{X}_R)\Delta X + (\partial F_2 / \partial \tilde{Y}_R)\Delta Y + (\partial F_2 / \partial \tilde{Z}_R)\Delta Z - c.\Delta dT = l_2 - F_2(\tilde{x}) + v_2 \\
&\vdots \\
&(\partial F_n / \partial \tilde{X}_R)\Delta X + (\partial F_n / \partial \tilde{Y}_R)\Delta Y + (\partial F_n / \partial \tilde{Z}_R)\Delta Z - c.\Delta dT = l_n - F_n(\tilde{x}) + v_n
\end{aligned}
\tag{3.17}
$$

These now form a set of linear observation equations where :

$$
\mathbf{A} = \begin{bmatrix}
\partial F_1/\partial \tilde{X}_R & \partial F_1/\partial \tilde{Y}_R & \partial F_1/\partial \tilde{Z}_R & -c \\
\partial F_2/\partial \tilde{X}_R & \partial F_2/\partial \tilde{Y}_R & \partial F_2/\partial \tilde{Z}_R & -c \\
\vdots & \vdots & \vdots & \vdots \\
\partial F_n/\partial \tilde{X}_R & \partial F_n/\partial \tilde{Y}_R & \partial F_n/\partial \tilde{Z}_R & -c
\end{bmatrix}
$$

$$
\mathbf{x} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \\ \Delta dT \end{bmatrix}, \mathbf{b} = \begin{bmatrix} l_1 - F_1(\tilde{x}) \\ l_2 - F_2(\tilde{x}) \\ \vdots \\ l_n - F_n(\tilde{x}) \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}
\tag{3.18}
$$

Since there are four unknowns, a minimum of four satellites need be observed to compute the complete solution. When using RINEXPOS in the point positioning mode or the least squares differential mode, this approach is taken and the coordinates are computed in the cartesian system.

3.3.4.3 Solving for Geographical Coordinates

It is sometimes desirable to separate the parameters into horizontal and vertical components, in order to include other equations, or observation types, into the system - for example, knowledge of velocities over the ground, knowledge of the

course over ground, and external height information may be readily included in a model solving for the parameters in their horizontal and vertical directions.

To solve for latitude, longitude and height it is necessary to redefine function (3.9) so that it now consists of $\bar{\phi}_R, \bar{\lambda}_R, \bar{h}_R$. The conversion between cartesian and geographical coordinates can be found in many texts, for instance Heiskanen and Moritz (1967)and Cross et al (1981), and leads to the functional model :

$$\left(\left((N+\bar{h})\cos\bar{\phi}_R \cos\bar{\lambda}_R - X_S\right)^2 + \left((N+\bar{h})\cos\bar{\phi}_R \sin\bar{\lambda}_R - Y_S\right)^2 + \left(\left(\frac{b^2}{a^2}N+\bar{h}\right)\sin\bar{\phi}_R - Z_S\right)^2\right)^{1/2} - d\bar{T}_R \times c = \bar{l} \qquad (3.19)$$

where N is the radius of curvature in the prime vertical.

This equation needs to be partially differentiated with respect to the geographical coordinates which is best done in a numerical manner. This leads to the formation of the **A** matrix and a system of equations similar to (3.18). The Kalman filter within RINEXPOS uses this approach for the formulation of the observation equations.

3.3.5 Dilution of Precision

The dilution of precision (DOP) is an indication of the quality of the results that can be expected from a GPS point position. It is a measure based solely on the geometry of the satellites and therefore can be computed without any pseudorange observations being recored.

DOP values are often expressed in different terms relating to the propagation of the satellite configuration into the position fix in its different components. The usual figures are :

GDOP	geometrical dilution of precision
PDOP	positional dilution of precision
TDOP	time dilution of precision
HDOP	horizontal dilution of precision
VDOP	vertical dilution of precision

The values are obtained from the inverse of the normal equation matrix of the solution. Suppose the design matrix is as in (3.18), then the cofactor matrix, **Cx**, is a symmetric 4x4 matrix that can be represented as (Hofmann-Wellenhof et al, 1992)

$$\mathbf{Cx} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1} = \begin{bmatrix} c_{XX} & c_{XY} & c_{XZ} & c_{XT} \\ & c_{YY} & c_{YZ} & c_{YT} \\ & & c_{ZZ} & c_{ZT} \\ & & & c_{TT} \end{bmatrix} \qquad (3.20)$$

The diagonal element are used for the following DOP definitions :

$$\begin{aligned} GDOP &= \sqrt{c_{XX} + c_{YY} + c_{ZZ} + c_{TT}} \\ PDOP &= \sqrt{c_{XX} + c_{YY} + c_{ZZ}} \\ TDOP &= \sqrt{c_{TT}} \end{aligned} \tag{3.21}$$

The horizontal and vertical DOPs can be computed from similar elements from the covariance matrix by either rotating **Cx** into a topocentric coordinate system (Drewett, 1989) or directly by computing the system in geographical coordinates. presently, RINEXPOS only compute the PDOP value.

It is most critical that DOP values should be used as an indication of when GPS is likely to produce good positioning results, and not as a measure that describes the quality of positioning that has actually taken place. There are a number of reasons why a DOP value, measuring geometry, may be misleading if used to describe resultant positions :

1. There may be outliers in the pseudorange observations resulting in a poor fix. This would not be picked up by a DOP value.
2. Low elevation satellites will improve the geometrical configuration. However, ranges observed to these satellites will have large atmospheric errors and again will lead to poor positioning.
3. There is no indication of the level and rate of Selective Availability which introduces errors on all observations.

3.3.6 Height Aided Solutions

The knowledge of the receiver's height above the ellipsoid is often used as an extra equation within the positioning solution. Originally this was carried out to reduce the number of pseudorange observations required to obtain a fix (3 are now needed instead of 4), but it is now used to add further redundancy into the solution especially in times of poor geometric configurations. Essentially the height observation can be written as distance from an imaginary point (p), so that:

$$\sqrt{\left(\left(\overline{X}_R - X_P\right)^2 + \left(\overline{Y}_R - Y_P\right)^2 + \left(\overline{Z}_R - Z_P\right)^2\right)} = \overline{h} \qquad (3.22)$$

Once the coordinates of the point have been found with the relevant height estimation, the function can be differentiated with respect to the receiver coordinates and these values can be included as an extra line within the observation equations. There are three separate approaches for implementing height aiding, all of which relate to Fig 3.8.

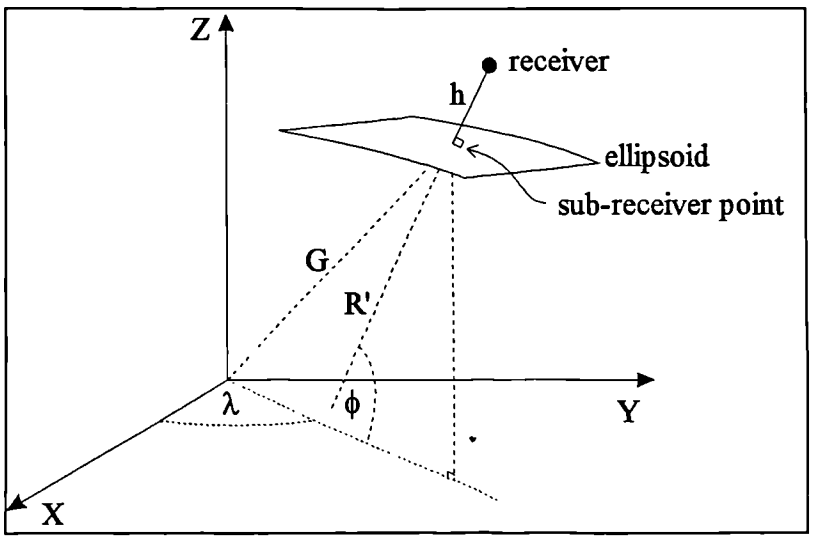


Fig 3.8 Height aiding determination

3.3.6.1 Height above the ellipsoid

The height above the ellipsoid (h) is the required distance. The coordinates of the sub-receiver point (ie the point from which the distance is measured) are $(\phi_R, \lambda_R, 0)$ with the coordinates of the receiver being (ϕ_R, λ_R, h_R). The provisional coordinates of the receiver can therefore be used to deduce the coordinates of the sub-receiver point, which can then be converted into the cartesian system. These coordinates and the ellipsoidal height can then go directly in equation (3.22). This approach is currently implemented within RINEXPOS.

3.3.6.2 Distance from the centre of the earth

The distance from the centre of the earth to the receiver point can be approximated from Fig 3.8 as

$$d = G + h \tag{3.23}$$

Again for this technique, the cartesian coordinates of the sub-receiver point must be determined in order to allow the distance G to be calculated. Suppose the cartesian coordinates of the sub-receiver point are (X_{SR}, Y_{SR}, Z_{SR}), then G can be determined by

$$G = \sqrt{Z_{SR}^2 + X_{SR}^2 + Y_{SR}^2} \tag{3.24}$$

The distance d measured from coordinates (0,0,0) can now be used directly in equation (3.22).

3.3.6.3 The direction of the vertical

Instead of finding the cartesian coordinates of a point from which a "height distance" can be derived, equations for the direction of the vertical (given by R' in Fig 3.8) can be used and numerically differentiated to form part of the observation equations. The normal radius of curvature along the vertical is often denoted as N, and can be used with the ellipsoidal height to derive a distance equation. Hence (Heiskanen and Moritz, 1967)

$$N+h=\frac{c}{\sqrt{1+e'^2\cos^2\phi}}+h \tag{3.25}$$

where $c=\frac{a^2}{b^2}=6399593.6258\text{m}$ for WGS84 (Anon, 1987)

$$e'^2=\frac{\sqrt{a^2-b^2}}{b}=0.00673949674227 \text{ for WGS84}$$

3.3.7 Pseudorange Corrections

In DGPS, pseudorange corrections are generated at the reference station (which has known coordinates) and are then applied to the appropriate pseudoranges observed at the mobile station. The corrections are found by:

$$lc_i = d_i - l_i + dT_{Si} - dT_R \tag{3.26}$$

where

lc_i	is the pseudorange correction to satellite i
d_i	is the geometric distance computed between the coordinates of satellite i and the true coordinates of the reference station
l_i	is the observed pseudorange to satellite i
dT_{Si}	is the clock correction for satellite i in metres

dT_R is the receiver clock correction at the reference station in metres.

Therefore the RINEXPOS program computes a DGPS solution at the mobile station by performing the following steps.

1. All the pseudoranges observed at the reference station are used in a least squares solution to determine the receiver clock offset and the station coordinates at that epoch.
2. The pseudorange corrections are found based on the known coordinates of the reference station, and the clock offsets.
3. These corrections are added to the observed pseudoranges at the mobile station.
4. The corrected pseudoranges at the mobile are processed in the same way as step 1 to produce the three-dimensional coordinates at the mobile station along with its receiver clock offset.

The solution at the mobile must only be carried out using pseudorange observations that have corrections available. This is the reason that reference stations have multi-channel receivers and will record with a relatively low elevation mask setting, thus allowing corrections to be generated for all available satellites. The clock offset that has been computed at the mobile station will now be contaminated by any errors that have been introduced at the reference station when computing its receiver clock offset (atmospheric errors, receiver noise, multipath, etc). Indeed it is not necessary to compute the clock offset at the reference station and many earlier DGPS services did not do this. Instead an estimate for the clock offset is made, either by using a polynomial approximation or by computing the clock offset based on an assumption that the correction for one particular satellite (the base satellite) is zero. This reduces the computational requirements at the reference station end as there is no longer a need for a least squares calculation (with matrix multiplication

and inversion). In these circumstances, a combination of the two receiver clock offsets and other errors are computed within the four parameter least squares solution at the mobile station. This type of approach is no longer recommended as the pseudorange corrections may appear extremely large (especially if a base satellite is used) and may suddenly "jump" if there has been a change in base satellite or if the approximation is reset. Whilst having no effect on the results at the mobile station, this does little to install confidence with the operators of the system.

3.4 RINEXPOS PERFORMANCE AND OUTPUT

3.4.1 General Screen Output

When carrying out the positioning computations, the following output is displayed to the screen:

1. The satellite numbers for which data has been obtained from the input ephemeris file.
2. The epoch being processed at the reference station and, if in differential mode, at the mobile station. Epoch number one is specified as the first time-tagged data within each observation file, independent on whether it was used for positioning (ie due to lack of data, poor observations, etc).
3. The current WGS-84 latitude, longitude and height position for the required station - ie the reference station in point positioning mode and the mobile for differential.
4. When processing in differential mode, the present age of correction between the two data sets.

An example of the screen output for differential positioning is shown in Fig 3.9.

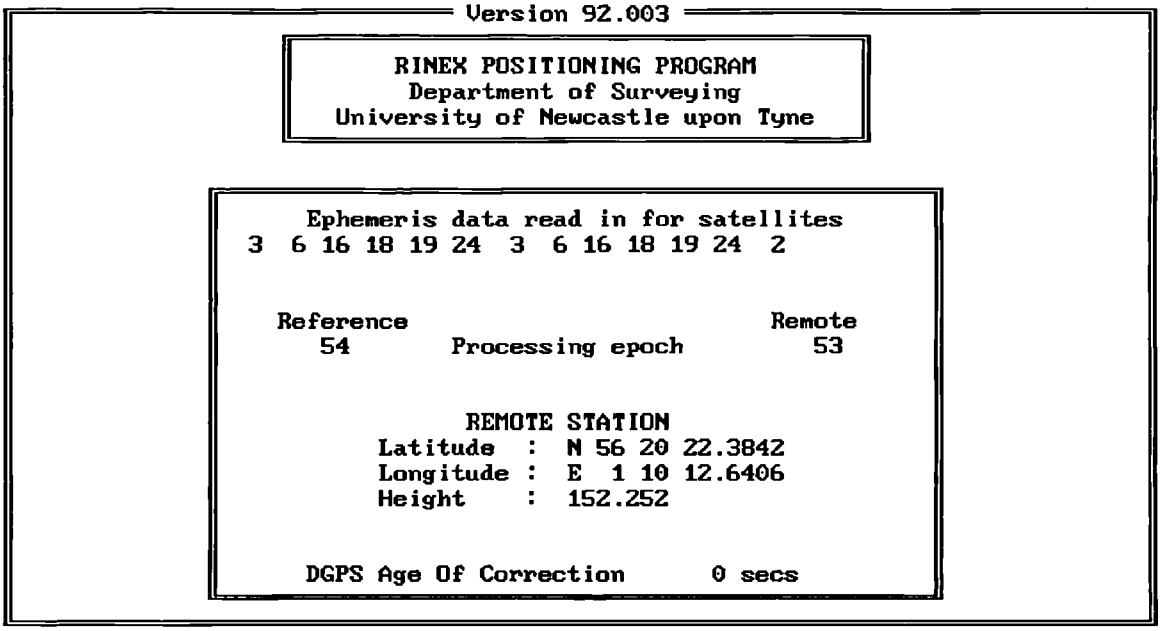


Fig 3.9 Differential mode screen output

When carrying out the computations of simply the satellite positions over a defined time interval, the screen output is somewhat different. In this case the satellite being processed is shown, as is the percentage that has been completed for that satellite.

3.4.2 File Output

Similar coordinate output is written to file during the positioning computations. In addition, there is a time-tag for each epoch and the PDOP value is also given. Thorough information on the data that has been processed is also provided via a header. This includes the data files that have been used, the positioning mode, and (if differential) the reference station details and height aiding information. Details of this output can be found in Appendix B along with examples of the output obtained when the "pseudorange corrections" option is selected.

CHAPTER FOUR
THE ANALYSIS OF TIME SERIES

4.1 INTRODUCTION TO TIME SERIES ANALYSIS

A time series is a collection of observations made sequentially in time (Chatfield, 1989). It is often very useful to analyse a time series, and this can be carried out in two general ways.

The first, is to look at the time series in the time domain - this is essentially an analysis of the observations against time. Different time series have different characteristics which are associated with different probability models. On analysising a particular series, it is usual to adopt a certain probabilty model and to calculate a number of statistical measures that describe the series. The statistic which is most associated with time domain analysis is the autocorrelation function describing the evolution of the process through time.

The second approach for analysing a time series is to look at the process in the frequency domain. This process breaks the series down into different cyclic components with varying frequencies. It is easiest to think of this type of analysis as describing the original series as a summation of a number of sinusoidal components, each with a different frequency and amplitude. By selecting the frequencies which have the largest amplitudes, it is possible to determine how often the overwhelming trends in the series occur. For instance, if the highest amplitude is associated with a frequency of 1/60 Hertz (cycles per second), then the trend within the series repeats every one minute.

This chapter begins by looking at time domain analysis and the different probability models that can be used. Frequency domain analysis is then introduced, and finally a visual interpretation of computer generated series is given.

4.2 ANALYSIS IN THE TIME DOMAIN

4.2.1 Stochastic Processes

A stochastic process is a set of random variables which are ordered in time and can be continuous or discrete. If time, t, is continuous (usually $-\infty < t < \infty$) then the random variable is denoted by X(t). If time is discrete (usually $t = 0, \pm 1, \pm 2, ...$) then the random variable is denoted by X_t. A useful way of describing such a process is by its moments, particularly the first and second moments. These are the mean and the variance and covariance of the process. The third moment is the skewness of a process (the degree of asymmetry of a distribution about it's mean), whilst the fourth (the kurtosis) measures the flatness of a distribution relative to a normal distribution. The first two moments of a stochastic process are described for continuous time with similar definitions applying in discrete time.

The mean function

$$\mu(t) = E[X(t)] \tag{4.1}$$

The variance function

$$\sigma^2(t) = Var[X(t)] \tag{4.2}$$

The covariance function

$$\gamma(t_i, t_j) = E\left\{\left[X(t_i) - \mu(t_i)\right]\left[Y(t_j) - \mu(t_j)\right]\right\} \tag{4.3}$$

The term autocovariance is used instead of covariance when X and Y are refer to the same stochastic process.

4.2.2 Stationary Processes

A time series is said to be strictly stationary if the joint probability distribution of $X(t_1), .., X(t_n)$ is the same as the joint probability distribution of $X(t_1 + \tau), .., X(t_n + \tau)$ for all $t_1, .., t_n, \tau$.

If n = 1 it implies that the distribution of X(t) must be the same for all t, so that

$$\mu(t) = \mu \tag{4.4}$$

$$\sigma^2(t) = \sigma^2 \tag{4.5}$$

ie the mean and variance do not depend on time.

By putting n = 2 the joint distribution of $X(t_1)$ and $X(t_2)$ depends only on the lag, $(t_2 - t_1)$. Thus the autocovariance function $\gamma(t_1, t_2)$ only depends on $(t_2 - t_1)$ and may be written as $\gamma(\tau)$ where $\tau = t_2 - t_1$ and

$$\begin{aligned} \gamma(\tau) &= E\{[X(t)-\mu][X(t+\tau)-\mu]\} \\ &= Cov[X(t), X(t+\tau)] \end{aligned} \tag{4.6}$$

The equations for the mean and autocovariance are often used to define a second-order stationary (weakly stationary) process with no assumptions being made about higher moments than those of second order. The value of the autocovariance depends on the units used to describe X(t) and can be difficult to interpret. It is sometimes useful to divide the autocovariance by the variance function to produce the

autocorrelation function, $\rho(\tau)$, which can lie between ± 1. Thus

$$\rho(\tau) = \gamma(\tau) - \gamma(0) \tag{4.7}$$

4.2.3 Useful Stochastic Processes

There are 4 different types of stochastic process, with different probability models, which are sometimes useful in setting up a model for a time series.

4.2.3.1 Purely random

A discrete-time process is purely random if it consists of a sequence of random variables $\{Z_t\}$ which are mutually independent and identically distributed. The process has a constant mean and variance and an autocovariance of zero, and is strictly stationary. A purely random process is often described as white noise.

4.2.3.2 Random walk

A process is said to random walk if the value at one particular instant is totally dependent on the previous value, with some added noise. Suppose $\{Z_t\}$ is a discrete, purely random process with mean μ and variance σ^2. A process $\{X_t\}$ is said to random walk if

$$X_t = X_{t-1} + Z_t \tag{4.8}$$

which gives

$$\begin{aligned} E(X_t) &= t\mu \\ Var(X_t) &= t\sigma^2 \end{aligned} \tag{4.9}$$

As the first two moments change with time, this process is non-stationary.

4.2.3.3 Moving average

A moving average process $\{X_t\}$ is a linear combination of a purely random process $\{Z_t\}$. Suppose $\{Z_t\}$ has a mean zero and variance σ_Z^2 then

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \tag{4.10}$$

where X_t is a moving average process of order q

$\{\beta_i\}$ are constants

Since the process is expressed in terms of purely random processes, it is second order stationary for all values of the $\{\beta_i\}$. The mean and variance can be expressed as

$$E(X_t) = 0 \tag{4.11}$$

$$Var(X_t) = \sum_{i=0}^{q} \beta_i^2 \tag{4.12}$$

If the purely random process $\{Z_t\}$ is normally (or Gaussian) distributed then so is the moving average process $\{X_t\}$ and it is strictly stationary. Usually in a least squares estimation a set of parameters (Xs) are described in terms of the observations by some linear combination. The observations, or measurements, contain different types of

errors (Zs) which are often assumed to be normally distributed. If this is the case, the linear combination can be described as a moving average process and the resultant parameters will also be normally distributed.

4.2.3.4 Autoregressive

An autoregressive, or Markovian, process is a process that is regressed on past values of itself. In other words, if $\{Z_t\}$ is a purely random process with mean zero and variance σ_Z^2 then an autoregressive process $\{X_t\}$ of order p can be described as

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \tag{4.13}$$

where $\{\alpha_i\}$ are constants

A process of order 0 is a representation of a purely random process as

$$X_t = Z_t \tag{4.14}$$

4.2.3.4.1 First order autoregressive

Typically a first order autoregressive process is used to model stochastic behaviour. This can be represented as

$$X_t = \alpha X_{t-1} + Z_t \tag{4.15}$$

If $\alpha = 1$ then this is a random walk process. If $|\alpha| < 1$ then initially the process is non-stationary and the variance experiences bounded growth. Noise is added to the system until it has reached a steady state, and the process then becomes second order stationary. With this condition ($|\alpha| < 1$), the first two moments can be expressed as

$$E(X_t) = 0 \tag{4.16}$$

$$\begin{aligned} Var(X_t) &= \sigma_Z^2/(1-\alpha^2) \\ &= \sigma_X^2 \end{aligned} \tag{4.17}$$

$$\begin{aligned} \gamma(k) &= E\left[X_t X_{t+k}\right] \\ &= \alpha^k \sigma_X^2 \end{aligned} \tag{4.18}$$

and the autocorrelation function as

$$\rho(k) = \alpha^{|k|} \qquad k = \pm 1, \pm 2, \ldots \tag{4.19}$$

4.2.3.5 Combination of processes

It can be seen that the above processes are not independent of each other. For instance, an autoregressive process can be described in terms of an infinite order moving average process. The autocovariance function used later in this report to describe pseudorange residuals is first order autoregressive, but other processes could equally be used. Another useful class of models for time series is a combined autoregressive and moving average process (ARMA). Thus an ARMA process $\{X_t\}$ is described by past values of itself combined with several purely random processes.

4.3 ANALYSIS IN THE FREQUENCY DOMAIN

4.3.1 A Sinusoidal Model

The analysis of a time series in the frequency domain is widely used in fields such as electrical engineering and geophysics, but seldom applied to surveying applications. For this reason, the following section intuitively discusses frequency domain analysis and associated terms, without attempting a rigorous approach.

A time series, X_t, that contains a sinusoidal component could be described as

$$X_t = R\cos(\omega t + \theta) \tag{4.20}$$

where ω is the frequency,
R is the amplitude, and
θ is the phase of the sinusoidal variation.

It would be usual to expect that X_t contains a summation of different sinusoidal components, and therefore could be represented as

$$X_t = \sum_{j=1}^{k} R_j \cos(\omega_j t + \theta_j) \tag{4.21}$$

where R_j is the amplitude at frequency ω_j.

By applying the cosine rule of cos(A + B) = cosAcosB - sinAsinB, equation (4.21) can be described as a combination of cosine and sine terms :

$$X_t = \sum_{j=1}^{k} a_j \cos\omega_j t + b_j \sin\omega_j t \tag{4.22}$$

where $a_j = R_j \cos\theta_j$, and

$b_j = -R_j \sin\theta_j$.

This rather simplistic approach shows how a time series can be represented by means of a multitude of sinusoidal amplitudes and frequencies.

4.3.2 Spectral Representation

The term "spectral" is associated with frequency domain analyses and is often teamed with "representation", "distribution function", and "density function". Each of these terms are discussed, but for full details refer to texts such as Chatfield (1989), Koopmans (1974) and Bloomfield (1976).

In (4.22) the summation is made from a limited number of different frequencies (j = 1 to k). In fact all possible frequncies can contribute to the series, and to show this mathematically it is neccesary to integrate (4.22). The limits of the integral will be for a discrete process and for a continuous process. The integral is known as a spectral representation of the time series.

The spectral distribution function, $F(\omega)$, arises from the Wiener-Khintchine theorem that states that for any stationary stochastic process with autocovariance function $\gamma(k)$, there exists a consistently increasing function $F(\omega)$ such that (Chatfield, 1989)

$$\gamma(k) = \int_0^{\pi} \cos(\omega k) dF(\omega) \qquad (4.23)$$

Equation (4.23) is called the spectral representation of the covariance function, and $F(\omega)$ can be shown to be the contribution to the variance of the process made by the frequency range 0 to ω.

The spectral density function is obtained by differentiating $F(\omega)$ with respect to ω and is often denoted by $f(\omega)$, ie

$$f(\omega) = \frac{dF(\omega)}{d\omega} \tag{4.24}$$

The function now repesents the contribution to the variance of components with frequencies ω to $\omega+d\omega$, and when the complete spectrum is drawn the area underneath the curve represents the variance of the process. The spectral density function is, in fact, the Fourier transform of the autocovariance function.

The word "power" is often used with both the spectral distribution and density functions. This originates from the engineer's use of the term describing the passage of an electric current through a resistance. For sinusoidal input, the power is proportional to the squared amplitude of the oscillation, and the total power is equal to the variance of the process (Chatfield, 1989).

4.3.3 Fourier Analysis

Fourier analysis is simply a way in modelling the time series by means of sinusoidal frequencies and amplitudes. The method uses a series developed by Jean Baptiste Fourier which describes a well-behaved function, x(t), as

$$x(t) = \frac{a_0}{2} + \sum_{r=1}^{k} (a_r \cos rt + b_r \sin rt) \tag{4.25}$$

where
$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)dt$$
$$a_r = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\cos(rt)dt \qquad r = 1,2,\ldots$$
$$b_r = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\sin(rt)dt \qquad r = 1,2,\ldots$$

The amplitude of the *r*th harmonic is given by

$$R_r = \sqrt{(a_r^2 + b_r^2)} \tag{4.26}$$

4.3.3.1 FFT

The Fast Fourier Transform (FFT) is a technique that finds the Fourier series in order to model a process and significantly reduces the number of computations that are required. This method is even more computationally efficient if the number of data samples within the process is a power of 2, and if this is not the case it is usual to pad the data with zeros or remove some of the final samples.

4.3.4 Sampling Frequencies

When converting the time series into the frequency domain, it is necessary to model the series with a wide range of different frequencies. The value of the limiting frequencies are of particular interest as represent the lowest and highest frequency that the data set can be sampled. The lowest frequency is simply the length of the data set. In other words if a data set is collected for one hour, it is possible to fit a sinusoidal component with a frequency of one hour (1/3600 Hz). Obviously there is only one set of data that can contribute to this frequency, so there is no redundancy when modelling at this frequency.

The highest frequency is known as the Nyquist and is related to the sampling, or

recording, interval of the data set and is given by (Koopmans, 1974)

$$\frac{1}{2\Delta t} \tag{4.27}$$

where Δt is the sampling interval.

This means that if the data set is recorded at one second intervals, it is impossible to obtain information at frequencies higher than 1/2 Hertz - one cycle every two seconds. Therefore if a particular phenomena is believed to occur at a high frequency, then the data set recorded to look at this must be at a frequency at least twice as high. Accordingly, if the phenomena is believed to be at a low frequency it is necessary to extend the period over which the data set is recorded.

4.4 THE ANALYSIS OF EXAMPLE TIME SERIES

4.4.1 Description of the Analysis

Three different time series processes have been generated using the Matlab package on the computer, in order to visually describe some of the aforementioned processes and analyses. In each case the time series , the autocorrelation function and the Fourier analysis is plotted. The Fourier analysis uses the FFT techniques and is a plot of amplitude against frequency. The variance of the process is determined using the time domain analysis (the autocovariance at zero lag before scaling into the autocorrelation function) and not by using frequency domain analyis (ie by the power spectral density function). In each case the data set represents a time series recorded at one second intervals for fifteen minutes giving 900 observations in total. The different time series given here are examples of first-order autoregressive process

with the values of α being 0, 0.8 and 1 respectively.

4.4.2 A White Noise Process

A purely random process (4.14) was generated and scaled to have a variance of approximately 9 m^2. The autocovariance function was determined and scaled into the autocorrelation function and the FFT was also computed. The analysis is shown in Fig 4.1. The points to notice from the diagram are

1. The time series plot represents a completely random, non-deterministic, process with a mean of zero a variance of 9 m^2.
2. The autocorrelation function shows a vertical line at zero lag, and thereafter noise. This represents the random process since, on average, there is no correlation between observations which are recorded at different times.
3. The FFT shows that the series is made up of sinusoidal components at all frequencies with no overwhelming part since they all have similar amplitudes.

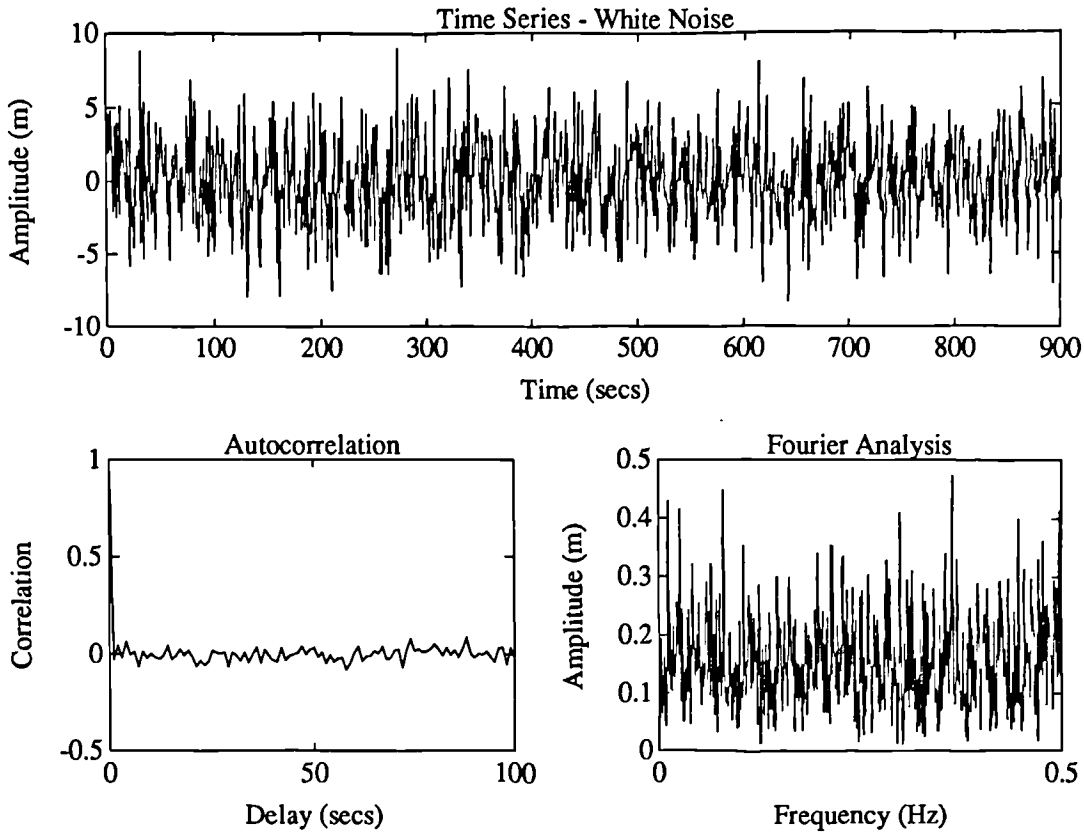


Fig 4.1 White noise analysis

4.4.3 A General First-order Autoregressive Process

The first-order autoregressive process (4.15) was generated for a value of $\alpha = 0.8$, and scaled so that the process had a variance of approximately 9 m^2. The analyses is shown in Fig 4.2 with the points of interest being

1. The time series shows how the observations at one instance are related to those at previous instances.
2. The autocorrelation function steadily declines from a value of one at zero lag to approximately a value of zero. This again represents the relationship between the observations.
3. The FFT shows that the process is built predominently out of low frequency components. This indicates that the process is slow to change, which can be verified by the time series plot. If the value of α was negative, then the time

series would oscillate between positive and negative values resulting in a FFT analysis with high frequency components.

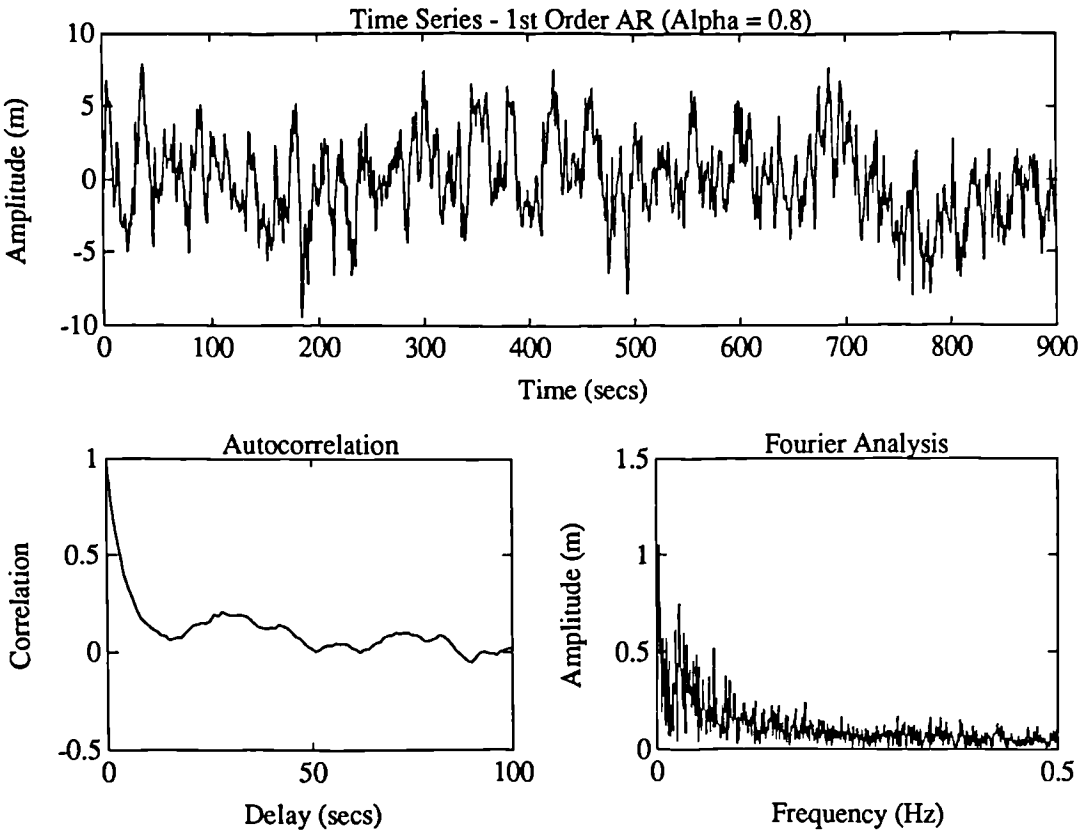


Fig 4.2 General first-order autoregressive analysis

4.4.4 A Random Walk Process

A random walk process (4.8) was generated and analysed as before. Since this is a non-stationary process, there has been no scaling of the variance which is time dependent. The points to notice for this analyses, shown in Fig 4.3, are

1. The time series clearly shows the "wandering" nature of the process which has a mean and variance which are both time dependent.
2. The autocorrelation plot has a shallow gradient representing the high correlation between data observations which have been taken at a considerable time interval apart. The curve will eventually tend towards zero.

3. The FFT plot again shows a concentration of low frequency components relating to the high correlation.

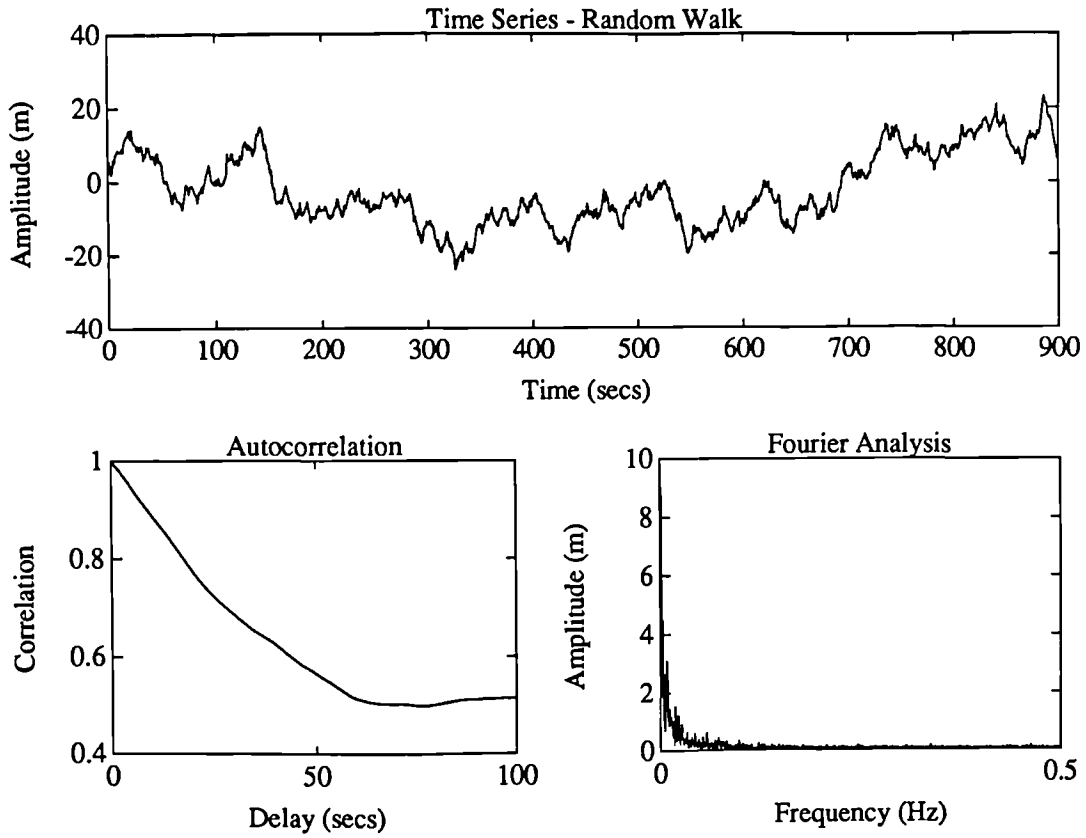


Fig 4.3 Random walk analysis

CHAPTER FIVE
GPS PSEUDORANGE ANALYSIS

5.1 MODELLING TEMPORAL CORRELATION

In order to investigate the actual correlation that exists within GPS pseudorange observations, as used for offshore positioning, it is necessary to select a certain time series that represents the data. The pseudorange measurements themselves are fast-changing due to the velocities of the satellites, and therefore their errors were analysed. The actual errors were obtained through the computation of the pseudorange residuals - the residual and error have the same magnitude, but a different sign,

5.1.1 Computation Of Residuals

Essentially there are two separate approaches to obtain the pseudorange residuals. The residuals from a least squares estimation would be heavily correlated amongst themselves and a large raw error for one range would be "dampened" and "spread" into the residuals for all of the ranges. This correlation is known since it can be obtained from the covariance matrix of the residuals (see 7.1.3 for further details):

$$\mathbf{C}_{\hat{\mathbf{v}}} = \mathbf{W}^{-1} - \mathbf{A}\mathbf{C}_{\hat{\mathbf{x}}}\mathbf{A}^{T} \qquad (7.6)$$

where **A** is the design matrix, and

W is the weight matix of the observations

So that this type of correlation is kept to a minimum, the following method was used to compute the residuals of individual pseudorange measurements to the satellites:

$$v_i = l_i - dT_{Si} + dT_R - d_i \qquad (5.1)$$

where v_i is the pseudorange correction to satellite i

d_i is the geometric distance computed between the coordinates of satellite i and the true coordinates of the reference station

l_i is the observed pseudorange to satellite i

dT_{Si} is the clock correction for satellite i in metres

dT_R is the receiver clock correction in metres

Equation (5.1) is similar to that for the computation of the pseudorange corrections (3.26), the only difference being the reversal of the sign. It is clear from (5.1) that to produce a true estimate of the individual residuals, it is not possible to avoid using a the least squares process as the pseudorange measurements include a receiver clock offset. To convert the pseudorange into an observed distance, it is necessary to compute the receiver clock error and the satellite clock offset at each epoch.

5.1.2 Autocovariance Analysis

The residuals for all the satellites at every epoch are divided into time-tagged residuals for each satellite. For each set, the autocovariance of the residuals is calculated by the product of two residuals at a certain time interval (δt) apart. For different values of δt (0, 2, 4, ... secs) the autocovariance is computed by taking the mean of the products of all residuals that are δt apart. This can be represented by:

$$\gamma(\delta t) = \frac{1}{n}\sum_{t=0}^{\infty} v_t . v_{t+\delta t} \qquad (5.2)$$

where $\gamma(\delta t)$ is the autocovariance at lag δt

v_t is the pseudorange residual at time t

$v_{t+\delta t}$ is the residual at time $t+\delta t$, and

n is the total number of $v_{t+\delta t}$.

Having generated the correlation coefficients at different lags/delays for a certain time series, it is then possible to fit a certain autocovariance function to the data. An important consideration for the choice of autocovariance function is that it will be used to generate a covariance matrix. Most mathematical procedures require this matrix to be inverted at some stage, and it must therefore be positive-definite. The translation of this property to the autocovariance function is not straight forward but, for certain classes of positive-definite function, the Fourier transform is everywhere positive or zero (Cooper and Cross, 1991).

The positive-definite first order autoregressive, AR(1) (Gauss-Markov) autocovariance function (see 4.2.3.4.1) was used to model the time correlation of the pseudorange residuals. This is given by:

$$\gamma(\delta t) = \sigma^2 e^{-|\delta t|/T} \qquad (5.3)$$

This function is characterised by two main properties which are illustrated in Fig 5.1 (Moritz, 1980).

- The variance, σ^2, which is the autocovariance when $\delta t=0$.
- The correlation time, T, which is the value of the time lag corresponding to an autocovariance of $\gamma(0)/2$.

A third property, curvature, is sometimes included which is a dimensionless measure combining the gradient and rate of change of gradient of $\gamma(\delta t)$.

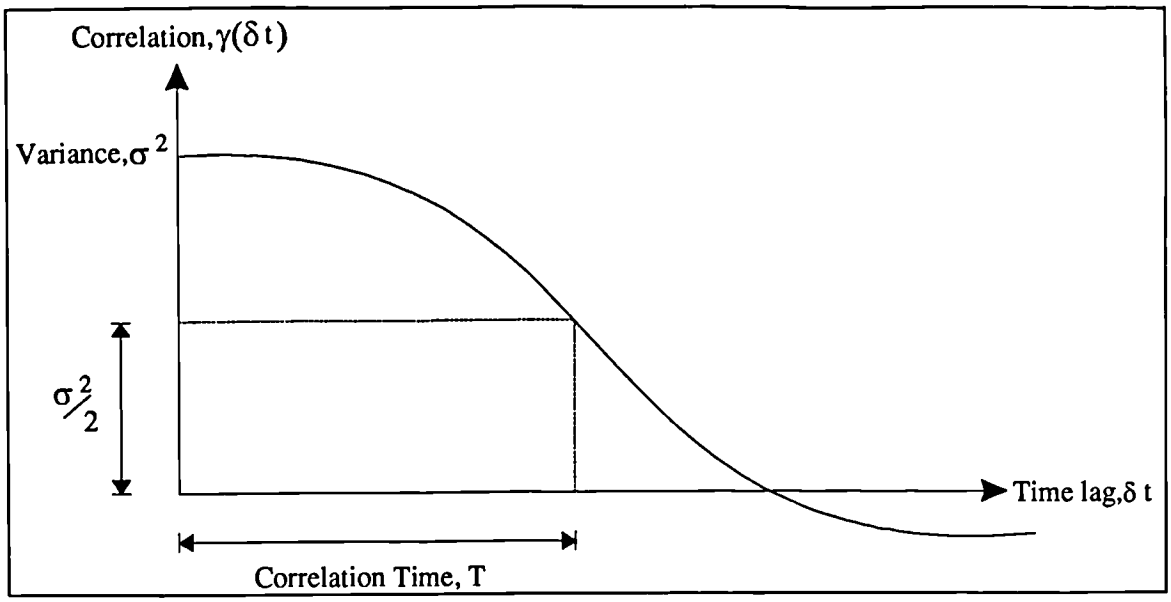


Fig 5.1 Properties of a local covariance function (Moritz, 1980)

By using this function, the pseudorange residuals are being modelled in the following fashion (see 4.2.3.4.1 for further details):

$$X_t = \alpha X_{t-1} + Z_t \tag{4.15}$$

where $\{X_t\}$ is the pseudorange residuals process

$\{Z_t\}$ is the system noise

α is a constant

The autocovariance function represented by equation (5.3) is the same as that for the AR(1) process in (4.18), from:

$$\gamma(k) = \alpha^k \sigma_x^2 \tag{4.18}$$

where $k = \delta t$

$$\therefore \quad \gamma(\delta t) = \alpha^{\delta t}\sigma_x^2$$

as $\gamma(\delta t) = -\gamma(\delta t)$

$$\therefore \quad \gamma(\delta t) = \alpha^{|\delta t|}\sigma_x^2$$

and $\alpha = e^{-1/T}$

$$\gamma(\delta t) = \sigma^2 e^{-|\delta t|/T} \tag{5.3}$$

5.1.3 Fourier Analyses

In order to search for trends, or cyclic patterns, in the time series a Fast Fourier Transform (FFT) was computed (see 4.3.3). This technique splits the time series into different sinusoids at a range of frequencies and with varying amplitudes.

5.2 TEMPORAL CORRELATION OF GPS PSEUDORANGE RESIDUALS

5.2.1 Categorising Temporal Correlation

Different data sets were selected to allow different causes of temporal correlation to be investigated. Three broad categories were identified, these being:

1. Correlation in the GPS system which the users can do little about. Causes of this include satellite clock errors, satellite orbit errors and atmospheric refraction. Selective Availability (SA), the intentional degradation of the pseudorange accuracy, by means of dithering the satellite clock and satellite position epsilon (1.4.1.3) is also in this category. Obviously these causes are time variant and will therefore be different for both data sets recorded at different times, but will be similar for all simultaneous data sets.

2. Correlation introduced by the computation methods that have been used. This is kept to a minimum but cannot be completely avoided. Calculating the receiver clock correction will introduce some spatial correlation between satellites at each epoch. Temporal correlation will result from any positional errors (for both the satellites and the known ground station) when computing the residuals. Positional errors at the ground station have been reduced by taking a mean position for all data sets

3. Correlation caused by the receiving equipment used. This will be due to different receiver/antenna design and data flow within the receiver, and from the set up location due to multipath. This is dependent on the particular type of receiver, the location of the antenna, and the geometry of the satellites.

5.2.2 Description of the Data Sets

Various GPS pseudorange data sets were recorded in order to analyse possible causes of temporal correlation. This included recording data when SA was both off and on, using different receiver types and different smoothing algorithms, thus enabling the following to be analysed:

- The effect and characteristics of SA on the pseudorange corrections.
- The reduction of noise levels and the increase in temporal correlation when smoothing raw data.
- The difference between different smoothing algorithms that manufacturers provide within the receiver.
- Differences between receiver types.
- Differences between individual receivers of the same type.

Three different receiver types were used for the collection of the data sets: the Ashtech LMXII with P code upgrade, the Trimble 4000 SST, and the Leica System 200. All processing was carried out using the C/A code pseudoranges as produced by the particular receiver, except for the earliest Ashtech files for which doppler measurements were included for smoothing algorithms in order to test the difference between the raw and smoothed data. All of the sets comprised of approximately one hour of data recorded every second. Table 5.1 summarises the data sets.

File	Date	Receiver	SA
STN10731	14/3/91	Ashtech LMXII	Off
STN30731	14/3/91	Ashtech LMXII	Off
GTYH1921	10/7/92	Trimble 4000SST	On
COMP2021	20/7/92	Leica SR299	On
SAMP2021	20/7/92	Leica SR299	On
PP923421	7/12/92	Ashtech LMXII	On
WH930611	2/3/93	Ashtech LMXII	On
PP930611	2/3/93	Ashtech LMXII	On

Table 5.1 Pseudorange data sets

The data sets that were recorded simultaneously were on the dates 14/3/91 when SA was not being implemented and on 2/3/93 when SA was officially turned on.

5.3 RESULTS OF THE PSEUDORANGE ANALYSES

5.3.1 The Analysis of Pre SA Data

The data sets used to analyse the characteristics of the GPS performance before the implementation of SA were recorded in early 1991 in Newcastle. Two different data sets were simultaneously recorded and were processed using a smoothing algorithm. One of the data sets (STN10731) was also processed using "raw" pseudoranges to examine the noise characteristics of early geodetic receivers, and the improvements that may be obtained with smoothing operations. The algorithm that was used to smooth the pseudoranges is given by:

$$PRS(t_n) = PRS(t_{n-1}) + ID(t_n) + 2I \qquad (5.4)$$

where

PRS	is the smoothed pseudorange,
t_n	is the epoch at time t_n,
t_{n-1}	is the previous epoch,
ID	is the scaled integrated doppler observable, and
2I	is an empirically derived ionospheric correction.

The integrated doppler measurement represents the rate of change of phase of the signal and is measured in cycles/second, or Hz. Multiplying the doppler figure by the time interval between the two epochs produces the change of phase, and then scaling this by the wavelength (19 cm for L1) results in an accurate indication of the change in distance that should have occurred. The ionospheric correction is used to account for the phase advance and group delay phenomena. This process must be initialised for each satellite by setting the smoothed pseudorange to be equal to the observed raw value and is reinitialised every time a cycle slip or loss of lock on the doppler measurement is detected.

The results in terms of satellite (SV) correlation times and variance for the two data sets using the smoothing algorithms are given in Tables 5.2 and 5.3. Full graphical output can be found in Appendix C.

SV	T (s)	$\sigma^2(m^2)$
2	104.1	16.2
6	276.9	47.4
11	52.2	2.3
16	226.6	30.7
17	337.7	34.3
19	601.1	36.8

Table 5.2 STN10731 - smoothed

SV	T (s)	$\sigma^2(m^2)$
2	297.0	8.8
6	1539.0	53.8
11	47.3	1.7
16	1674.0	38.2
17	2624.0	37.4
19	897.0	38.5

Table 5.3 STN30731 - smoothed

The general trend for both data sets is that all residuals lie within a ± 15m band, which is confirmed by the variance values (maximum of 53.8 m^2). The autocorrelation analysis shows long correlation times (except for SV 11) for both data sets, particularly for STN30731. The analysis for both files showed very little deterioration of the correlation over the maximum lag interval used (180 s) and for some of the

satellites the autocorrelation function value was 0.9 even at this lag. The difference between the two data sets can be put down to the doppler tracking within the different receivers, and also that the initial residuals for all satellites in the STN10731 file were in the region of ± 30m. The Fourier analyses again showed very similar patterns for both data sets. There were no substantial peaks at any frequency, except for at 0 Hz which represents a general offset from zero which most probably resulted from unchanging atmospheric conditions. The fact that no other substantial peaks were found means that no repeatable cyclic patterns occured. An example of the time series and the autocorrelation functions are shown in Fig 5.2 and 5.3.

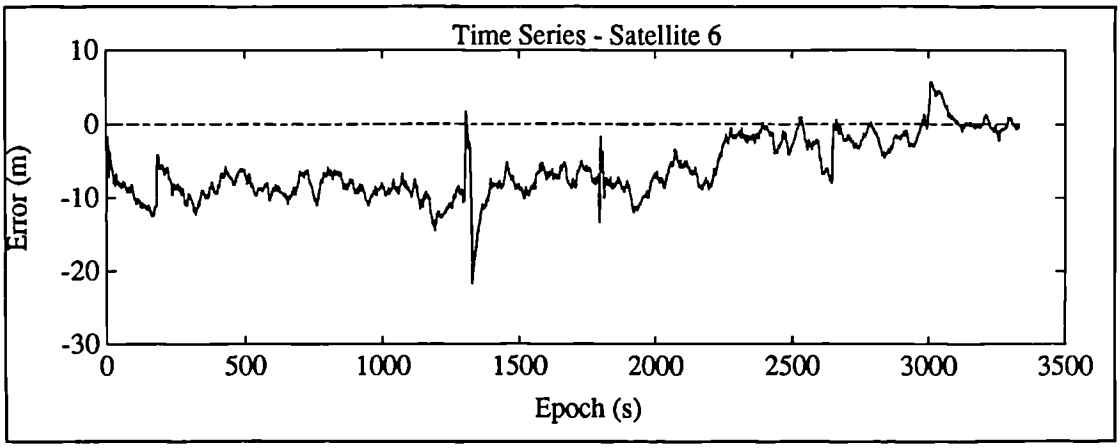


Fig 5.2 Pre SA time series

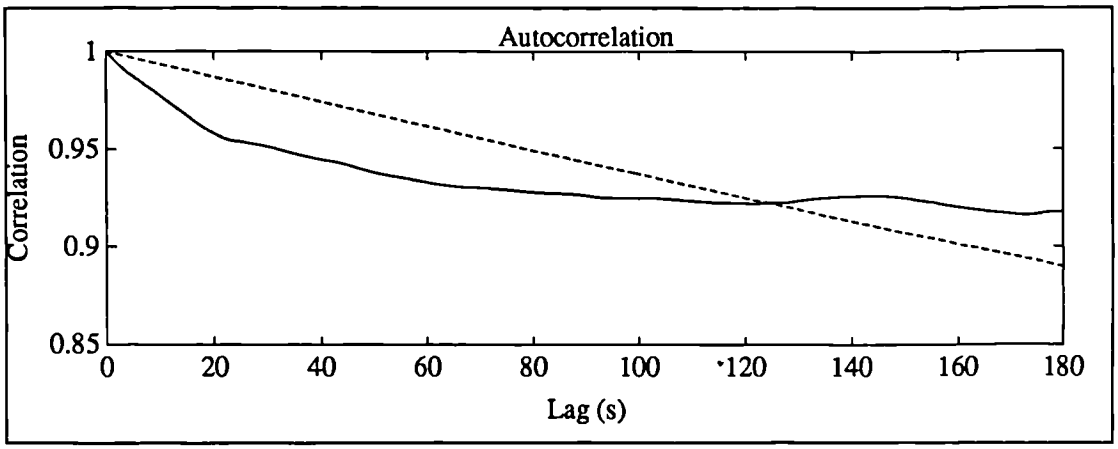


Fig 5.3 Pre SA autocorrelation

The data set for STN10731 was processed for a second time, but with no smoothing algoritms applied. The results are given in Table 5.4.

SV	T (s)	$\sigma^2(m^2)$
2	59.1	35.9
6	158.8	85.4
11	16.1	10.0
16	128.2	56.6
17	216.7	52.3
19	240.5	87.4

Table 5.4 STN10731 - raw

The differences shown in Tables 5.4 and 5.2 are those expected from introducing smoothing algorithms. The variances are greater, representing more noise in the system, and the correlation times are shorter, again a result from the increase of noise. The Fourier analyses (see Appendix C) are exactly the same for all satellites, indicating that the same patterns are in both the smoothed and raw data. This proves that the smoothing algorithms simply reduce the noise level of the pseudorange data and do not change the general characteristics of the series.

5.3.2 Post SA data

5.3.2.1 Trimble data sets

The Trimble data set was recorded with a 4000SST receiver near Merthyr Tydfil, Wales, in July 1992. The pseudorange data is initially smoothed within the receiver, and no further algorithms were implemented. SA was on, which is indicated in the time series plots in Appendix D, since the residuals are frequently changing and are in

the range ± 150m. The noise characteristics of this data set cannot be examined in the same way as for the pre-SA data since the overwhelming cause of the high variances is that of SA. The autocorrelation plots show the traditional pattern represented in Fig 5.1 with the values for all satellites being below 0.5 at the maximum lag (180 s). The exponential curve, representing the first-order autoregressive process (see Fig 5.4) does not appear to fit this data (this is the same for most of the data collected with SA on) and a more appropriate process may have been a second or third order autoregressive, or a combined process. The Fourier analysis shows cyclic trends with large amplitudes at frequencies ranging from 0.00025 Hz (a pattern that repeats every 4000 s) to approximately 0.0035 Hz (repeating every 280 s). For many of the satellites (SVs 24, 20, 12, 3) the maximum amplitude is at a frequency of 0.0012 Hz, which is one cycle every 830 s. The variance and correlation time results for this data set are given in Table 5.5.

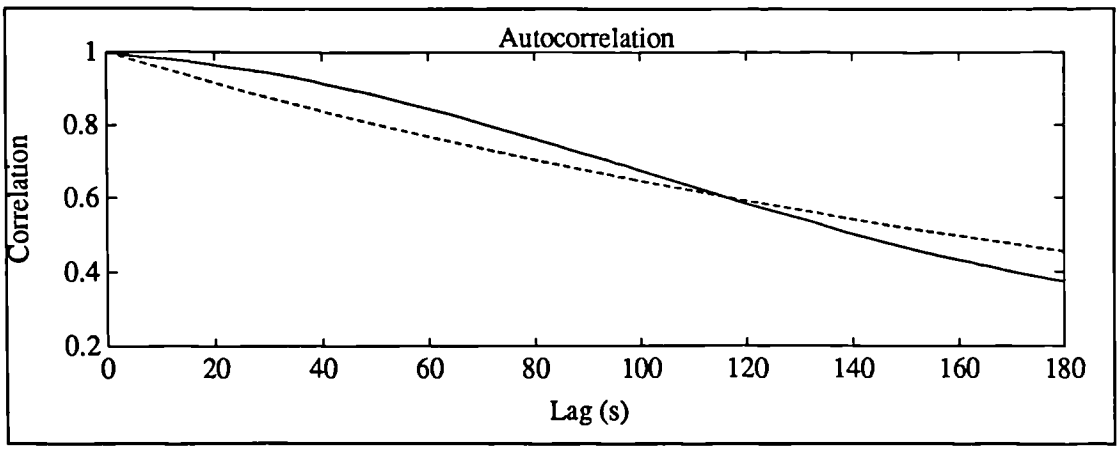


Fig 5.4 Trimble example autocorrelation

SV	T (s)	$\sigma^2(m^2)$
3	128.9	1068.0
12	171.0	1517.0
13	232.1	1774.0
20	170.0	3425.0
24	169.7	1808.0
25	228.4	981.6

Table 5.5 GTYH1922

5.3.2.2 Leica data sets

Two data sets were collected with the Leica SR299 receivers which have the capability to perform different smoothing operations within the receiver. One data set (COMP2021) was recorded with compacted pseudoranges which are generated by taking an average of many high frequency ranges, for instance by recording pseudorange measurements every 0.1 s and averaging to get a value for every second. The other data set (SAMP2021) was using sampled measurements which uses the observation recorded at the particular instant. Both were recorded in Newcastle in July 1992 when SA was on. There is no noticeable difference between the two approaches in terms of correlation times, variances (see Tables 5.6 and 5.7) and the Fourier analyses, and since the data sets were not recorded simultaneously no conclusions about the smoothing processes can be made. The Fourier analyses (Fig 5.5 is an example) showed cyclic trends ranging from the lowest possible frequency, 0.0005 Hz or one cycle every 2000 s, to a higher frequency of 0.004 Hz, or one cycle every 250 s. A substantial peak was noticeable for all satellites at a frequency of approximately 0.0025 Hz, or every 400 s. Full output is provided in Appendix E.

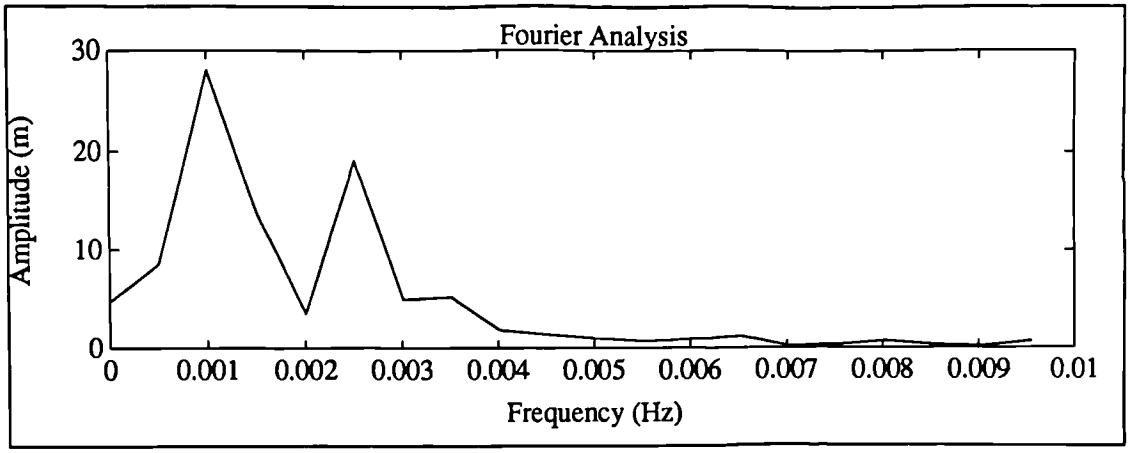


Fig 5.5 Leica example Fourier analysis

SV	T (s)	$\sigma^2(m^2)$
11	196.7	1095.0
17	94.6	1046.0
19	110.1	876.9
21	145.5	756.2
23	147.2	1072.0
28	125.1	2788.0

Table 5.6 COMP2021

SV	T (s)	σ2(m2)
11	138.0	1020.0
17	192.0	1865.0
19	131.4	610.7
21	207.1	1118.0
23	105.6	1242.0
28	218.6	2973.0

Table 5.7 SAMP2021

5.3.2.3 Ashtech data sets

Data was recorded with the Ashtech LMXII receivers which had been upgraded to measure the P code and to phase smooth the C/A code pseudoranges. The sets chosen for analysis were observed in December 1992 (PP923421) and March 1993 (PP/WH 930611) which would allow any changes in the high frequency characteristics of SA over that period to be identified. For the March data, two simultaneous data sets recorded with the same instrument type at a distance apart of approximately 10 km. Similar atmospheric conditions could be assumed, and therefore this analysis would allow for differences between individual receivers to be examined. The variance and correlation results for PP923421 are given in Table 5.8 which clearly show the high variances and relatively short correlation times expected under SA conditions. The correlation functions again follow the classical shape, which is not best fitted with an AR(1) process, indictating that the estimated correlation times are all longer than has actually occured. The errors in all of the pseudoranges are between ± 150 m and have a maximum amplitude (except for SV 16) for the frequency representing one cycle every hour - this is the lowest frequency that can be tested. Full graphical details for this data set can be found in Appendix F, and an example of the time series is shown in Fig 5.6.

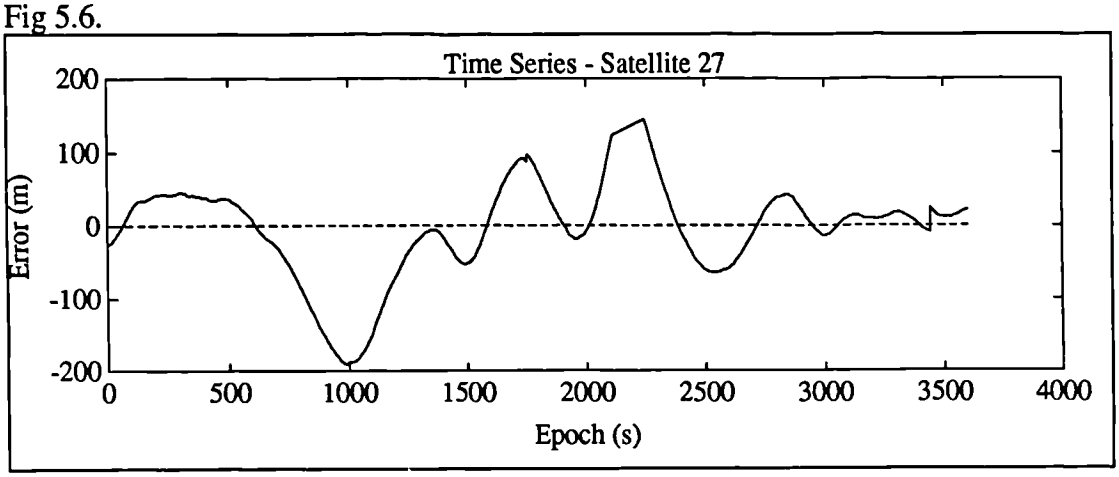


Fig 5.6 Ashtech example time series

SV	T (s)	$\sigma^2(m^2)$
3	150.6	681.5
16	149.9	1518.0
18	308.5	708.2
19	235.1	1629.0
24	151.4	644.1
27	219.4	1498.0

Table 5.8 PP923421

The results of the data sets recorded in March 1993 are shown in Tables 5.9, 5.10 and Appendix G. The errors are in the range ± 200m which, again, is confirmed by their variances and the amplitude of the lowest frequency within the Fourier analyses. The exponential function appears to fit the correlation values for most of the satellites with the estimated correlation times ranging from 132.0 to 325.5 s.

SV	T (s)	$\sigma^2(m^2)$
2	160.7	1892.0
16	214.9	3357.0
17	171.6	1782.0
26	132.0	3119.0
27	325.5	4067.0

Table 5.9 PP930611

SV	T (s)	$\sigma^2(m^2)$
2	225.9	1292.0
16	274.3	2481.0
17	182.2	1506.0
26	196.7	2360.0
27	231.9	3508.0

Table 5.10 WH930611

The Fourier analyses identified cyclic patterns ranging from 0.0003 Hz to 0.004 Hz and these were similar for the same satellites at the two stations.

5.4 CONCLUSIONS

The results given in 5.3 allowed a number of conclusions to be made on the effects of smoothing algorithms, receiver design and SA.

5.4.1 Smoothing Routines

The data recorded when SA was not activated (STN10731) allowed for the analysis of smoothing routines since the major error source within the data was user dependent - a combination of receiver noise and multipath. The results show a decrease in the average standard deviation of the series from 7.1 m for the raw data (Table 5.4) to 5.9 m for the smoothed data (Table 5.2). The differences in the correlation times is also as expected with the smoothing process lengthening the values by a factor of between 2 and 3. This is also visually apparent since the raw data shows an immediate temporal decorrelation to one half due to the excess noise. The similarities between

the Fourier analyses prove that both the raw and smoothed data is the same time series.

The differences between the Leica compacted and sampled data sets proved inconclusive as there was no common trend for the correlation times or variances. This can be explained by initial phase or doppler smoothing carried out within the receiver and that the data was not recorded simultaneously and may have experienced differing atmospheric conditions or levels of SA.

5.4.2 Receiver Types

Comparing the output from the different receiver types showed little difference due mainly to the overwhelming SA error sources present in most of the data. However, the tests do show the advancement in receiver design/technology in terms of producing higher precision C/A code measurement. By looking at the data sets in their chronological order, a decrease in the noise can be seen from the earliest Ashtech files (Appendix C) to the Trimble files (Appendix D) to Leica (Appendix E) and finally the later Ashtech files (Appendix F and G).

Data recorded over the same periods with the same receiver types showed noticeable differences. For instance, the pre SA data for STN10731 had shorter correlation times than for STN30731: an average of 266 against 1180 seconds. Differences in the Doppler tracking abilities, maybe due to localised multipath, can account for these differences. The post SA Ashtech files PP930611 and WH930611 showed similarities within the estimated quantities but, again, this was due to the overwhelming SA effects. PP930611 had an average correlation time of 201 s and standard deviation of 53 m, and WH930611 resulted in 222 s and 46 m respectively. The Fourier analyses are also similar showing that there was no significant atmospheric or multipath differences over the short baseline.

5.4.3 Characterising SA

The most obvious difference between the pre and post SA results is that of variances. The pre SA data produced standard deviation values of 5 - 8 m against the post SA data which all averaged out to over 40 m. This is caused by the intentional dithering of the satellite clock frequency and the epsilon orbit movement. The correlation times are also very different, with SA resulting in a decrease of correlation time of almost 20 minutes against values around 3 minutes for the post SA data. This is the reason why the age of correction for DGPS operations is so critical.

The choice of time series model to fit the data is also an important conclusion from the analyses carried out in this thesis. The AR(1) process appears to fit the pre SA data (STN10731, STN30731) and therefore gives a good indication of the actual correlation times. The autocorrelation plots for the post SA data, however, showed the more classical curve (as of Fig 5.1) for all satellites in the Trimble, Leica, Ashtech PP923421 data sets. This type of function would more appropriately be modelled using a higher order AR process (see 4.2.3.4) or a combined process (4.2.3.5). It should also be noted that the autocorrelations for the most recent Ashtech files PP930611 and WH930611 reverted back to the AR(1) exponential function - possibly a change in the SA creating algorithms.

The Fourier analyses for the SA data shows that SA consists of numerous cyclic patterns at varying amplitudes. It is extremely difficult to identify changes occurring over the period July 92 (Trimble and Leica data) to March 93 (Ashtech) although some general characteristics can be seen. The frequencies over which the FFT could sample the time series were very much dependent on the recording interval and period for the data sets. All the data was recorded at second intervals which means that the highest possible frequency was 0.5 Hz or one cycle every 2 s, (see 4.3.4) although no

significant patterns were found at these high frequencies. Since most data sets were approximately one hour in length, the lowest frequency is 1 cycle every 3600 s. Table 5.11 attempts to characterise the analyses in to amplitude and frequency bands.

Amplitude (m)	Frequency (Hz)	Time of cycle (s)
	0.0003	3333
20 -40	↓	↓
	0.0015	666
10 - 20	↓	↓
	0.0021	476
5 - 10	↓	↓
	0.0039	256

Table 5.11 SA frequencies

It should be noted that no distinction can be made between the Block I (which are incapable of implementing SA) and Block II satellites. This is because at each epoch all the satellites are used to compute the receiver clock offset (5.1) which is used to compute the residuals. The same reasoning means that individual satellites within a data set should not be separately analysed.

CHAPTER SIX
THE KALMAN FILTER

6.1 INTRODUCTION

The Kalman filter was originally developed in the field of electrical engineering (Kalman, 1960). It is now readily used for processing information in many different fields including surveying, navigation and GPS processing. The Kalman filter is often perceived as a "black box" often producing excellent results, but occasionally disastrous ones. This chapter presents the standard filter algorithms, reliability algorithms and explaining the meaning of terms commonly associated with the Kalman filter.

The Kalman filter is an optimal mathematical filter. This means that it follows a certain criterion (the least squares criterion) when sequentially producing the "best fit" results for a particular problem. The difference between the Kalman filter and standard least squares is that the filter is specifically designed for dealing with systems which have a known behaviour with time, by modelling that behaviour and incorporating it with the observables. One of the simplest examples of such a system is that of a moving vessel with the desired results being its position and velocity. This movement along with the observations from the positioning system can be used to provide an optimal solution and is often processed at every instant that the observations come into the system, thus continuously providing the navigator with the vessels position.

6.1.1 Predicting, Filtering and Smoothing

The filter is often split into three sections known as prediction, filtering and smoothing. The same scenario of the moving vessel will be used to represent the workings of these operations and they occur within the filter in the following order

1. Using the vessel's present results (positions, velocities and their precisions) and the model of the vessels movement, the results at the next epoch (in the future) are **predicted.**

2. At the next epoch (the present), new position data is now available and is incorporated with the predicted results to obtain the **filtered** solution.

3. Starting from this last filtered solution, the filtered and predicted results from previous epochs can then be **smoothed** to obtain better results for these epochs.

The smoothed solution is often carried out post-mission using all the available solutions. Cross (1983) explains these terms showing a vessel's track with its positions obtained at different epochs (t_1, t_2, ...,t_i). The epoch t_i is considered the present and t_j is the time at which the vessel's position is to be estimated. Prediction, filtering , and smoothing can then be defined as

$t_i < t_j$	predicting,
$t_i = t_j$	filtering
$t_i > t_j$	smoothing

and are represented in Fig 6.1.

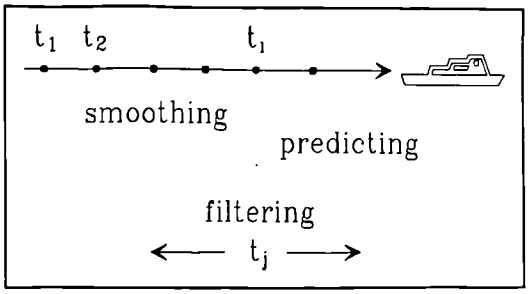


Fig 6.1 Predicting, filtering and smoothing

6.1.2 GPS Processing using the Kalman Filter

There are many different examples in GPS processing for which the Kalman filter is the ideal tool. For more details of these examples, the reader should refer to Merminod (1989).

- **Navigation.** Kalman filters have been used for navigation on land, sea and in space, and have been developed to incorporate most positioning systems. This situation uses the example above, with the vessels motion being modelled and combined with the positioning observations. Models for this application will be discussed later in this thesis.

- **Pseudorange smoothing**. The GPS pseudorange values at future epochs can be predicted, as can phase observations. The actual pseudorange values can then be combined with the phase observations within the filtered solution to create the so-called "phase-smoothed pseudoranges". Similar operations can be carried out using GPS doppler and pseudorange observations.

- **Orbit computations.** The Keplerian elements contained within the GPS navigation message are predicted values. These values are updated, or filtered, at the GPS Control Segment using tracking data from the monitor stations.

More predictions based on these values are then made, which are subsequently uploaded to the satellites.

- **Integer ambiguity resolution.** The Kalman filter may be used to help resolve the integer number of cycles ambiguity for phase processing. Filtering could be particularly important for maintaining a small search area for the single epoch FARS phase processing techniques.

- **Performance monitoring.** The Kalman filter is an ideal tool for the integrity monitoring of the GPS system. Outages in individual pseudorange measurements can be readily identified, as can phase cycle slips, by using associated reliability test statistics. Such testing procedures will be discussed later .

6.2 THE KALMAN FILTER MATHEMATICAL MODELS

The Kalman filter incorporates information from both the incoming observations and the perceived dynamics of the system. It therefore uses two mathematical models

1. the **measurement**, **functional**, or **primary** model that relates the parameters to the observations, and
2. the **dynamic**, **kinematic**, or **secondary** model that relates the parameters at epoch t_i to those at epoch t_{i-1}.

Before describing the two models, it is necessary to introduce the state vector containing the desired parameters.

6.2.1 The State Vector

The **state vector** within the Kalman filter algorithms is a vector containing the parameters to be estimated along with additional parameters that model the behaviour of the system. If the system is that of a moving vessel with the positioning system ultimately providing its easting and northing position, then the state vector, $\mathbf{x}_i$, may comprise of

$$\mathbf{x}_i = \begin{bmatrix} E_i & N_i & \dot{E}_i & \dot{N}_i & \ddot{E}_i & \ddot{N}_i \end{bmatrix}^T \tag{6.1}$$

where

E_i, N_i are the position components in easting and northing

$\dot{E}_i, \dot{N}_i$ are the velocity components

$\ddot{E}_i, \ddot{N}_i$ are the acceleration components

and the subscript i denotes the state at time i.

Along with the state vector, the precision of these parameters held as variances and their covariances are also computed. These are contained within the non-singular covariance matrix of the parameters, $\mathbf{Cx}_i$.

6.2.2 The Measurement Model

This model relates the measurements to the parameters within the state vector. Suppose that $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_i$ are the true values of the parameters at epochs t_1, t_2, ..., t_i and that $\mathbf{l}_1$, $\mathbf{l}_2$, ..., $\mathbf{l}_i$ are vectors of the observations at the corresponding epochs (Cross, 1983). The measurement model can be written as

$F_1(\bar{\mathbf{x}}_1) = \bar{\mathbf{l}}_1$ at epoch t_1

$F_2(\bar{\mathbf{x}}_2) = \bar{\mathbf{l}}_2$ at epoch t_2,

or more generally

$$F_i(\bar{x}_i) = \bar{l}_i \qquad \text{at epoch } t_i. \tag{6.2}$$

if the relationship between the observations and the parameters is non-linear, then the model must be linearised. The form of the linear measurement model that is used within the Kalman filter algorithms is given by

$$\mathbf{A}_i\mathbf{x}_i = \mathbf{b}_i + \mathbf{v}_i \tag{6.3}$$

where

$\mathbf{A}_i$	is the design, or Jacobian, matrix given by $\partial F_i/\partial \mathbf{x}_i$
$\mathbf{x}_i$	is the estimate of the state vector
$\mathbf{b}_i$	is the "observed - computed" quantities
$\mathbf{v}_i$	is the vector of residuals.

Typically, the measurement model will not solve for all the parameters within the state vector at any one epoch. For example, in (6.1) the positioning system may only be used to determine the easting and northing coordinates of the vessel.

6.2.3 The Dynamic Model

The dynamic model represents the behaviour of the system as it varies with time. The general functional relationship therefore relates the state vector between two subsequent epochs ($F_{i-1,i}$), and can be expressed as

$$F_{i-1,i}(\bar{x}_{i-1}, \bar{x}_i, t_{i-1}, t_i) \tag{6.4}$$

where

$\bar{x}_i$	is the true state vector at time t_i, and
$\bar{x}_{i-1}$	is the true state vector at time t_{i-1}.

This is reduced to the discrete linearised form

$$\mathbf{x}_i = \mathbf{M}_{i-1}\mathbf{x}_{i-1} + \mathbf{y}_{i-1} \tag{6.5}$$

where $\mathbf{M}_{i-1}$ is the transition matrix from time t_{i-1} to time t_i,

$\mathbf{y}_{i-1}$ is the dynamic model noise (residuals) from time t_{i-1} to time t_i.

6.2.4 Example of a Dynamic Model

The transition matrix, $\mathbf{M}_{i-1}$, describes the behaviour of the system between the two time epochs and will depend on how that behaviour has been modelled. It is often (especially for navigation systems) sufficient to use a very simple polynomial model to describe the behaviour. Consider y(t) to be a continuous process, which is then expanded using Taylor's theorem about the point t to determine the process at a later time ($t + \delta t$). This expansion is given by (Cross, 1987b)

$$y(t+\delta t) = y + \dot{y}\delta t + \ddot{y}\delta t^2/2 + \dddot{y}\delta t^3/6 + \cdots \tag{6.6}$$

where $\dot{y}$ represents the first time differential of the process.

If the differentials are also considered continuous then two further Taylor's theorem expansions can be written. If all terms higher than the third differential are said to be negligible, the two expansions become

$$\dot{y}(t+\delta t) = \dot{y} + \ddot{y}\delta t + \dddot{y}\delta t^2/2 \tag{6.7}$$

$$\ddot{y}(t+\delta t) = \ddot{y} + \dddot{y}\delta t \tag{6.8}$$

If two time epochs t_{i-1} and t_i are considered to be an interval δt apart, then equations (6.6), (6.7), and (6.8) can be combined and expressed in matrix notation as

$$\begin{bmatrix} y \\ \dot{y} \\ \ddot{y} \end{bmatrix}_i = \begin{bmatrix} 1 & \delta t & \delta t^2/2 \\ 0 & 1 & \delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ \dot{y} \\ \ddot{y} \end{bmatrix}_{i-1} + \begin{bmatrix} \delta t^3/6 \\ \delta t^2/2 \\ \delta t \end{bmatrix} [\dddot{y}]_{i-1} \tag{6.9}$$

To represent a positioning scenario, let the process y(t) be the easting and northing coordinates of a vessel. Equation (6.9) then contains the state vector (6.1) and can becomes

$$\begin{bmatrix} E \\ N \\ \dot{E} \\ \dot{N} \\ \ddot{E} \\ \ddot{N} \end{bmatrix}_i = \begin{bmatrix} 1 & 0 & \delta t & 0 & \delta t^2/2 & 0 \\ 0 & 1 & 0 & \delta t & 0 & \delta t^2/2 \\ 0 & 0 & 1 & 0 & \delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} E \\ N \\ \dot{E} \\ \dot{N} \\ \ddot{E} \\ \ddot{N} \end{bmatrix}_{i-1} + \begin{bmatrix} \delta t^3/6 & 0 \\ 0 & \delta t^3/6 \\ \delta t^2/2 & 0 \\ 0 & \delta t^2/2 \\ \delta t & 0 \\ 0 & \delta t \end{bmatrix} \begin{bmatrix} \dddot{E} \\ \dddot{N} \end{bmatrix}_{i-1} \tag{6.10}$$

which is of the dynamic model form

$$\mathbf{x}_i = \mathbf{M}_{i-1}\mathbf{x}_{i-1} + \mathbf{T}\mathbf{g} \tag{6.11}$$

The vector $\mathbf{y}_{i-1}$ in (4.5) is now represented by the product of a matrix **T** by a random noise vector **g**. The vector **g**, representing the rate of change of the vessel's acceleration, is often referred to as the **driving noise** or **forcing function.**

6.2.5 The Stochastic Models

The Kalman filter algorithms combine the dynamic and measurement models together to create an optimum solution. In order to do this, both models must be given an appropriate weighting which is expressed by their stochastic models - their covariance matrices.

The stochastic model of the measurement model is expressed as the covariance matrix of the observations, $\mathbf{Cl}_i$ ($= \mathbf{W}_i^{-1}$). This matrix contains the variances of the individual observations from the particular sensor(s) and any covariances among them.

The stochastic model of the dynamic model, **Cy**, requires a complicated numerical integration process for most applications (Cross, 1987b). For the polynomial model example in (6.10), however, the stochastic model can be obtained by applying Gauss' propagation of error law. The driving function, **g**, representing the rate of change of the vessel's acceleration can be considered to be a random process with a diagonal covariance matrix, **Cg**. From the error propagation law, the covariance matrix of the dynamic noise can be found by

$$\mathbf{Cy} = \mathbf{TCgT}^T \tag{6.12}$$

For the dynamic model equations in (6.10), the stochastic model can be obtained from the matrix products

$$\begin{bmatrix} \delta t^3/6 & 0 \\ 0 & \delta t^3/6 \\ \delta t^2/2 & 0 \\ 0 & \delta t^2/2 \\ \delta t & 0 \\ 0 & \delta t \end{bmatrix} \begin{bmatrix} \sigma^2_{\ddot{E}} & 0 \\ 0 & \sigma^2_{\ddot{N}} \end{bmatrix} \begin{bmatrix} \delta t^3/6 & 0 & \delta t^2/2 & 0 & \delta t & 0 \\ 0 & \delta t^3/6 & 0 & \delta t^2/2 & 0 & \delta t \end{bmatrix} \tag{6.13}$$

where $\sigma_{\ddot{E}/\ddot{N}}$ is the standard deviation of the rate of change of the vessel's acceleration in either direction.

Cross (1987b) gives two reasons for assigning the correct values for the two stochastic models.

1. The covariances matrix of the state vector (computed from the Kalman filter algorithms) depends on the two stochastic models. If these models are incorrect, the resultant precisions obtained from the state covariance will also be wrong. This is most serious if the models are over optimistic and the positioning results appear to be of a higher quality than they actually are.

2. The second reason depends not on the actual sizes of the covariance matrices, but on the ratio between them. If the values of **Cl** are too over optimistic then the Kalman filter solution will tend to follow the solution from the measurement model with the dynamic model having very little effect. If the position of a moving vessel was plotted, it would follow a rugged track which would be a cause of the errors and noise within the positioning system. This is known as **under-filtering**. On the other hand, if the values of **Cy** are too optimistic in relationship to **Cl**, then the vessel's track will appear extremely smooth and may not follow the vessel's true track. This is **over-filtering**.

It should be noted that both these circumstances can be overcome by applying the correct statistical tests.

6.3 THE KALMAN FILTER ALGORITHMS

6.3.1 The Kalman Filter Assumptions

The Kalman filter is numerical optimal estimator. This can be defined as an algorithm which processes all of the available information to produce an estimate of the state of a system whilst satisfying some predefined optimality criterion (Cross, 1987b). The Kalman filter satisfies the "least squares" criterion by minimising the quadratic form (Cross, 1987a)

$$\mathbf{v}_{i-1}^{T}\mathbf{C}\hat{\mathbf{x}}_{i-1}^{-1}(+)\mathbf{v}_{i-1} + \mathbf{v}_{i}^{T}\mathbf{Cl}_{i}^{-1}\mathbf{v}_{i} + \mathbf{y}_{i-1}^{T}\mathbf{Cy}_{i-1}^{-1}\mathbf{y}_{i-1} \qquad (6.14)$$

where $\mathbf{C}\hat{\mathbf{x}}_{i-1}^{-1}(+)$ is the covariance of the filtered state (6.22)

Least squares estimators are often referred to as best linear unbiased estimates (or BLUEs) and their justification can be found in many texts, for example Cross (1983).

A basic assumption in the formulation of the standard Kalman filter equations is that there is no physical correlation between the observations at epoch t_i and at a different epoch, t_k. There is also no equivalent correlation within the dynamic model, nor any correlation between the two models. Following these assumptions, the statistical models for the Kalman filter can be described as (Gao et al, 1992)

$$E(\mathbf{y}_k\mathbf{y}_i)=\begin{cases}\mathbf{Cy} & i=k\\ 0 & i\neq k\end{cases} \tag{6.15}$$

$$E(\mathbf{v}_k\mathbf{v}_i)=\begin{cases}\mathbf{Cl} & i=k\\ 0 & i\neq k\end{cases} \tag{6.16}$$

$$E(\mathbf{y}_k\mathbf{v}_i)=0 \tag{6.17}$$

6.3.2 The Prediction Equations

The prediction, or time update, algorithms represent the effects of the dynamic model on the state and its covariance. Since the vector **y** is not actually known, it is assumed to be zero and the state vector is given by

$$\hat{\mathbf{x}}_i(-)=\mathbf{M}_{i-1}\hat{\mathbf{x}}_{i-1}(+) \tag{6.18}$$

The algorithms and notation follow that of Cross (1987a). The symbol ^ denotes an estimated quantity. The symbols (-) and (+) after a matrix or vector refer to instants immediately before and after measurement updates respectively, and are referred to as the predicted and filtered solutions. Appendix H gives other notation commonly found within Kalman filter texts.

The predicted covariance matrix of the state can be obtained from

$$\mathbf{C}\hat{\mathbf{x}}_i(-)=\mathbf{M}_{i-1}\mathbf{C}\hat{\mathbf{x}}_{i-1}(+)\mathbf{M}_{i-1}^{T}+\mathbf{Cy} \tag{6.19}$$

It is clear that equations (6.17) and (6.18) require starting values for filtered state vector and its covariance matrix at the very first epoch. These values should be as near as possible to the "truth", in order for the Kalman filter solutions to quickly

tend towards optimality. The further the initial starting values from the actual values, the more epochs it takes the filter to reach this situation.

6.3.3 The Filtering Equations

The algorithms for filtering, or the measurement update process, provide a best estimate for the state and its covariance by incorporating the observations with the predicted values from (6.17) and (6.18). The filtered solutions are given by

$$\mathbf{G}_i = \mathbf{C}\hat{\mathbf{x}}_i(-)\mathbf{A}_i^T(\mathbf{Cl}_i + \mathbf{A}_i\mathbf{C}\hat{\mathbf{x}}_i(-)\mathbf{A}_i^T)^{-1} \tag{6.20}$$

$$\hat{\mathbf{x}}_i(+) = \hat{\mathbf{x}}_i(-) + \mathbf{G}_i(\mathbf{b}_i + \mathbf{A}_i\hat{\mathbf{x}}_i(-)) \tag{6.21}$$

$$\mathbf{C}\hat{\mathbf{x}}_i(+) = (\mathbf{I} - \mathbf{G}_i\mathbf{A}_i)\mathbf{C}\hat{\mathbf{x}}_i(-) \tag{6.22}$$

G is the Kalman filter **gain** matrix which controls the amount the predicted state vector with its covariance matrix is affected by the measurements.

A more stable solution for the filtered covariance matrix, which should always maintain symmetry, is given by

$$\mathbf{C}\hat{\mathbf{x}}_i(+) = (\mathbf{I} - \mathbf{G}_i\mathbf{A}_i)\mathbf{C}\hat{\mathbf{x}}_i(-)(\mathbf{I} - \mathbf{G}_i\mathbf{A}_i)^T + \mathbf{G}_i\mathbf{Cl}_i\mathbf{G}_i^T \tag{6.23}$$

6.3.4 The Smoothing Equations

The smoothing process is a backward operator that is often carried out after the completion of the system to fully optimise the state vector and its covariance matrix. If there are a total of n epochs resulting in n state vectors (ie i=n), then the smoothing solution will produce results at epochs n-1, n-2, ..., 1. The smoothing

algorithms require a starting position, for which the filtered solutions at the final epoch is used - thus for this epoch the filtered and smoothed solutions are the same. Using the notation (s) to indicate a smoothed solution, the process can be written as

Initially letting

$$\hat{\mathbf{x}}_n(s) = \hat{\mathbf{x}}_n(+) \quad \& \quad \mathbf{C}\hat{\mathbf{x}}_n(s) = \mathbf{C}\hat{\mathbf{x}}_n(+) \tag{6.24}$$

then (for i = n, n - 1, n - 2, ... etc)

$$\mathbf{S}_i = \mathbf{C}\hat{\mathbf{x}}_{i-1}(+)\mathbf{M}_{i-1}^T\mathbf{C}\hat{\mathbf{x}}_i^{-1}(-) \tag{6.25}$$

$$\hat{\mathbf{x}}_{i-1}(s) = \hat{\mathbf{x}}_{i-1}(+) + \mathbf{S}_i(\hat{\mathbf{x}}_i(s) - \hat{\mathbf{x}}_i(-)) \tag{6.26}$$

$$\mathbf{C}\hat{\mathbf{x}}_{i-1}(s) = \mathbf{C}\hat{\mathbf{x}}_{i-1}(+) + \mathbf{S}_i(\mathbf{C}\hat{\mathbf{x}}_i(s) - \mathbf{C}\hat{\mathbf{x}}_i(-))\mathbf{S}_i^T \tag{6.27}$$

Equations (6.25) to (6.27) are used recursively to smooth back as many epochs as desired and the process is known as fixed-interval smoothing, or as the R-T-S method (Rauch, Tung and Striebel (1965) were first responsible for its description). These algorithms do have many restrictions in the amount of data that has to be stored for all of the epochs that require smoothing. This includes the predicted and filtered state vectors with their appropriate covariance matrices at all epochs, and the transition matrices at all update points. Merminod (1989) gives a thorough selection of other smoothing algorithms that may be used and that require less storage necessities.

6.3.5 Non-linear Models

It is quite rare to have a practical solution for which both the measurement and dynamic models are linear. Iterative solutions are then required within the Kalman filter to obtain the predicted and filtered states.

For a *non-linear measurement model*, the linearised model as in (6.3) is used to solve for the corrections to provisional values, thus in

$$\mathbf{A}_i\mathbf{x}_i = \mathbf{b}_i + \mathbf{v}_i \tag{6.3}$$

where

$\mathbf{A}_i$	is given by $\partial \mathrm{F}_i / \partial \tilde{\mathbf{x}}_i(+)$
$\mathbf{b}_i$	is the "observed - computed" given by $\mathbf{l}_i - \mathrm{F}_i(\tilde{\mathbf{x}}_i(+))$
$\mathbf{x}_i$	is the corrections to the provisional value of the filtered state vector, $\tilde{\mathbf{x}}_i(+)$

by letting $\delta\mathbf{x}_i = \hat{\mathbf{x}}_i(-) - \tilde{\mathbf{x}}_i(+)$, the filtered state is calculated iteratively by

$$\tilde{\mathbf{x}}_i(+) = \hat{\mathbf{x}}_i(-) + \mathbf{G}_i(\mathbf{b}_i - \mathbf{A}_i\delta\mathbf{x}_i) \tag{6.28}$$

Equation (6.28) is solved iteratively by firstly letting $\tilde{\mathbf{x}}_i(+) = \hat{\mathbf{x}}_i(-)$, repeating until there is no significant change in $\tilde{\mathbf{x}}_i(+)$, and ultimately setting the filtered state vector to this final value, ie $\hat{\mathbf{x}}_i(+) = \tilde{\mathbf{x}}_i(+)$. Since the design matrix ($\mathbf{A}_i$) depends on the provisional values, so will the gain matrix ($\mathbf{G}_i$), and these, along with the $\mathbf{b}_i$ vector should be recomputed during each iteration. In some cases, dependent on model design, the design and gain matrices will not significantly alter and therefore need not be recomputed.

For a *non-linear dynamic model*, the predicted state vector, the transition matrix, and the dynamic model noise needs to be computed at every recursion of the filter. Cross (1987a) develops the formula for the case of the dynamic model being a first order non-linear differential equation of the form

$$\dot{\mathbf{x}}_i = \mathbf{F}(\mathbf{x}_i, t) \tag{6.29}$$

The predicted state vector can then be found from

$$\hat{\mathbf{x}}_i(-) = \hat{\mathbf{x}}_{i-1}(+) + \int_0^{\Delta t} \dot{\mathbf{x}} dt \tag{6.30}$$

and the transition matrix via the relationship

$$\frac{\partial \mathbf{F}}{\partial \mathbf{x}} \mathbf{M} = \dot{\mathbf{M}} \tag{6.31}$$

and finally the covariance matrix of the dynamic noise by

$$\mathbf{Cy}_i = \int_{t_{i-1}}^{t_i} \mathbf{M}_{i-1} \mathbf{Cy}_{i-1} \mathbf{M}_{i-1}^T dt \tag{6.32}$$

6.4 STATISTICAL TESTING OF THE KALMAN FILTER

The approach given here for the statistical testing of Kalman filters is that developed by the Delft Geodetic Computing Centre. The procedures are for the detection, identification and adaptation (DIA) of the overall model. Local tests are carried out on information at a particular epoch (the predicted state and the values of the

incoming observations) and test for model errors at that epoch, ie blunders. Global tests are used to test for unmodelled global trends that may build up and are not detected by the local tests. This section will deal only with the overall model and slippage tests (detection and identification of biases) and for details of the adaptation procedures (eliminating the presence of biases), the reader should refer to Teunissen (1990).

6.4.1 The Predicted Residuals

Both the global and local tests are carried out on the **predicted residuals**, or the **innovation sequence**, which is defined "as the difference between the actual system output and the predicted output based on the predicted state." (Teunissen 1990). Thus at epoch t_i, the predicted residuals, $\hat{\mathbf{v}}_i(-)$, can be written as

$$\hat{\mathbf{v}}_i(-) = \mathbf{b}_i - \mathbf{A}_i\hat{\mathbf{x}}_i(-) \tag{6.33}$$

with the covariance matrix as

$$\mathbf{C}\hat{\mathbf{v}}_i(-) = \mathbf{Cl}_i + \mathbf{A}_i\mathbf{C}\hat{\mathbf{x}}_i(-)\mathbf{A}_i^T \tag{6.34}$$

In the case of a non-linear measurement model, the predicted residuals are given by the initial "observed - computed" values held in the $\mathbf{b}_i$ vector.

6.4.2 Local Model Tests

The detection of local model errors is carried out by the **local overall model** (LOM) test statistic and then the identification of the observation(s) which contain model errors (blunders) is carried out via a one-dimensional **local slippage** (LS) test.

If the dynamic model is valid, then the predicted residuals should be a zero mean white noise (stationary random) process with known covariance, and this is used to detect the overall validity of the model. The two local hypotheses to detect model misspecification at epoch t_i are

$$\begin{aligned} H_o^i \quad & \hat{\mathbf{v}}_i(-) \approx N(0, \mathbf{C}\hat{\mathbf{v}}_i(-)) \\ H_o^i \quad & \hat{\mathbf{v}}_i(-) \approx N(\mathbf{K}\nabla, \mathbf{C}\hat{\mathbf{v}}_i(-)) \end{aligned} \tag{6.35}$$

where $\mathbf{K}$ is a m_i-by-b matrix specifying the type of error that is being sought (assumed known),

∇ is a b-by-1 vector specifying the whereabouts of the error (unknown),

m_i is the degrees of freedom (the number of observations) at epoch t_i.

The size of b depends on the type of model error within the alternative hypothesis and can range from 1 to m_i. Teunissen (1990) explains that if a particular sensor failure (which is being tested for) requires additional parameters within the state vector to model it, then b is equal to the number of these extra parameters. This section will deal entirely on the one-dimensional case (b = 1), for which the vector ∇ becomes a scalar and the matrix $\mathbf{K}$ becomes a vector denoted by $\mathbf{c}$.

The uniformly-most-powerful-invariant (UMPI) test statistic for detecting model errors in the null hypothesis is

$$T^i = \mathbf{v}_i(-)^T \mathbf{C}\mathbf{v}_i(-)^{-1} \mathbf{v}_i(-) \tag{6.36}$$

$$T^i_{LOM} = T^i / m_i \tag{6.37}$$

This value is then tested for a local model error, which is said to have occurred if

$$T^i_{LOM} \geq F_\alpha(m_i, \infty, 0) \tag{6.38}$$

where $F_\alpha(m_i, \infty, 0)$ is the α percentile central F-distribution value with m_i and ∞ degrees of freedom.

Once a model misspecification has occurred, the individual predicted residuals are analysed to determine the observation with the most likely model error. This uses the data snooping LS test statistic

$$t^i_{LS} = \frac{\mathbf{c}^T \mathbf{C}\hat{\mathbf{v}}_i(-)^{-1} \hat{\mathbf{v}}_i(-)}{(\mathbf{c}^T \mathbf{C}\hat{\mathbf{v}}_i(-)^{-1} \mathbf{c})^{1/2}} \tag{6.39}$$

where $\mathbf{c}$ is a null vector except for unity at the corresponding observation being tested, ie if, at epoch t_i, there were a total of 5 observations and observation 3 was being tested then

$$\mathbf{c} = [0 \quad 0 \quad 1 \quad 0 \quad 0]^T \tag{6.40}$$

It should be stressed that the c vector does not necessarily have just one element of unity, but is dependent on the test being carried out.

The slippage test is carried out for all observations at epoch t_i, and the most likely blunder is said to have occurred at the observation for which $|t^i_{LS}|$ is at a maximum and this observation can be rejected. The LOM test of (6.37) can now be recomputed to test for model errors within the remaining data. The tests are

therefore repeated recursively until no further blunders are detected. Various alternative hypotheses can also be tested for.

6.4.3 Global Model Errors

The local model tests may prove insensitive to unmodelled trends within the mathematical model which can build up slowly with time. The test statistics used to detect such biases need to incorporate information from previous epochs within the filter, and therefore require memory capabilities. The **global overall model** (GOM) test statistic along with the **global slippage** (GS) test are for the detection and identification of such phenomena.

The following two hypotheses are considered for the detection of global model errors occurring between epochs t_l and t_k.

$$\begin{aligned} H_o^{l,k} \qquad & \hat{\mathbf{v}}_{l,k}(-) = N(0, \mathbf{C}\hat{\mathbf{v}}_{l,k}(-)) \\ H_o^{l,k} \qquad & \hat{\mathbf{v}}_{l,k}(-) = N(\mathbf{K}_{l,k}\nabla, \mathbf{C}\hat{\mathbf{v}}_{l,k}(-)) \end{aligned} \tag{6.41}$$

where

$$\hat{\mathbf{v}}_{l,k}(-) = \left[\hat{\mathbf{v}}_l(-)^T \quad \hat{\mathbf{v}}_{l+1}(-)^T \quad \cdots \quad \hat{\mathbf{v}}_k(-)^T\right]^T \tag{6.42}$$

The matrix $\mathbf{K}_{l,k}$ (now a $\sum_{i=l}^{k} m_i$-by-b matrix) is again specifying the type of model error likely to occur, but now ∇ is specifying the time period in which it occurs.

The UMPI test statistic is given by

$$T_{GOM}^{l,k} = \frac{\sum_{i=l}^{k} \mathbf{v}_i(-)^T \mathbf{C}\mathbf{v}_i(-)^{-1} \mathbf{v}_i(-)}{\sum_{i=l}^{k} m_i} \tag{6.43}$$

which can be written in the recursive form of

$$T_{GOM}^{1,k} = T_{GOM}^{1,k-1} + \frac{T^k - m_k T_{GOM}^{1,k-1}}{\sum_{i=1}^{k} m_i} \tag{6.44}$$

A global model error is said to be present within the system if

$$T_{GOM}^{1,k} \geq F_\alpha(\sum_{i=1}^{k} m_i, \infty, 0) \tag{6.45}$$

The global slippage test is more difficult to perceive and depends explicitly on the alternative hypothesis (Teunissen, 1990). For the one-dimensional case $\mathbf{K}_{1,k}$ reduces to the vector $\mathbf{c}_{1,k}$, and the global slippage test is given by

$$t_{GS}^{1,k} = \frac{\sum_{i=1}^{k} \mathbf{c}_{1,k_i}{}^T \mathbf{C}\hat{\mathbf{v}}_i(-)^{-1} \hat{\mathbf{v}}_i(-)}{\left[\sum_{i=1}^{k} \mathbf{c}_{1,k_i}{}^T \mathbf{C}\hat{\mathbf{v}}_i(-)^{-1} \mathbf{c}_{1,k_i}\right]^{1/2}} \tag{6.46}$$

The vector $\mathbf{c}_{1,k_i}$ depends on how the particular global error propagates through the Kalman filter and must be determined within the filter itself. Teunissen and Salzmann (1989) give three different examples, dependent on model error, of how to compute this vector for circumstances that may commonly arise when using the Kalman filter to process an integrated navigation system. The misspecifications that are covered are

1. a permanent slip in the state vector that starts at time t_1,
2. a single slip in the vector of observables that starts at time t_1, and
3. a sensor failure that starts at time t_1.

6.5 OTHER FILTERS AND FILTER TERMINOLOGY

6.5.1 The Bayes Filter

The filtering algorithms derived by Kalman are not the only least squares optimal estimation procedures for obtaining the states of a time varying system. Another such method is the **Bayes filter** which produces identical results as the Kalman form.

The only difference between both filters is in the manner the filtered covariance matrix of the state vector and the gain matrix are computed. For the Bayes filter these are computed as follows (Gelb, 1974).

$$\mathbf{C}\hat{\mathbf{x}}_i(+) = \left[\mathbf{C}\hat{\mathbf{x}}_i(-)^{-1} + \mathbf{A}_i^T\mathbf{Cl}_i^{-1}\mathbf{A}_i\right]^{-1} \tag{6.47}$$

$$\mathbf{G}_i = \mathbf{C}\mathbf{x}_i(+)\mathbf{A}_i^T\mathbf{Cl}^{-1} \tag{6.48}$$

The Bayes form therefore requires two main matrix inversions during each epoch of the filter (the matrix **Cl** only has to be held in its inverted form). Both these matrices are of the size of the number of parameters within the state vector. There is only one matrix inversion necessary in the Kalman form (6.20) and this is a matrix of the size of the number of observations coming into the filter at the particular epoch. It is therefore computationally more efficient to use the Kalman form when there are more parameters than observations within the system - this is typical for the navigation scenario. On the other hand, if a large number of observations were contributing to the same parameters then the Bayes form of the filter would be more efficient. Such an example might be when using the filter to model the shape of a seismic streamer with observations form GPS, laser tracking, transponders, magnetic compasses etc, being available.

6.5.2 Filter Types and Terminology

The Kalman or Bayes filter is presently being used to process data from an extensive range of different applications. This has led to advances in the use of single filters, and the use of multiple filters during the estimation which, in turn, have led to new terminology associated with the filter. Some of these attributes are listed below (Merminod, 1989).

6.5.2.1 Single filters

The following are some of the terms often used when describing a single filter.

- **Extended.** This is used to describe non-linearities between the measurements and/or the parameters by linearising the measurement and dynamic models. If the performance can be improved via iteration (as in [6.28] and [6.30]), then this is known as an **iterated extended** filter.

- **Augmented.** The size of the state vector has been increased to model an additional relationship that is common to several observations. A bias within the positioning system being used is such an example.

- **Constrained.** Relations involving only the parameters are included within the filter. A typical example is a height constrained solution within a navigation system, where the value for height is a priori knowledge from previous height values or some assumptions being made (eg the sea surface is the same as the geoid). Although the height value is not an actual observation it can be modelled in the same fashion and entered as a pseudo-measurement.

- **Adaptive.** The idea of an adaptive Kalman filter has already been implied in section 6.3. The DIA local and global model tests based on the predicted residuals are used to detect errors within the measurement model. Once a misspecification has been identified the filter has to be adapted by, for example, eliminating outliers detected by local tests or by augmenting the state vector to model some previously unknown bias detected by the global tests.

6.5.2.2 Combined filters

Several filters can be combined in three different ways. Firstly, **horizontal** filtering is where the results from one filter are used as the input for another filter. Secondly, **vertical** filtering is where different data sets are passed through different filters - for instance, to monitor the performance of one positioning system against another. The final type of combined filter is where the results from a vertical structure are processed together within another filter. This is therefore a combination of the vertical and horizontal structures.

Different terminology is often associated with filters which use either of these structures.

- **Cascaded.** This is a horizontal structure where the initial filter is used to merely compress the data before it is processed by the main filter.

- **Distributed.** This is a vertical structure with two, or more, filters independently processing data from different sensors. An example is the processing of Syledis and DGPS as primary and secondary navigation systems on a vessel.

- **Federated.** This uses the combined structure to process data from different sensors. An example of a federated filter is the integration of GPS and INS where both sensors are initially processed using independent local filters. The results are then passed to the main filter which estimates an optimal solution for the state based on the two sets of results. The results from the main filter can then be used to benefit the performance within the local filters. For this example, the short term stability of INS can be used to aid the ambiguity finding for GPS phase processing, and the long term stability of GPS can be used as updates to correct the drift within the INS.

- **Parallel.** The concept of parallel filtering is to test different hypotheses when processing the same data. There is no interaction between the different filters, although their different precision measures would be analysed to determine the best hypothesis for the system. Merminod (1989) gives the example of selecting the set of integer ambiguities via parallel filtering.

The aforementioned terminology in shown in Table 6.1 relating the processing of one or more sensors with either one or more filters (Merminod, 1989).

	single filter	*multi filter*
single sensor	classical	cascaded
multi-sensor	distributed	federated

Table 6.1 Filter terminology

CHAPTER SEVEN
THE EFFECT OF TEMPORAL CORRELATION ON POSITION, PRECISION, AND RELIABILITY

7.1 LEAST SQUARES REPRESENTATION OF THE KALMAN FILTER

One way to determine the effect of temporal correlation on position and precision results is to use the least squares observation approach and to account for the correlation within the covariance matrix. The following chapter describes how this is done, and then examines the ability of the standard Kalman filter algorithms to detect blunders within correlated data sets.

7.1.1 The Kalman Filter Statistical Assumptions

The Kalman filter combines two models (the measurement and dynamic models) to compute both the components of a state vector, and their precisions, at particular moments in time, or epochs. The two models can be described as (Cross, 1987a):

The measurement model

$$\mathbf{A}_i\mathbf{x}_i = \mathbf{b}_i + \mathbf{v}_i \qquad (6.3)$$

with the covariance matrix of the set of observations

$$\mathbf{Cl}_i$$

and the dynamic model

$$\mathbf{x}_i = \mathbf{M}_{i-1}\mathbf{x}_{i-1} + \mathbf{y}_{i-1} \tag{6.5}$$

with the covariance matrix of the dynamic noise

$\mathbf{Cy}_{i-1}$

These models are incorporated in the standard prediction, filtering and smoothing equations as described in chapter six. The derivation of the equations (Cross, 1987a) depends on the statistical assumptions

$$E(\mathbf{y}_k\mathbf{y}_i) = \begin{cases} \mathbf{Cy}_k & i = k \\ 0 & i \neq k \end{cases} \tag{6.15}$$

$$E(\mathbf{v}_k\mathbf{v}_i) = \begin{cases} \mathbf{Cl}_k & i = k \\ 0 & i \neq k \end{cases} \tag{6.16}$$

This indicates that both the measurement and system (dynamic) noise are a white noise process and are therefore uncorrelated between epochs. That is, the magnitude of the observations being received at one epoch are not related to those received at the previous epoch, nor will have any bearing on what will be received at the next epoch. This is rarely true in surveying applications. The following sections will introduce some possible causes for temporal correlation between observation epochs and will analyse its effect on the position and corresponding precisions as output from a Kalman filter. This is carried out by representing the Kalman filter in the general linear least squares observation equation form, and by introducing the statistical model:

$$E(v_k v_i) \neq 0 \qquad i \neq k \tag{7.1}$$

7.1.2 Correlation within GPS Observations

Most types of measurements used for positioning in navigation are correlated in time for a number of reasons. For instance GPS pseudorange observations will be temporally correlated due to atmospheric refractions, hardware designs in both the satellite and the receiver, any smoothing of the raw observables (integrated doppler or phase aiding), other error sources, and, in particular, selective availability. The amount of temporal correlation within actual pseudorange observations, recorded between March 1991 and March 1993, has been described in chapter five.

In DGPS operations, the reason for the rapid transfer of the pseudorange correction message between the reference and the mobile stations is that the errors in the measurements are assumed to be correlated over a short period of time. The differencing technique of DGPS is only "valid" if the errors at the time of generation at the reference station are "similar" to the errors experienced at the mobile station when the corrections are being applied. The validity will depend mainly on the age of the corrections as a ratio of the frequency of the satellite clock dither being applied through SA, although other error sources (orbit, atmospheric, and multipath errors) will also have an effect.

7.1.3 The Least Squares Observation Equations

The least squares observation equations can be expressed in their linear form as: (Cross, 1983)

$$\mathbf{Ax} = \mathbf{b} + \mathbf{v} \tag{7.2}$$

where

A	design matrix relating the observations to the parameters
x	vector of parameters
b	vector of observations
v	vector of residuals

with a covariance matrix of the set of observations

$\mathbf{W}^{-1}$

By definition, the least squares solution for the vector of parameters, $\hat{\mathbf{x}}$, is the solution that minimises the sum of the squares of the weighted residuals, $\mathbf{v}^T\mathbf{W}\mathbf{v}$. The usual approach for computing $\hat{\mathbf{x}}$ is via the normal equations, where

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{W}\mathbf{b}) \tag{7.3}$$

$$\hat{\mathbf{v}} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{b} \tag{7.4}$$

with the a posteriori covariance matrices

$$\mathbf{C}_{\hat{\mathbf{x}}} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1} \tag{7.5}$$

$$\mathbf{C}_{\hat{\mathbf{v}}} = \mathbf{W}^{-1} - \mathbf{A}\mathbf{C}_{\hat{\mathbf{x}}}\mathbf{A}^T \tag{7.6}$$

For a full derivation of these equations refer to Cross (1983).

7.1.4 The Kalman Filter as Least Squares Observation Equations

The least squares equations can, theoretically, be used to represent the Kalman filter. Consider a moving vessel with its position and velocity components in the state vector. The Kalman filter processes each epoch at a time. When data comes into the system it will produce estimates for the states and their covariances for that particular instant. By combining the states at all epochs into one vector and incorporating the Kalman filter observation and dynamic models into a design matrix, then the solution for all epochs can be computed simultaneously via least squares. This approach has an advantage in that observations at different epochs are being combined and any temporal correlation between the observations can be accounted for within the a priori covariance matrix, $\mathbf{W}^{-1}$. Adversely the approach greatly expands the size of the parameters and is, of course, computationally inefficient.

The Kalman filter models can be combined into a least squares solution in the following manner. All vectors and matrices with a subscript (representing an epoch) refer to the Kalman filter models and those without subscripts are the least squares notation.

The observation equations

$$\mathbf{Ax} = \mathbf{b} + \mathbf{v}$$

$$\begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 & 0 & 0 \\ -\mathbf{M}_1 & \mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 & 0 & 0 \\ 0 & -\mathbf{M}_2 & \mathbf{I} & 0 & 0 & 0 \\ & & \ddots & & & \\ 0 & 0 & 0 & 0 & 0 & \mathbf{A}_m \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \\ \mathbf{x}_m \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ 0 \\ \mathbf{b}_2 \\ 0 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{y}_1 \\ \mathbf{v}_2 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{v}_m \end{bmatrix} \tag{7.7}$$

with covariance matrix

$$\mathbf{W}^{-1} \begin{bmatrix} \mathbf{Cl}_1 & 0 & \gamma(\delta t_{12}) & 0 & \cdots & \gamma(\delta t_{1m}) \\ 0 & \mathbf{Cy}_1 & 0 & 0 & \cdots & 0 \\ \gamma(\delta t_{21}) & 0 & \mathbf{Cl}_2 & 0 & \cdots & \gamma(\delta t_{2m}) \\ 0 & 0 & 0 & \mathbf{Cy}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(\delta t_{m1}) & 0 & \gamma(\delta t_{m2}) & 0 & \cdots & \mathbf{Cl}_m \end{bmatrix} \tag{7.8}$$

where $\gamma(\delta t_{ij})$ is the autocovariance function for a first order autoregressive / Gauss Markov process. This represents the temporal correlation within the observations since

$$\gamma(\delta t_{ij}) = E(\mathbf{v}_i \mathbf{v}_{i+\delta t_{ij}}) \tag{7.9}$$

and for the aforementioned case, $\gamma(\delta t_{ij})$ can be expressed as

$$\gamma(\delta t_{ij}) = \sigma_1^2 e^{-|\delta t_{ij}|/T} \tag{5.}$$

where

δt_{ij}	the time interval between epochs i and j
σ_l^2	the variance of the observations
T	the correlation time of the observations

7.2 A SIMPLE KALMAN FILTER MODEL

For this part of the analysis, a very simple model was used so as to keep computational demands to a minimum. A scenario of a vessel travelling in a straight line with the observed quantities being its easting and northing position at every epoch was set up. The state vector contained four elements : the vessel's position and velocity components in easting and northing. As the motion of the vessel was assumed to be in a straight line, the dynamic model was one of constant velocity. The two models can be shown in the usual Kalman filter matrix notation as follows.

The measurement model

$$\mathbf{A}_i\mathbf{x}_i = \mathbf{b}_i + \mathbf{v}_i$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}_i \begin{bmatrix} E \\ N \\ \dot{E} \\ \dot{N} \end{bmatrix}_i = \begin{bmatrix} E^\circ \\ N^\circ \end{bmatrix}_i + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_i \qquad (7.10)$$

where E° and N° are the observed quantities.

The dynamic model

$$\mathbf{x}_i = \mathbf{M}_{i-1}\mathbf{x}_{i-1} + \mathbf{y}_{i-1}$$

$$\begin{bmatrix} E \\ N \\ \dot{E} \\ \dot{N} \end{bmatrix}_i = \begin{bmatrix} 1 & 0 & \delta t & 0 \\ 0 & 1 & 0 & \delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{i-1} \begin{bmatrix} E \\ N \\ \dot{E} \\ \dot{N} \end{bmatrix}_{i-1} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}_{i-1} \tag{7.11}$$

where δt is the time interval between epochs i and i-1.

The covariance matrix for the observation model, $\mathbf{Cl}_i$, was taken to be diagonal with its elements representing the variances of the observations. The typical way to compute the covariance matrix for the dynamic model, $\mathbf{Cy}_{i-1}$ is by using Gauss's propagation of error law

$$\mathbf{Cy} = \mathbf{TCgT}^T \tag{6.12}$$

where **Cg** is the diagonal covariance of the vessel's driving noise. The noise is assumed to be white and therefore have a random distribution and in this case is the vessel's acceleration. The matrix **T** models the effect of the noise on the state vector and its elements will consist of the components of the Taylor's series. Using this approach leads to a singular $\mathbf{Cy}_{i-1}$, a problem that could not be overcome as the matrix has to be inverted to form the least squares a priori weight matrix. This matrix was also taken to be diagonal with its elements representing the certainty of the dynamic model.

7.2.1 Kalman Filter Straight Line Data

Data representing a vessel travelling in a straight line was created with the observables simply being the easting and northing position of the vessel at a particular epoch. A starting position, bearing and velocity is initially specified and the straight line positions are computed. The time interval between the epochs was set to one second and positions for 100 epochs in all were computed resulting in a data set comprising of 200 observations.

A set of normally (Gaussian) distributed errors with zero mean and unit variance was generated using the Box - Muller method (Press et al 1989). Errors taken from this set were added to each observation. The track of the vessel is shown in Fig 7.1.

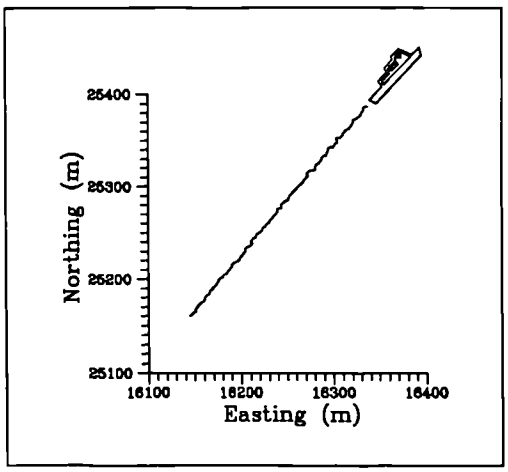


Fig 7.1 Simple Kalman filter data

7.2.2 Verification of the Least Squares Approach

A set of data comprising of straight line position components with random errors imposed on them (as described in 7.2.1) was passed through the Kalman filter models and the least squares representation. The whole data set was processed by the Kalman

filter approach firstly using the two forward operations (prediction and filtering) with the following being computed

$$\hat{\mathbf{x}}_i(+) \qquad C\hat{\mathbf{x}}_i(+)$$

$$\hat{\mathbf{x}}_{i+1}(-) \qquad C\hat{\mathbf{x}}_{i+1}(-)$$

portraying the filtered and predicted states respectively. Each state comprised of the position and the velocity components, ie $\left[E \quad N \quad \dot{E} \quad \dot{N}\right]_i^T$. The backward smoothing operation was then used to provide an optimal solution for all epochs (except for the final epoch) with the resulting output for later comparison

$$\hat{\mathbf{x}}_{i-1}(s) \qquad C\hat{\mathbf{x}}_{i-1}(s)$$

In all 100 epochs were used and thus 99 smoothed states were estimated. Unit weights were used for both Kalman filter models.

The data set was also processed using the least squares approach with the same models and weighting. No temporal correlation was applied within the a priori covariance matrix. With 100 epochs, the vector of parameters comprised of 400 elements which were estimated from 596 observations. This vector and the elements from the corresponding a posteriori covariance matrix was directly compared to the smoothed Kalman filter results.

The two sets of results were identical (< 0.00001 m) after the first twelve epochs. The initial discrepancies are due to the time it takes for the Kalman filter to settle down and are caused by the start state vector (and covariance matrix) and the ratio between the dynamic and observation weights. The start state vector is the state at the epoch before the Kalman filter starts operating, ie it is the state that is used to

computed the predicted state, $\hat{\mathbf{x}}_1(-)$. A start covariance matrix must also be specified. The ratio of the weighting between the two models affects how quickly a poor starting state and covariance can be pulled into their optimum solutions. If the weighting of the dynamic model is initially much greater than that of the observation model then the system will take a long time to converge. This can be seen through the gain matrix which controls the effect the observations have on the system when producing the filtered solutions. Indeed the same data set was processed with different weighting - the standard deviations for the observations was kept at 1 m, but that for the positions from the dynamic model was reduced to 5 cm. Using the same starting state and covariance, the two resulting solutions failed to compare to the same level of precision over all 99 epochs.

7.3 THE EFFECT OF CORRELATED MEASUREMENTS ON POSITION AND PRECISION

To discover the effect of temporal correlation on the state and its covariance matrix, a data set with correlated errors imposed onto it can be processed by the least squares representation of the Kalman filter. By ignoring the temporal correlation, ie setting $\gamma(\delta t_{ij})$ to be zero in all cases, the Kalman filter result can be obtained. By processing the data set again, but calculating the correct values for $\gamma(\delta t_{ij})$ will give an optimum solution for the data (since the correlation is now accounted for). Comparing the two results produces the effect of ignoring temporal correlation.

7.3.1 Generating Correlated Measurements

A data set depicting a vessel travelling in a straight line was again created and consisted of 100 epochs, each one second apart (see 7.2.1). Instead of adding random

errors with a normal distribution on to the observations, autoregressive errors with a known correlation time and variance were placed on to them. A set of autoregressive (Gauss / Markov) errors, $\{X_t\}$, can be generated from the set of normalised random errors, $\{Z_t\}$, in the following way :

From chapter four, the first order autoregressive process can be written as: (Chatfield, 1989)

$$X_t = \alpha X_{t-1} + Z_t \tag{4.15}$$

By back substituting this gives :

$$X_t = \alpha(\alpha X_{t-2} + Z_{t-1}) + Z_t \tag{7.12}$$

and by continuing this process, we finally get :

$$X_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \cdots + \alpha^t Z_0 \tag{7.13}$$

where

$\alpha = e^{-1/T}$

and T is the correlation time of $\{X_t\}$.

The variance of the process is given by:

$$\sigma_x^2 = \sigma_z^2 / 1-\alpha^2 \tag{7.14}$$

This is, in fact, a representation of the first order autoregressive process as a moving average process.

Five data sets were produced with the characteristics given in Table 7.1.

T(secs)	σ_x^2(m)
5	1
10	1
20	1
30	1
60	1

Table 7.1 Characteristics of the correlated data sets

7.3.2 Differences in Position

All five data sets were processed in the aforementioned manner. The a priori standard deviations were set to 1m for the observation model and 5cm for the dynamic model's position components. The velocity components of the dynamic model were set to one fifth of the position variance (in this case 0.5mm). No a priori covariances were used in either model.

The differences in positions for each data set were obtained by subtracting the position component obtained when the temporal correlation was ignored from the corresponding component when the correlation was accounted for. This was done for both eastings and northings and the resulting data sets were analysed for maximum difference (modula), mean difference and their standard deviation. The results are given in Table 7.2.

T (secs)	5	10	20	30	60
Max ΔE (m)	0.174	0.799	0.286	0.642	0.534
Max ΔN	0.454	0.536	0.502	0.275	0.344
Mean ΔE	-0.013	0.099	-0.002	-0.040	0.030
Mean ΔN	-0.061	0.019	0.046	0.018	0.032
Sd ΔE	0.100	0.293	0.153	0.235	0.134
Sd ΔN	0.218	0.284	0.179	0.137	0.083

Table 7.2 Summary of the differences in position

These position difference results show that position errors will result when using correlated data within the standard Kalman filter equations. The mean and the standard deviations of the sets are small indicating that the track of the vessel will appear to be similar to its optimum track with no apparent bias in the solution. The positional standard deviations given in Table 7.3 are, on the most part, greater than the maximum differences in eastings in northings (except for ΔE, T = 10 secs). This reiterates that there are no significant positional differences since it is expected that approximately only 68% of the positional errors will lie within the 1σ value. It can also be observed that the size of the differences does not appear to be dependent on the correlation time of the measurements.

7.3.3 Differences in the A Posteriori Covariance

The a posteriori standard deviations of the parameters at each epoch can be obtained by isolating the relevant part of the global a posteriori covariance matrix from the least squares computation. A mean standard deviation for eastings and northings was obtained for both when the correlation was accounted for and when it was ignored. Due to the initial weighting for the Kalman filter models, the standard deviations for

eastings is the same as that for northings. The results when accounting for the temporal correlation are shown in Table 7.3.

T (secs)	5	10	20	30	60
σ East (m)	0.581	0.710	0.830	0.886	0.950
σ North	0.581	0.710	0.830	0.886	0.950

Table 7.3 A posteriori positional standard deviations from correlated measurements

The standard deviations when correlation is ignored are shown in Table 7.4. Since these are obtained assuming that the errors on the data are normally distributed, the values are the same for all data sets.

σ East (m)	0.259
σ North	0.259

Table 7.4 A posteriori positional standard deviations ignoring correlation

The differences between the two tables are very significant. Even with a short correlation time, the precision with which a Kalman filter will estimate the position has been obtained is very over optimistic. As the correlation time increases, so this precision becomes worse. The scale of this can be seen graphically by drawing the error ellipses (in this case circles) for a selection of the above data.

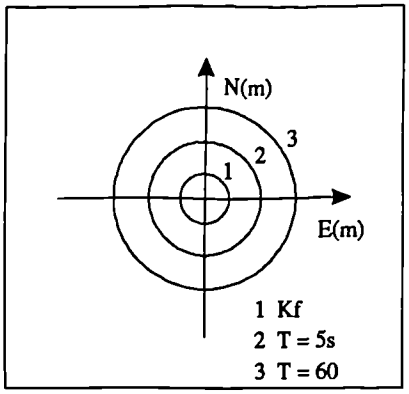


Fig 7.2 Mean positional error ellipses for correlated data

In Fig 7.2, the inner circle (labelled 1) represents the error ellipse produced from a standard Kalman filter - ie the temporal correlation has been ignored. The two other ellipses are for the data set with correlation times of 5 and 60 seconds. The error ellipse represents the precision for a two-dimensional position fix. There is a 39.4% probability that the least squares estimate of the vessel's position lies within the error ellipse centred at its true position (Cross, 1983).

7.3.4 Unit Variance Tests

The unit variance is a statistic that can be used to assess the a priori variances and covariances used within the adjustment. It is computed by (Cross, 1983)

$$\hat{\sigma}^2 = \mathbf{v}^T\mathbf{W}\mathbf{v}/n - m \qquad (7.15)$$

where $n - m$ are the degrees of freedom.

The expectation of the unit variance is unity and this is used as the null hypothesis within a central F-distribution test. Assuming no blunders within the observations, there are two possible reasons for the unit variance to fail the test. The first is that the a priori covariance matrix needs to be multiplied by $\hat{\sigma}^2$, and the second is that the

model used for the least squares computation is incorrect or incomplete (Cross, 1983). On failing the test, it is usual to scale all a posteriori covariance matrices by the factor $\hat{\sigma}^2$.

The unit variance was computed for all the data sets both when correlation was accounted for and when it was ignored. The results are shown in Table 7.5.

T (secs)	$\hat{\sigma}^2$ with $\gamma(\delta t_{ij})$ applied	$\hat{\sigma}^2$ with $\gamma(\delta t_{ij})$ ignored
0	0.985	N/A
5	0.967	0.551
10	1.035	0.459
20	0.997	0.207
30	1.066	0.215
60	0.832	0.057

Table 7.5 Unit variance values

The 0.95 F-distribution percentile with 196 and ∞ degrees of freedom is a value of 1.21. When correlation has been applied the unit variance passes the test, although this is only just the case for a correlation time of 60 seconds. When correlation has been ignored, the unit variance fails the test for all data sets. Any resultant scaling of the a posteriori covariance matrix will further enhance the differences between the standard deviations presented in Tables 7.3 and 7.4.

7.4 DGPS BLUNDER DETECTION WITHIN THE KALMAN FILTER

The reliability of the Kalman filter, the ability to detect blunders, for a typical DGPS scenario relies not only on the correlation characteristics of the data, but also on GPS

geometry. A Kalman filter to represent vessel motion at a DGPS mobile station was established with the observations being "corrected" GPS pseudoranges. By processing temporally correlated corrected pseudoranges with known blunders, the reliability of the system could be tested.

7.4.1 GPS Pseudorange Observation

The GPS pseudorange measurement can be modelled as

$$\rho = \rho_0 + c(dt_R - dt^S) + c\tau_d + \varepsilon + n \tag{7.16}$$

where

ρ	= pseudorange measurement
ρ_0	= true distance between satellite and receiver
dt_R	= receiver clock offset
dt^S	= satellite clock offset
τ_d	= atmospheric refraction time delay
ε	= other error sources (eg SA epsilon and dither)
n	= receiver noise
c	= speed of light

ρ_0 relates the receiver and satellite coordinates as a straight line vector. The satellite coordinates and the clock offset are obtained from the navigation message. The refraction delay and the other error sources are assumed to cancel for two stations in the same locality (the DGPS scenario) and is therefore ignored, as is the receiver noise. The resulting unknowns are the receiver clock offset and the receiver position $(\phi \quad \lambda \quad h)$ on the WGS84 ellipsoid which can thus be solved for with a minimum of four pseudorange measurements.

7.4.2 DGPS Kalman Filter

The principle behind DGPS is that the magnitudes of τ_d and ε will be similar for stations within close proximity of each other and differencing should almost nullify their effects. A least squares solution at the reference station provides an approximation of τ_d and ε within the dt_R bias parameter and corrections for each pseudorange observation can be computed. These corrections are then applied to the observations at the mobile station and the corrected observations are incorporated within the observation model of a standard Kalman filter estimation.

The dynamic model of the Kalman filter links the accelerations and velocities of the observation parameters via the Taylor's series expansion to represent a moving vessel. The full state vector can be shown as

$$\left[\phi \quad \lambda \quad h \quad dt_R \quad \dot{dt}_R \quad \dot{\phi} \quad \dot{\lambda} \quad \dot{h} \quad \ddot{\phi} \quad \ddot{\lambda}\right]^T \tag{7.17}$$

where $\dot{\phi}$ and $\ddot{\phi}$ indicates the velocity and acceleration of the parameter respectively.

7.4.3 Correlated Pseudorange Observations

Pseudoranges transmitted from different satellites to the receiver at a certain epoch can be generated from information contained in the GPS navigation message along with knowledge of the receiver clock offset. Correlated errors can then be placed on the ranges according to the AR(1) Gauss-Markov process. In all, seven data sets were created with different correlation characteristics at both the reference and mobile stations. Each data set consisted of 360 epochs of 1 second observations with five satellite coverage throughout, and with an average PDOP of 4.4. The errors of the

corrected pseudoranges at the mobile station had a 1σ value of approximately 3 m and correlation times of 0, 5, 10, 20, 30, 60, 120 seconds respectively. Two graphical examples of the errors in the corrected ranges as seen at the mobile station (for satellite 18) are shown in Figs 7.3 and 7.4.

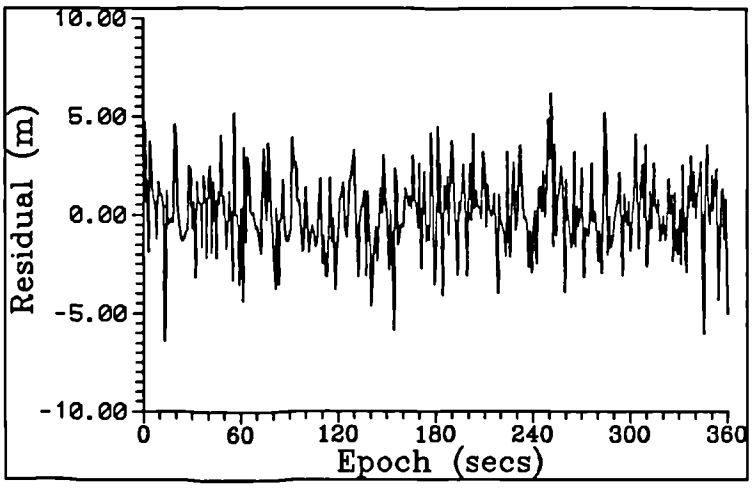


Fig 7.3 Errors in DGPS corrected pseudoranges (T = 0 secs)

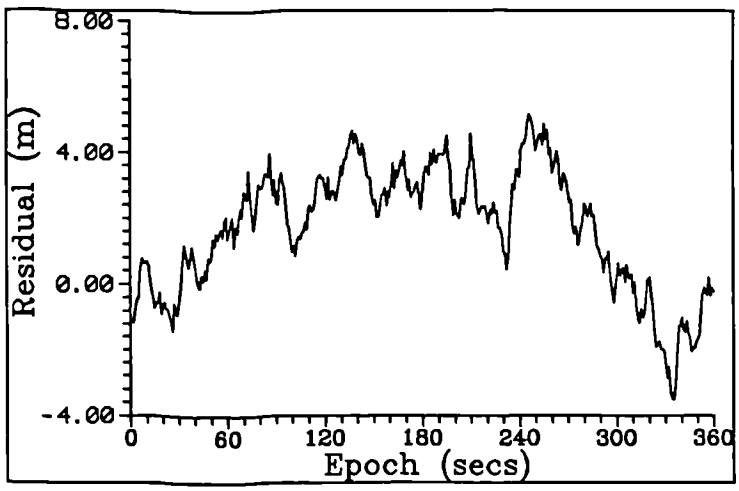


Fig 7.4 Errors in DGPS corrected pseudoranges (T = 60 secs)

7.4.4 Kalman Filter Blunder Detection

The methods used to detect outliers within pseudorange observations were the local overall model (LOM) test and the one-dimensional local slippage (LS) test described

in chapter six. The following details concerning these tests have been adapted from Teunissen and Salzmann (1988), Teunissen (1990), and Salzmann (1993).

The local tests are carried out on the predicted residuals, or the innovation sequence, which is defined (Teunissen, 1990)

$$\hat{\mathbf{v}}_i(-) = \mathbf{b}_i - \mathbf{A}_i\hat{\mathbf{x}}_i(-) \tag{6.33}$$

with the covariance matrix as

$$\mathbf{C}\hat{\mathbf{v}}_i(-) = \mathbf{C}\mathbf{l}_i + \mathbf{A}_i\mathbf{C}\hat{\mathbf{x}}_i(-)\mathbf{A}_i^T \tag{6.34}$$

where

$\hat{\mathbf{x}}_i(-)$, $\mathbf{C}\hat{\mathbf{x}}_i(-)$ are the predicted state vector and its covariance matrix respectively

$\mathbf{A}_i$, $\mathbf{b}_i$ are from equation (6.3).

If the dynamic model is valid, then the predicted residuals should be a zero mean white noise (stationary random) process with known covariance, and this is used to detect the overall validity of the model. The two local hypotheses to detect model misspecification at epoch t_i are

$$\begin{aligned} H_o^i \quad & \hat{\mathbf{v}}_i(-) \approx N(0, \mathbf{C}\hat{\mathbf{v}}_i(-)) \\ H_o^i \quad & \hat{\mathbf{v}}_i(-) \approx N(\mathbf{K}\nabla, \mathbf{C}\hat{\mathbf{v}}_i(-)) \end{aligned} \tag{6.35}$$

The test statistic used for detecting model errors in the null hypothesis was:

$$T^i = \hat{\mathbf{v}}_i(-)^T \mathbf{C}\hat{\mathbf{v}}_i(-)^{-1}\hat{\mathbf{v}}_i(-) \tag{6.36}$$

$$T^i_{LOM} = T^i / m_i \tag{6.37}$$

A model error was detected only if:

$$T^i_{LOM} \geq F_\alpha(m_i, \infty, 0) \tag{6.38}$$

These formulae (see section 6.4 for more comprehensive details) were used to detect a misspecification in the local model. If one has been detected, then the individual observations are searched for the most likely measurement error. This was carried out using the data snooping LS test statistic

$$t^i_{LS} = \frac{\mathbf{c}^T \mathbf{C}\hat{\mathbf{v}}_i(-)^{-1} \hat{\mathbf{v}}_i(-)}{(\mathbf{c}^T \mathbf{C}\hat{\mathbf{v}}_i(-)^{-1} \mathbf{c})^{1/2}} \tag{6.39}$$

where **c**, in this case, is a null vector except for unity at the corresponding observation being tested. The slippage test is carried out for all observations at epoch t_i, and the most likely blunder is said to have occurred at the observation for which $\left|t^i_{LS}\right|$ is at a maximum.

7.4.5 Detection of Blunders within Correlated Data

A set of correlated pseudorange data, with added blunders, was passed through the Kalman filter and the predicted residuals were tested using the UMPI LOM/LS methods (with a 0.95 F-distribution percentile). One blunder was placed randomly within the data set for each run of the filter. In all, 900 blunders were placed within the observations (100 each of size 4, 5, 6, 7, 8, 9, 10, 11, 12 metres) and percentages for correctly detected blunders against size was obtained. This process was repeated

for all the correlated pseudorange data sets (correlation times of 0, 5, 10, 20, 30, 60,120 seconds). The results are given in Table 7.6.

T(secs) Size(m)	0	5	10	20	30	60	120
4	1	1	0	0	0	0	0
5	1	3	1	0	0	0	0
6 (2σ)	7	4	8	0	0	0	0
7	8	10	8	0	0	1	2
8	26	23	13	13	8	5	6
9 (3σ)	32	28	25	24	18	20	24
10	45	47	27	35	35	47	27
11	59	59	51	54	55	60	50
12 (4σ)	78	73	75	77	77	78	82

Table 7.6 Percentage of detected blunders for differently correlated data

These results clearly show a trend for small blunders (<=3σ) when the ability to detect becomes less with increasing correlation time. Data with a long correlation time failed to identify these blunders, whereas approximately 10% were detected using data with shorter correlation times. For larger size of blunder (>3σ and <=4σ) the performance of the LOM and LS test statistics are similar independent of correlation time with all data sets showing around an 80% detection rate for 4σ outliers. It should also be noted that for data with short periods of correlation (T=0, 5 seconds) and for small blunder sizes (4, 5 m), the correct epoch containing the blunder was sometimes identified but the LS test statistic failed to identify the correct observation. These results were not included in the table.

7.5 CONCLUSIONS

1. The standard Kalman filter equations can be expressed in the least squares observation equation form. Identical results between this least squares approach and the smoothed estimates from a Kalman filter can be obtained.

2. Errors in position are obtained when processing temporally correlated data sets and ignoring the correlation. These errors can be determined by processing the data set twice as a least squares computation, once ignoring the correlation and the other accounting for the correlation with appropriate covariances. On average the discrepancies between the two solutions are not large, with a small mean and with no apparent bias. The maximum discrepancies for different correlation times are smaller than the computed positional standard deviations indicating that the position errors are not significant. The errors are not dependent on correlation time.

3. Large differences in the a posteriori covariance matrix are obtained between data for which temporal correlation has been accounted for and for when it has been ignored. The differences are dependent on the correlation time of the observations with the longer the correlation time, the greater the difference. This suggests that the Kalman filter produces an extremely over optimistic estimate of the precision of a position fix.

4. When ignoring correlation, the least squares solution invariably failed the unit variance test. Any resultant scaling of the a posteriori covariance matrix will further enhance the over optimistic estimates of the standard deviations of the position fix.

5. The ability to detect blunders within a DGPS Kalman filter is affected by the correlation time of the observations. For small blunders ($<3\sigma$), the system is less reliable the longer the correlation time of the pseudorange observations. For larger blunders, there does not appear to be a significant reduction in reliability for different correlation times.

CHAPTER EIGHT
CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

8.1 CONCLUSIONS

DGPS is a proven positioning technique and is widely used within the offshore oil industry. When processing DGPS observations using the Kalman filter algorithms, certain assumptions concerning temporal correlation are ignored. This thesis has identified that pseudorange measurements are temporally correlated and that this significantly affects the output from the filter. The conclusions can be divided into two sections: the first describing the correlation within recorded data sets, and the second examining its implications when using the Kalman filter as a processing technique.

8.1.1 Pseudorange Analysis

A first-order autoregressive, AR(1), process has been selected to model the errors within GPS pseudorange observations. This has allowed the variances and correlation times of different data sets recorded between March 1991 and March 1993 to be estimated and compared. The data has also been processed using the Fast Fourier Transform (FFT) technique to identify different cyclic patterns, due to various error sources, that may appear in the data. The GPS data sets have been selected to identify the effects of different error sources, including the atmosphere, different receiver types, different smoothing routines and, most importantly, the intentional selective availability (SA). Conclusions from this analysis are:

1. Routines adopted for the smoothing of pseudorange observables with carrier phase or doppler measurements change the statistics of the data sets. The only data set that was recorded with pseudoranges that were not smoothed internally within the receiver were those from March 1991. One of these was processed in its "raw" state and with external smoothing applied. The results showed an expected decrease in noise which was represented in a decrease in the variance for the smoothed data set. Similarly, smoothing the data resulted in an increase in the correlation time and, in particular, altered the short term correlation properties. A reduction was seen with the average standard deviations dropping from 7.1 to 5.9 m and correlation times from 136 to 266 s when smoothing had been applied. Some of the correlation times after smoothing were as long as 10 minutes, and even three times this length for a similar data set recorded simultaneously. The fact that the Fourier analyses are the same for all satellites both with and without smoothing, indicates that the errors present both occur on the carrier and code measurements.

2. Few conclusions could be made when comparing the output from different receiver types, mainly due to the difference in the technologies at the time the observations were taken, and the overwhelming effects of SA. The pre SA data did show that differences occur even when using the same data set. Results show that for one data set, the correlation times were more than three times that of the other set, although their variances were similar. This was due to cycle slips and loss of lock within one of the receivers, thus reducing the impact of the smoothing algorithm that was used. The other data sets showed a visual chronological decrease in noise. This is almost certainly due to receiver enhancements and internal processing that was carried out. The best example of this is seen when comparing the raw Ashtech March 1991 to the

Ashtech March 1993 data sets - exactly the same receivers were used, except that they had been "upgraded" several times during this period.

3. The effects of SA were characterised by examining data both before and after SA was officially implemented. The most obvious conclusion was from the increase in standard deviation from approximately 5 m to 50 m, and this therefore represents the largest error source within the pseudorange observations. SA is implemented through the dithering of the satellite clock frequency and the incorrect description of the orbit (epsilon). The Fourier analyses for all data sets under SA showed similar patterns with large cyclic patterns occurring every hour and smaller trends at higher frequencies. These are almost certainly due mainly to the dither component, since epsilon would only have a long term effect. Epsilon can change as frequently as new navigation messages are transmitted (once every hour), and therefore only the lowest frequencies detected within the analyses could be due to epsilon. However, the orbit description will be changed on the hour, and if this has occurred, a distinct step should be observed in the time series plots. Since this was never experienced within the data sets analysed, epsilon can be considered as to have a very small amplitude or be at a lower frequency than once every hour. The correlation times are shorter for the data recorded after SA has been implemented which, along with the high variance, is the reason why the age of correction is so important in DGPS operations. With short correlation times and high variances, the pseudorange corrections are only "valid" for a short period of time, and therefore must be applied at the mobile station as soon as possible after they have been generated at the reference.

8.1.2 Kalman Filter Performance

Chapter seven reported on the effects that correlated data had on the performance of the Kalman filter equations in terms of the resultant position, its precision and for blunder detection. The position and precision analysis was carried out by expressing the Kalman filter equations in a least squares observation equation form and accounting for the correlation within the a priori covariance matrix. Due to the inefficiency of this approach, a simplified model was used. Reliability analysis was carried out by processing correlated GPS observations using a Kalman filter set up for a DGPS scenario. Errors were placed within the observations and the actual performance of the blunder detection was assessed. The conclusions are summarised:

1. Using the least squares approach gave exactly the same results as with the Kalman filter, when a diagonal covariance matrix was used. Implementing a full matrix (thus accounting for temporal correlation) showed difference in both position and precision. On average, errors in position were not large, with a small mean, no apparent bias, and considered insignificant. The discrepancies were also not dependent on correlation time. The resultant differences in precision produced another story. Large errors in the a posteriori covariance matrix were obtained when ignoring the correlation and these could be further increased if unit variance tests are carried out and any a posteriori scaling occurs. These errors were dependent on the correlation time, with the longer the correlation time the greater the error. The Kalman filter has therefore been shown to produce an extremely over optimistic estimate of the precision of a position fix, and great care should be taken when assessing such measures. The standard deviation from the Kalman filter using observations with a correlation time of 60 s was almost a quarter of the value

when the correlation was accounted for. In other words, a quoted standard deviation of 25 cm derived from a Kalman filter is a far too optimistc precision, and a value of approximately 1 m is more realistic.

2. The standard Kalman filter blunder detection algorithms are also affected by temporally correlated measurements. This is most significant for small errors ($<3\sigma$), with the system becoming less reliable the longer the correlation time of the pseudorange observations. Data with a long correlation time (30 to 120 s) failed to detect these blunders, whereas approximately 10% were detected using data with shorter correlation times (0 to 20 s). For larger errors, there was no reduction in reliability performance for the differently correlated data sets, although this was still not optimal with approximately 80% of outliers being detected using a 95% percentile. It should be noted, however, that the test statistics were not used to detect any errors larger than 4σ. The corrected GPS pseudorange observations had a standard deviation of 3 m. It can therefore be concluded that the Kalman filter local model test statistics are sensitive to the correlation times of the observables if errors of 9 m or less are being sought. For detecting errors of 12 m, or more, the statistics are not dependent on correlation time.

8.2 SUGGESTIONS FOR FURTHER WORK

The following gives some ideas for further research into similar work.

1. A test bed processing of the modern receiver types to analyse the receiver performance. This would involve establishing many different receivers within the same regional location and taking simultaneous observations over various

sessions. Results from this would include receiver performance in terms of internal measurement precisions and/or improvements through smoothing routines, and also receiver compatibility which would conclude in whether or not different receivers should be used at DGPS reference and mobile stations.

2. Further time series models should be examined to test which best fits the pseudorange data for a particular session. For this thesis an AR(1) model was selected which proved not to be the best choice under some of the SA data and a higher order, or an integrated model, may have given a better representation. If this was carried out over a long period of time, the true characteristics and algorithms implementing SA could be deduced. In turn, this could lead to the real-time modelling of the SA error source. Examining data sets of different lengths, for instance a days observation, will allow for a more thorough examination of dither and epsilon, as will a different approach for modelling Block I and Block II satellites.

REFERENCES AND BIBLIOGRAPHY

ACKROYD N LORIMER R (1990) *Global Navigation: A GPS User's Guide*, Lloyd's of London Press, 202pp.

AL-NAKIB N (1992) *Selected INMARSAT Applications*, Presentation provided to the Netherlands Institute of Navigation, 9pp.

ANON *Global Positioning System: Making the World Better*, IBM Brochure.

ANON (1987) *Department of Defence World Geodetic System 1984*, DMA Technical Report, DMA TR 8350.2.

ANON (1989) *Introduction to the GPS System and to SERCEL GPS Receivers*, SERCEL, Issue November 1989.

ANON (1990) *RTCM Recommended Standards for Differential Navstar GPS Service Version 2.0.* RTCM Special Committee No 104, PO Box 19087, Washington 20036, 80pp.

ANON (1991) *ICD-GPS-200: Interface Control Document*, Arinc Research Corporation, 11770 Warner Ave, Suite 210, Fountain Valley, CA 92708, 115pp.

ASHJAEE J LORENZ R (1992) *Precision GPS Surveying after Y-code*, Proceedings of ION GPS-92, Albuquerque, New Mexico, 3pp

ASHKENAZI V MOORE T (1986) *The Navigation of Navigation Satellites*, Journal of Navigation, Vol 39(3), 16pp.

BAKER PJ (1986) *GPS Policy*, Proceedings of the Fourth International Symposium on Satellite Positioning, University of Texas, 13pp.

BAKER PJ (1987) *GPS in the Year 2000 and Beyond*, Journal of Navigation, Vol 40(2), 7pp.

BARBOUX JP (1993) Data Links for Real Time Differential GPS: New Technical Solutions, Proceedings of DSNS 93, Amsterdam, 9pp.

BESER J PARKINSON BW (1984) *The Application of NAVSTAR Differential GPS in the Civilian Community*, Global Positioning System, Institute of Navigation, Vol 2, 29pp.

BLANCHARD WF (1989) *Data Links for Differential GPS*, Proceedings of Nav 89: Satellite Navigation, Royal Institute of Navigation, 11pp.

BLOOMFIELD P (1976) *Fourier Analysis of Time Series: An Introduction*, Wiley, New York.

BURGESS A (1989) *GPS Program Status*, Proceedings of Nav 89: Satellite Navigation, Royal Institute of Navigation, 4pp.

CHATFIELD C (1989) *The Analysis of Time Series: An Introduction*, Chapman and Hall, Fourth Edition, 241pp.

CHEN W (1992) *Integration of GPS and INS for Precise Surveying Applications*, PhD Thesis, Department of Surveying, University of Newcastle upon Tyne,219pp.

CHIN GY (1987) *Two-dimensional Measures of Accuracy in Navigational Systems*, US Department of Transport, DOT-TSC-RSPA-87-1.

COHEN CE PERVAN B PARKINSON BW (1992) *Estimation of Absolute Ionospheric Delay Exclusively through Single-Frequency GPS Measurements*, Proceedings of ION GPS-92, Albuquerque, New Mexico, 6pp.

COOK GE (1983) *NAVSTAR GPS for Sea and Air Navigation: Principles and Present Status*, Journal of Navigation, Vol 36(3), 9pp.

COOPER MAR CROSS PA (1991) *Statistical Concepts and their Application in Photogrammetry and Surveying*, Photogrammetric Record, Vol 13(77), 33pp.

CROSS PA HOLLWEY JR SMALL LG (1981) *Geodetic Appreciation*, Working paper No 2, Department of Surveying, University of East London, 184pp.

CROSS PA (1983) *Advanced Least Squares Applied to Position Fixing*, Working Paper No 6, Department of Surveying, University of East London, 205pp.

CROSS PA PRITCHETT CH (1986) *A Kalman Filter for Real-time Positioning during Geophysical Surveys at Sea*, Proceedings of FIG 86, Toronto, Canada, 19pp.

CROSS PA (1987a) *Kalman Filter / Smoother Equations: Their Derivation and Implementation*, Paper presented at the Royal Institute of Chartered Surveyors Hydrographic Society seminar, University of Nottingham, 16pp.

CROSS PA (1987b) *Kalman Filtering and its Application to Offshore Position-fixing*, The Hydrographic Journal, No 44, 6pp.

CROSS PA (1989) *Position: Just what does it mean?*, Proceedings of Nav 89, The Royal Institute of Navigation, London, 10pp.

CROSS PA ROBERTS WDS (1990) *Differential Offshore Positioning using Block II GPS Satellites*, Proceedings of Hydro 90, The Hydrographic Society, University of Southampton, 10pp.

de JONG CD (1991) *GPS - Satellite Orbits and Atmospheric Effects*, Reports of the Faculty of Geodesy, Number 91.1, Delft University of Technology, The Netherlands, 112pp.

DREWETT AA (1989) *GPS: Aspects of Pseudorange Positioning Offshore*, PhD Thesis, Department of Surveying, University of Newcastle upon Tyne.

GAO Y KRAKIWSKY EJ LIU ZW (1992) *A new Algorithm for Filtering a Correlated Measurement Sequence*, Manuscripta Geodetica, Vol 17(2), 8pp.

GELB A (1974) *Applied Optimal Estimation*, MIT Press, Cambridge MA, 369pp.

GOAD CC (1990) *Optimal Filtering of Pseudoranges and Phases from Single-Frequency GPS Receivers*, Navigation, Journal of the Institute of Navigation, Vol 37(3), 12pp.

GRAVISS LP (1992) *GPS Development Program Status*, Proceedings of ION GPS-92, Albuquerque, New Mexico, 13pp.

GU X TIEMEYER B LIPP A (1993) *High Precision Navigation with Wide Area DGPS*, Proceedings of DSNS 93, Amsterdam, 10pp.

GURTNER W MADER G (1990) *Receiver Independent Exchange Format Version 2*, CSTG GPS Bulletin, Vol 3(3), NGS, 8pp.

GURTNER W (1993) *RINEX: The Receiver Independent Exchange Format Version 2*, Revision April 1993, Astronomical Institute, University of Berne, 12pp.

HALLMANN U (1993) *DGPS Radiobeacon Systems: Status, Progress, Technology and Performance*, Proceedings of DSNS 93, Amsterdam, 8pp.

HEISKANEN WA MORITZ H (1967) *Physical Geodesy*, Reprint, Institute of Physical Geodesy, Technical University, Graz, Austria, 364pp.

HOFMANN-WELLENHOF B LICHTENEGGER H COLLINS J (1992) *GPS: Theory and Practice*, Springer-Verlag Wien, 326pp.

JAZWINSKI AH (1970) *Stochastic Processes and Filtering Theory*, New York Academic Press, 376pp.

JENSEN MHB NICOLAI R (1990) *Some Practical examples of the use of Standardised Statistical testing of Positioning for Seismic and Engineering Surveys*, Proceedings of FIG 90, Helsinki, Finland, 12pp.

JENSEN MHB (1992) *Quality Control for Differential GPS in Offshore Oil and Gas Exploration*, GPS World, Vol 3(8), 7pp.

JONES SSD (1979) *GPS NAVSTAR: A Review*, Journal of Navigation, Vol32(3), 10pp.

KALAFUS RM CHIN CY (1986) *Measures of Accuracy in the NAVSTAR GPS: 2drms vs CEP*, Proceedings of the National Technical Meeting, Institute of Navigation, Longbeach CA, 6pp.

KALMAN RE (1960) *A new Approach to Linear Filtering and Prediction Problems*, ASME, Journal of Basic Engineering, Vol 82(D), 10pp.

KOOPMANS LH (1974) *The Spectral Analysis of Time Series*, New York Academic Press.

KREMER TG KALAFUS RM LOOMIS PVW REYNOLDS JC (1989) *The Effect of Selective Availability on Differential GPS Corrections*, Navigation, Vol 37(1), 13pp.

LAMBECK K (1988) *Geophysical Geodesy: The Slow Deformations of the Earth*, Clarendon Press, Oxford, 718pp.

LANGLEY RB (1991) *The GPS Receiver - An Introduction*, GPS World, Vol 2(1), 4pp.

LANGLEY RB (1990) *Why is the GPS Signal So Complex?*, GPS World, Vol 1(3), 4pp.

LASITER EM PARKINSON BW (1977) *The Opertational Status of NAVSTAR GPS*, Journal of Navigation, Vol 30(1), 12pp.

LEICK A (1990) *GPS Satellite Surveying*, John Wiley & Sons, 352pp.

LIU ZW KRAKIWSKY EJ GAO Y (1992) *An Analysis of Three Methods for Filtering a Correlated Measurement Sequence*, Manuscripta Geodetica, Vol 17(2), 9pp.

McNEFF J (1992) *NAVSTAR Global Positioning System (GPS) Signal Policy*, Proceedings of ION GPS-92, Albuquerque, New Mexico, 10pp.

MERMINOD B (1989) *The Use of Kalman Filters in GPS Navigation*, Unisurv S35, Reports from the School of Survey, The University of New South Wales, Kensington, NSW Australia, 203pp.

MONTGOMERY H (1993) *Nonprecision Approaches, GPS in IVLS, and New Federal Policy*, GPS World, Vol 4(3), 4pp.

MORITZ H (1980) *Advanced Physical Geodesy*, Abacus Press, 500pp.

NICOLAI R (1988) *Statistics in Positioning Quality Control*, Proceedings of Hydro 88, The Hydrographic Society, London, 10pp.

PAYNE Jr CR (1982) *Navstar Global Positioning System: 1982*, Proceedings of the Third International Symposium on Satellite Doppler Positioning, New Mexico State University, 28pp.

POPE AJ (1976) *The Statistics of Residuals and the Detection of Outliers*, NOAA Technical Report NOS 65 NGS 1, 131pp.

PRATT AR (1992) *The Electronics and Hardware of GPS Receivers*, Peek Traffic Ltd, UK, January 1992, 27pp.

PRESS WH FLANNERY BP TEUKOLSKY SA VETTERLING WT (1989) *Numerical Recipes in Pascal: The Art of Scientific Computing*, Cambridge University Press, 759pp.

RAUCH HE TUNG F STRIEBEL CT (1965) *Maximum Likelihood Estimates of Linear Dynamic Systems*, AAIA Journal, Vol 3(8), 6pp.

ROBERTS WDS CROSS PA (1993) *The Effect of Temporal Correlation within the Kalman Filter applied to Offshore Positioning*, The Hydrographic Journal, No 67, 7pp.

RODGERS C (1992) *Multipath Simulation Software Developed for the Design of a Low Multipath DGPS Antenna for the US Coast Guard*, Proceedings of ION GPS-92, Albuquerque, New Mexico, 8pp.

SALZMANN M (1993) *Least Squares Filtering and Testing for Geodetic Navigation Applications*, PhD Thesis, Delft University of Technology, The Netherlands, 209pp.

SCULL D (1984) *The US Federal Radionavigation Plan*, Journal of Navigation, Vol 37(3), 10pp.

SCULL D (1989) *The 1988 Federal Radionavigation Plan and the Civil GPS Service*, Proceedings of Nav 89: Satellite Navigation, Royal Institute of Navigation, 7pp.

SHARIF AB (1989) *Analysis of Broadcast, Precise and Integrated Orbits for GPS Satellites*, PhD Thesis, Department of Surveying, University of Newcastle upon Tyne, 321pp.

SHIRER HO (1991) *GPS and the US Federal Radionavigation Plan*, GPS World, Vol 2(2), 5pp.

SLACK ER (1990) *Differential GPS Service through the INMARSAT Global System*, COMSAT Mobile Communications, 950 L'Enfant Plaza SW, Washington, 20024, 10pp.

SPILKER Jr JJ (1980) *GPS Signal Structure and Performance Characteristics*, Global Positioning System, The Institute of Navigation, Vol 1, 25pp.

STANSELL Jr TA (1980) *Civil Marine Applications of GPS*, Global Positioning System, Institute of Navigation, Vol 1, 11pp.

STEIN WL (1986) *NAVSTAR GPS: 1986 Status and Plans*, Proceedings of the Fourth International Geodetic Symposium on Satellite Positioning, University of Texas, 12pp.

TALBOT NC (1992) *Efficient Surveying under GPS Encryption*, Trimble Navigation, Sunnyvale, CA.

TEUNISSEN PJG (1990) *An Integrity and Quality Control Procedure for use in Multi Sensor Integration*, Proceedings of ION GPS-90, Colorado Springs, 9pp.

TEUNISSEN PJG SALZMANN MA (1988) *Performance Analysis of Kalman Filters*, Reports of the Faculty of Geodesy, Number 88.2, Delft University of Technology, The Netherlands, 18pp.

VAN DIERENDONCK AJ RUSSELL SS KOPITZKE ER BIRNBAUM M (1980) *The GPS Navigation Message*, Global Positioning System, The Institute of Navigation, Vol 1, 18pp.

WELLS D BECK N DELIKARAOGLOU D KLEUSBERG A KRAKIWSKY EJ LACHAPELLE G LANGLEY RB NAKIBOGLU M SCHWARZ K TRANQUILLA JM VANICEK P (1986) *Guide to GPS Positioning*, Canadian GPS Associates.

APPENDIX A

EXAMPLES OF RINEX DATA FILES

The RINEX navigation message

```
     2                NAVIGATION DATA                         RINEX VERSION / TYPE
ASHTORIN                                 13 - AUG - 92 13:51 PGM / RUN BY / DATE
                                                             COMMENT
                                                             END OF HEADER
 3 92  8 10 13  0  0.0 -.519670546055D-03 -.550244294573D-10  .000000000000D+00
     .510000000000D+02 -.913437500000D+02  .120326440651D-08  .521478630844D+00
    -.515952706337D-05  .129857703578D-01  .566989183426D-05  .515367980957D+04
     .133200000000D+06  .109896063805D-06  .185361661610D+01 -.147148966789D-06
     .112231582471D+01  .349375000000D+03  .249062771154D+01 -.650741391705D-08
    -.472876840076D-09  .000000000000D+00  .657000000000D+03  .000000000000D+00
     .200000000000D+01  .000000000000D+00 -.419095158577D-08  .307000000000D+03
     .127920000000D+06  .000000000000D+00  .000000000000D+00  .000000000000D+00
17 92  8 10 12  0  0.0 -.521168112755D-05 -.909494701773D-12  .000000000000D+00
     .640000000000D+02 -.839687500000D+02  .406945522343D-08  .965234372743D-01
    -.439025461674D-05  .688534637447D-02  .993534922600D-05  .515367559814D+04
     .129600000000D+06 -.223517417908D-06  .264417122194D+01  .316649675369D-07
     .963508774556D+00  .193406250000D+03  .145833865121D+01 -.778032408173D-08
     .217866217860D-10  .000000000000D+00  .657000000000D+03  .000000000000D+00
     .700000000000D+01  .000000000000D+00  .139698386192D-08  .320000000000D+03
     .127890000000D+06  .000000000000D+00  .000000000000D+00  .000000000000D+00
28 92  8 10 12  0  0.0  .167074613273D-04  .227373675443D-11  .000000000000D+00
     .213000000000D+03 -.523437500000D+02  .467840916023D-08 -.150648617154D+01
    -.262074172497D-05  .763588387053D-02  .214017927647D-05  .515364113617D+04
     .129600000000D+06 -.223517417908D-07  .152296184899D+01 -.143423676491D-06
     .963243783039D+00  .336562500000D+03  .272892602163D+01 -.842999400023D-08
    -.144648882349D-09  .000000000000D+00  .657000000000D+03  .000000000000D+00
     .700000000000D+01  .000000000000D+00  .000000000000D+00  .213000000000D+03
     .127920000000D+06  .000000000000D+00  .000000000000D+00  .000000000000D+00
23 92  8 10 12  0  0.0  .117765739560D-05  .113686837722D-12  .000000000000D+00
     .400000000000D+02  .169062500000D+02  .447661504041D-08  .299548705758D+01
     .977888703346D-06  .611350964755D-02  .834092497826D-05  .515365083694D+04
     .129600000000D+06  .558793544769D-08 -.263525996326D+01 -.819563865662D-07
     .958363637544D+00  .218062500000D+03 -.249848060020D+01 -.779496754883D-08
     .621454457501D-09  .000000000000D+00  .657000000000D+03  .000000000000D+00
     .700000000000D+01  .000000000000D+00  .139698386192D-08  .296000000000D+03
     .128100000000D+06  .000000000000D+00  .000000000000D+00  .000000000000D+00
```

The RINEX observation file

```
     2              OBSERVATION DATA                         RINEX VERSION / TYPE
ASHTORIN                                 13 - AUG - 92 13:51 PGM / RUN BY / DATE
                                                             COMMENT
PP92                                                         MARKER NAME
                                                             MARKER NUMBER
WR_                                                          OBSERVER / AGENCY
773                 LM-XII2             6GP4                 REC # / TYPE / VERS
959                                                          ANT # / TYPE
  3667215.1700  -103470.1500  5200004.1000                  APPROX POSITION XYZ
        0.0001        0.0000        0.0000                   ANTENNA: DELTA H/E/N
     1     1                                                 WAVELENGTH FACT L1/2
     6    L1    L2    C1    P2    D1    D2                   # / TYPES OF OBSERV
     1                                                       INTERVAL
  1992     8    10    11    32   31.000000                  TIME OF FIRST OBS
  1992     8    10    12    33   46.000000                  TIME OF LAST OBS
                                                             END OF HEADER
 92  8 10 11 32 31.0000000  0  3  3 28 17                              0.000127710
    119783.81815          0.000 1  23064067.448           0.000        3304.876
         0.000
     46402.47617   5127430.16316  21520769.571    21520780.664       -2389.033
     -1861.639
     20653.10117  10951670.84116  20870124.603    20870137.464        1256.466
       979.046
 92  8 10 11 32 32.0000000  0  3  3 28 17                              0.000127730
    123089.045 5          0.000 1  23064692.965           0.000        3305.469
         0.000
     44013.812 6   5125568.863 5  21520312.148    21520326.808       -2388.329
     -1861.044
     21910.104 7  10952650.318 6  20870363.808    20870376.759        1257.514
       979.902
 92  8 10 11 32 33.0000000  0  3  3 28 17                              0.000127760
    126394.952 5          0.000 1  23065326.547           0.000        3306.065
         0.000
     41625.931 7   5123708.174 6  21519856.613    21519872.083       -2387.533
     -1860.422
     23168.178 7  10953630.638 6  20870602.562    20870615.933        1258.566
       980.766
 92  8 10 11 32 34.0000000  0  3  3 28 17                              0.000127780
    129701.454 5          0.000 1  23065950.205           0.000        3306.460
         0.000
     39238.753 6   5121848.036 6  21519402.188    21519417.867       -2386.998
     -1860.024
     24427.269 7  10954611.745 6  20870842.816    20870856.007        1259.403
       981.408
 92  8 10 11 32 35.0000000  0  3  3 28 17                              0.000127810
    133008.322 5          0.000 1  23066579.499           0.000        3306.951
         0.000
     36852.069 6   5119988.283 6  21518950.431    21518963.681       -2386.400
     -1859.477
     25687.167 7  10955593.485 6  20871081.601    20871095.961        1260.353
       982.061
```

APPENDIX B

EXAMPLES OF RINEXPOS OUTPUT

The Positions File

DIFFERENTIAL GPS POSITIONS
Reference Station 1 Newcastle
Latitude N 54 58 39.5827
Longitude W 1 36 58.5502
Height 64.5355

Reference file name 1 DATA\PP922231.92O
Remote file name DATA\SSSS2231.92O
Date 10 8 1992

Time	Latitude	Longitude	Height	PDOP
DGPS Age Of Correction	0 secs			
11 35 31.000	N 55 0 56.7008	W 1 45 19.4727	80.052	3.922
11 35 32.000	N 55 0 56.6218	W 1 45 19.4381	89.651	3.922
11 35 33.000	N 55 0 56.7468	W 1 45 19.4586	84.250	3.922
11 35 34.000	N 55 0 56.5290	W 1 45 19.1145	99.088	3.923
11 35 35.000	N 55 0 56.5696	W 1 45 19.3356	76.869	3.923
11 35 36.000	N 55 0 56.6295	W 1 45 19.1908	106.654	3.923
11 35 37.000	N 55 0 56.2453	W 1 45 19.1258	106.152	3.924
11 35 38.000	N 55 0 56.4704	W 1 45 19.1865	115.416	3.924
11 35 39.000	N 55 0 56.2877	W 1 45 19.3789	101.190	3.925
11 35 40.000	N 55 0 56.7107	W 1 45 19.4212	105.629	3.925
11 35 41.000	N 55 0 56.9255	W 1 45 19.4402	98.481	3.925
11 35 42.000	N 55 0 56.9990	W 1 45 19.4048	89.926	3.926
11 35 43.000	N 55 0 57.1010	W 1 45 19.2117	91.340	3.926
11 35 44.000	N 55 0 56.6074	W 1 45 19.2964	102.910	3.927
11 35 45.000	N 55 0 56.5620	W 1 45 19.0571	104.503	3.927
11 35 46.000	N 55 0 56.7877	W 1 45 19.2107	108.649	3.927
11 35 47.000	N 55 0 57.0011	W 1 45 19.3621	91.977	3.928
11 35 48.000	N 55 0 56.7675	W 1 45 19.4260	84.289	3.928
11 35 49.000	N 55 0 56.9427	W 1 45 19.6720	91.319	3.928
11 35 50.000	N 55 0 56.3547	W 1 45 19.3774	92.634	3.929
11 35 51.000	N 55 0 56.4562	W 1 45 19.1936	95.994	3.929
11 35 52.000	N 55 0 56.6609	W 1 45 19.2593	86.673	3.930
11 35 53.000	N 55 0 56.4846	W 1 45 19.1285	94.346	3.930

The Pseudorange Corrections File

REFERENCE STATION PSEUDORANGE CORRECTIONS
Reference Station Newcastle
Latitude N 54 58 40.2100
Longitude W 1 36 56.7500
Height 134.0000

Reference file name DATA\PP922231.92O

128131.0000	23	-30.711	1
128131.0000	3	-57.200	1
128131.0000	28	-24.624	1
128131.0000	17	-66.061	1
128132.0000	23	-45.616	1
128132.0000	3	-62.281	1
128132.0000	28	-36.017	1
128132.0000	17	-80.333	1
128133.0000	23	-31.349	1
128133.0000	3	-57.134	1
128133.0000	28	-25.820	1
128133.0000	17	-66.815	1
128134.0000	23	-48.505	1
128134.0000	3	-67.999	1
128134.0000	28	-36.735	1
128134.0000	17	-85.981	1
128135.0000	23	-35.809	1
128135.0000	3	-53.807	1
128135.0000	28	-19.446	1
128135.0000	17	-62.315	1
128136.0000	23	-50.744	1
128136.0000	3	-69.855	1
128136.0000	28	-42.920	1
128136.0000	17	-91.259	1
128137.0000	23	-52.618	1
128137.0000	3	-67.382	1
128137.0000	28	-41.971	1
128137.0000	17	-89.424	1

APPENDIX C

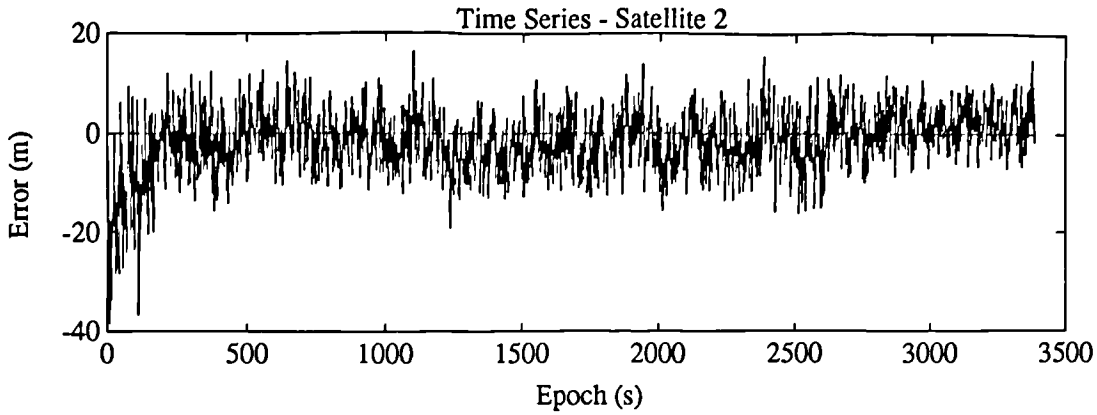
PRE SA OUTPUT

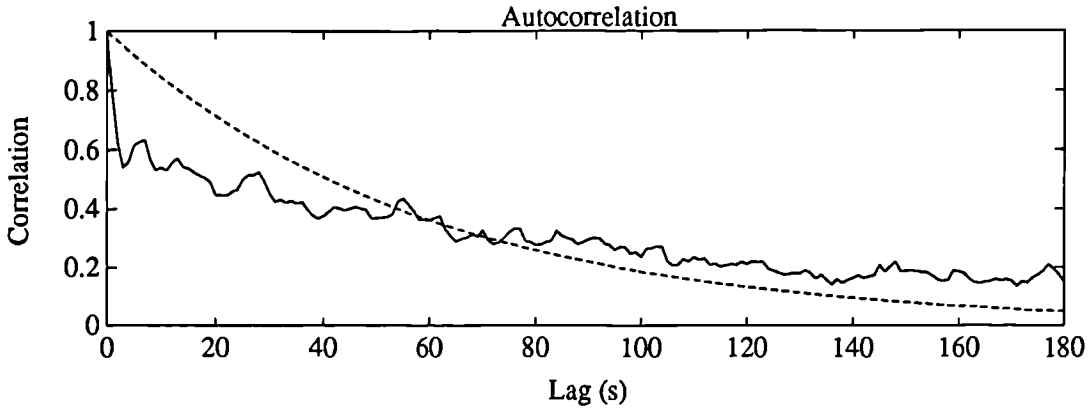
Time series, autocorrelation and Fourier analysis for data files:

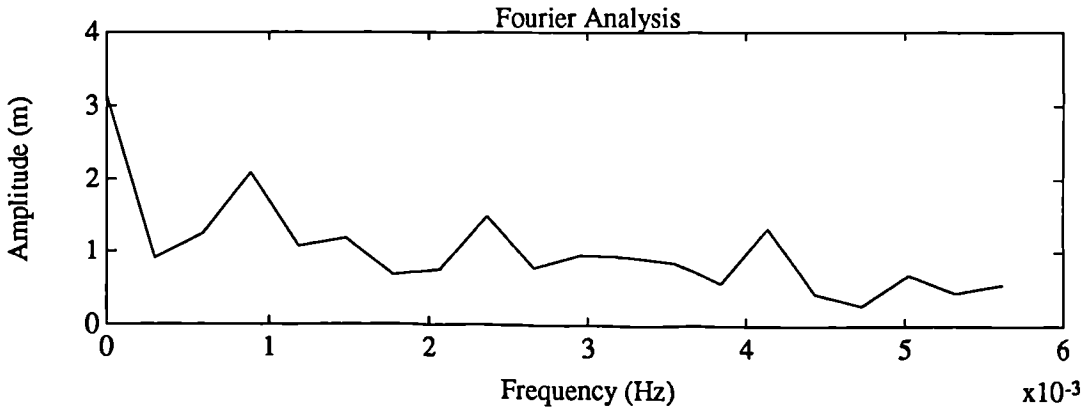
STN10731 - raw

STN10731 - smoothed

STN30731 - smoothed

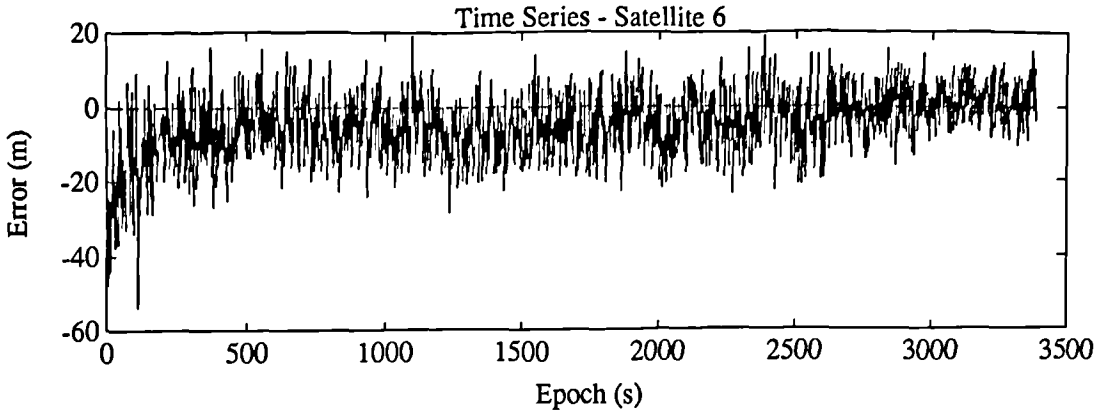


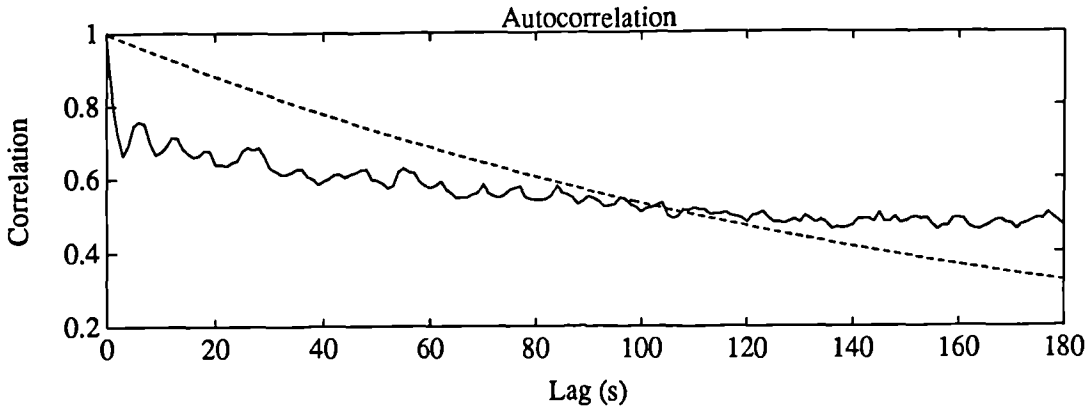


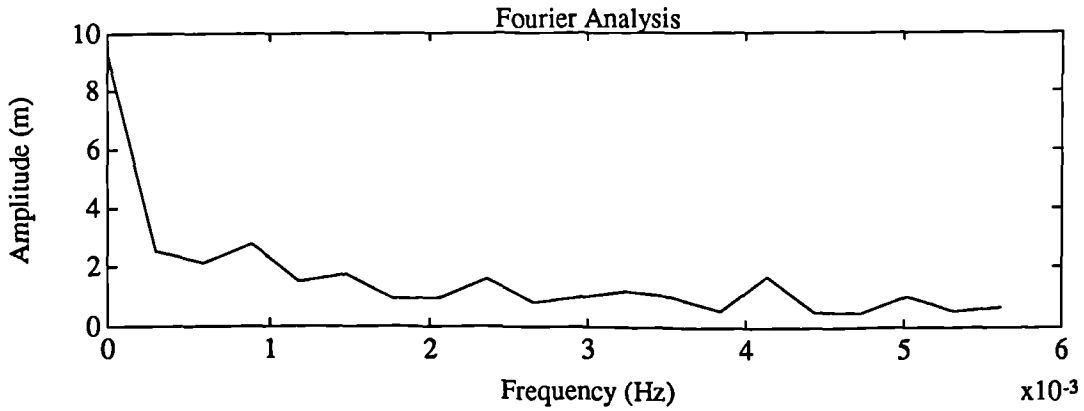


Variance = 35.91 m2

Correlation Time = 59.14 s

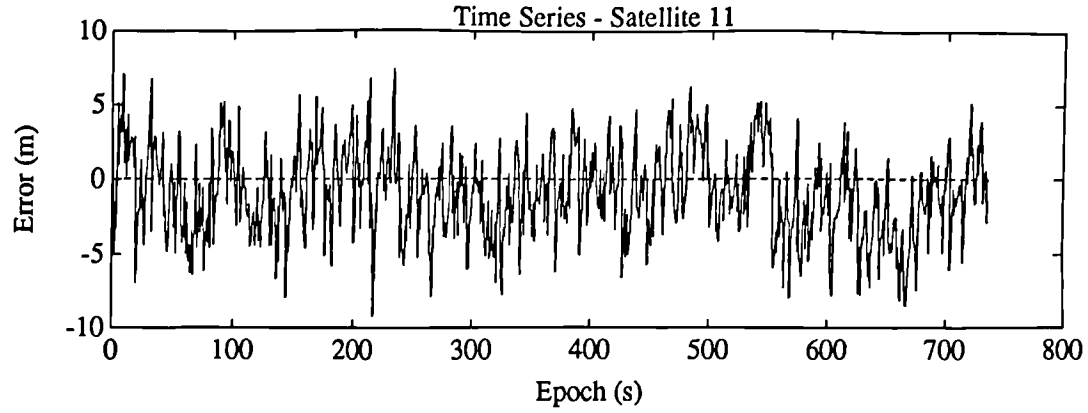


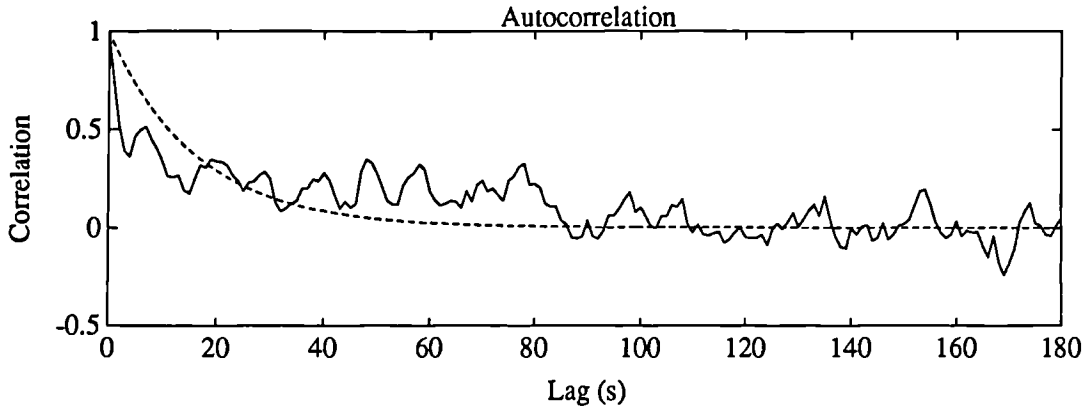


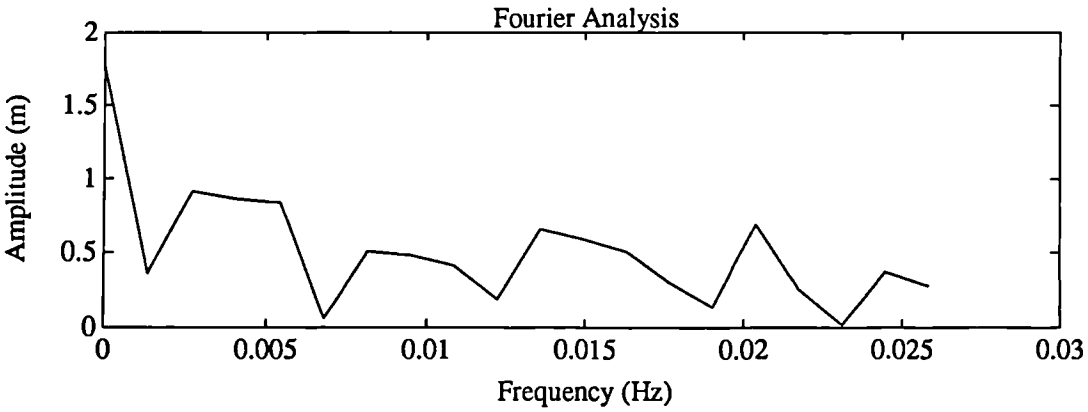


Variance = 85.38 m2

Correlation Time = 158.8 s

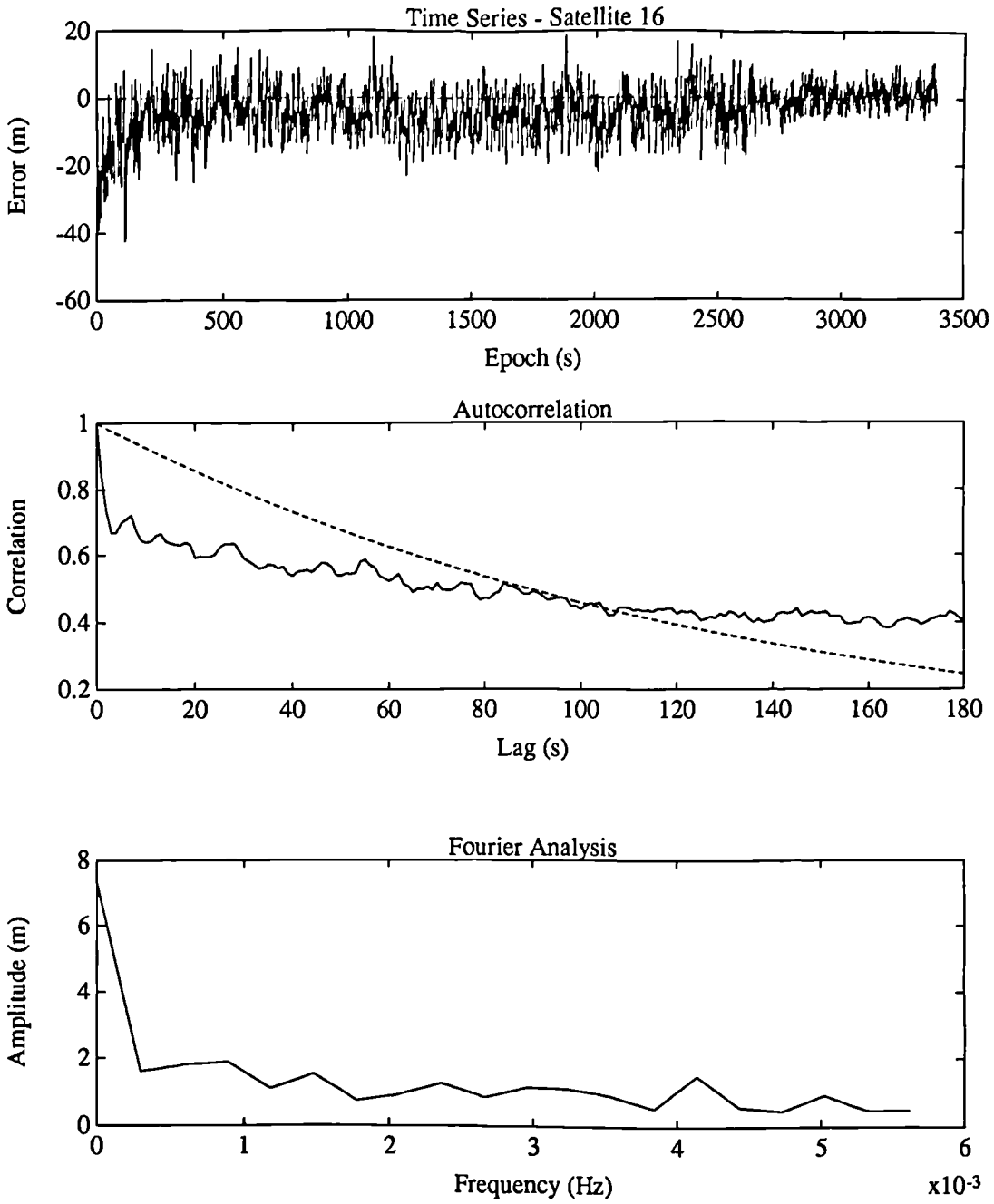






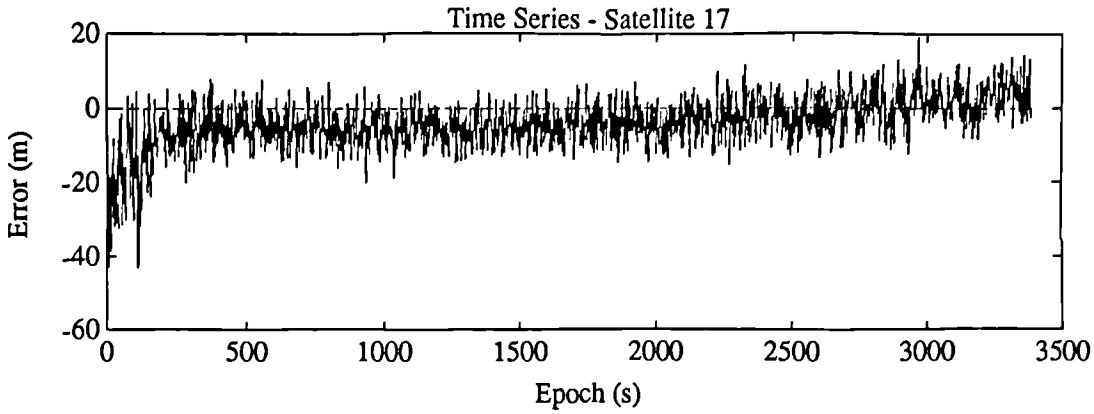
Variance = 9.979 m2

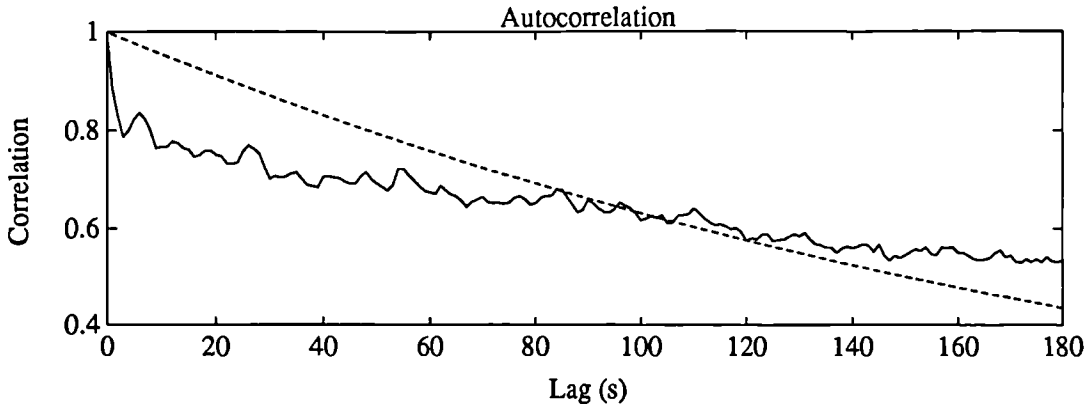
Correlation Time = 16.09 s

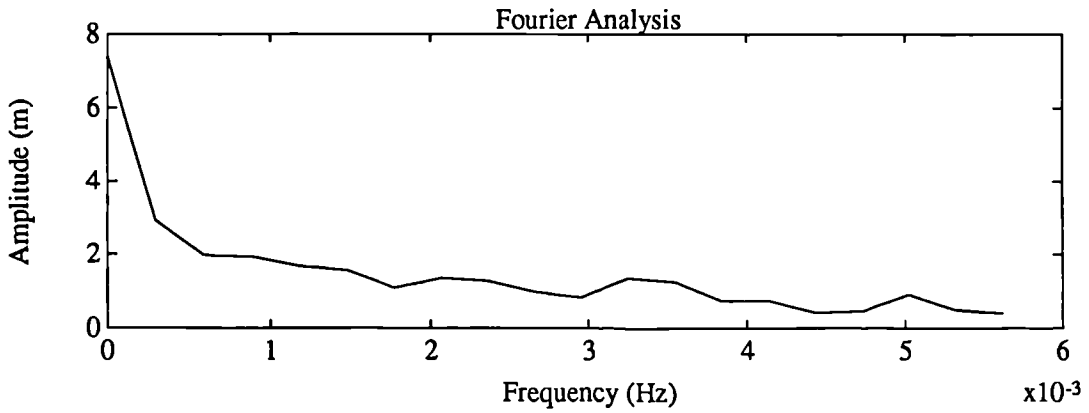


Variance = 56.55 m2

Correlation Time = 128.2 s

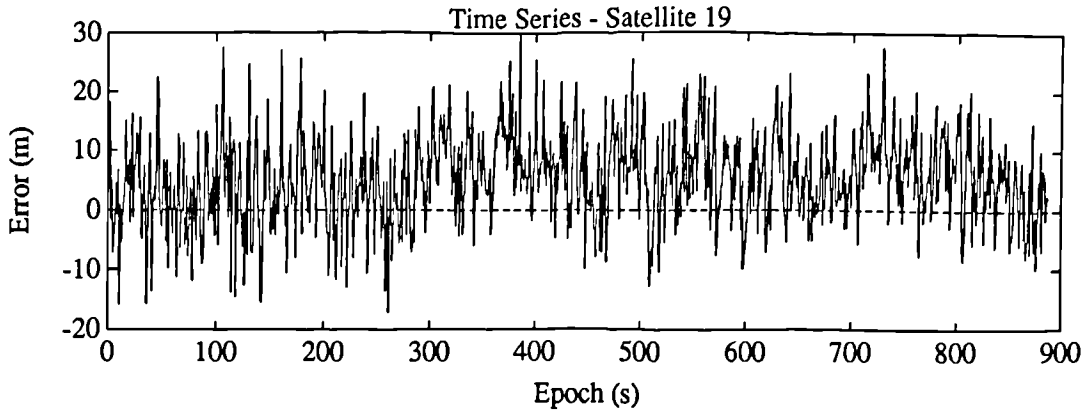


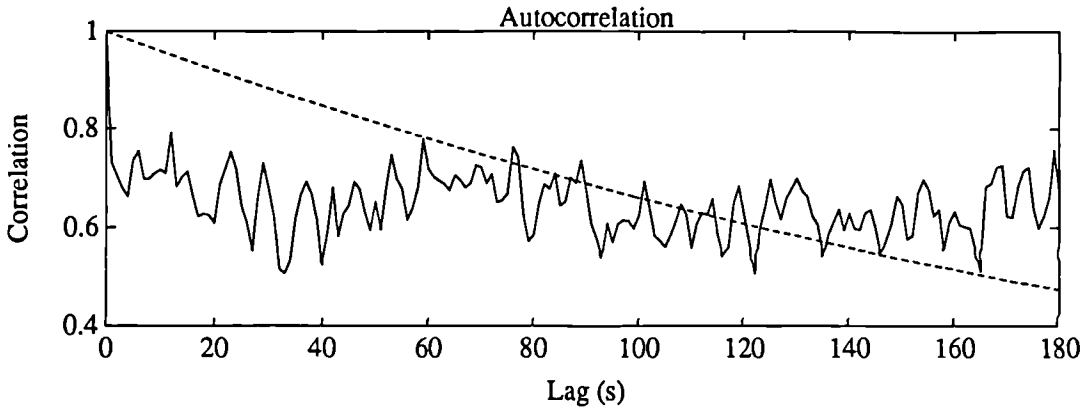


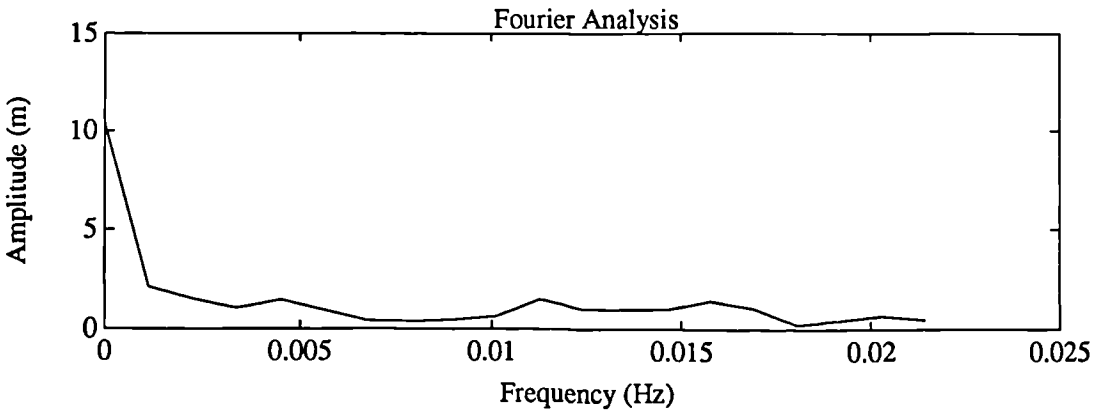


Variance = 52.31 m2

Correlation Time = 216.7 s

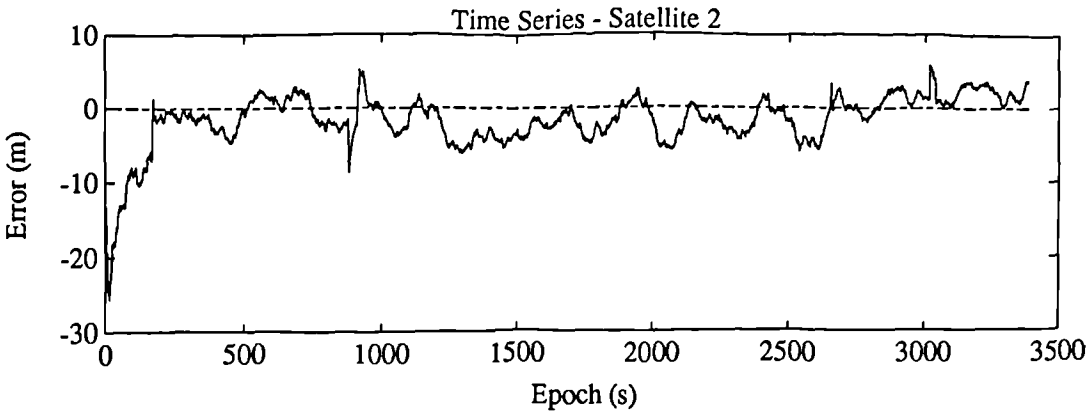


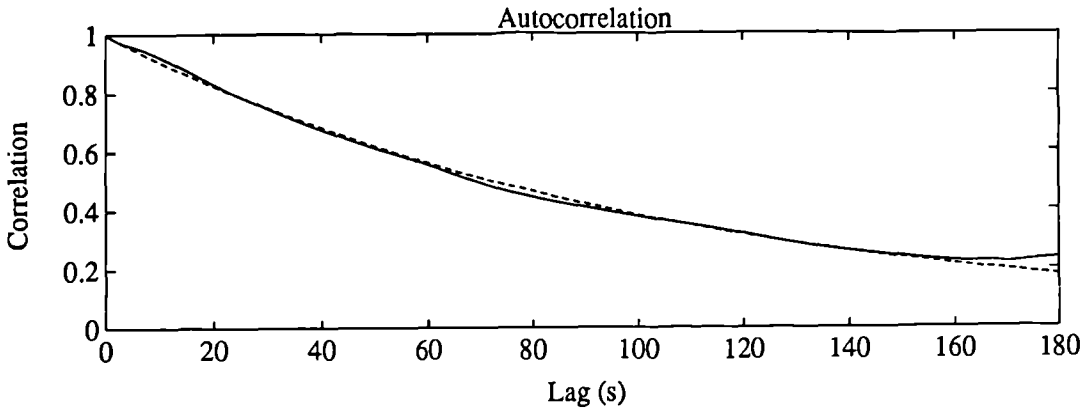


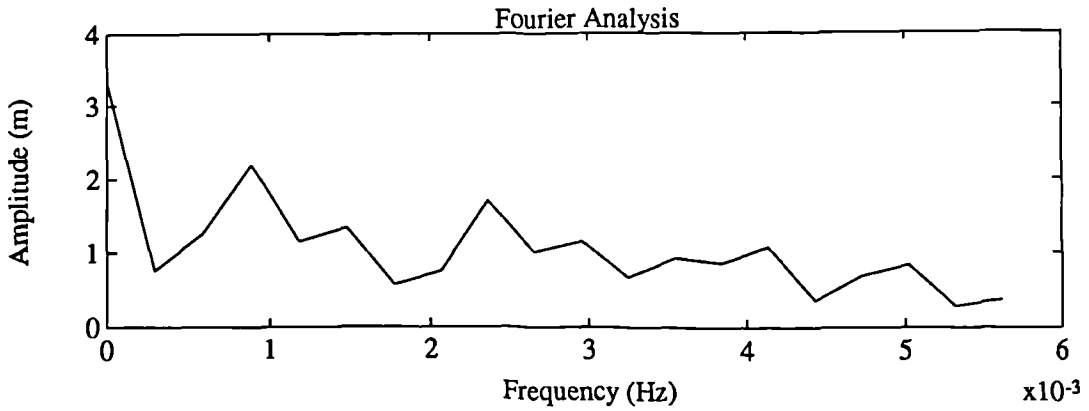


Variance = 87.44 m2

Correlation Time = 240.5 s

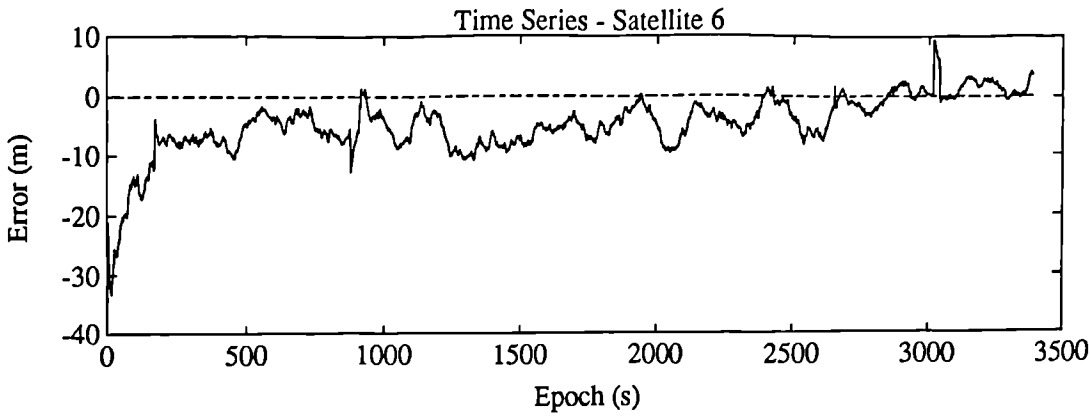


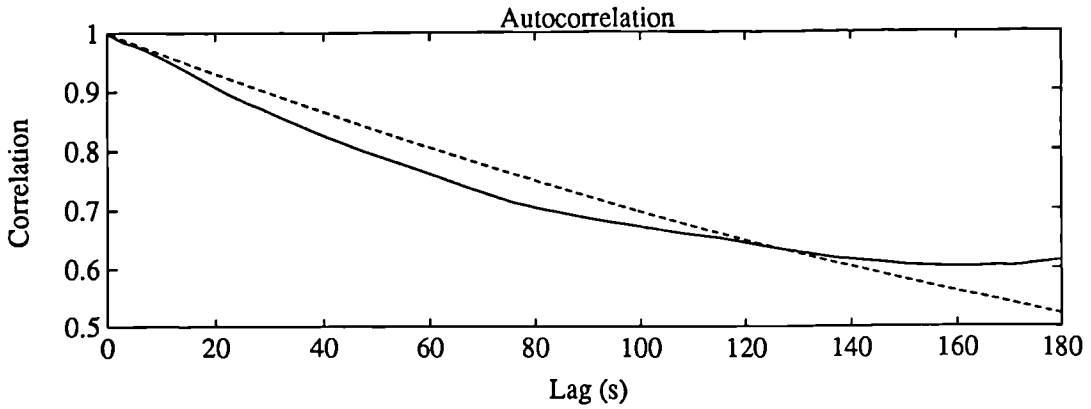


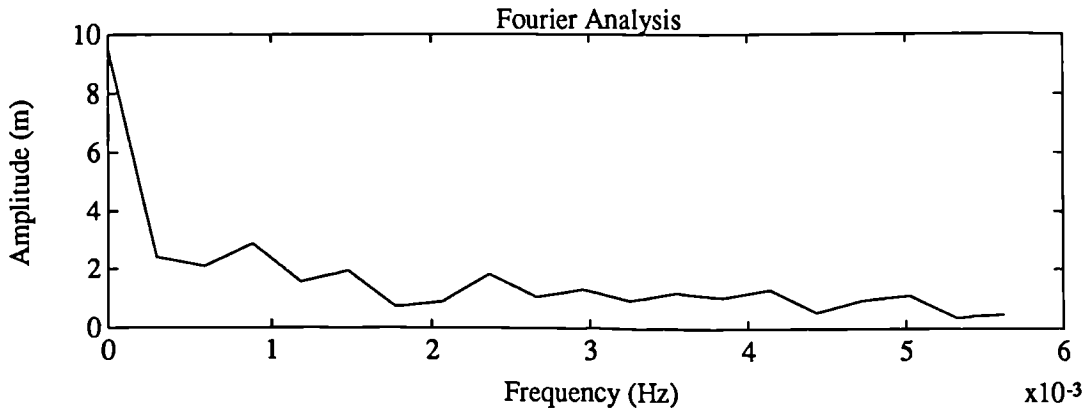


Variance = 16.24 m2

Correlation Time = 104.1 s

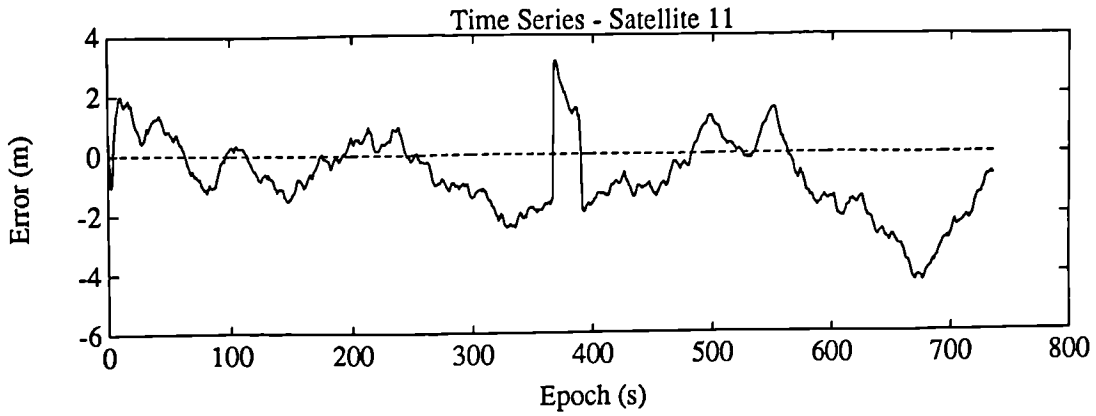


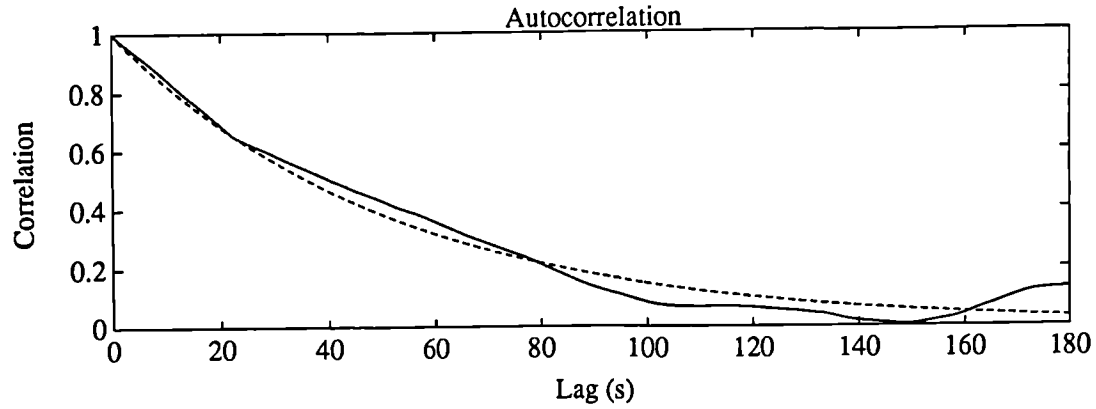


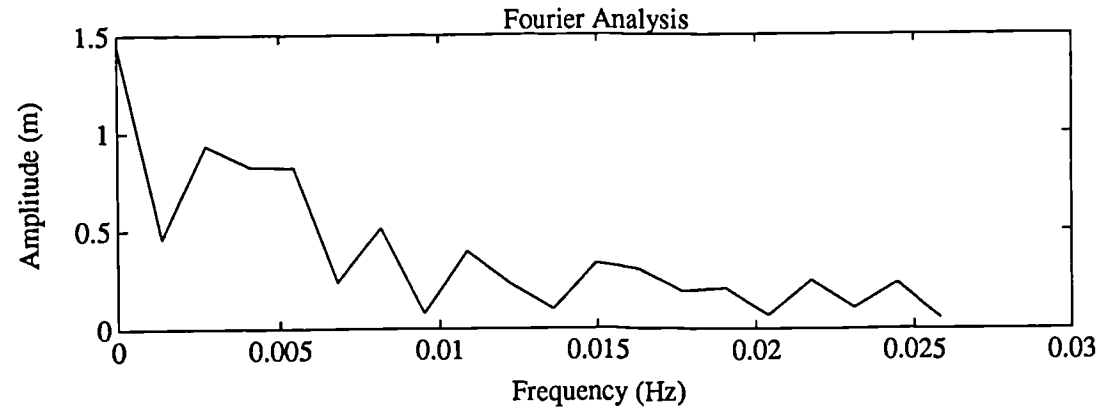


Variance = 47.42 m2

Correlation Time = 276.9 s

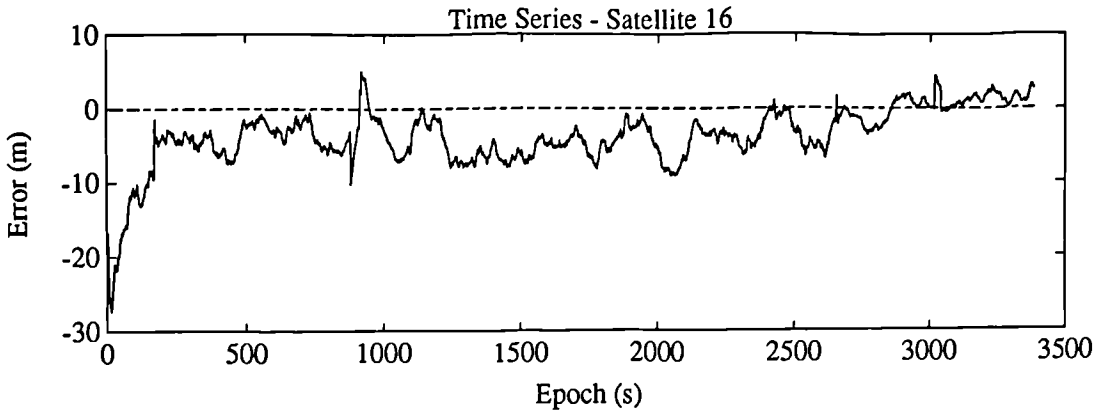


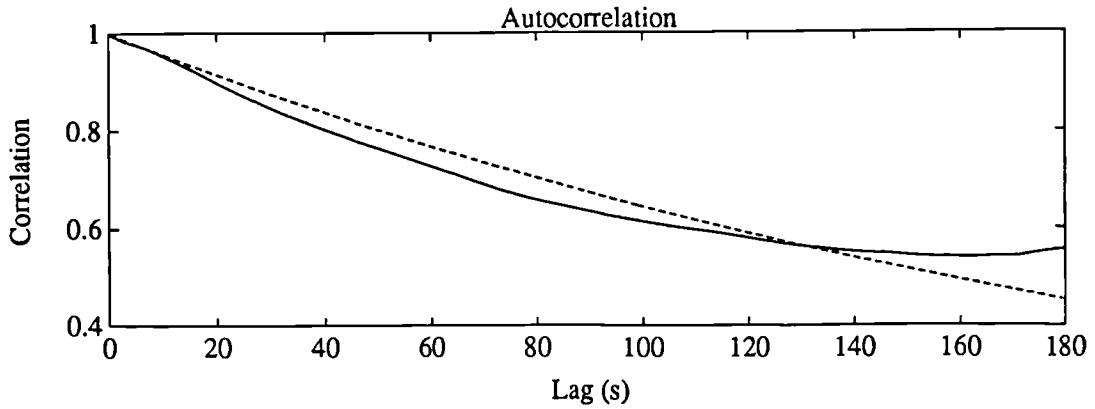


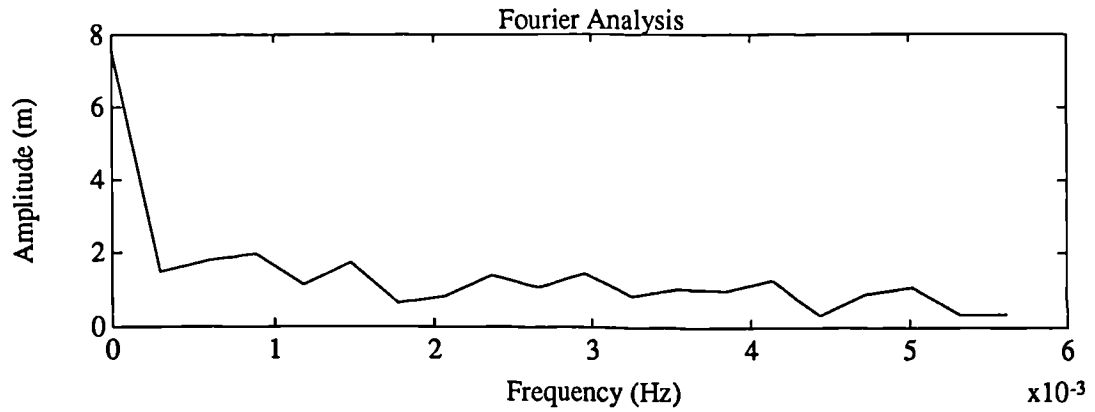


Variance = 2.346 m2

Correlation Time = 52.16 s

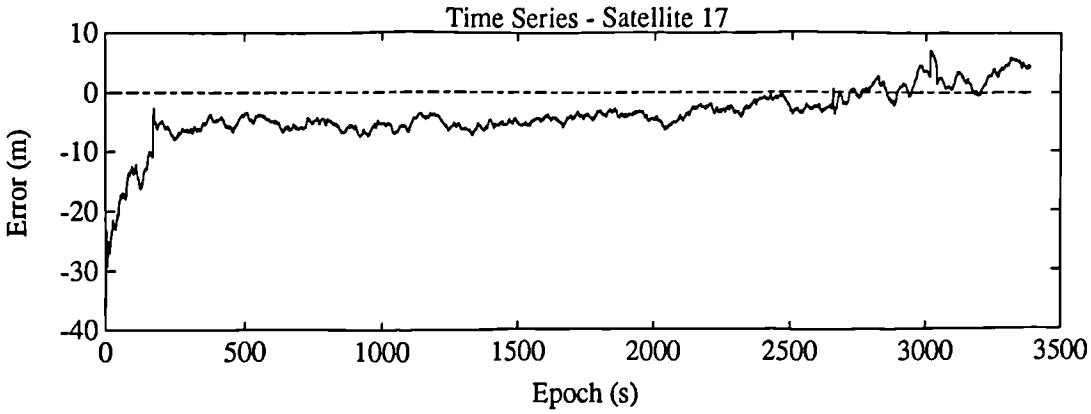


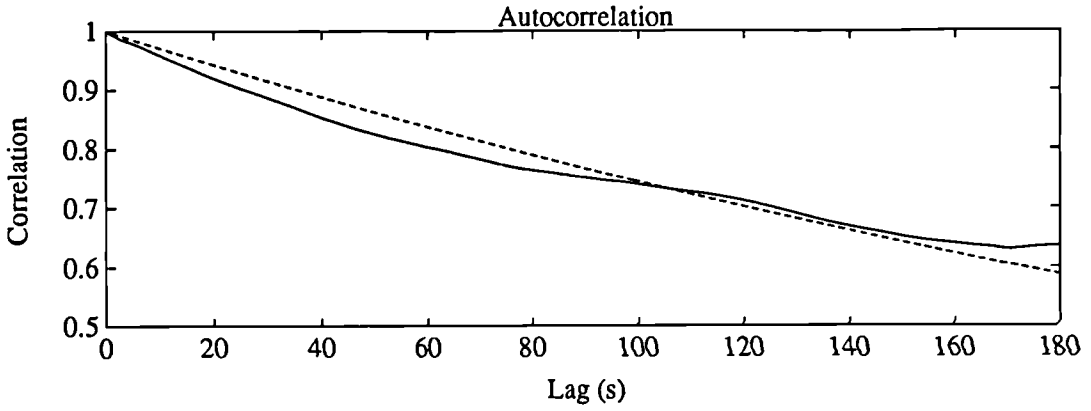


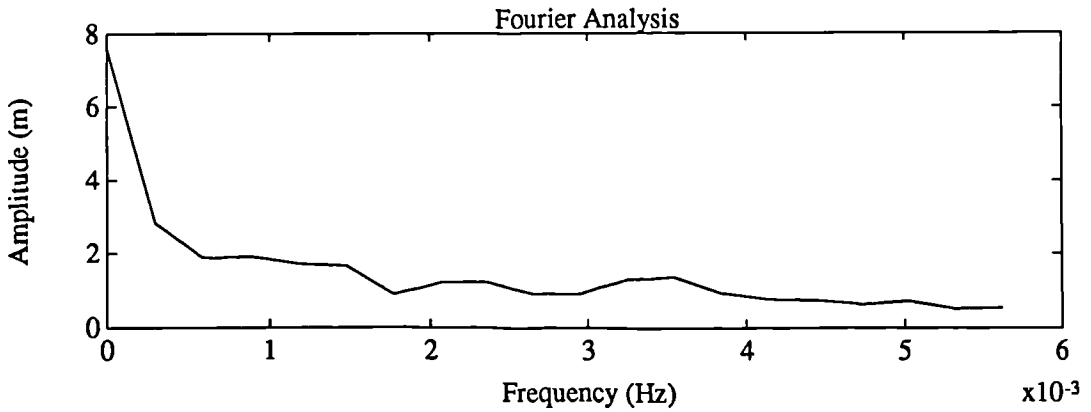


Variance = 30.7 m2

Correlation Time = 226.6 s

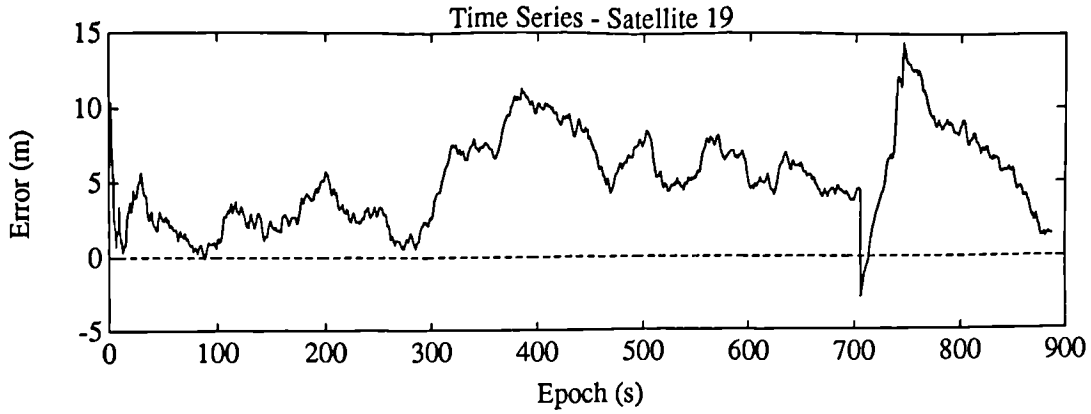


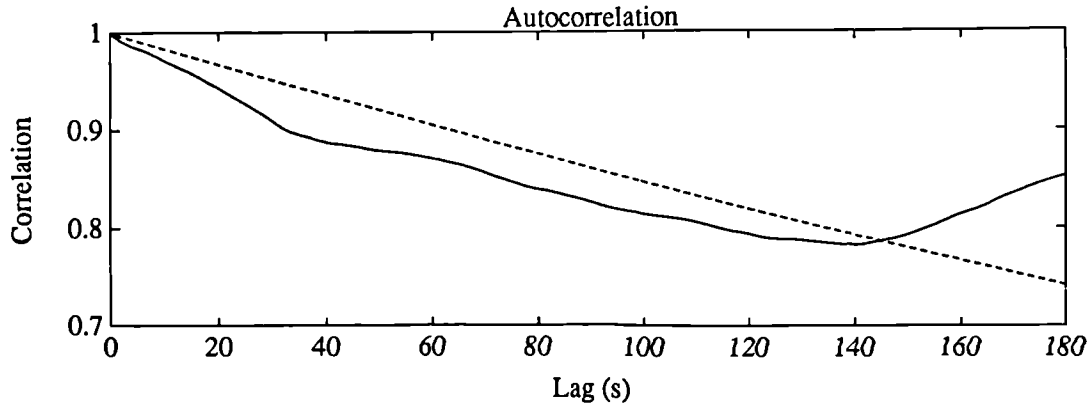


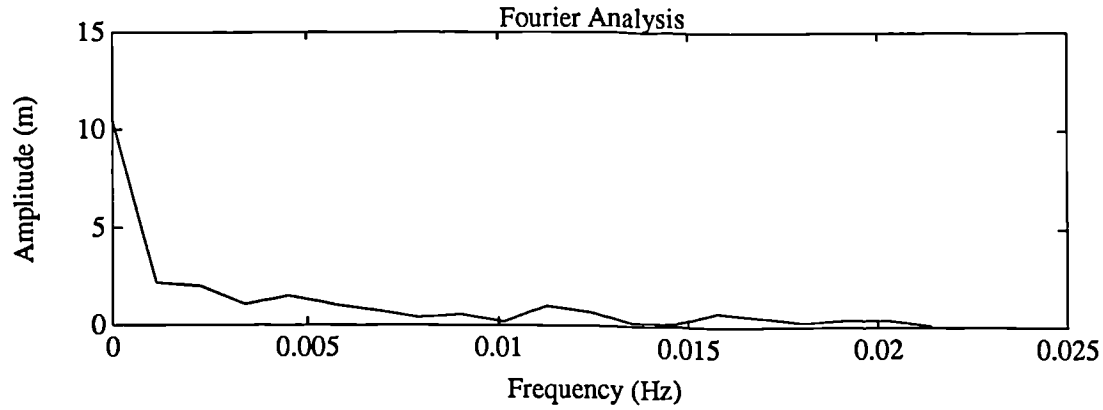


Variance = 34.32 m2

Correlation Time = 337.7 s

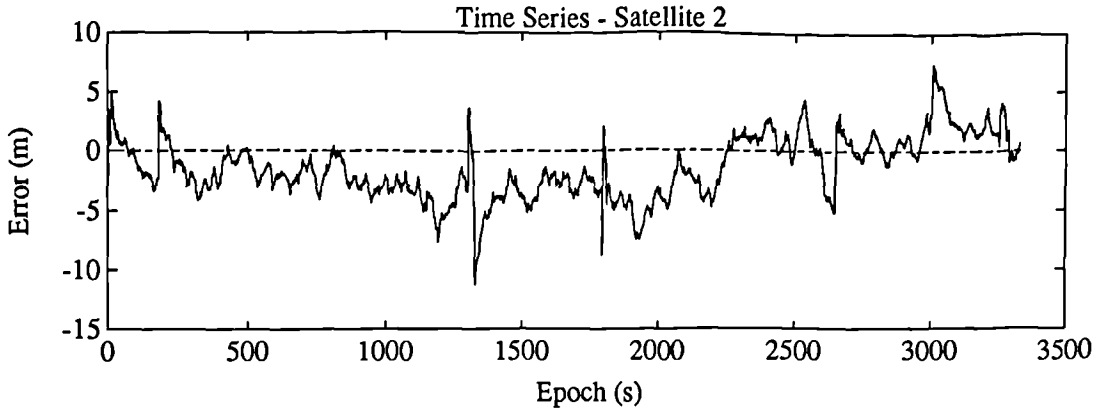


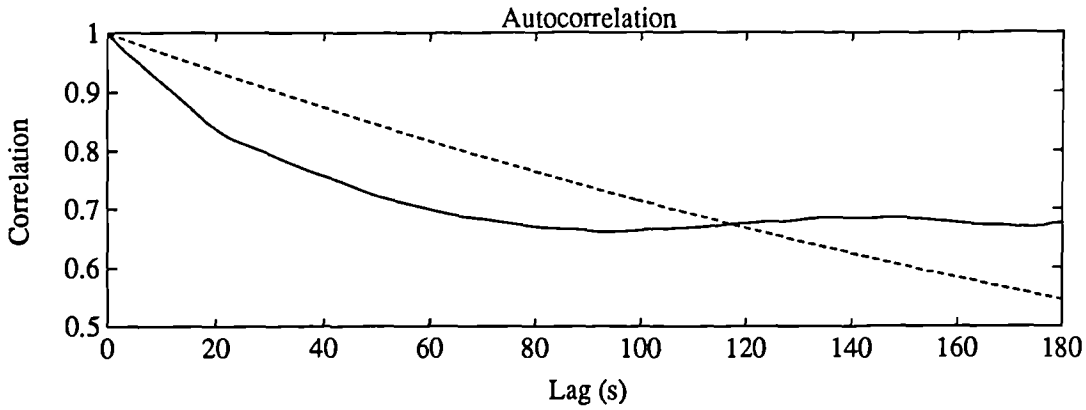


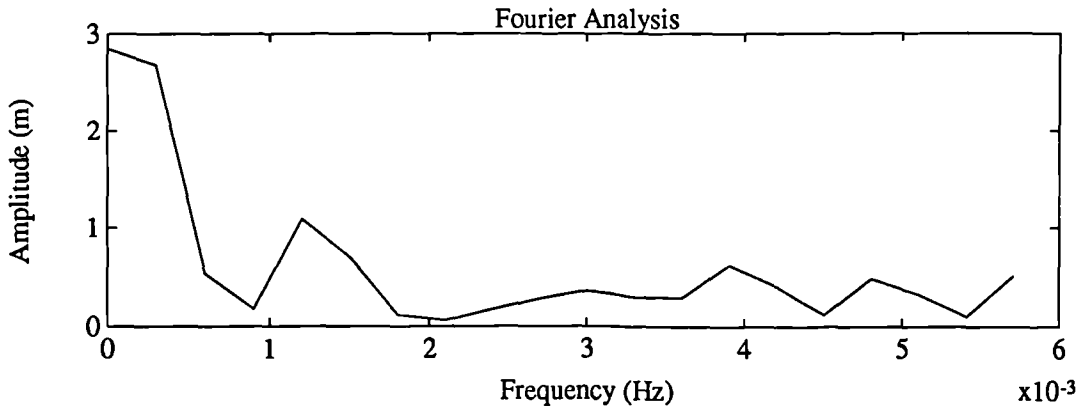


Variance = 36.75 m2

Correlation Time = 601.1 s

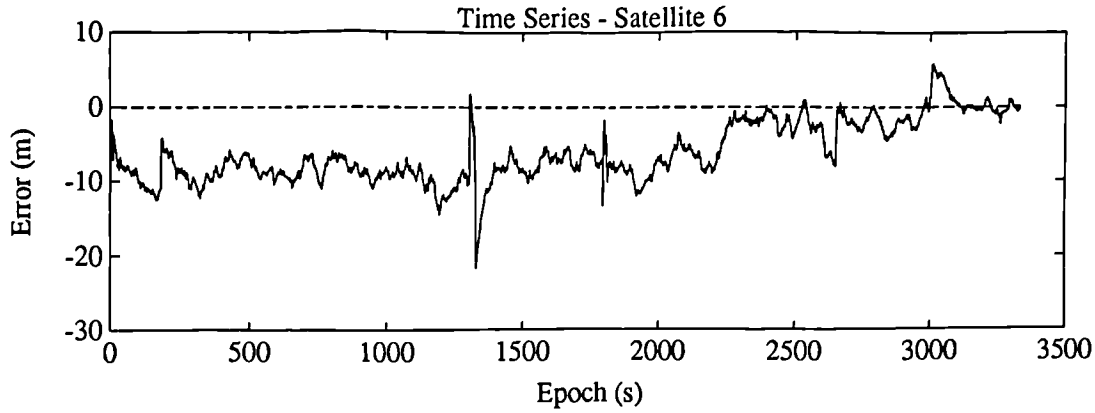


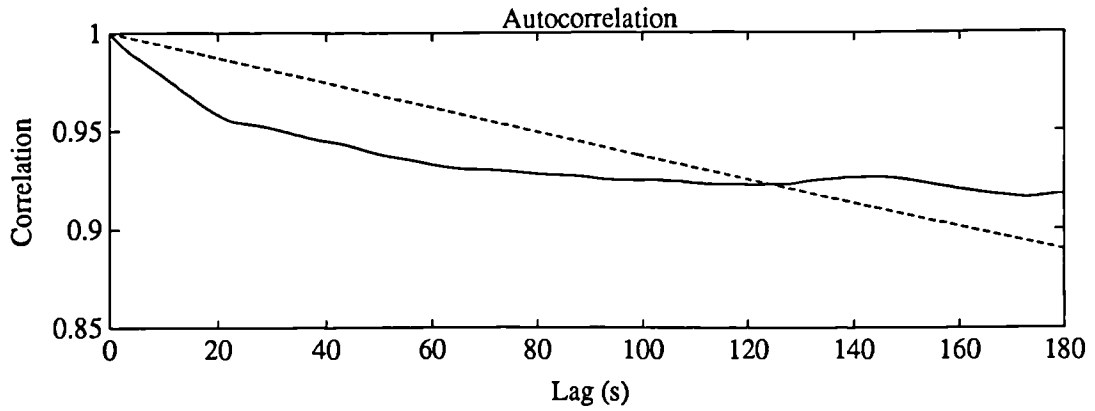


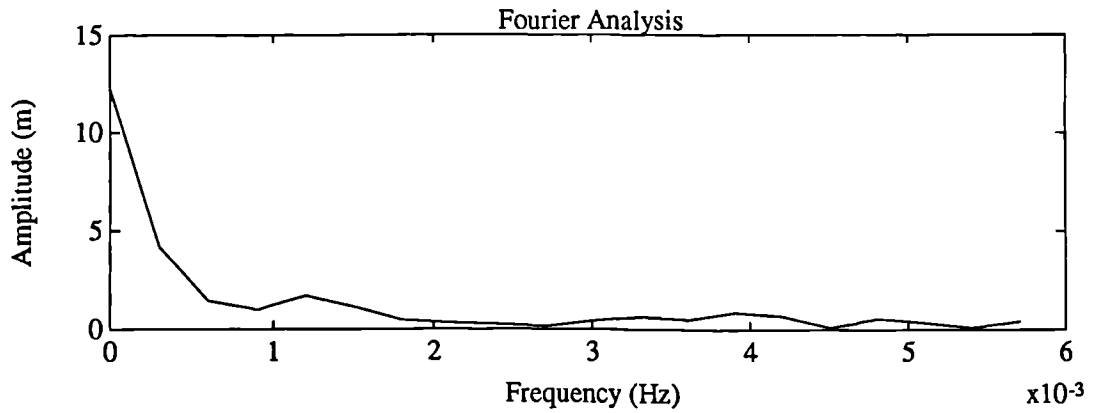


Variance = 8.817 m2

Correlation Time = 297 s

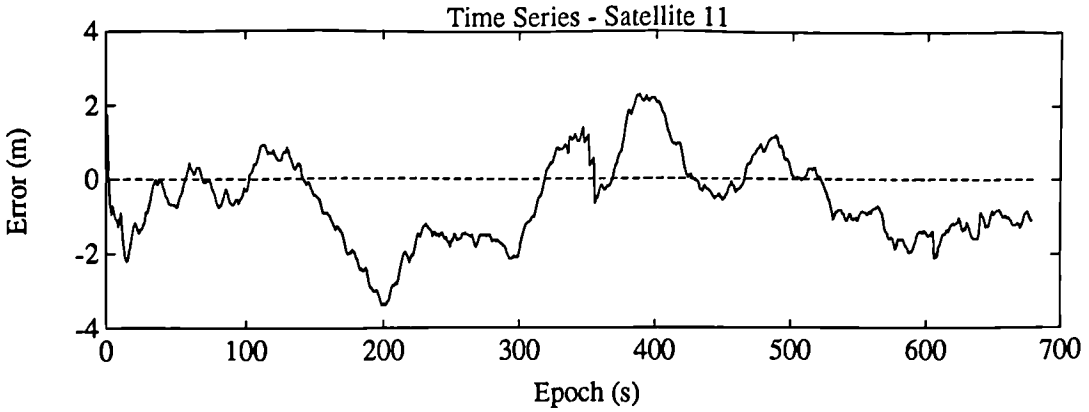


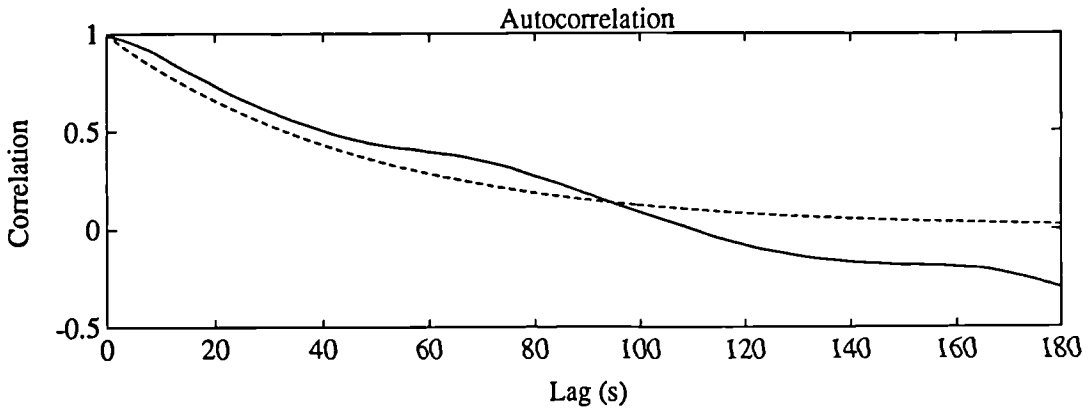


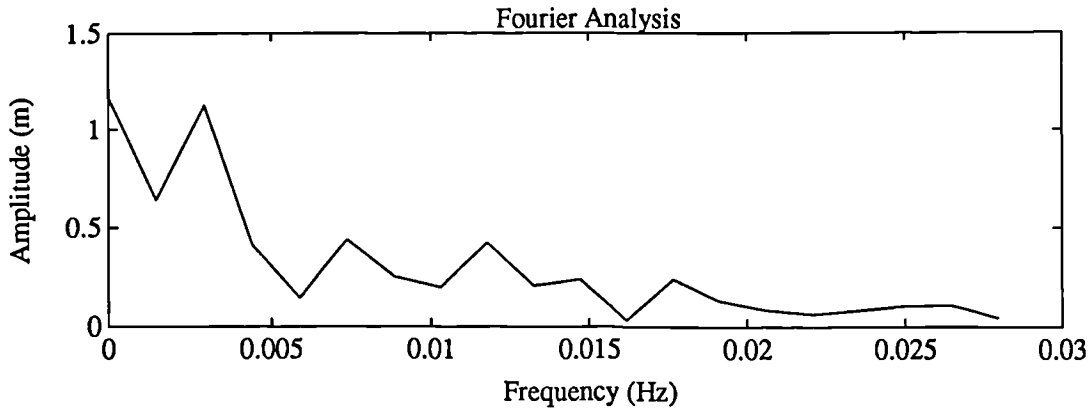


Variance = 53.84 m2

Correlation Time = 1539 s

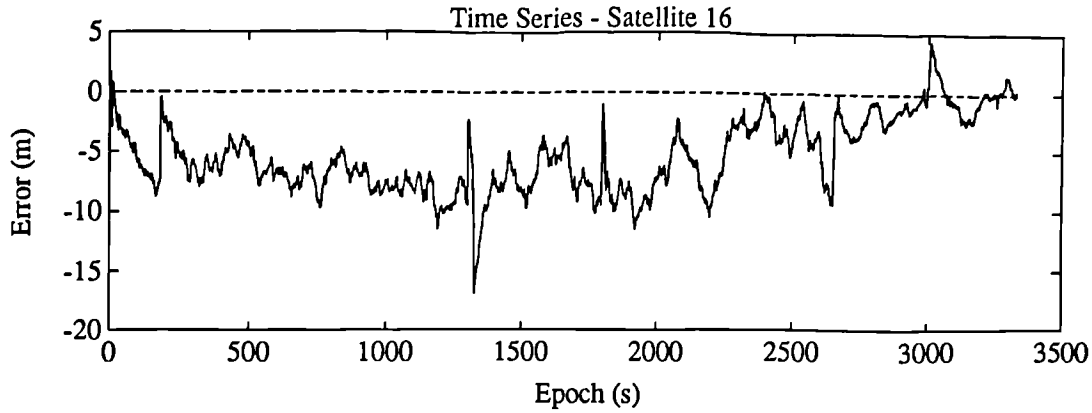


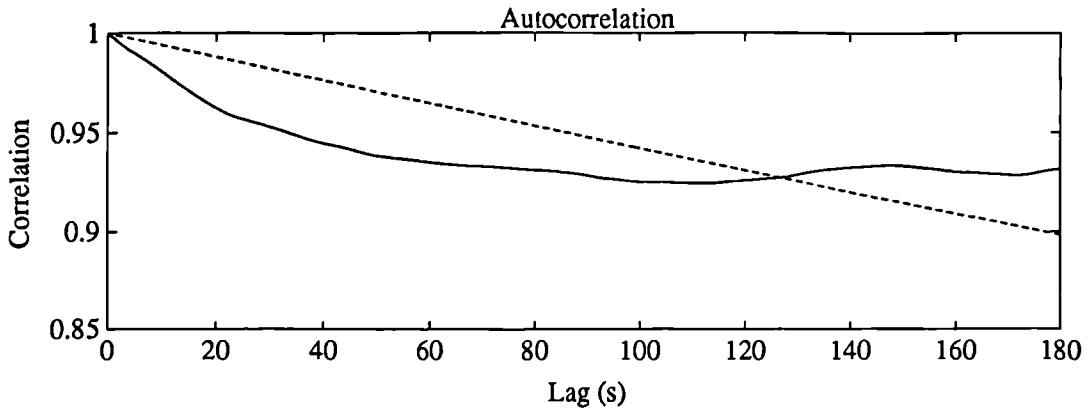


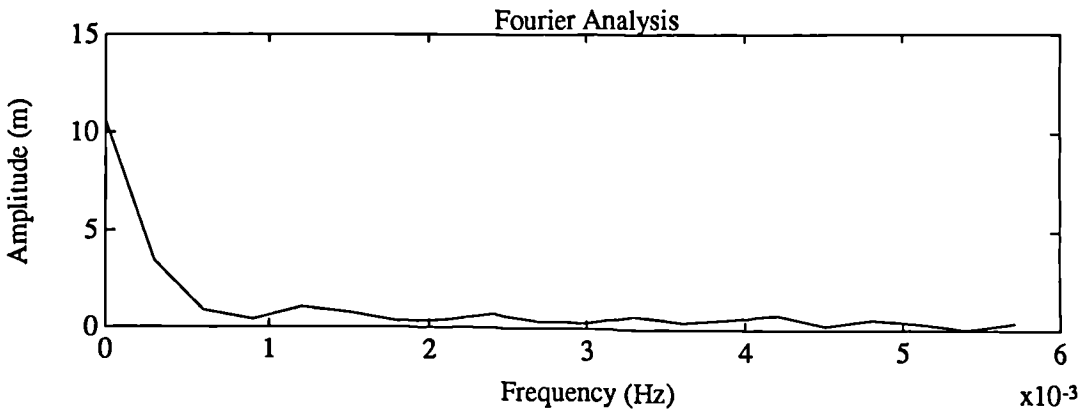


Variance = 1.675 m2

Correlation Time = 47.29 s

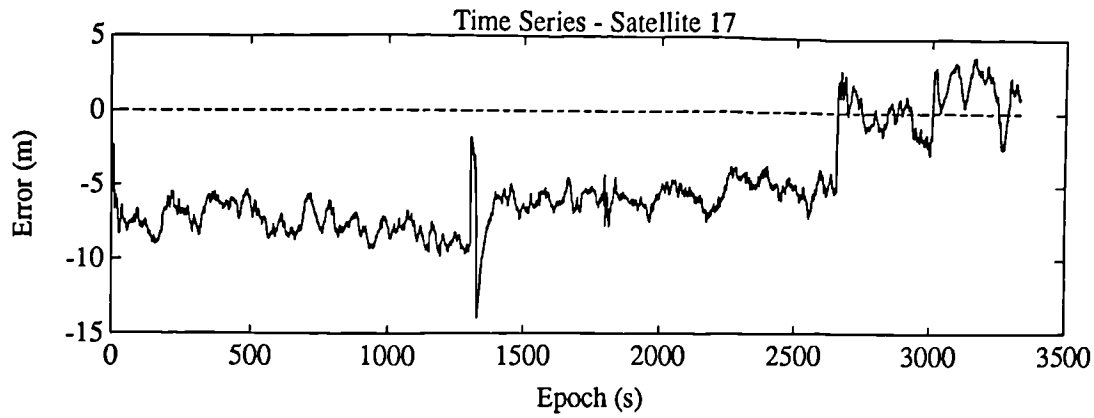


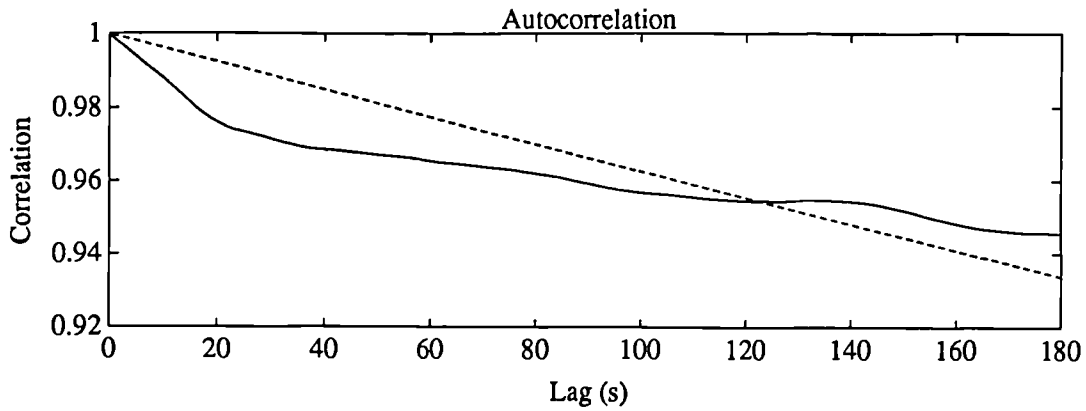


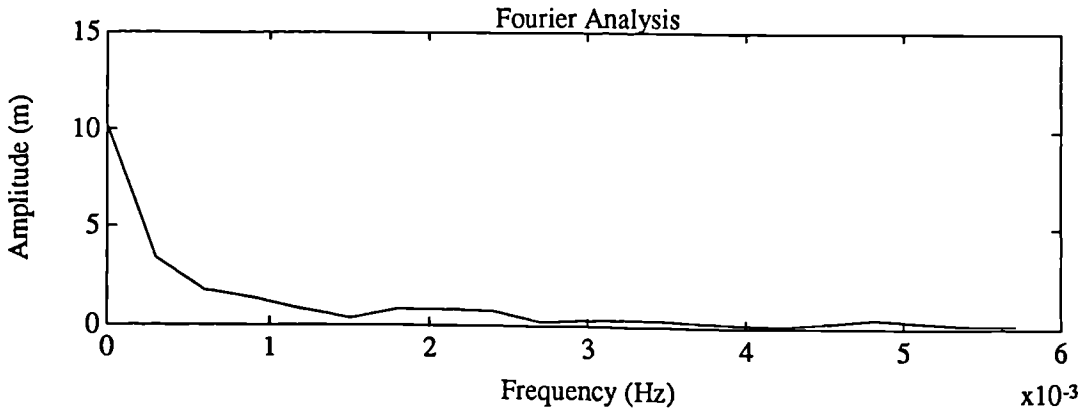


Variance = 38.24 m2

Correlation Time = 1674 s

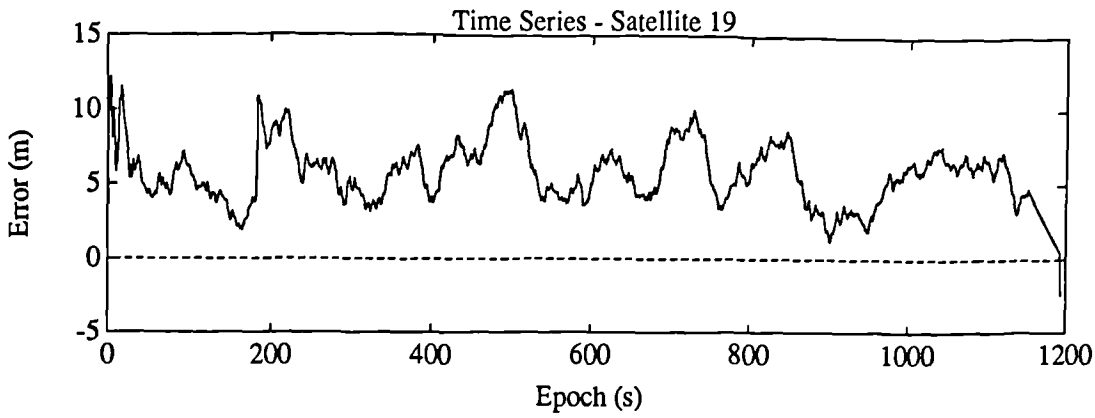


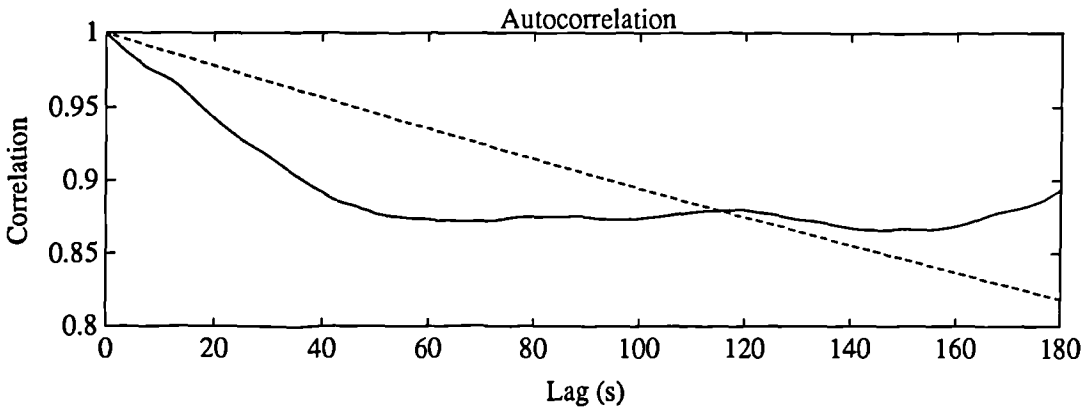


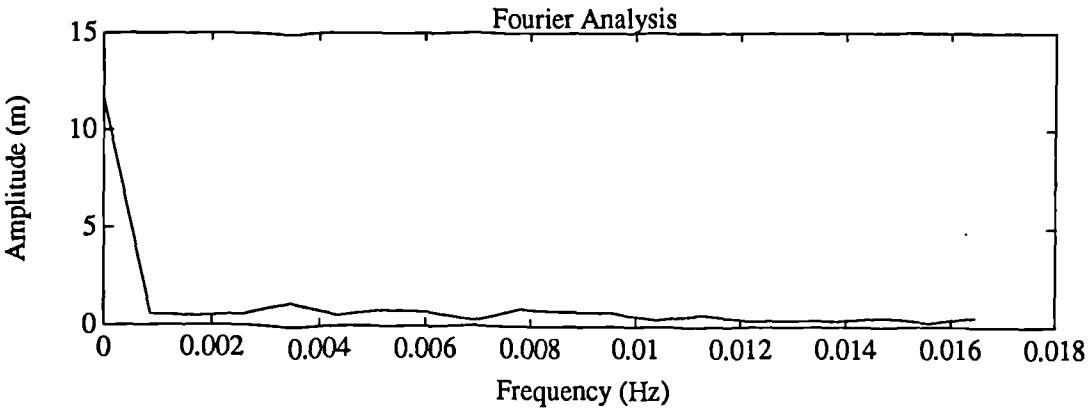


Variance = 37.39 m2

Correlation Time = 2624 s







Variance = 38.48 m2

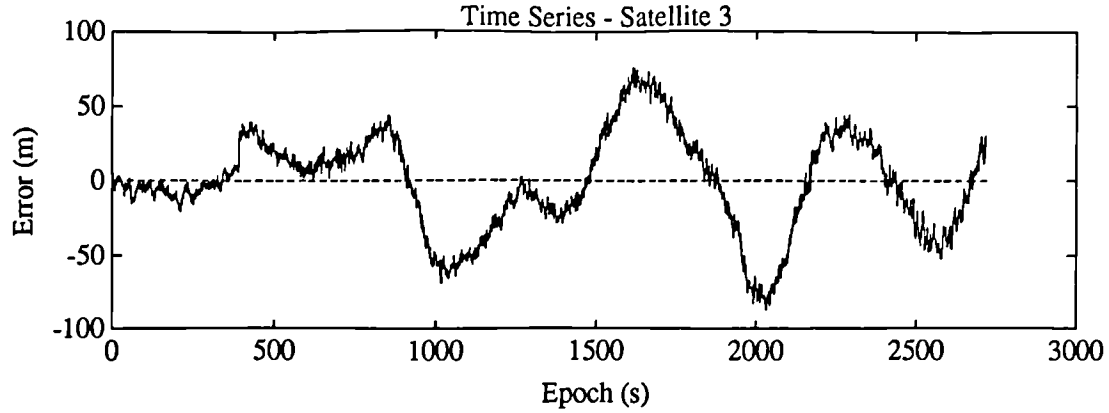
Correlation Time = 897 s

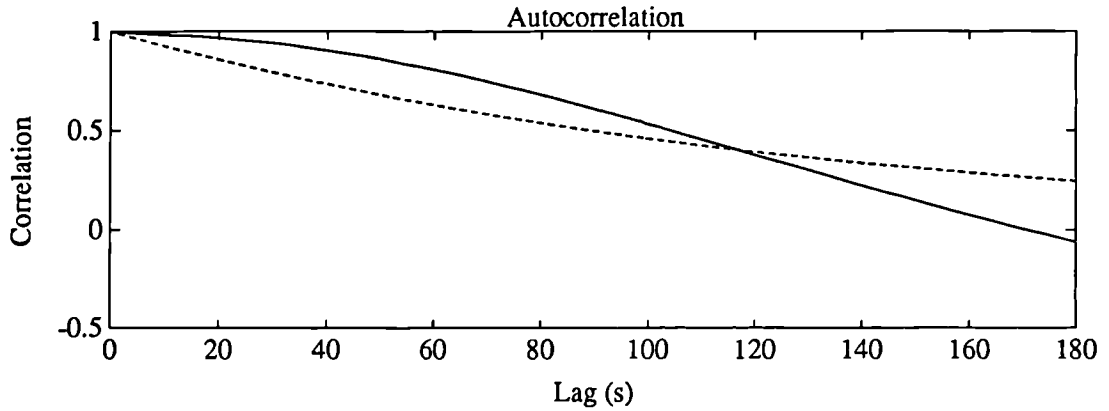
APPENDIX D

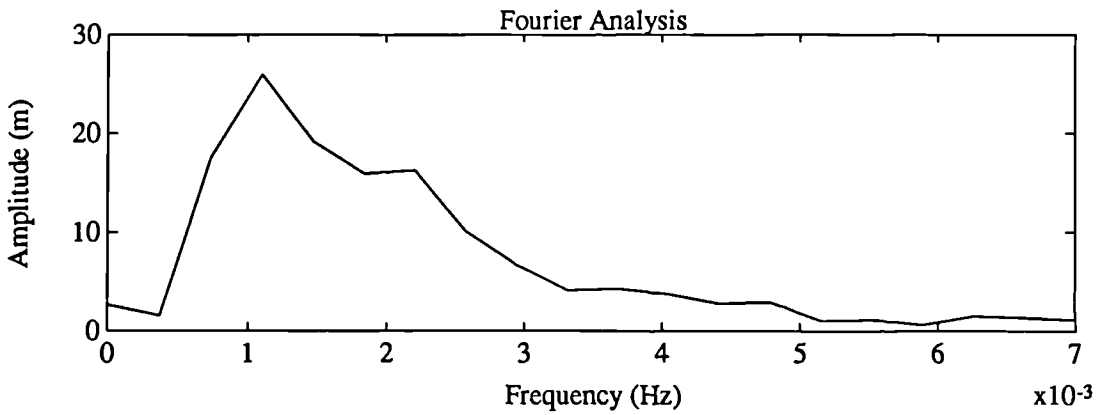
TRIMBLE OUTPUT

Time series, autocorrelation and Fourier analysis for data files:

GTYH1921

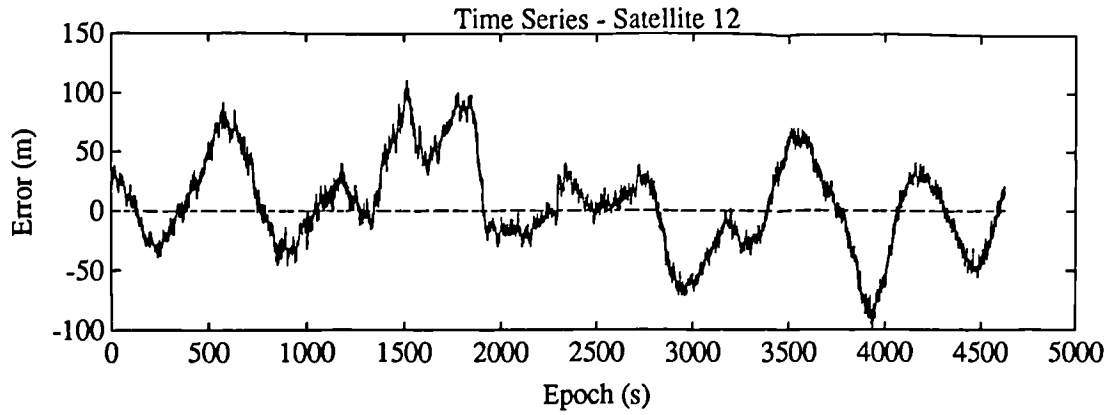


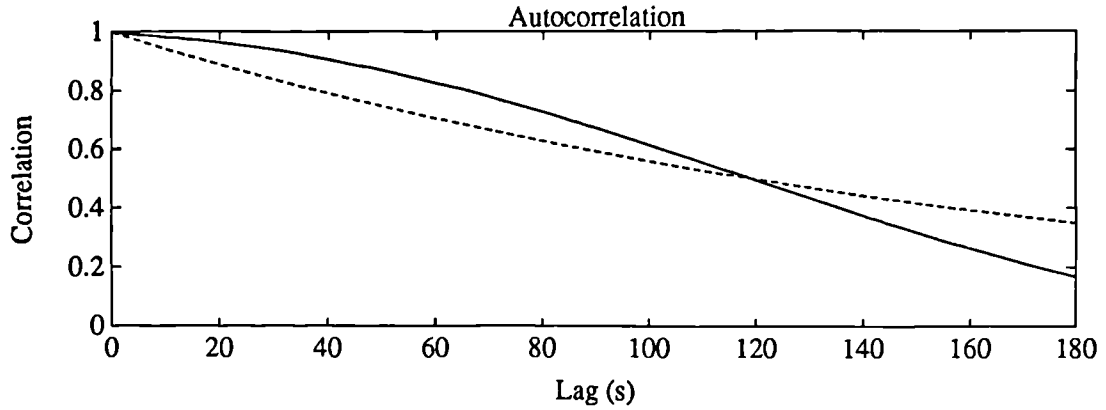


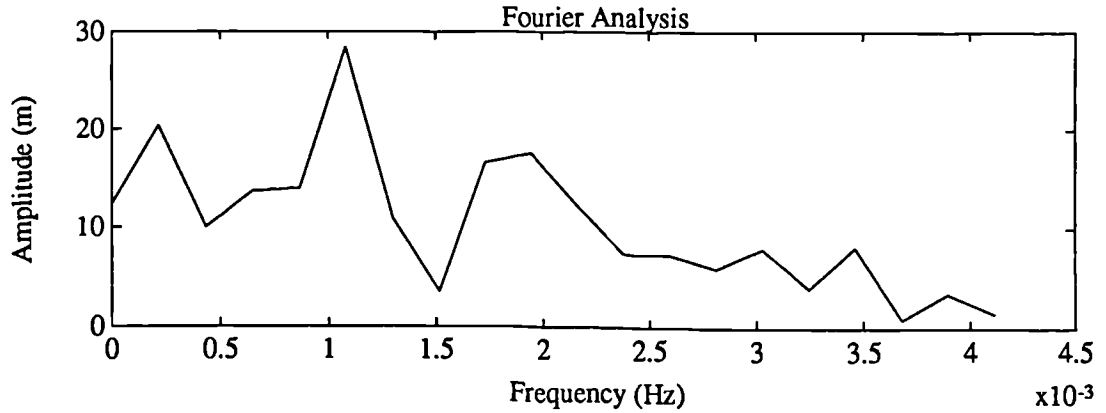


Variance = 1068 m2

Correlation Time = 128.9 s

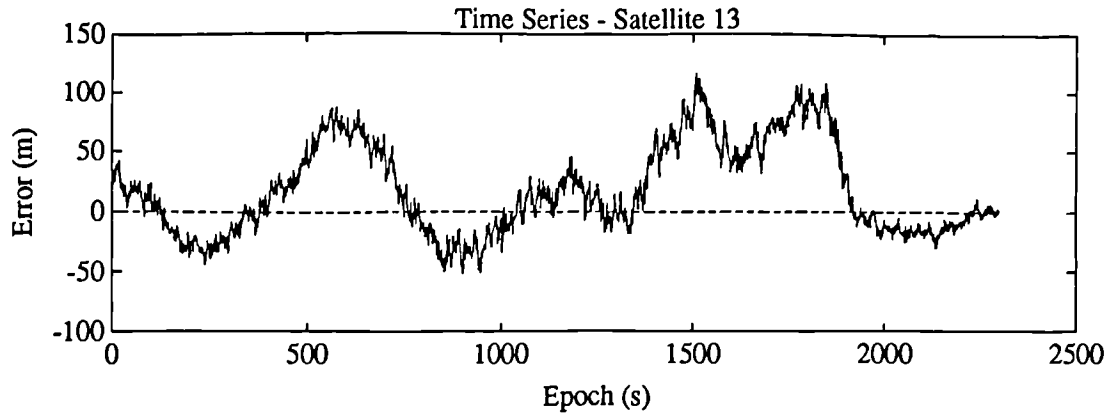


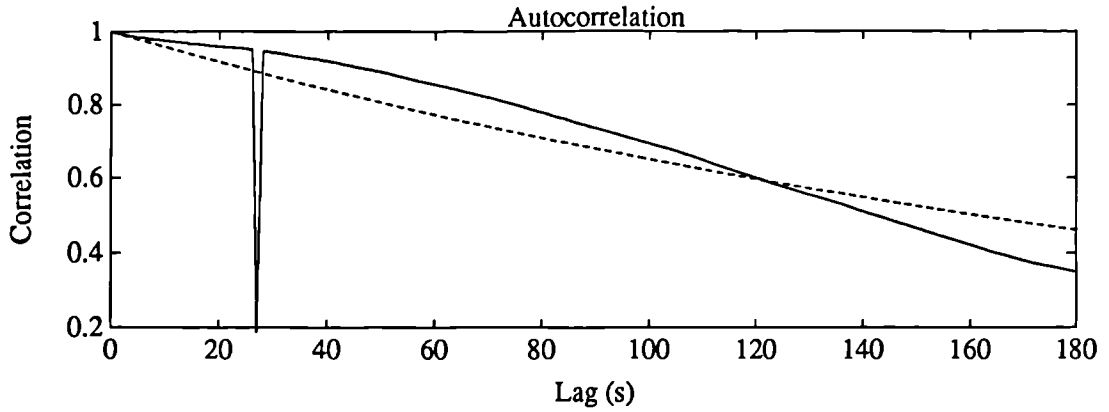


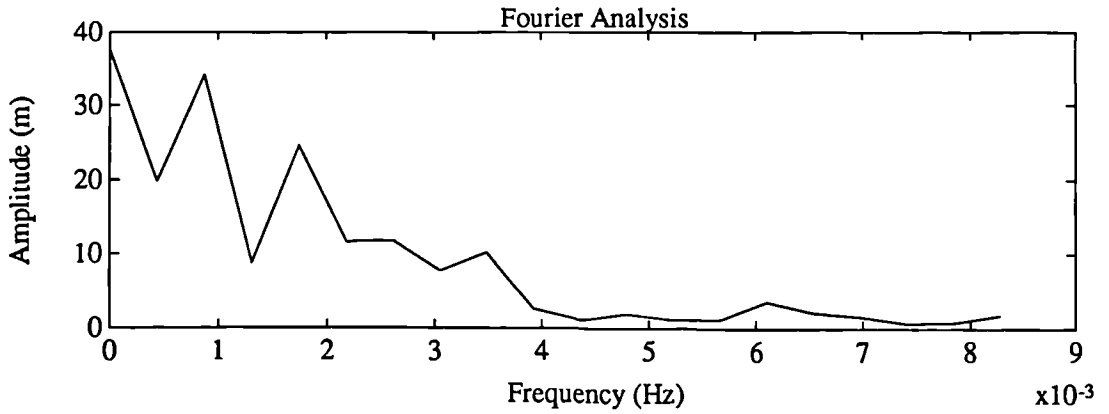


Variance = 1517 m2

Correlation Time = 171 s

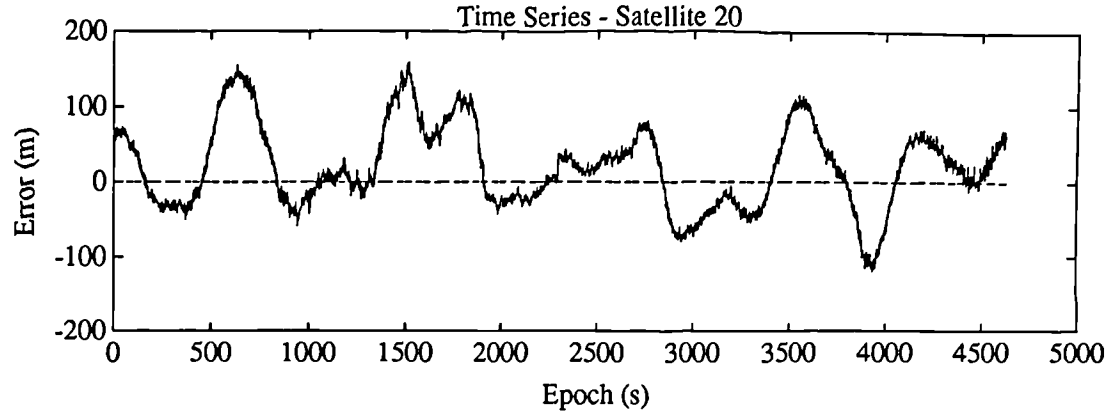


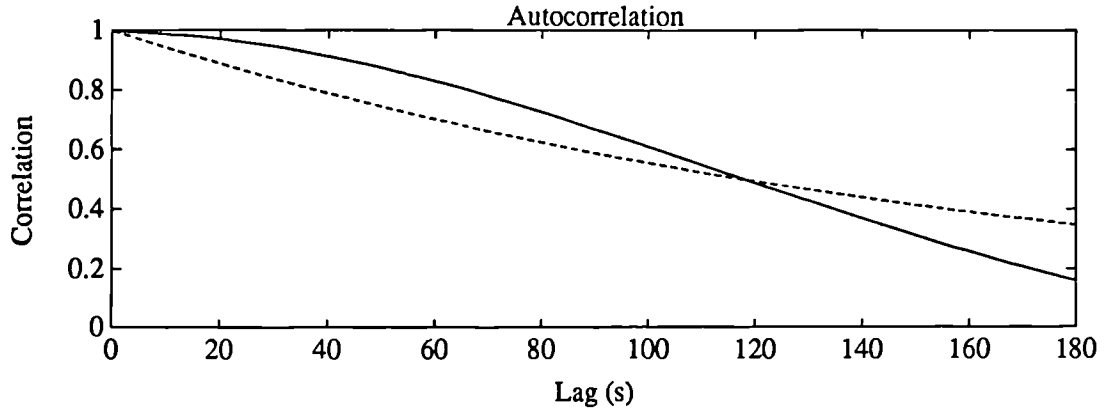


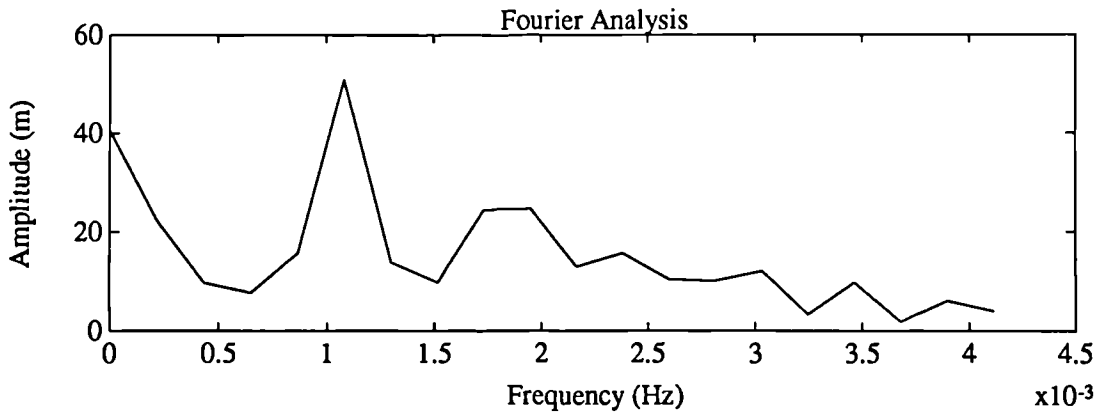


Variance = 1774 m2

Correlation Time = 232.1 s

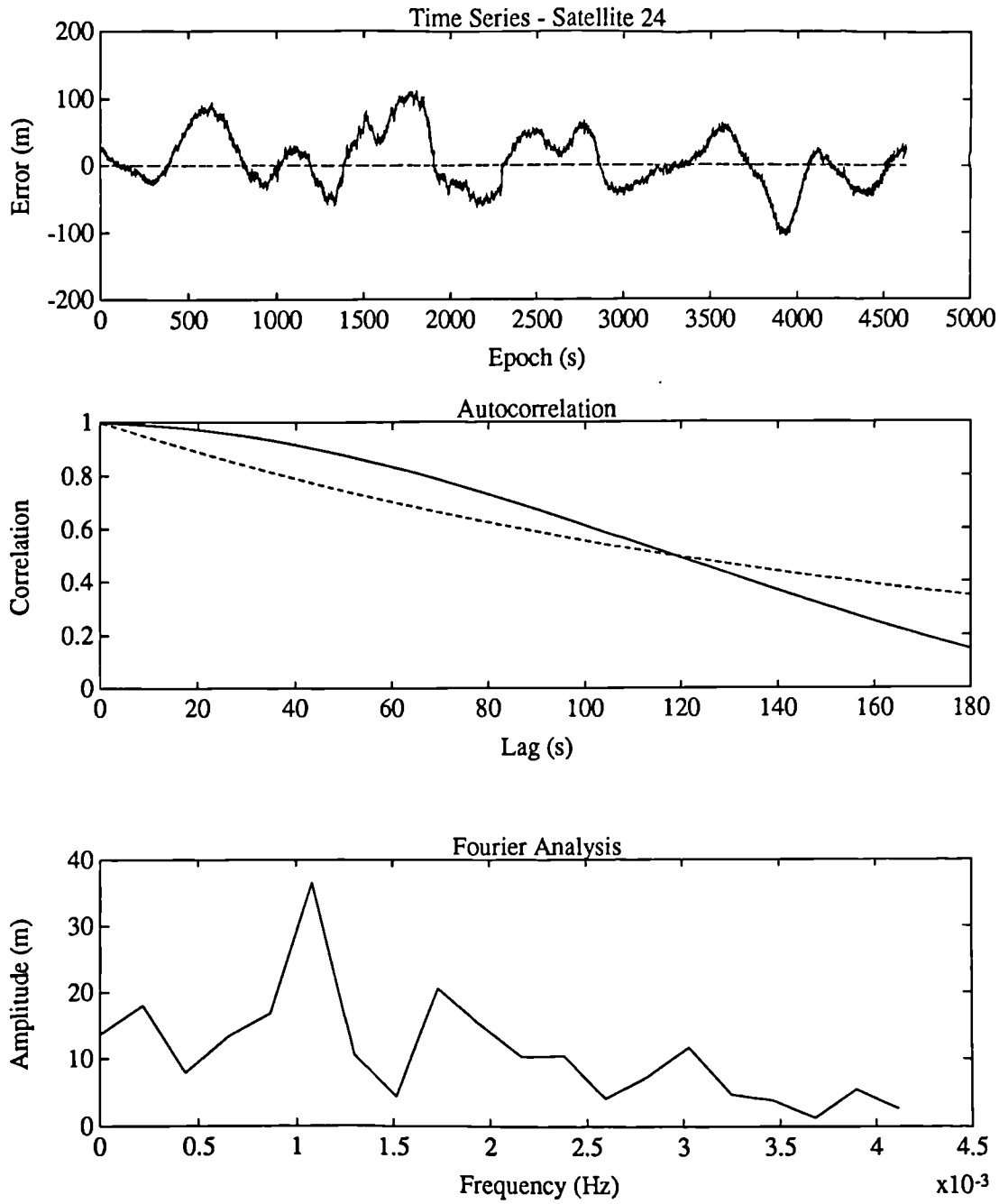






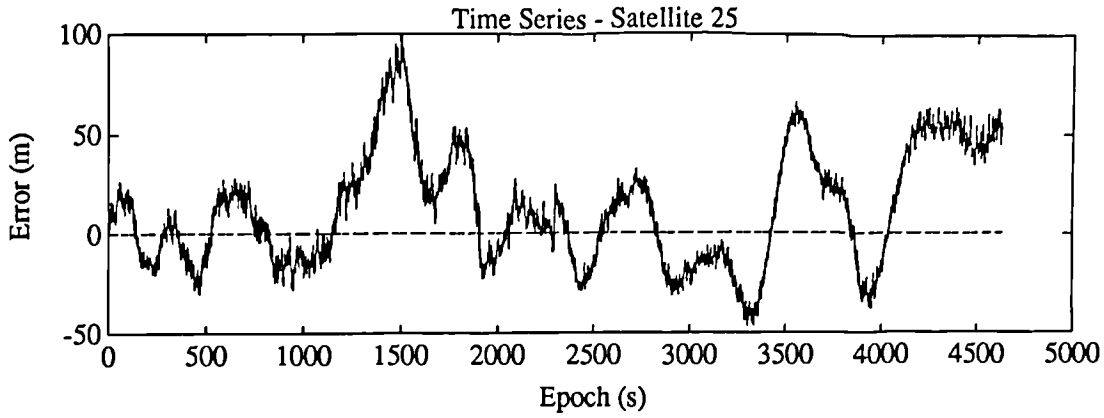
Variance = 3425 m2

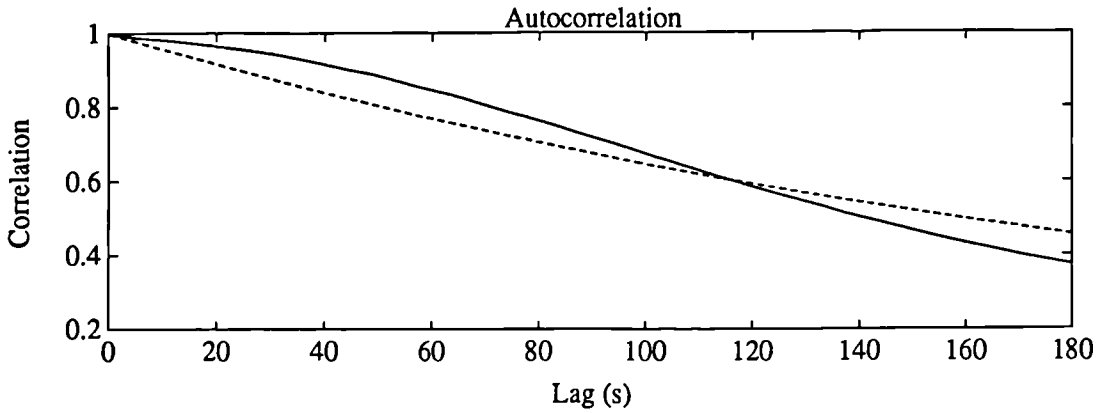
Correlation Time = 170 s

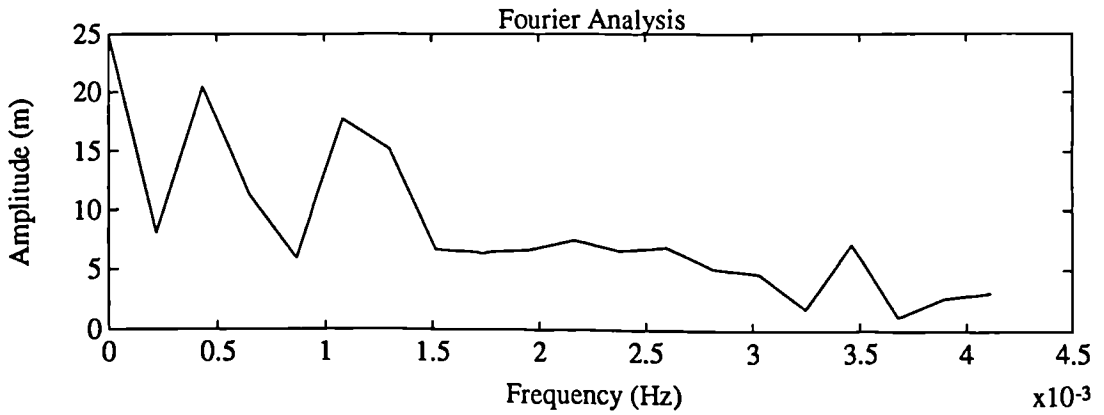


Variance = 1808 m2

Correlation Time = 169.7 s







Variance = 981.6 m2

Correlation Time = 228.4 s

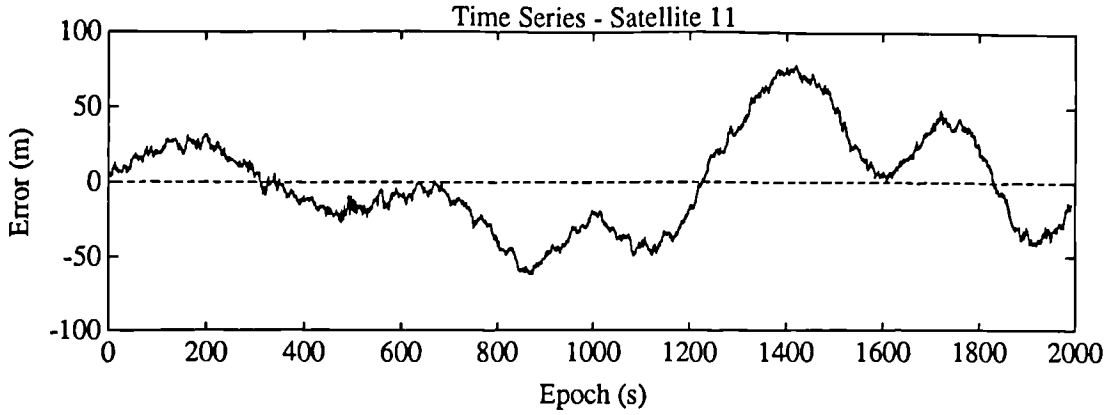
APPENDIX E

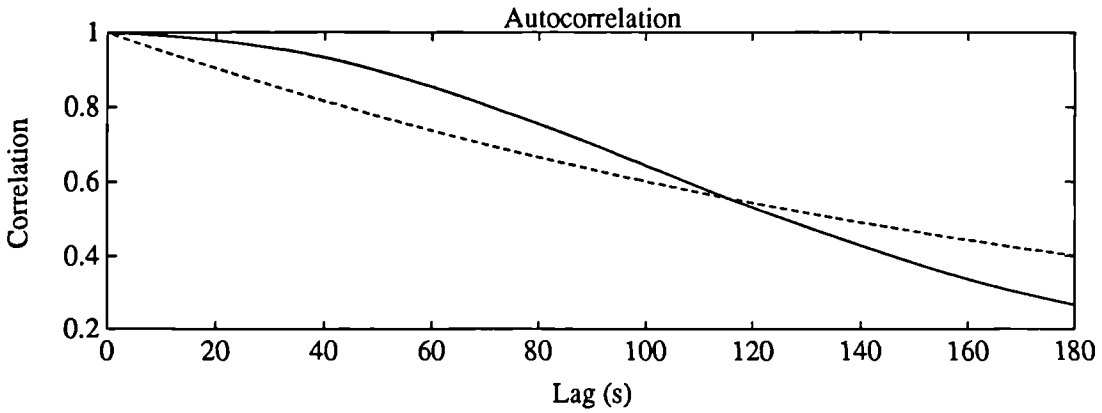
LEICA OUTPUT

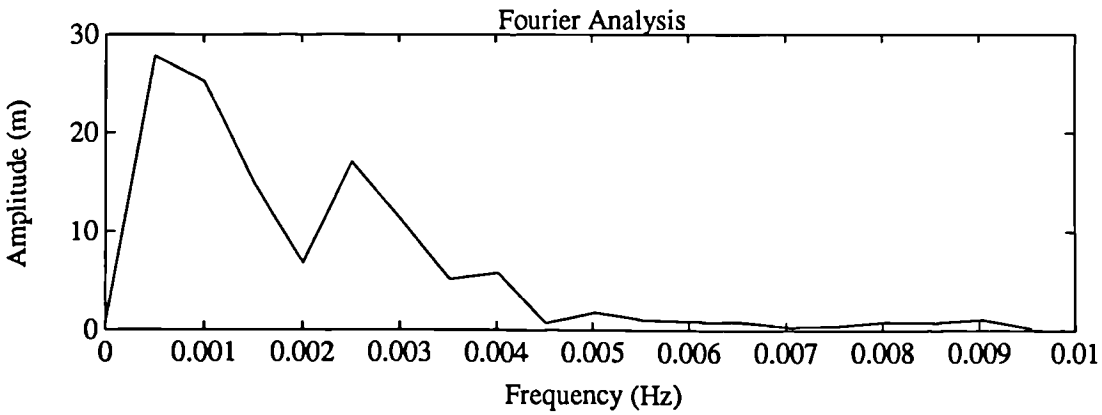
Time series, autocorrelation and Fourier analysis for data files:

COMP2021

SAMP2021

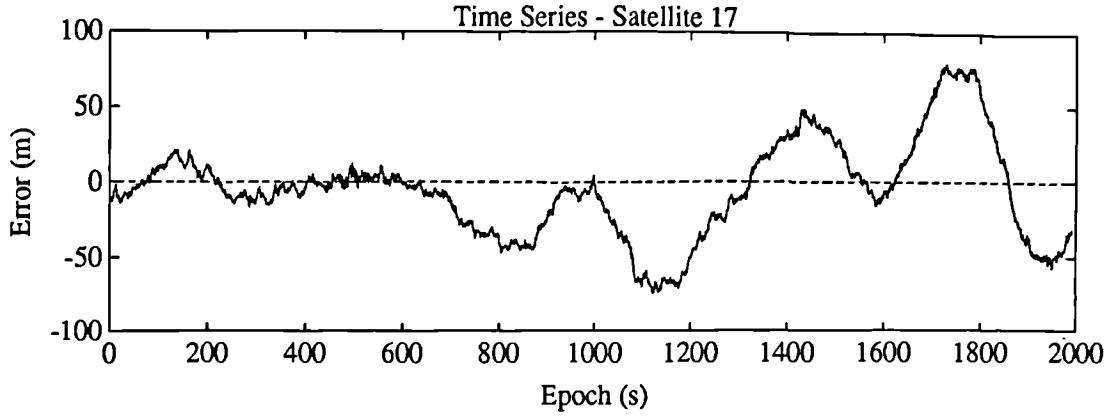


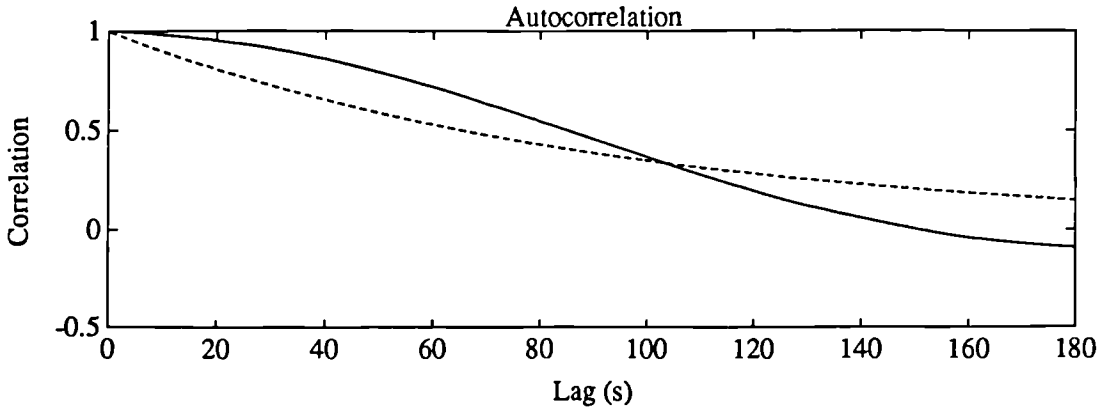


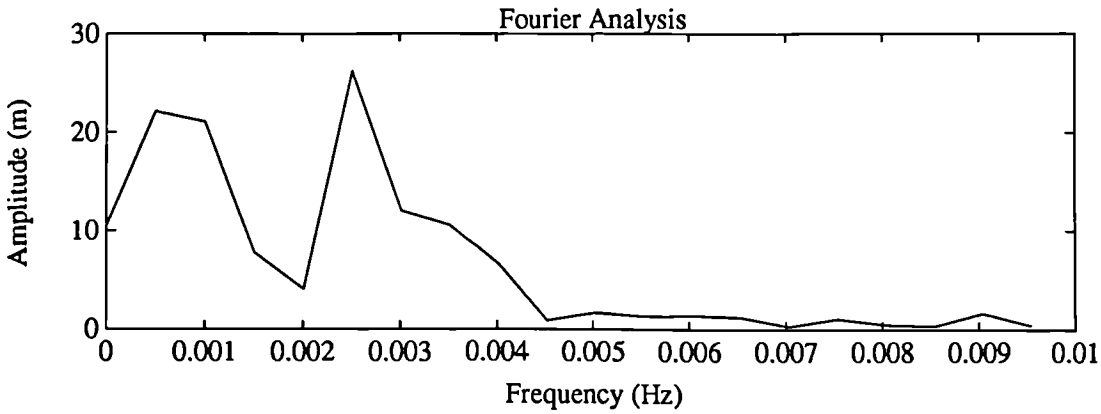


Variance = 1095 m2

Correlation Time = 196.7 s

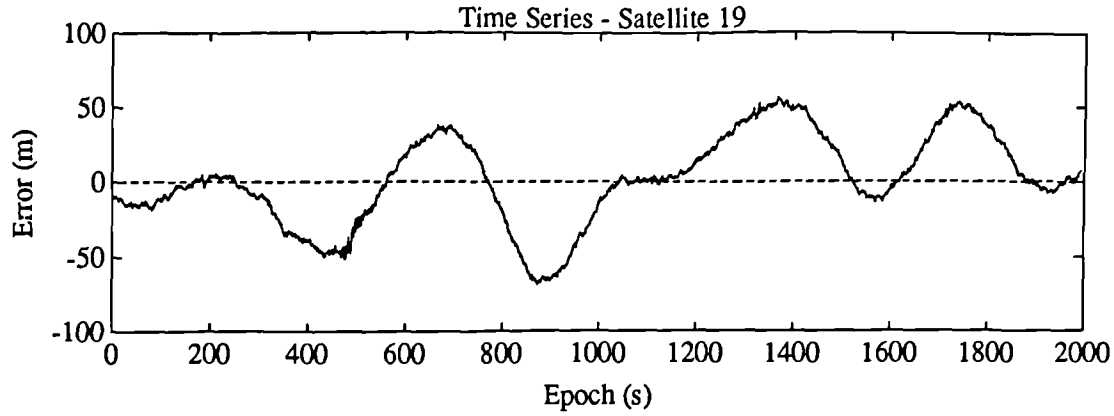


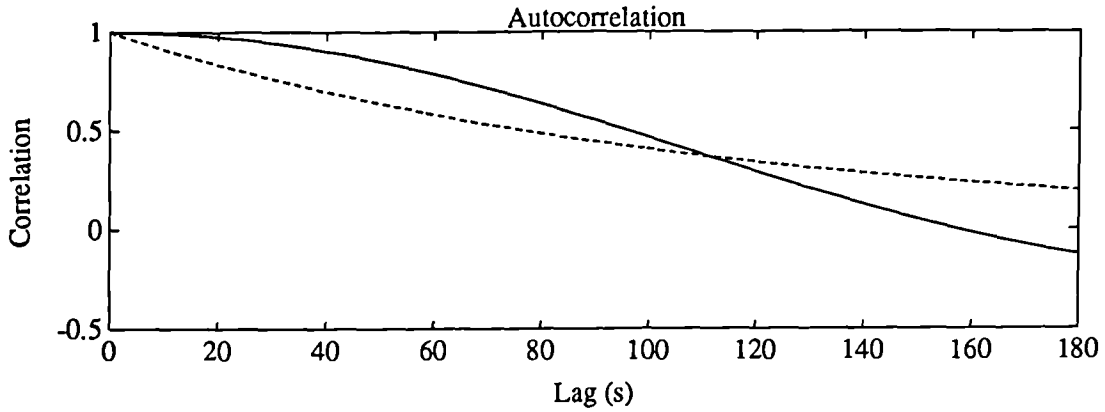


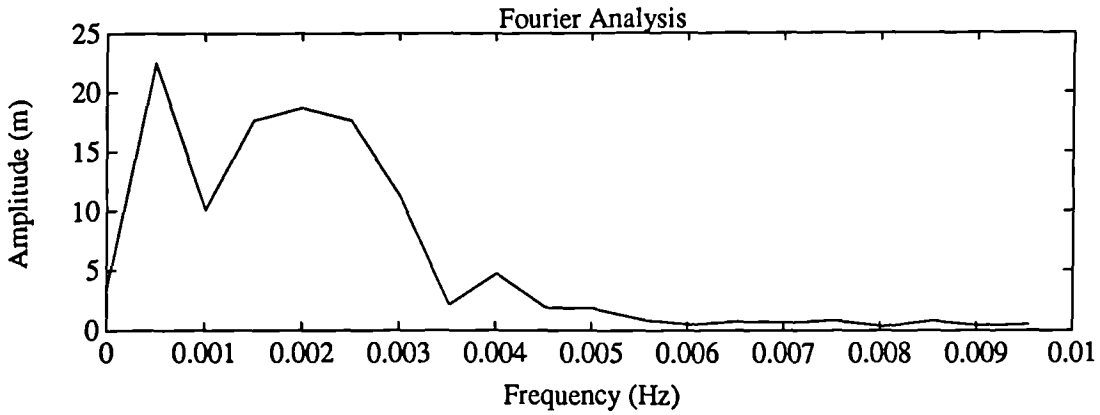


Variance = 1046 m2

Correlation Time = 94.6 s

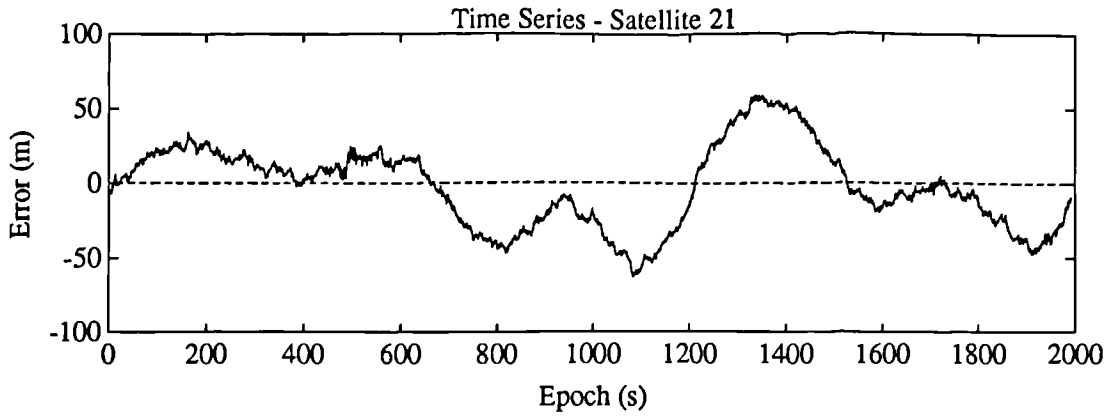


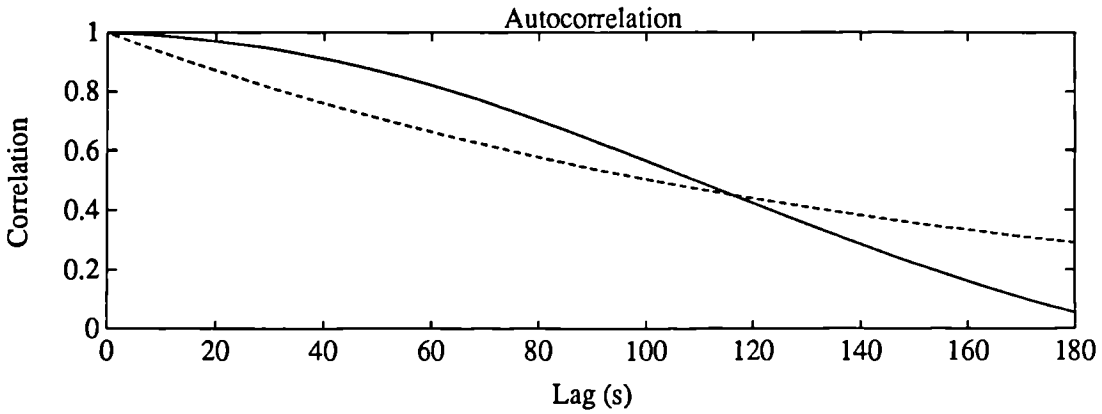


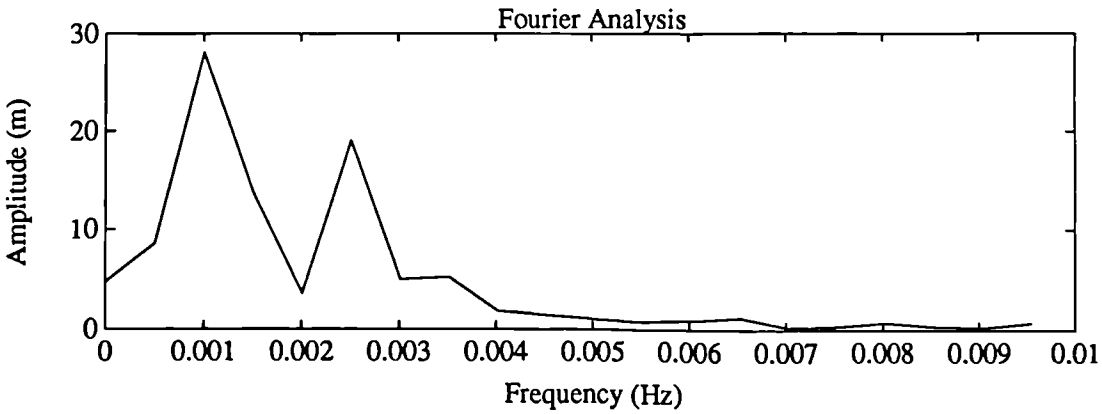


Variance = 876.9 m2

Correlation Time = 110.1 s

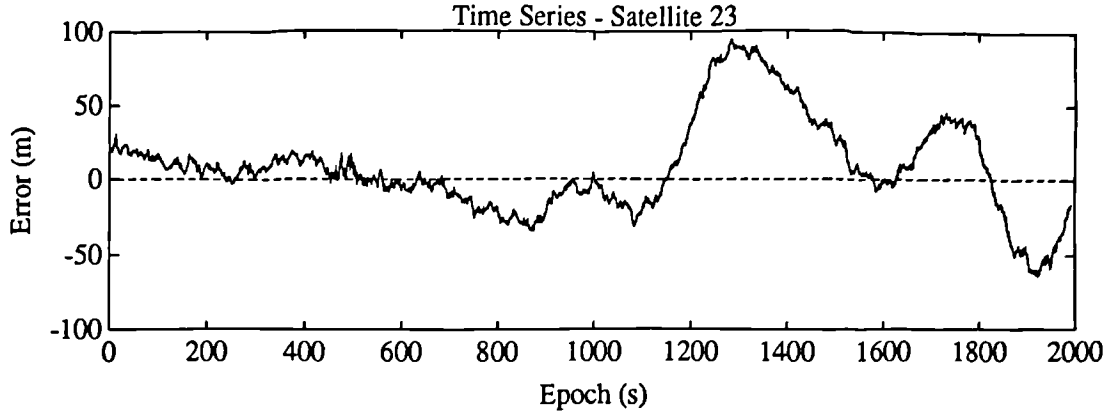


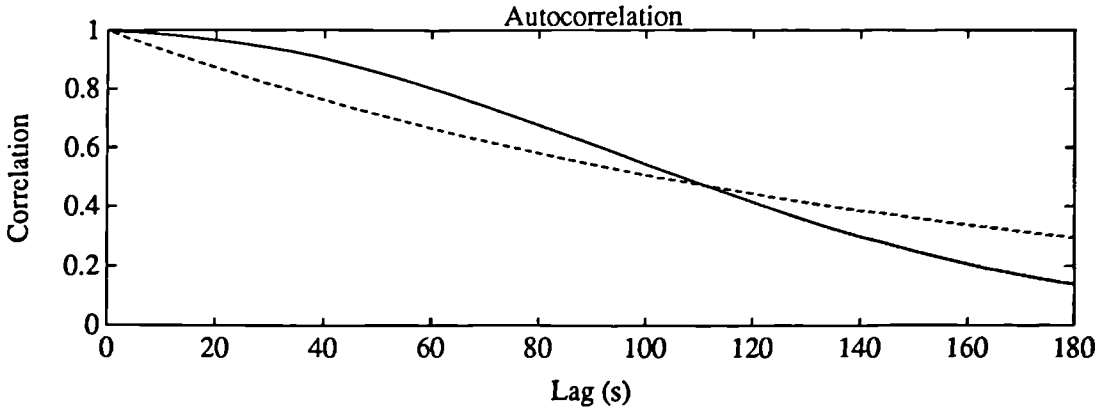


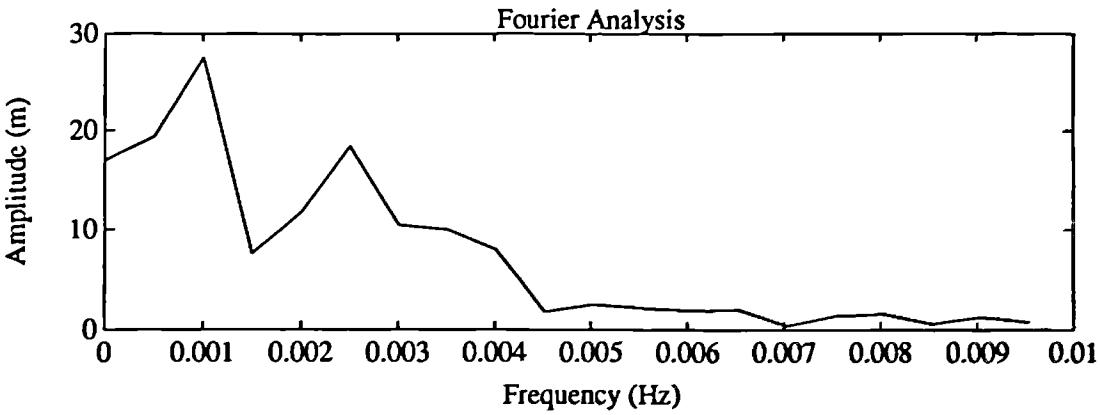


Variance = 756.2 m2

Correlation Time = 145.5 s

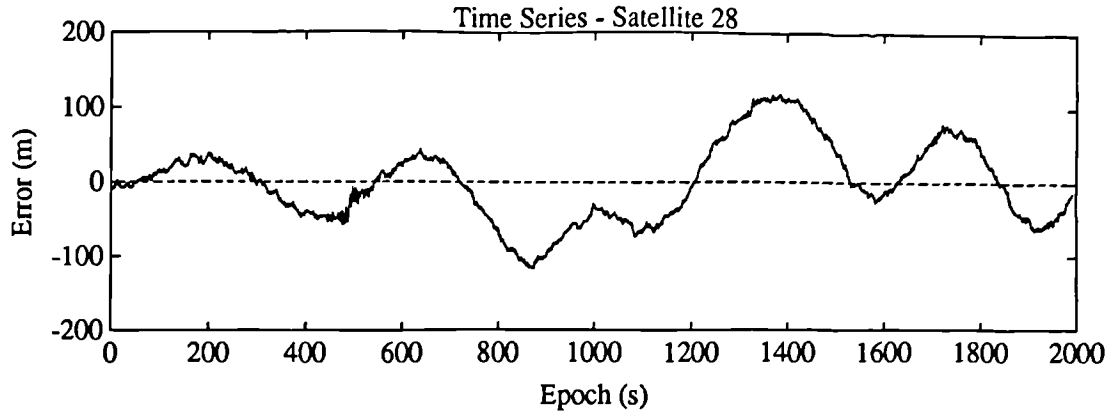


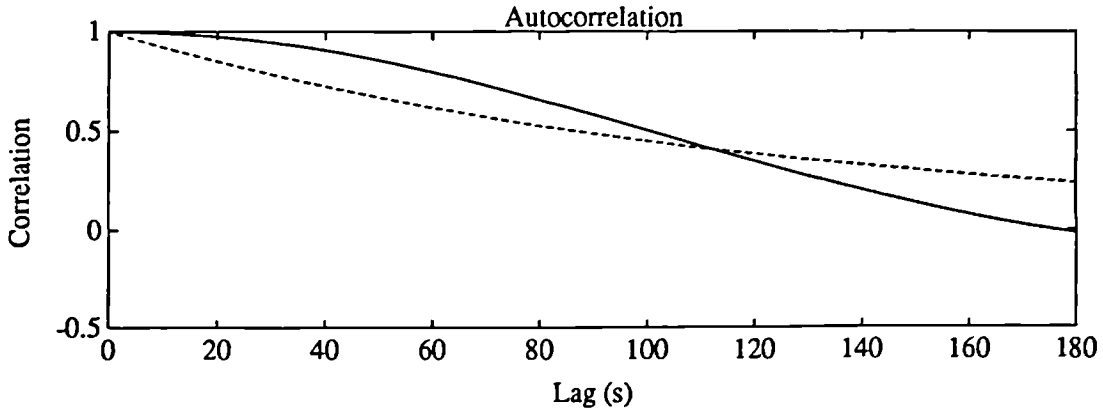


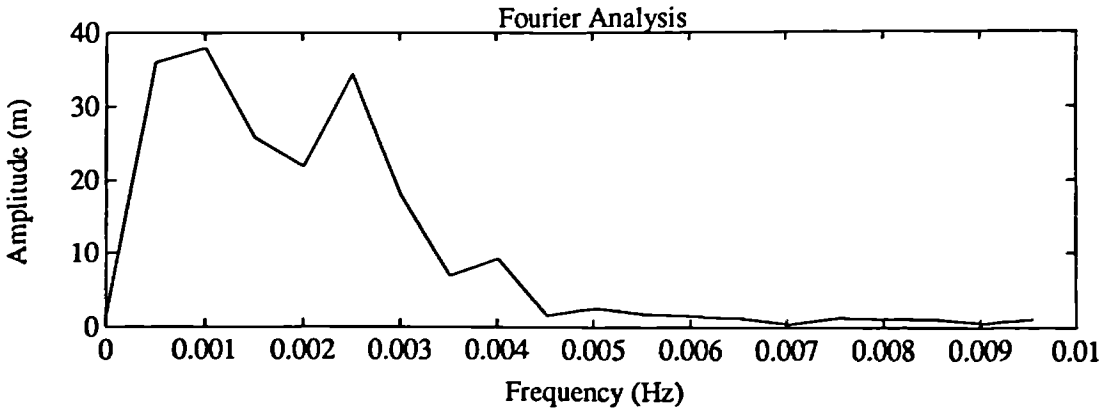


Variance = 1072 m2

Correlation Time = 147.2 s

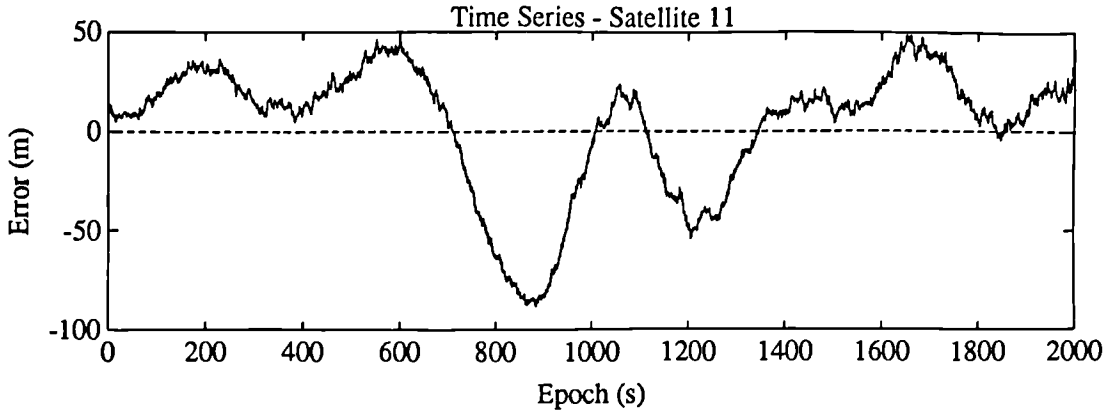


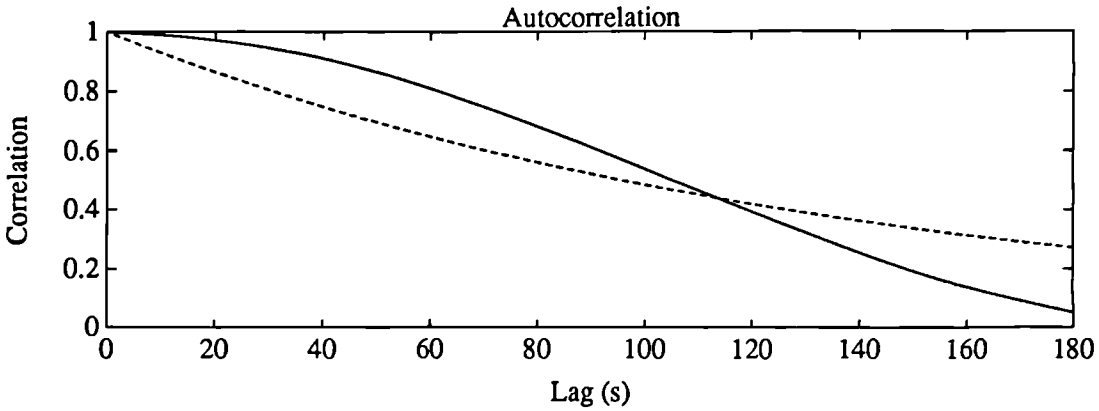


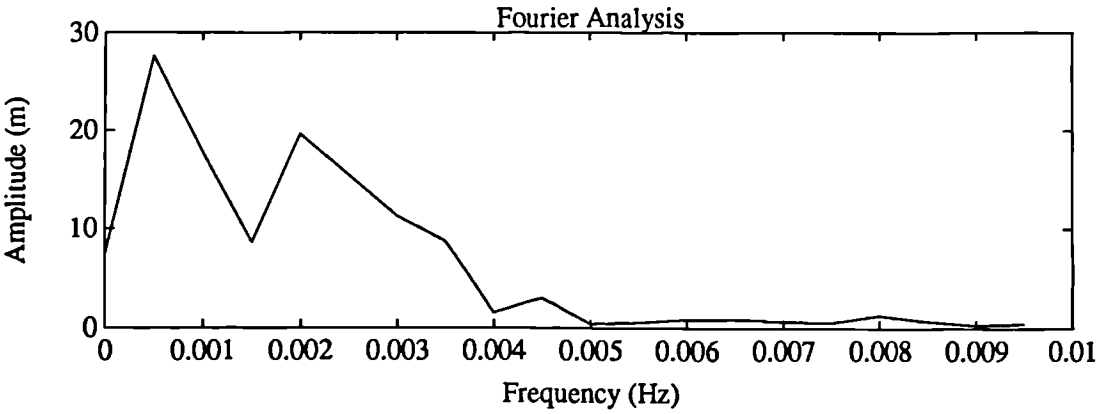


Variance = 2788 m2

Correlation Time = 125.1 s

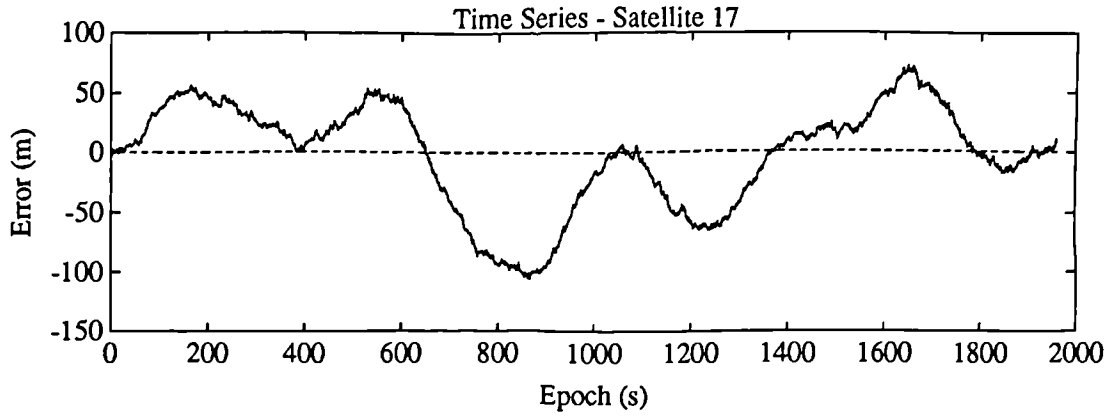


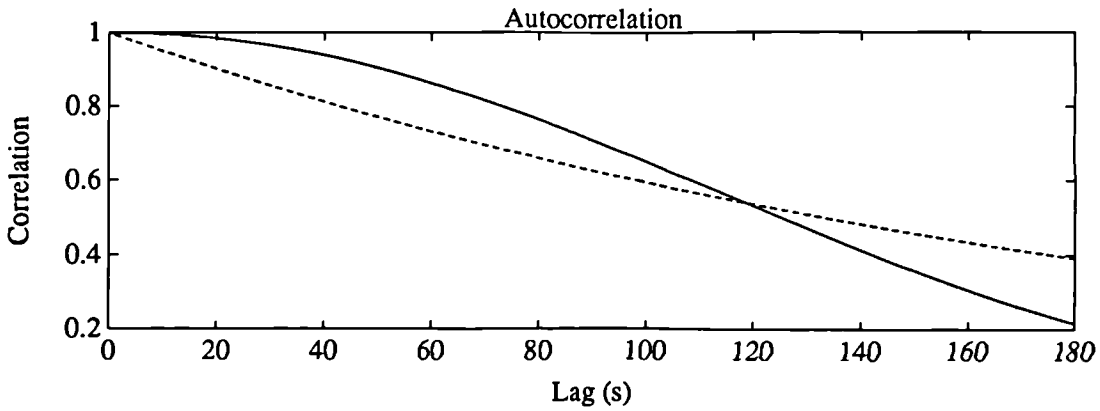


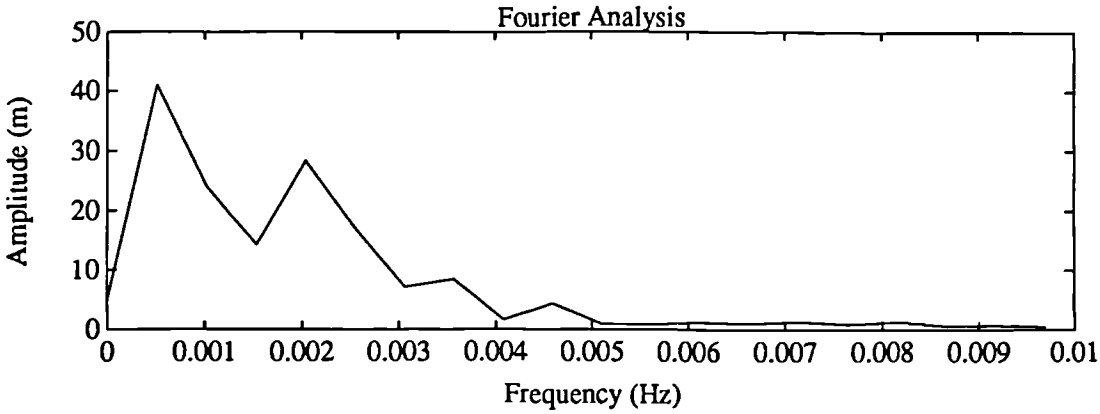


Variance = 1020 m2

Correlation Time = 138 s

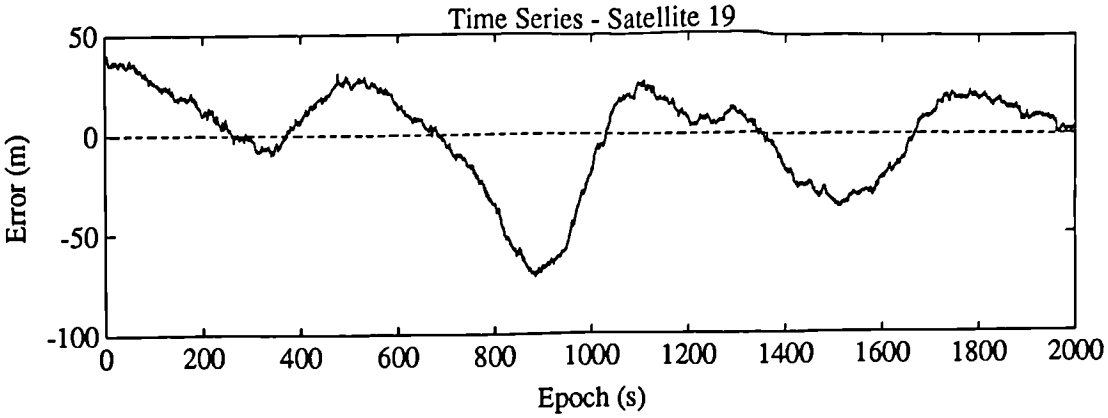


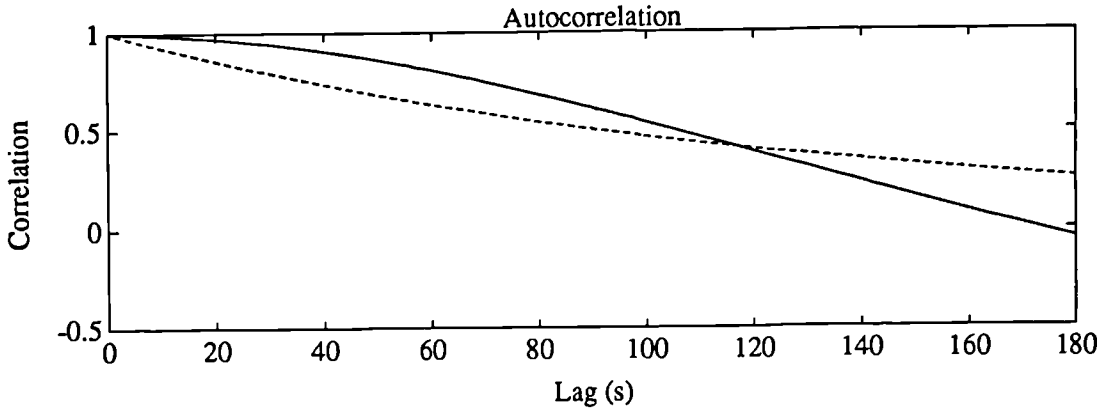


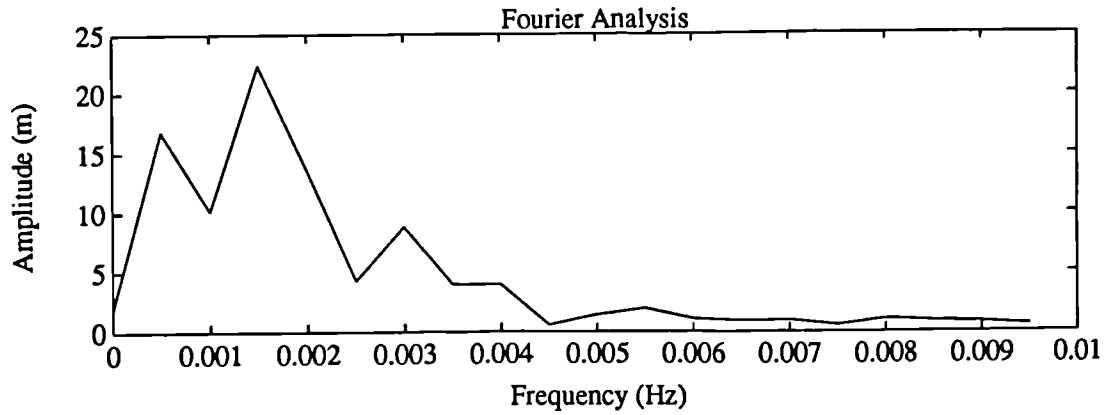


Variance = 1865 m2

Correlation Time = 192 s

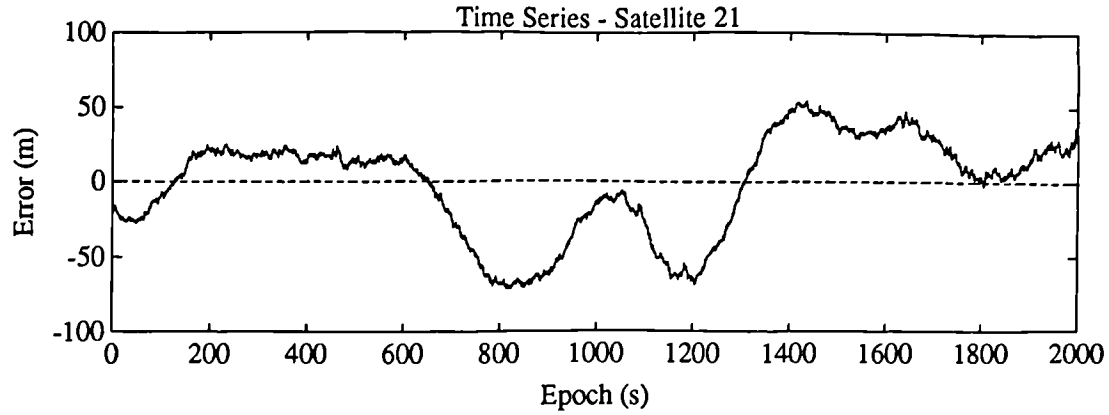


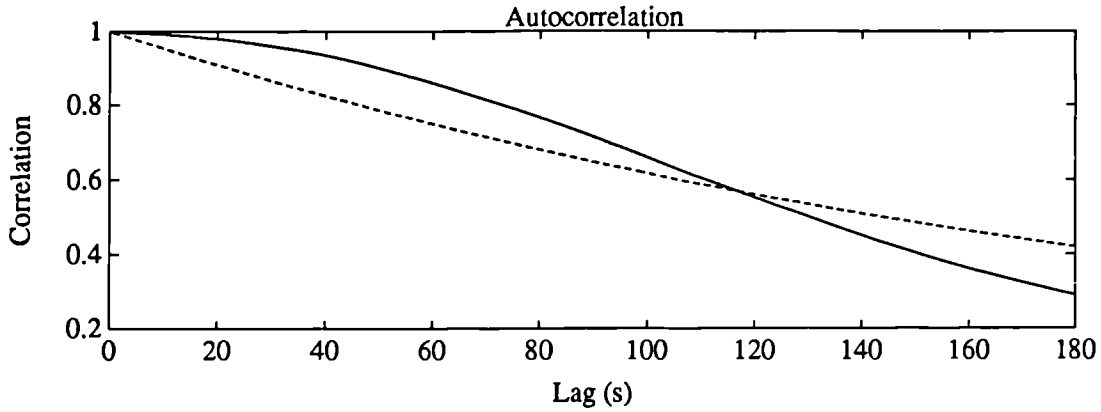


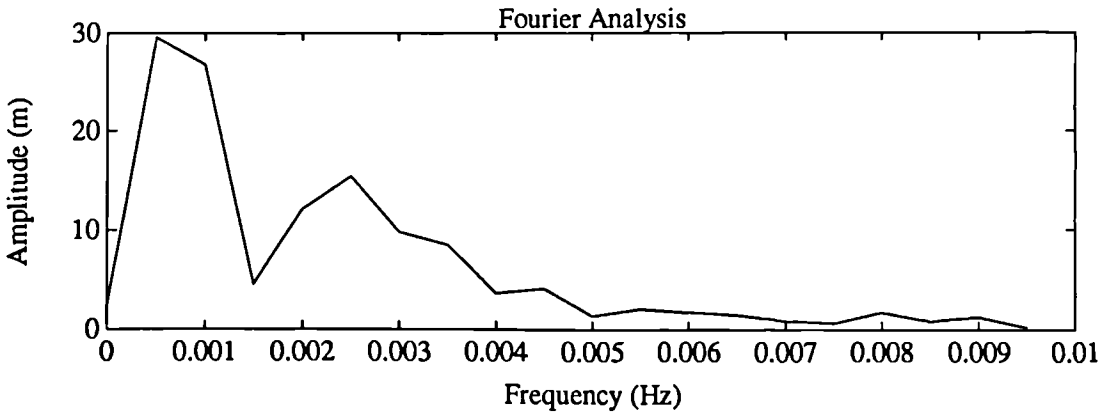


Variance = 610.7 m2

Correlation Time = 131.4 s

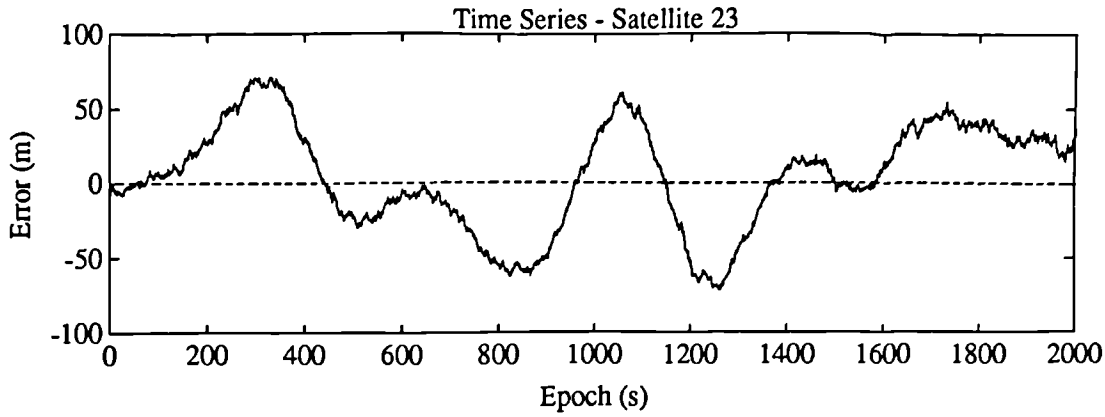


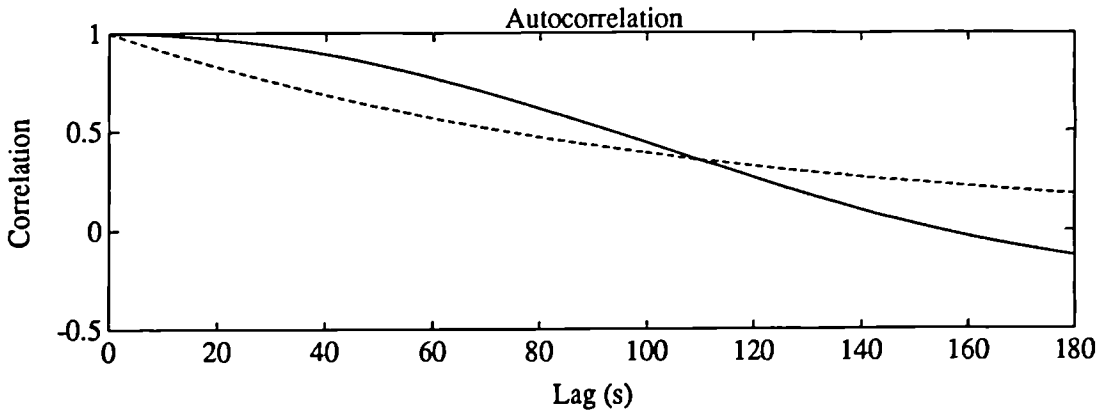


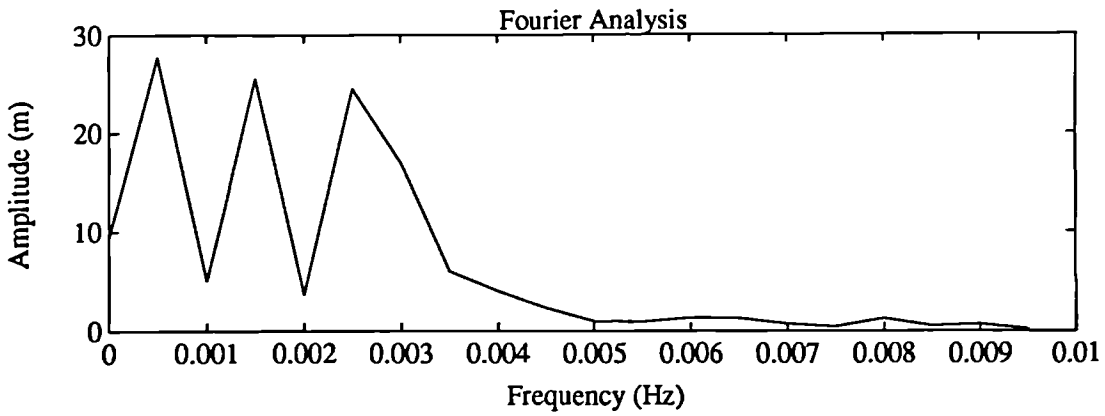


Variance = 1118 m2

Correlation Time = 207.1 s

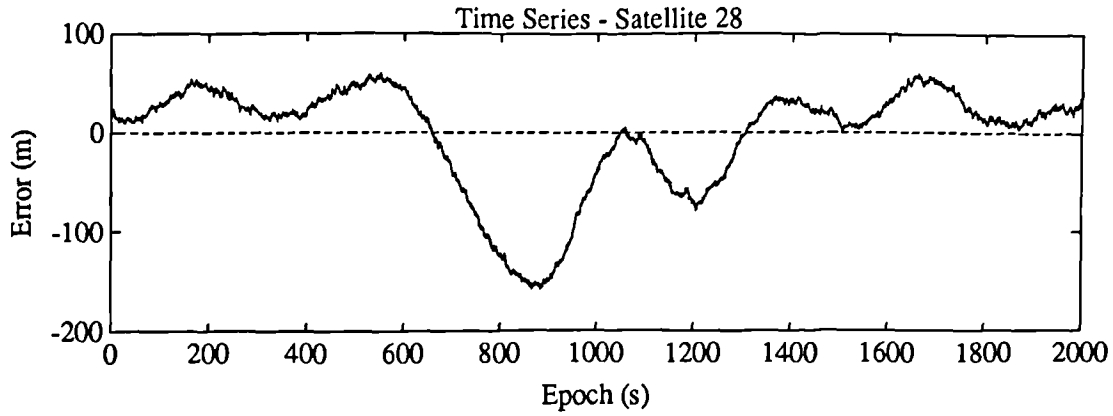


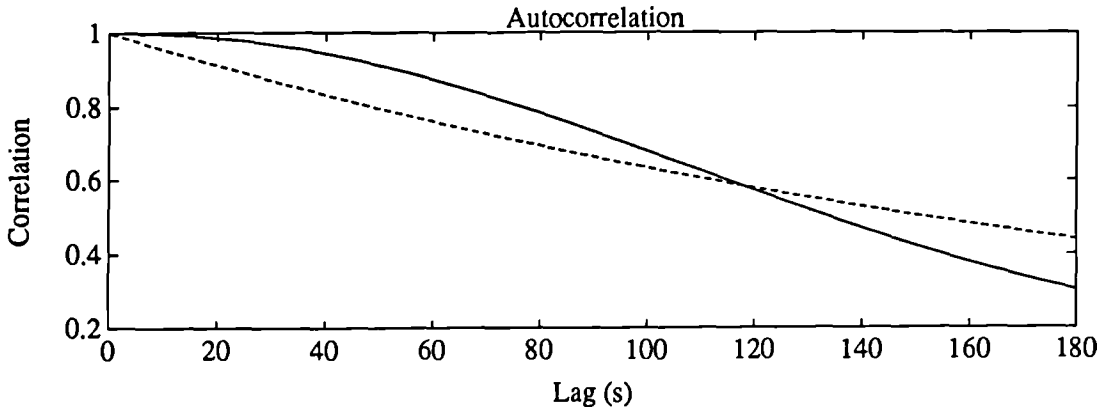


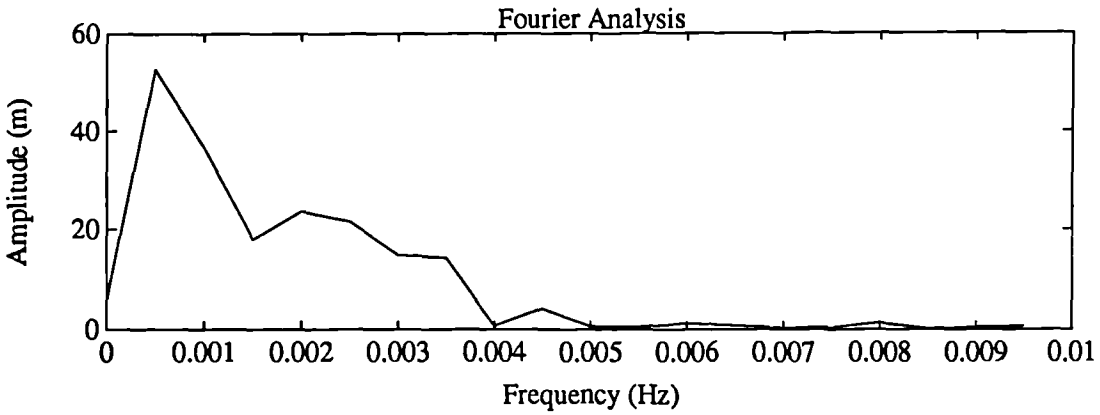


Variance = 1242 m2

Correlation Time = 105.6 s







Variance = 2973 m2

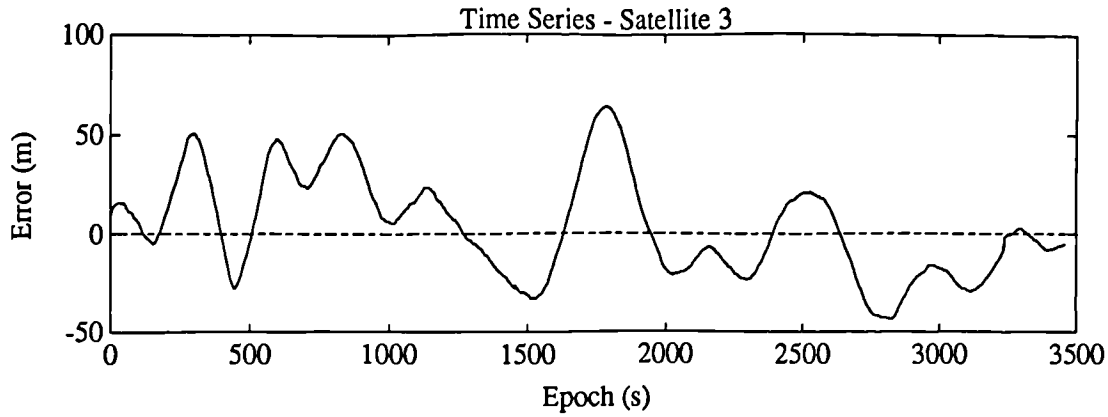
Correlation Time = 218.6 s

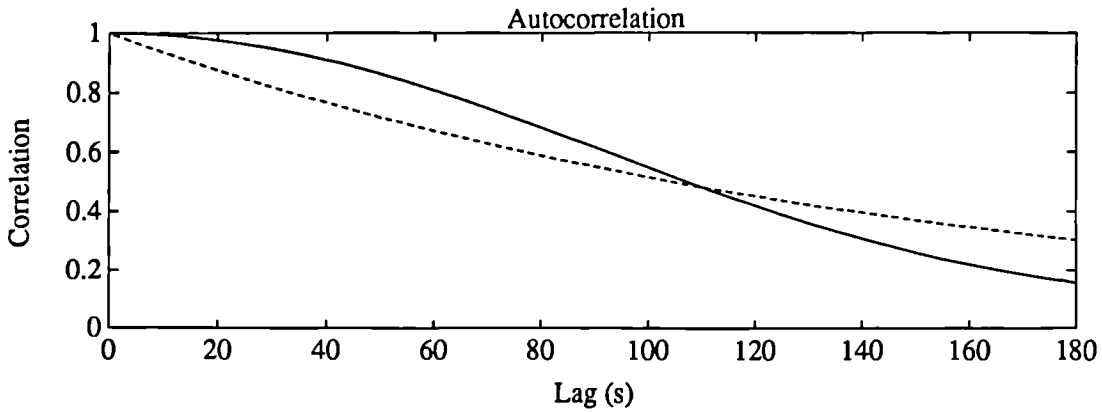
APPENDIX F

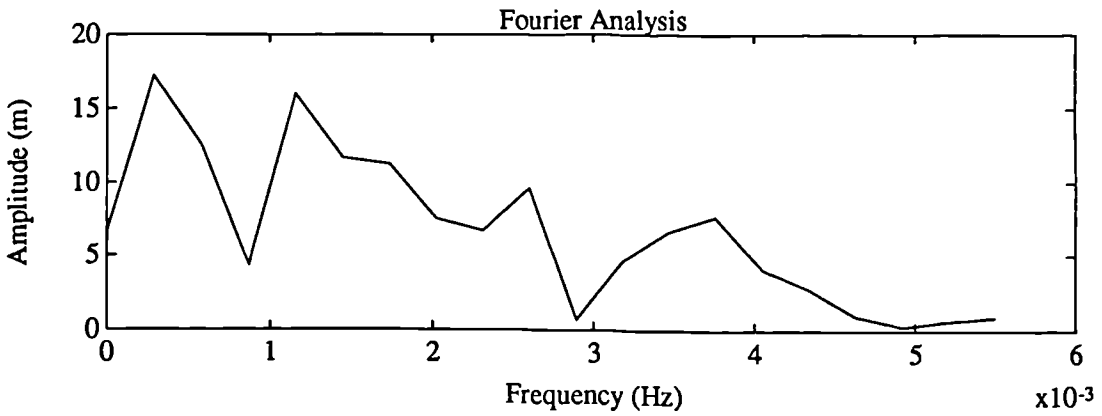
ASHTECH DECEMBER '92 OUTPUT

Time series, autocorrelation and Fourier analysis for data files:

PP923421

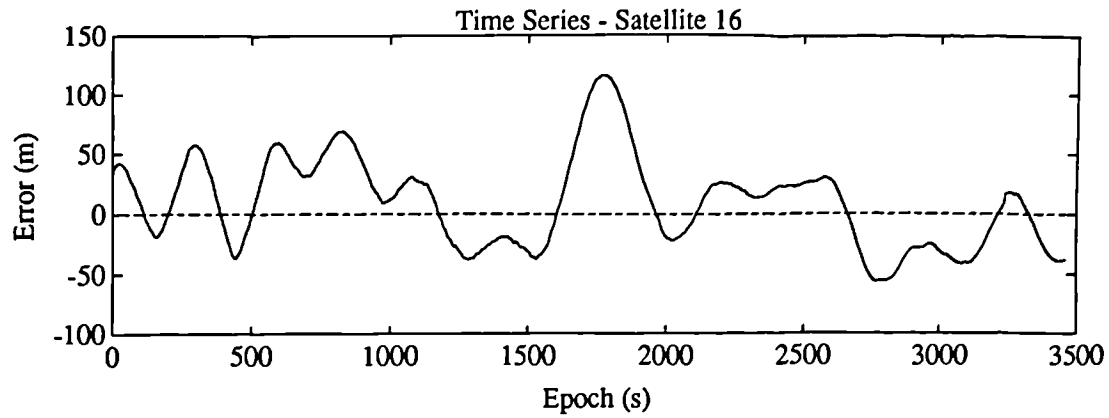


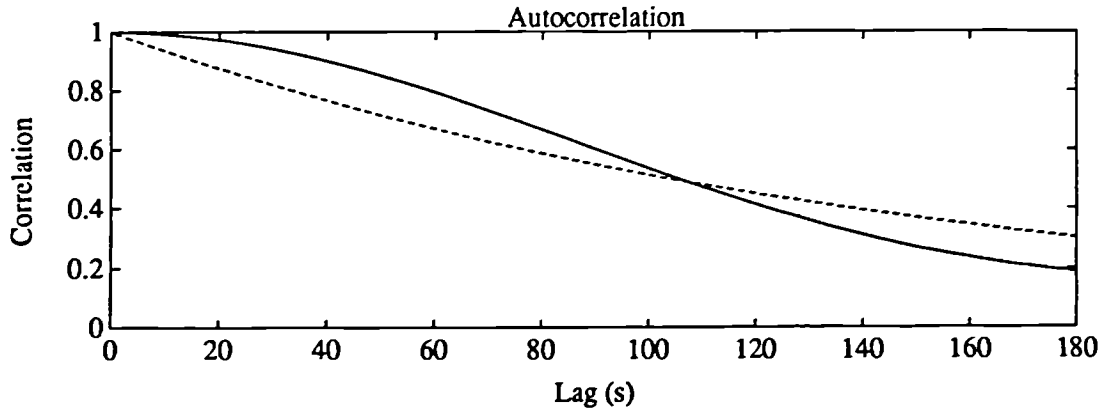


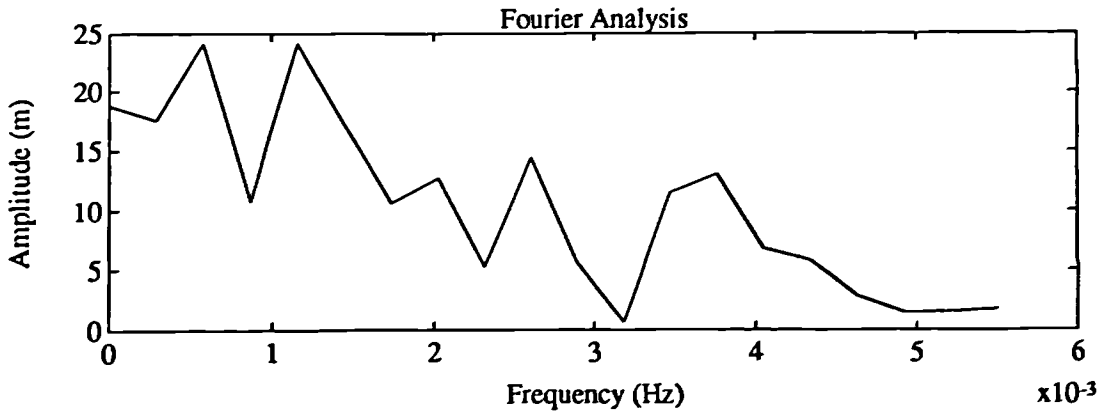


Variance = 681.5 m2

Correlation Time = 150.6 s

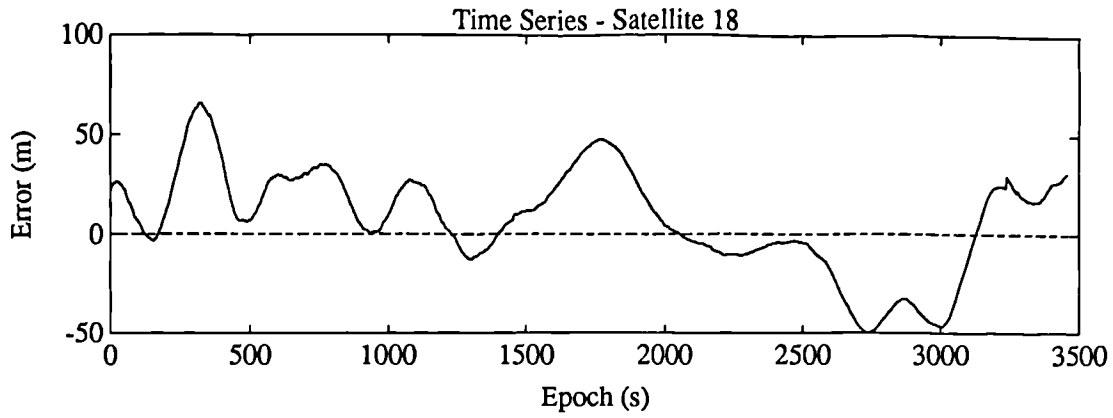


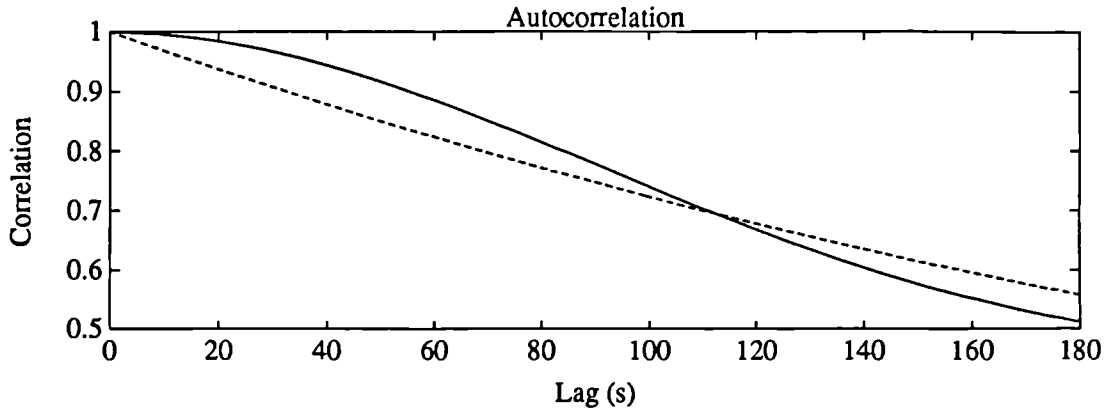


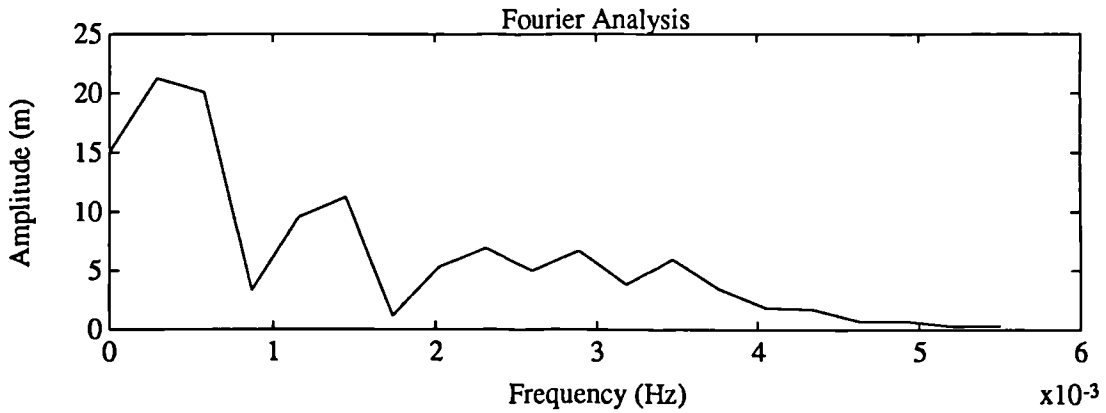


Variance = 1518 m2

Correlation Time = 149.9 s

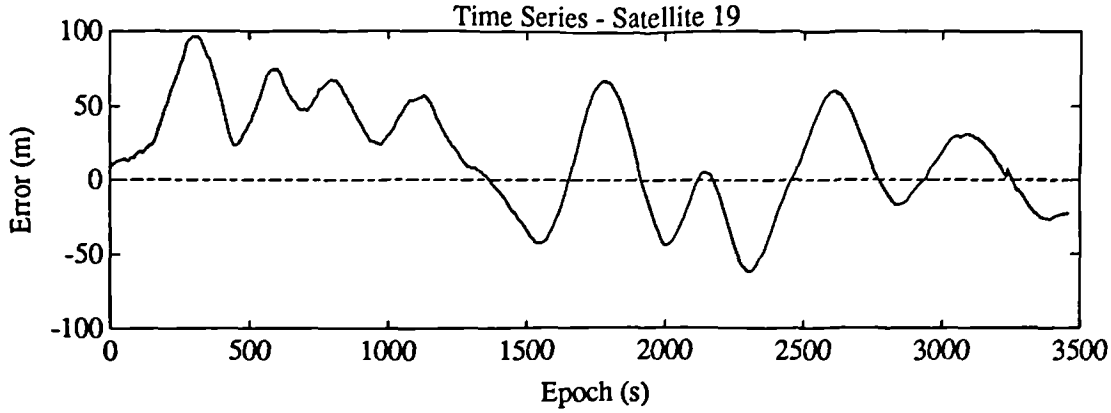


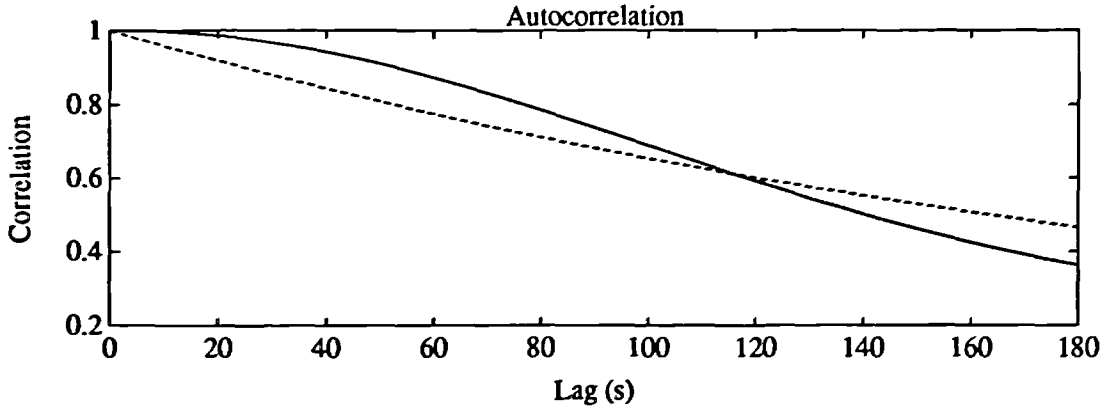


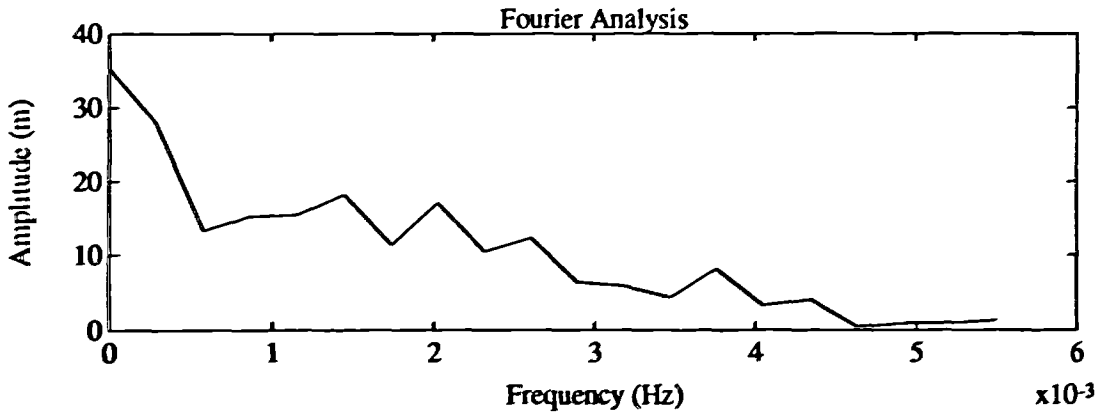


Variance = 708.2 m2

Correlation Time = 308.5 s

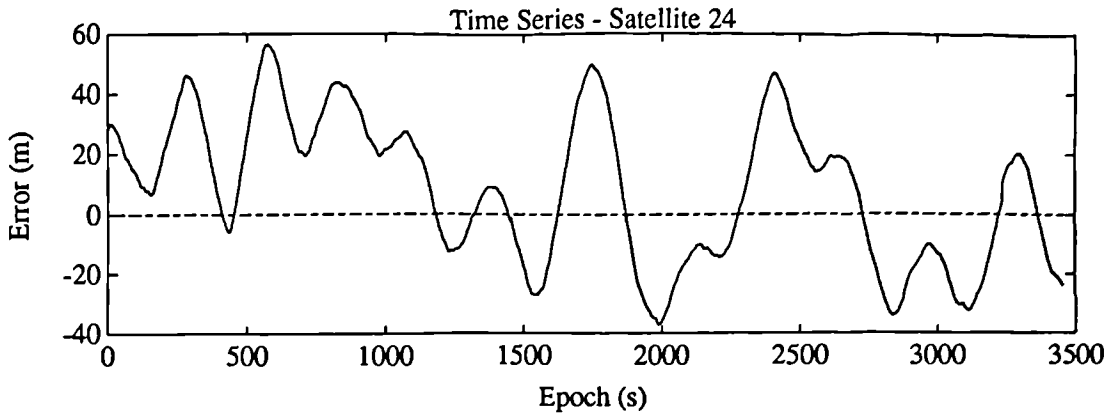


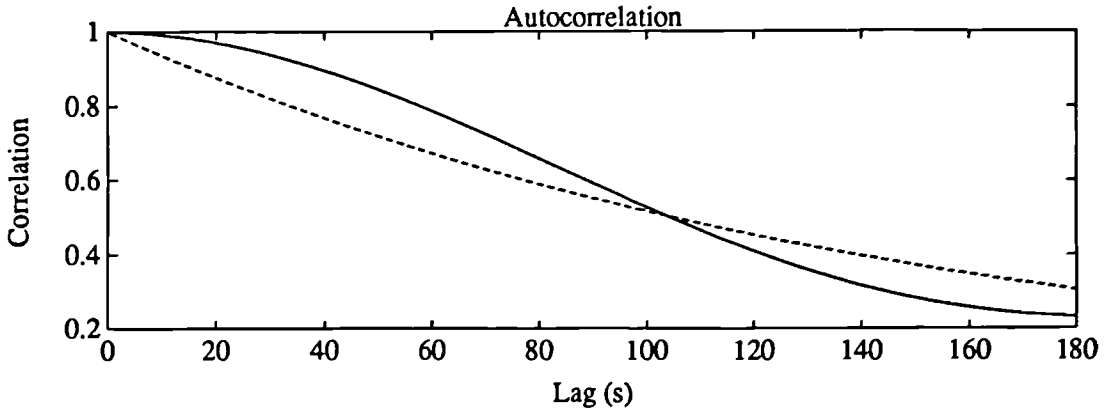


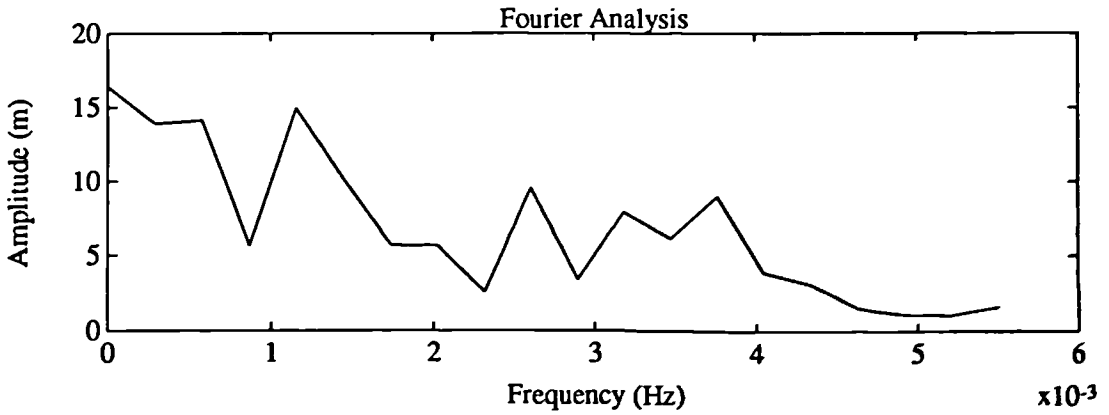


Variance = 1629 m2

Correlation Time = 235.1 s

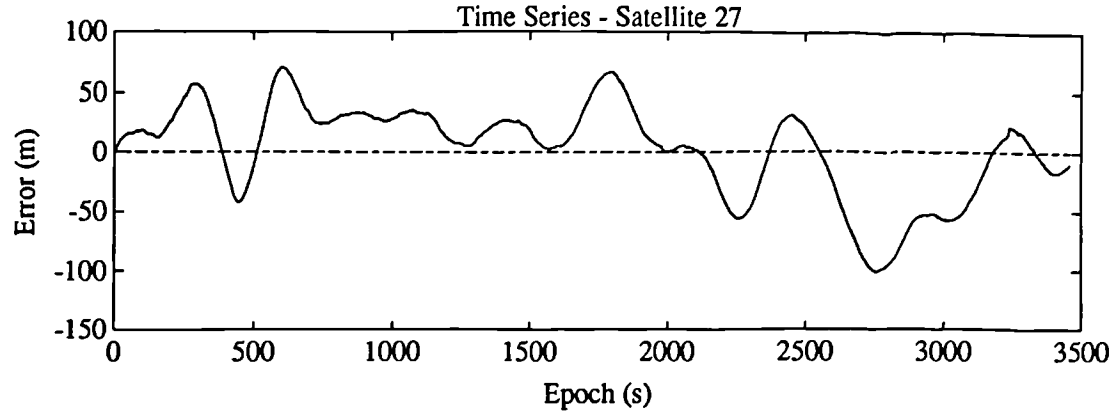


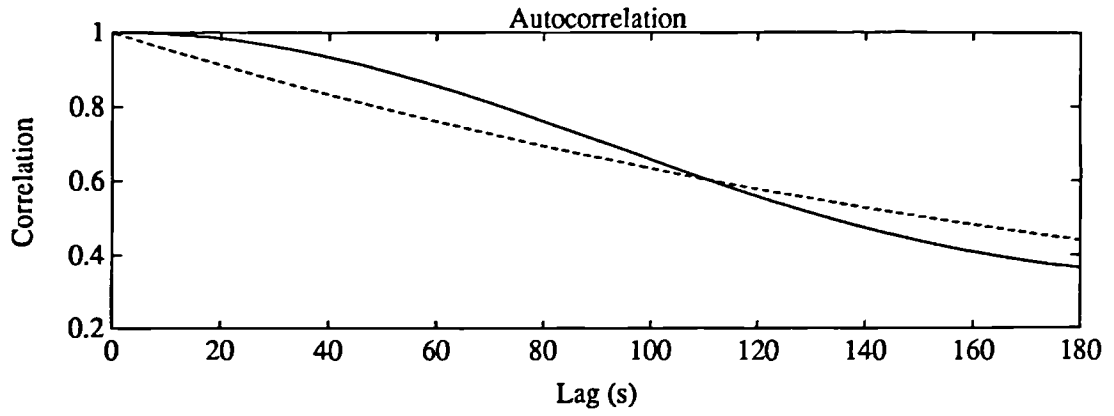


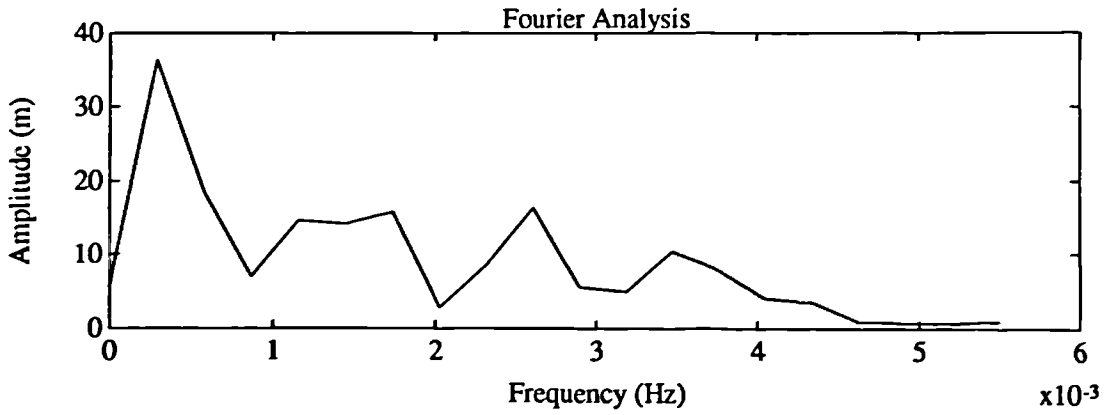


Variance = 644.1 m2

Correlation Time = 151.4 s







Variance = 1498 m2

Correlation Time = 219.4 s

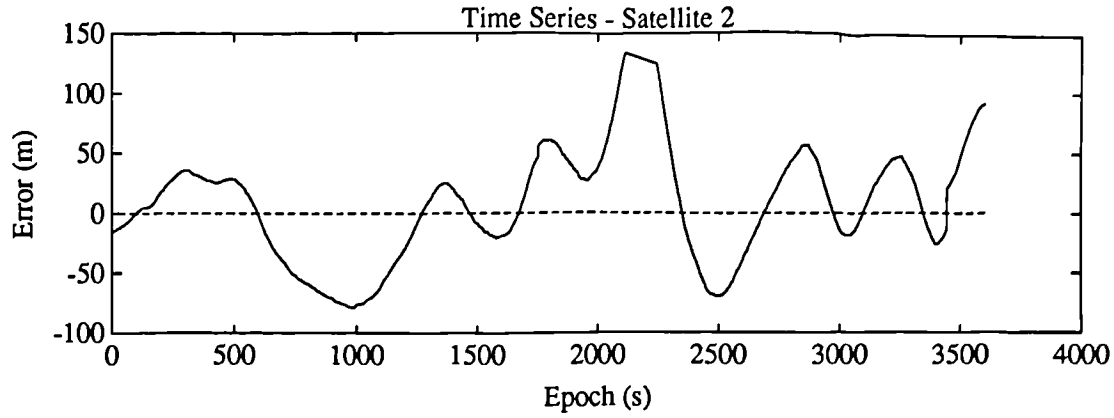
APPENDIX G

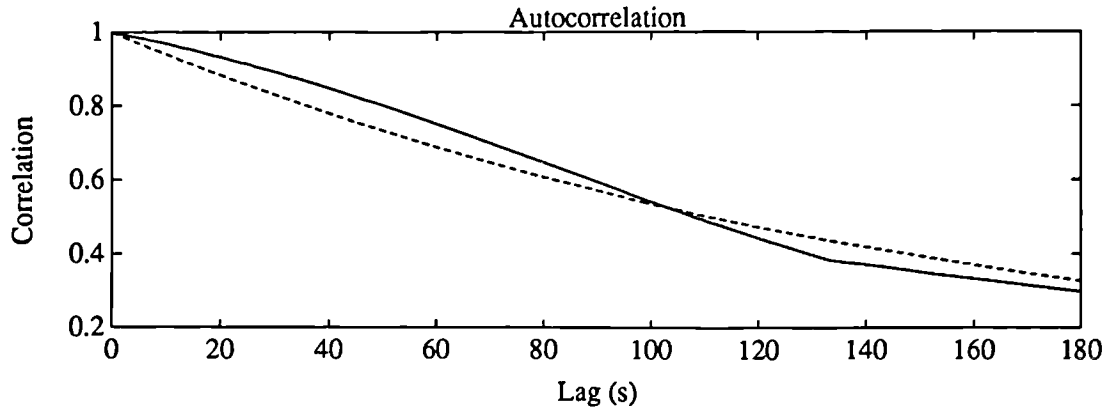
ASHTECH MARCH '93 OUTPUT

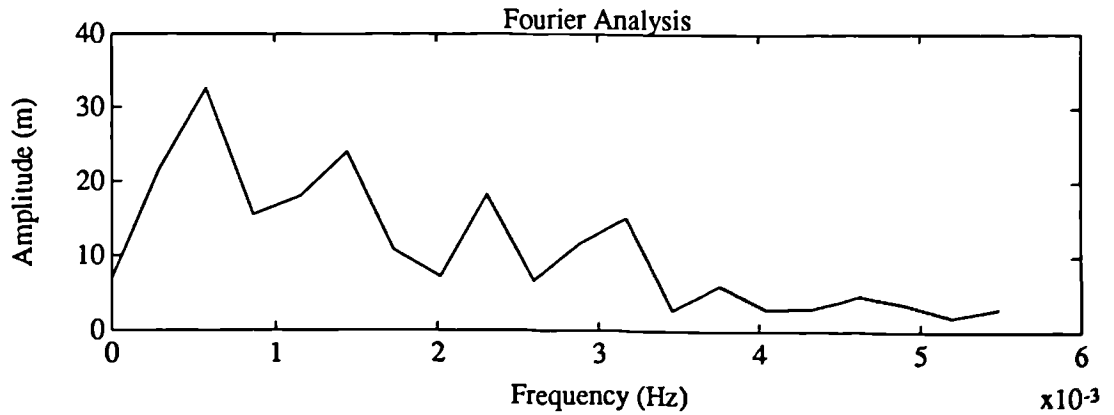
Time series, autocorrelation and Fourier analysis for data files:

PP930611

WH930611

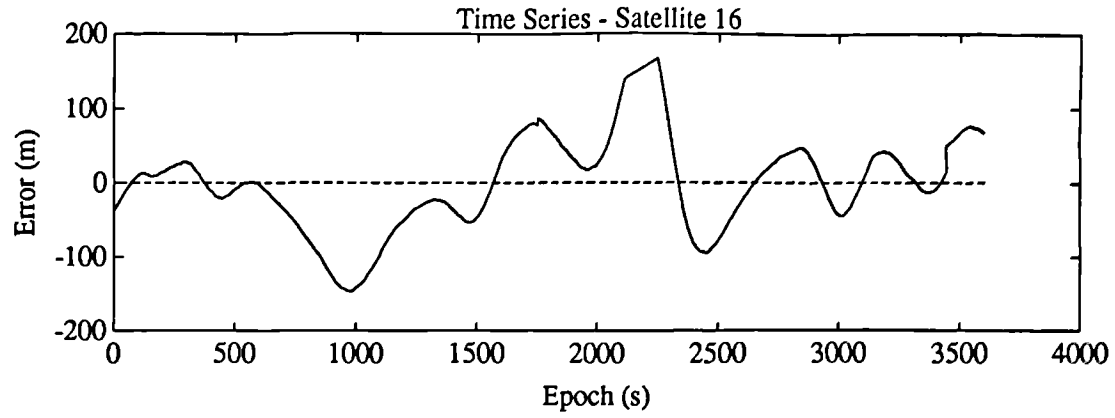


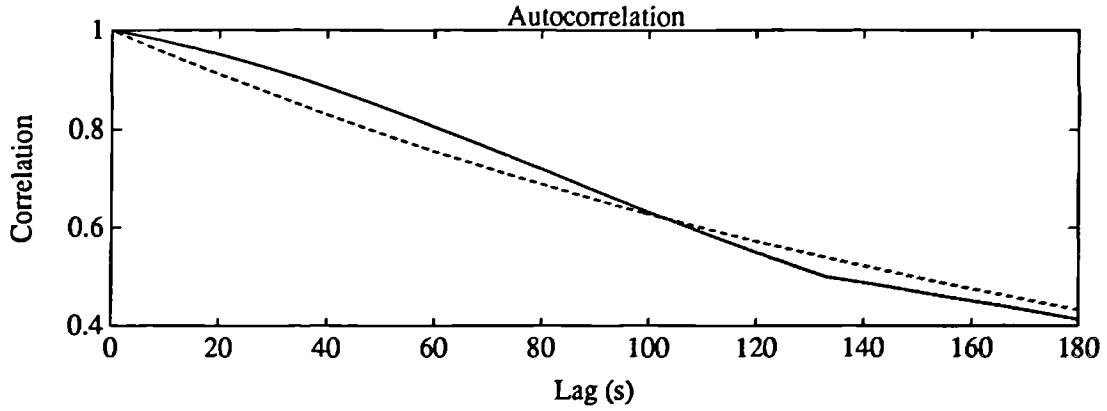


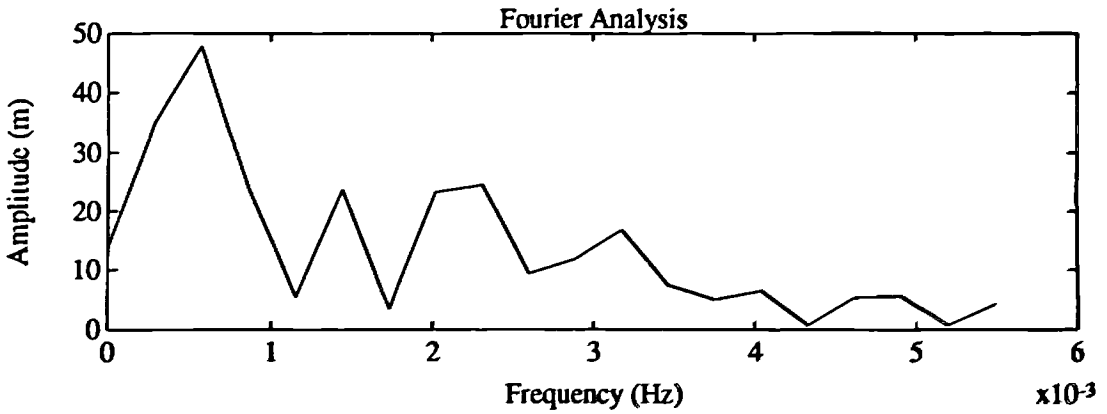


Variance = 1892 m2

Correlation Time = 160.7 s

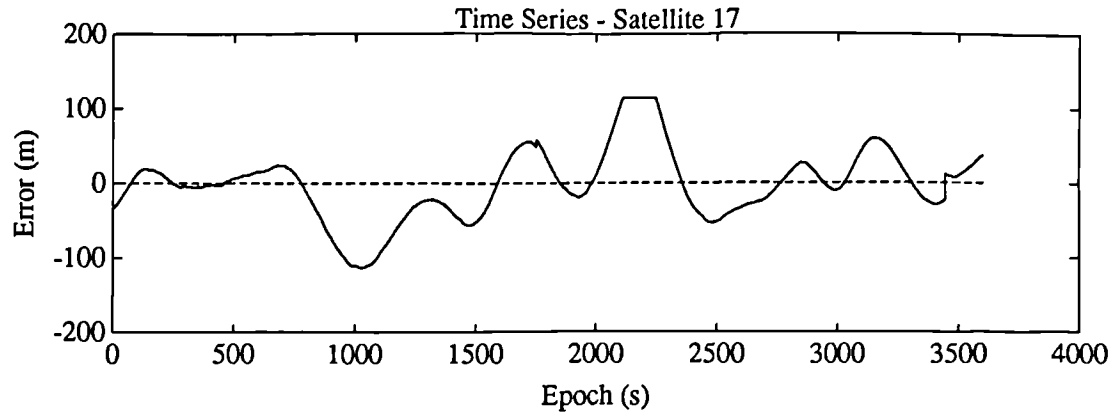


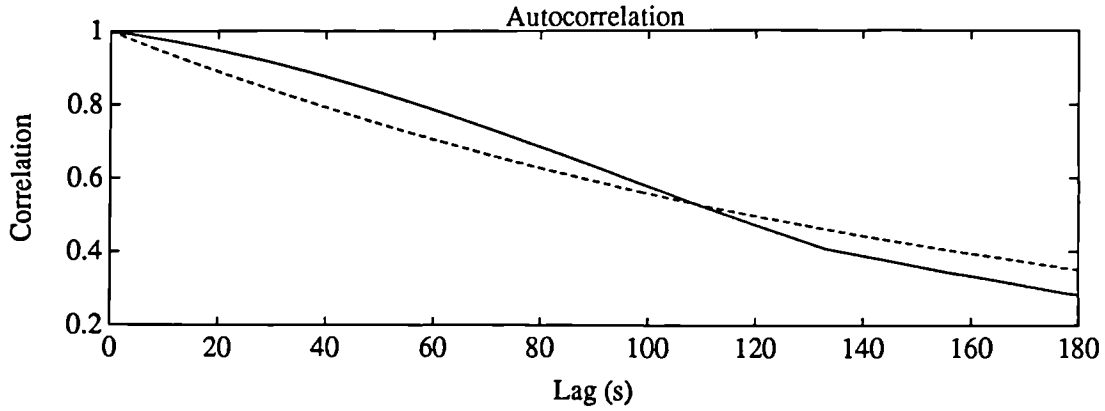


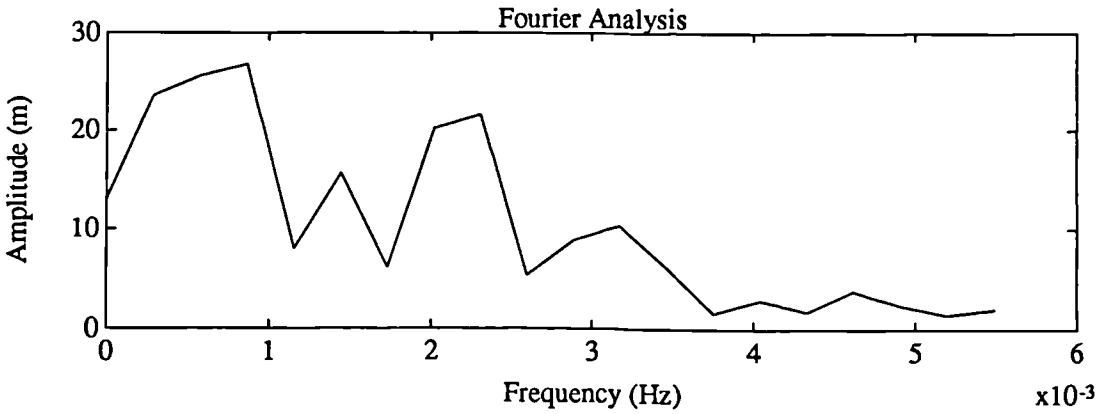


Variance = 3357 m2

Correlation Time = 214.9 s

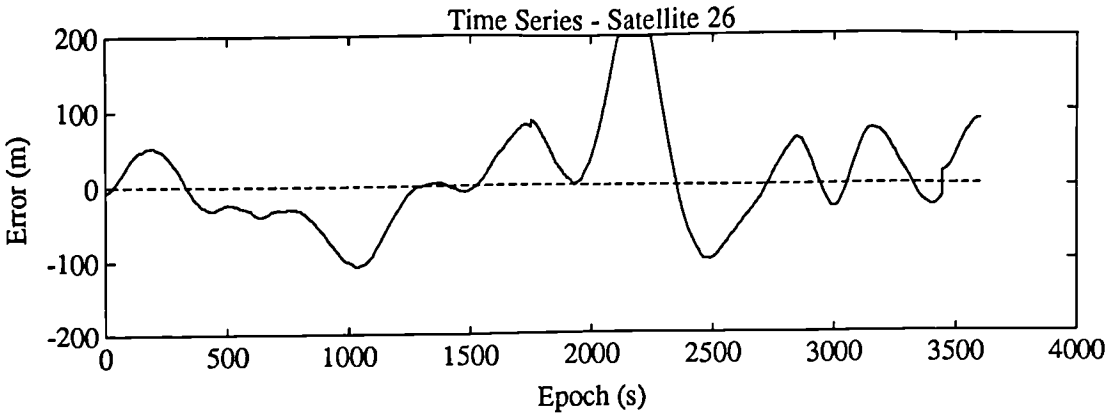


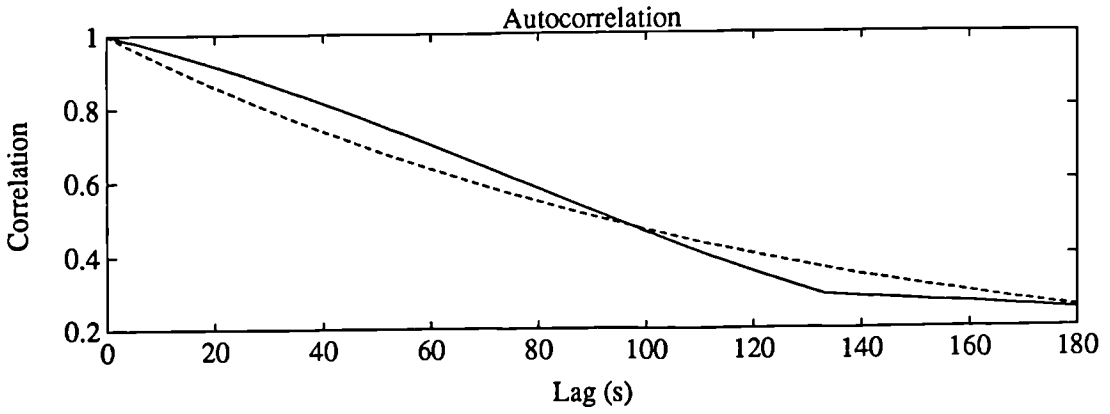


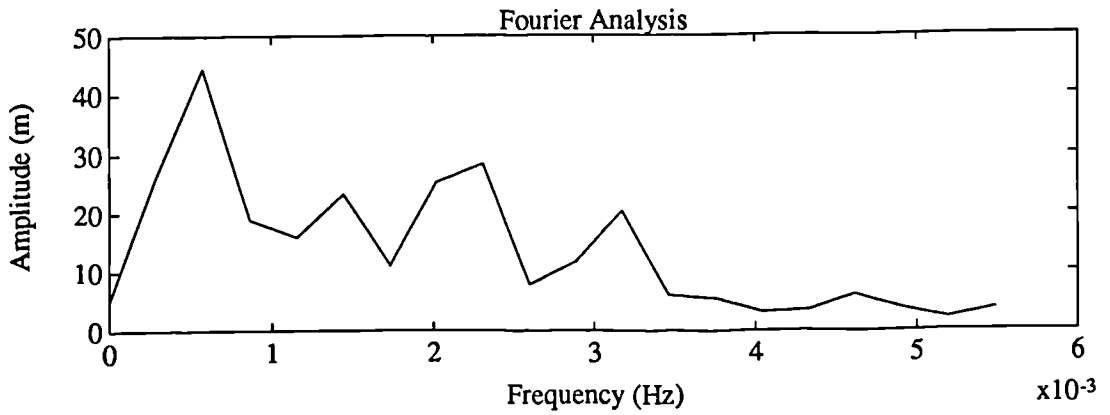


Variance = 1782 m2

Correlation Time = 171.6 s

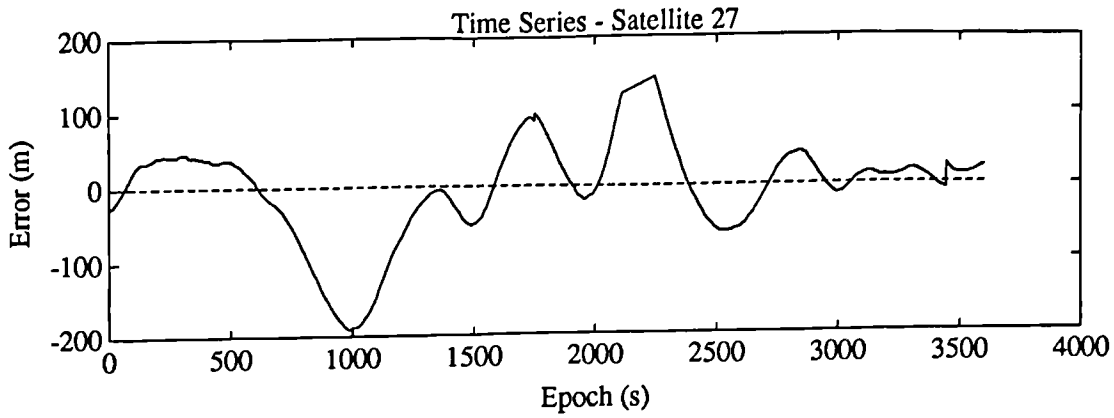


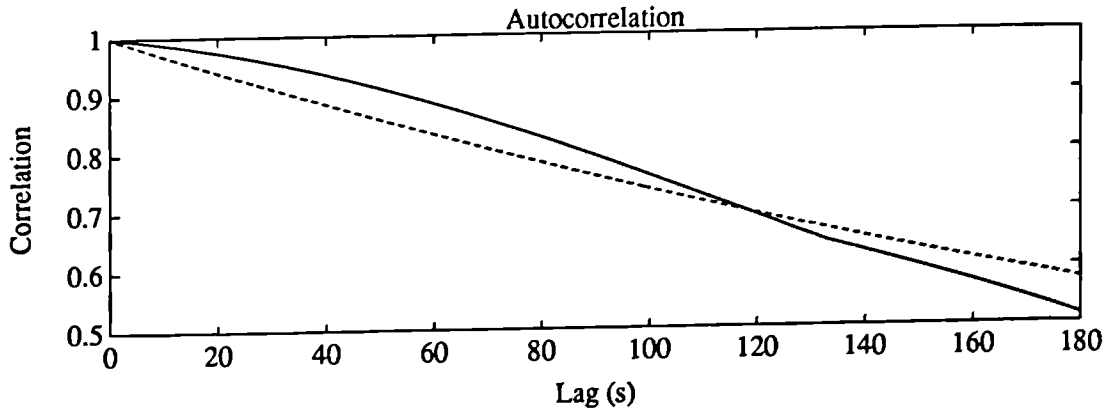


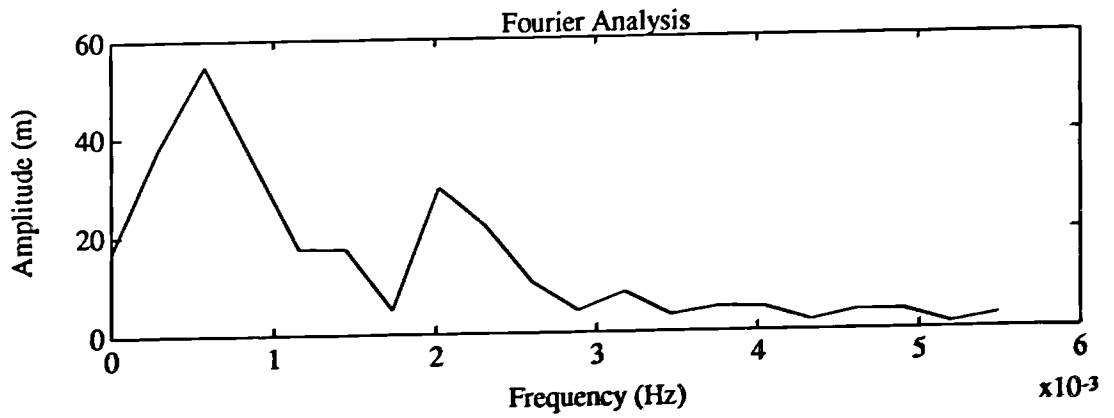


Variance = 3119 m2

Correlation Time = 132 s

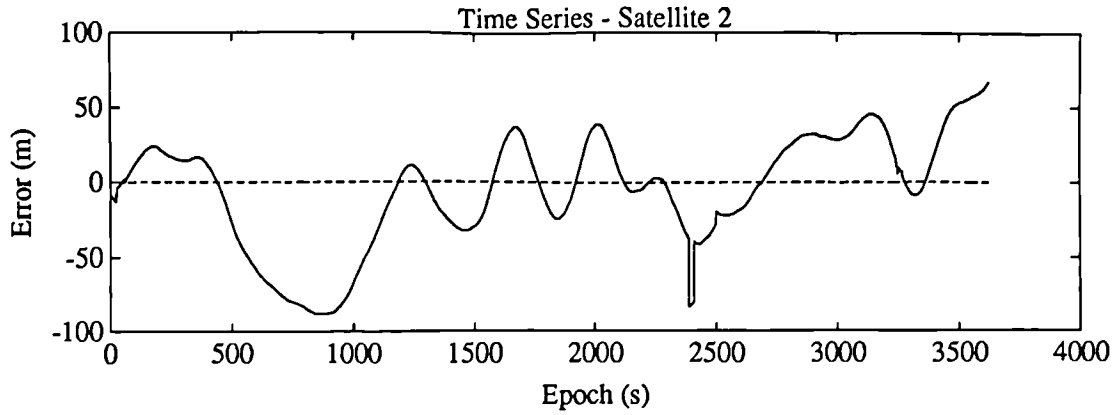


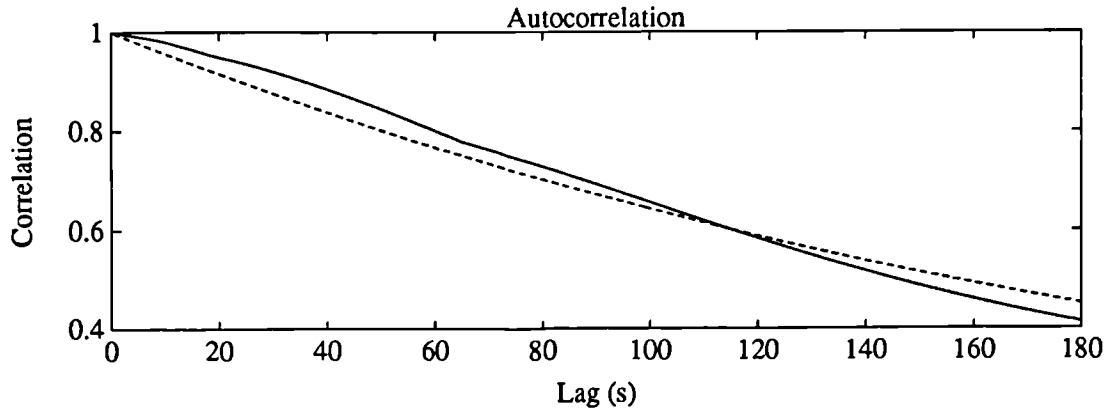


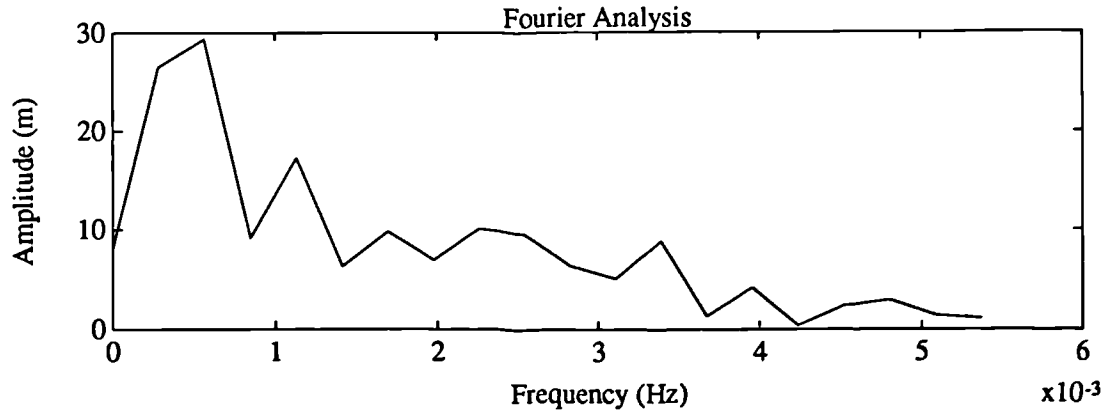


Variance = 4067 m2

Correlation Time = 325.5 s

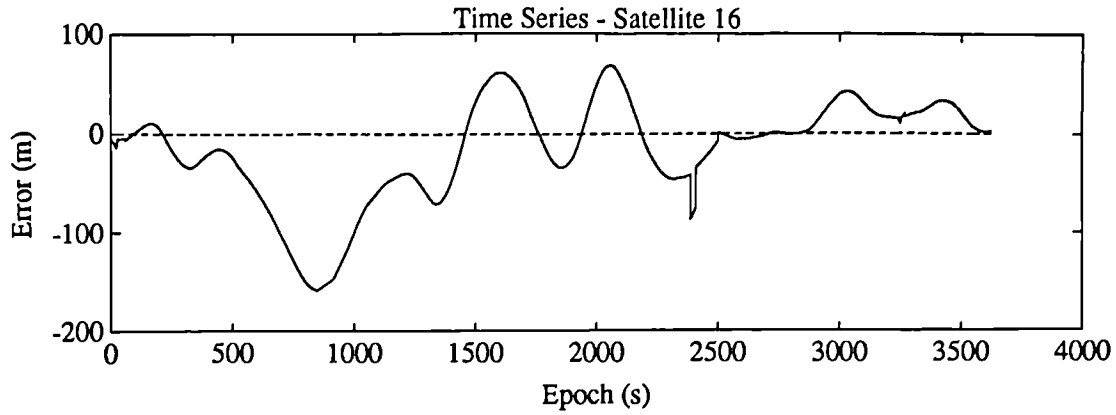


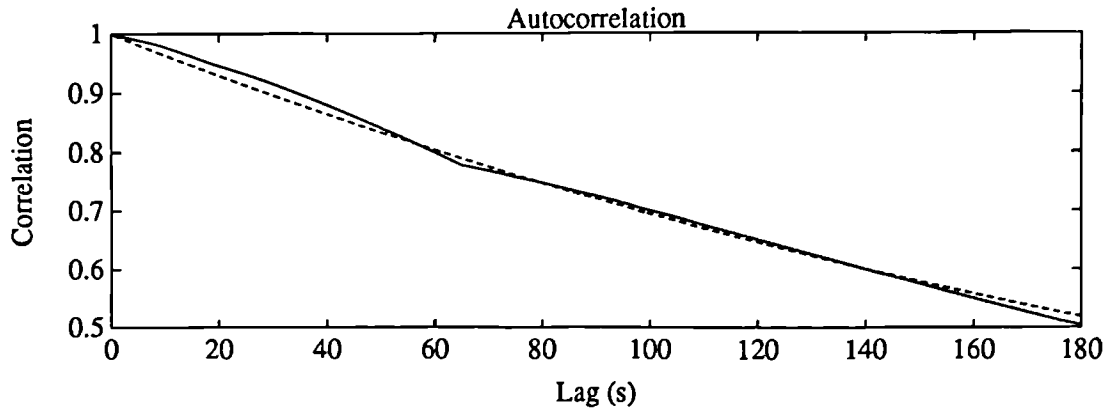


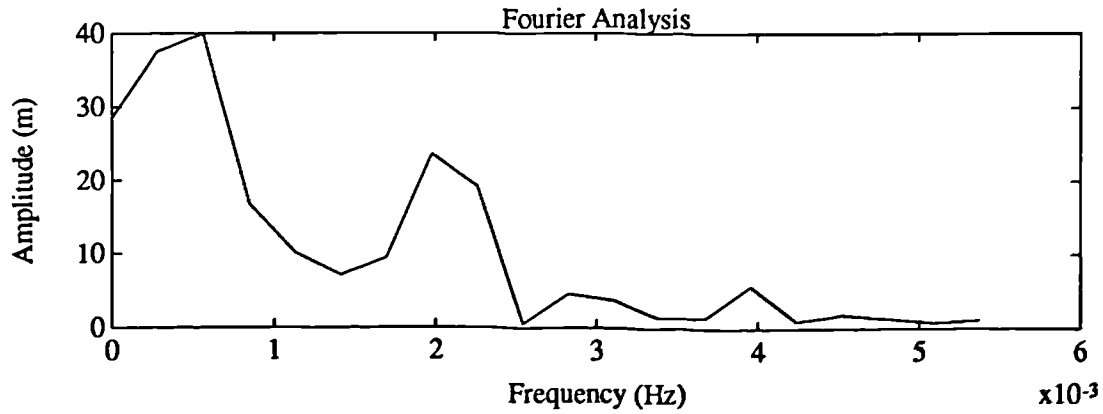


Variance = 1292 m2

Correlation Time = 225.9 s

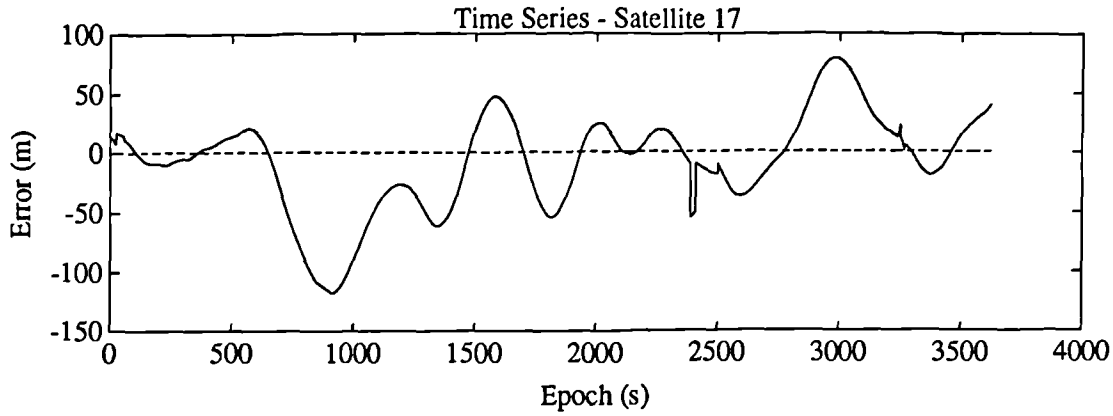


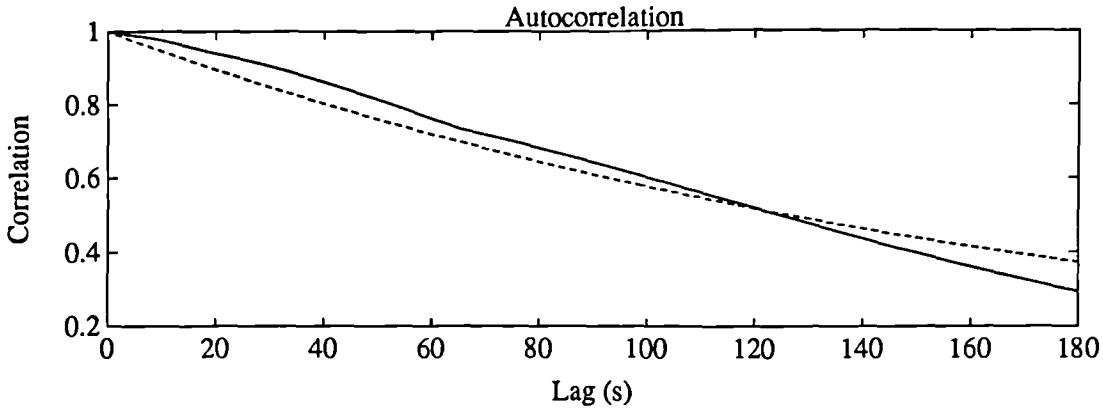


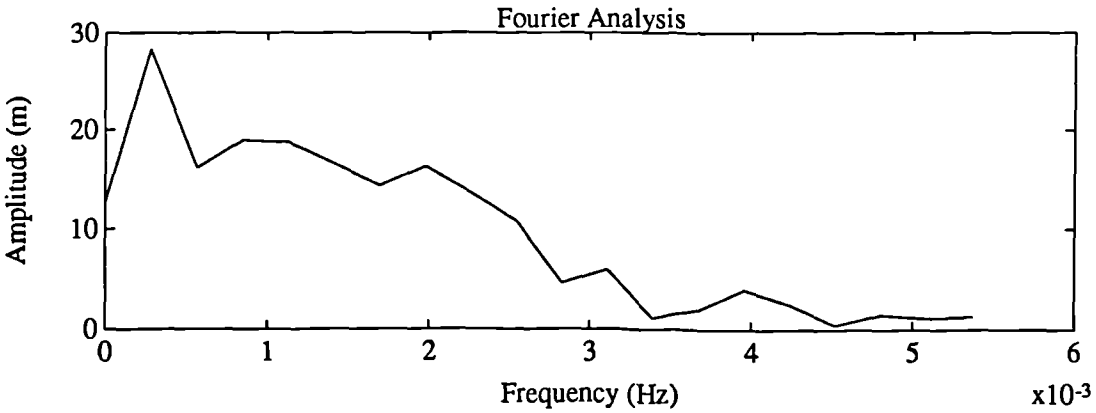


Variance = 2481 m2

Correlation Time = 274.3 s

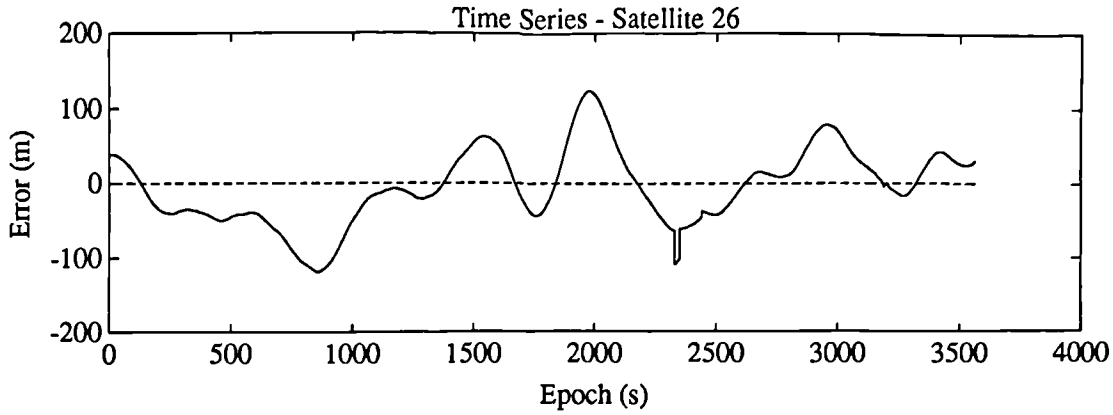


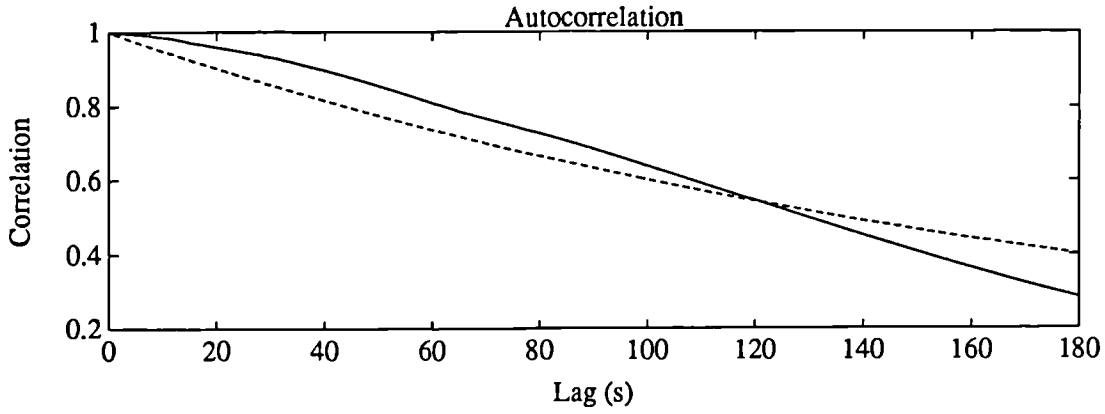


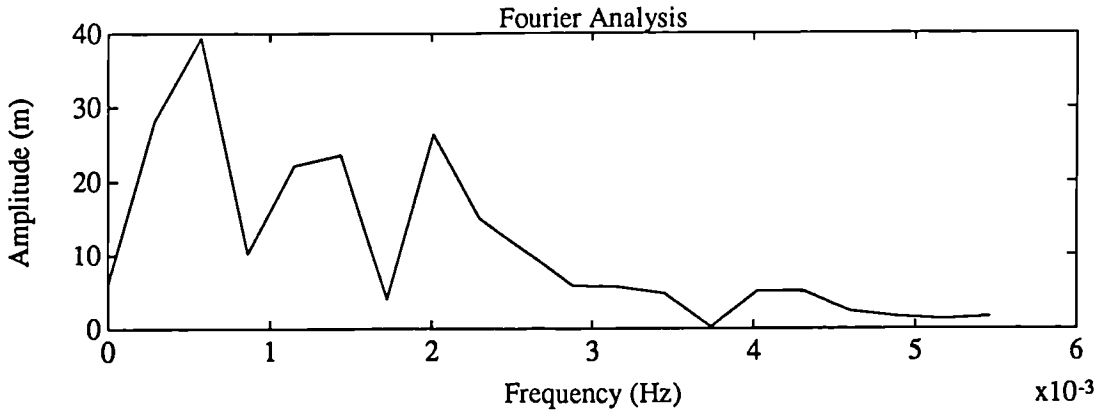


Variance = 1506 m2

Correlation Time = 182.2 s

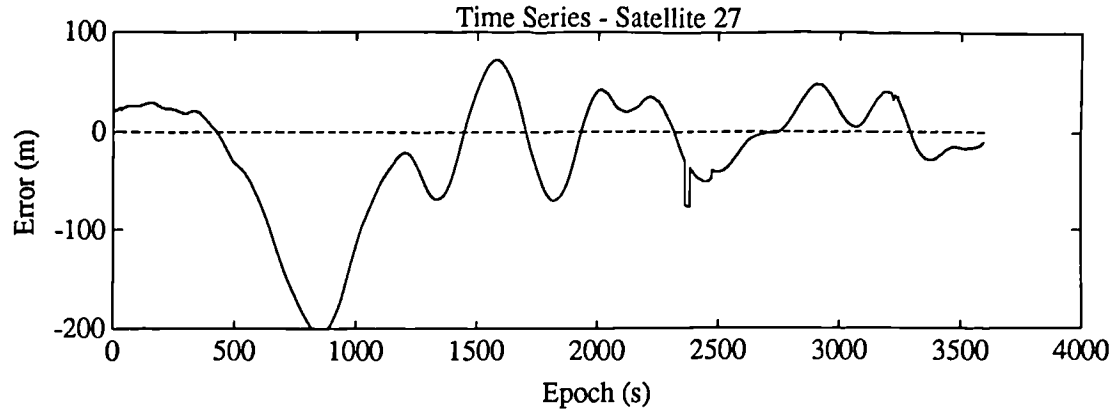


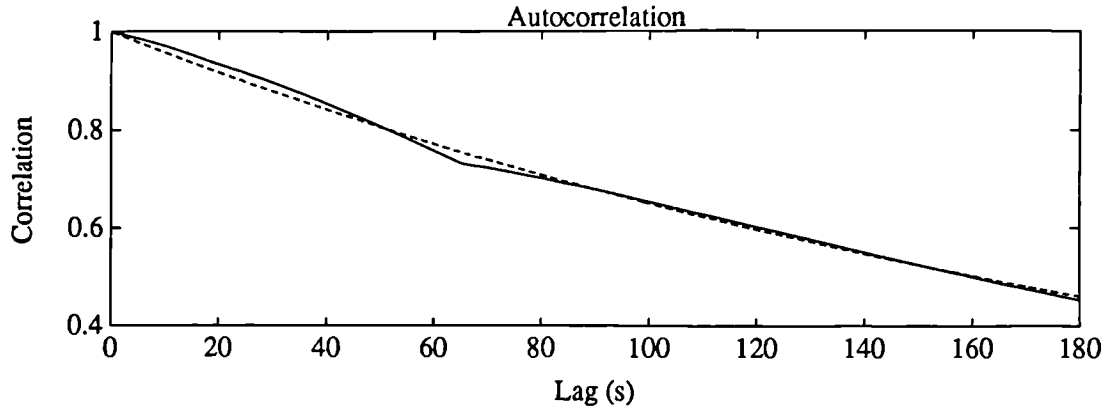


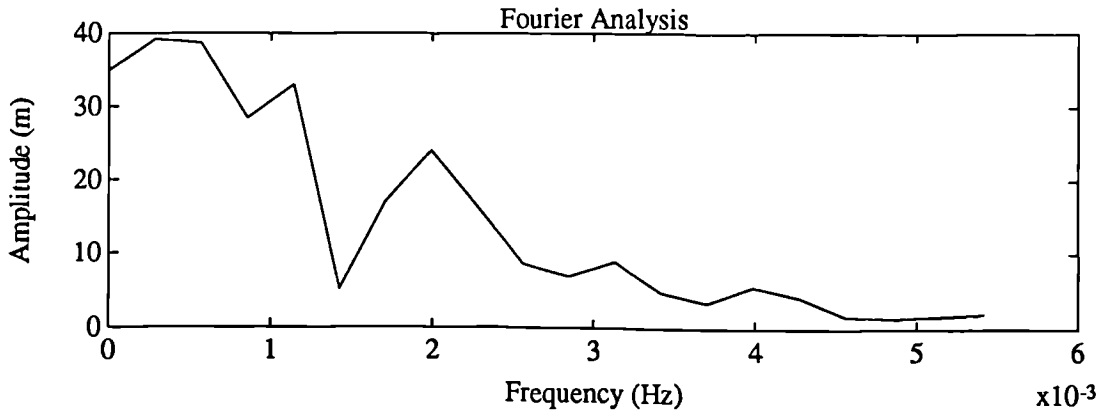


Variance = 2360 m2

Correlation Time = 196.7 s







Variance = 3508 m2

Correlation Time = 231.9 s

APPENDIX H

DIFFERENT KALMAN FILTER NOTATION

	Cross	Teunissen	Merminod	Gelb
Model Parameters				
Estimate of the state vector	$\mathbf{x}$	$\mathbf{x}$	$\overset{o}{\mathbf{x}}+\delta\mathbf{x}$	$\mathbf{x}$
Design / Jacobian matrix	$\mathbf{A}$	$\mathbf{A}$	$\mathbf{A}$	$\mathbf{H}$
Observed - computed vector	$\mathbf{b}$	$\mathbf{y}$	$\overset{o}{\mathbf{v}}$	$\mathbf{b}$
Vector of residuals	$\mathbf{v}$	$-\mathbf{e}$	$-\mathbf{v}$	$-\mathbf{v}$
Covariance of observations	$\mathbf{Cl}$	$\mathbf{R}$	$\mathbf{Q}_{ll}$	$\mathbf{R}$
Transition matrix	$\mathbf{M}$	Φ	Φ	Φ
Dynamic model noise	$\mathbf{q}$	$\mathbf{q}$	$\mathbf{w}$	$\mathbf{w}$
Covariance of dynamic noise	$\mathbf{Cy}$	$\mathbf{Q}$	$\mathbf{Q}_{ww}$	$\mathbf{D}$
From time t_{i-1} to time t_i				
Predicted state	$\hat{\mathbf{x}}_i(-)$	$\hat{\mathbf{x}}_{i\|i-1}$	$\bar{\mathbf{x}}$	$\hat{\mathbf{x}}_i(-)$
Predicted covariance	$\mathbf{C}\hat{\mathbf{x}}_i(-)$	$\mathbf{P}_{i\|i-1}$	$\mathbf{Q}_{\bar{x}\bar{x}}$	$\mathbf{P}_i(-)$
Filtered state	$\hat{\mathbf{x}}_i(+)$	$\hat{\mathbf{x}}_{i\|i}$	$\hat{\mathbf{x}}$	$\hat{\mathbf{x}}_i(+)$
Filtered covariance	$\mathbf{C}\hat{\mathbf{x}}_i(+)$	$\mathbf{P}_{i\|i}$	$\mathbf{Q}_{\hat{x}\hat{x}}$	$\mathbf{P}_i(+)$
Gain at t_i	$\mathbf{G}_i$	$\mathbf{K}_i$	$\mathbf{K}$	$\mathbf{K}_i$
From time t_i to time t_{i-1}				
Smoothed state	$\hat{\mathbf{x}}_{i-1}(s)$	$\hat{\mathbf{x}}_{i-1\|i}$	$\hat{\hat{\mathbf{x}}}$	
Smoothed covariance	$\mathbf{C}\hat{\mathbf{x}}_{i-1}(s)$	$\mathbf{P}_{i-1\|i}$	$\mathbf{Q}_{\hat{\hat{x}}\hat{\hat{x}}}$	
Smoothing matrix	$\mathbf{S}_i$		$\mathbf{J}$	