SATELLITE BORNE RADAR ALTIMETRY -

EMPIRICAL ORBIT REFINEMENT AND OCEAN SIGNAL RECOVERY TECHNIQUES

Gordon William Jolly

Doctor of Philosophy

The University of Aston in Birmingham

April 1995

Abstract

Measurements of the sea surface obtained by satellite borne radar altimetry are irregularly spaced and contaminated with various modelling and correction errors. The largest source of uncertainty for low Earth orbiting satellites such as ERS-1 and Geosat may be attributed to orbital modelling errors.

The empirical correction of such errors is investigated by examination of single and dual satellite crossovers, with a view to identifying the extent of any signal aliasing: either by removal of long wavelength ocean signals or introduction of additional error signals. From these studies, it was concluded that sinusoidal approximation of the dominant one cycle per revolution orbit error over arc lengths of 11,500 km did not remove a significant mesoscale ocean signal. The use of TOPEX/Poseidon dual crossovers with ERS-1 was shown to substantially improve the radial accuracy of ERS-1, except for some absorption of small TOPEX/Poseidon errors.

The extraction of marine geoid information is of great interest to the oceanographic community and was the subject of the second half of this thesis. Firstly through determination of regional mean sea surfaces using Geosat data, it was demonstrated that a dataset with 70cm orbit error contamination could produce a marine geoid map which compares to better than 12cm with an accurate regional high resolution gravimetric geoid. This study was then developed into Optimal Fourier Transform Interpolation, a technique capable of analysing complete altimeter datasets for the determination of consistent global high resolution geoid maps. This method exploits the regular nature of ascending and descending data subsets thus making possible the application of fast Fourier transform algorithms. Quantitative assessment of this method was limited by the lack of global ground truth gravity data, but qualitative results indicate good signal recovery from a single 35-day cycle.

Keywords: Satellite altimetry, Orbit error, Optimal interpolation, Mean sea surface, Fast Fourier Transform

To Leonie.

Acknowledgements

Contents

List of figures

List of tables

Chapter 1

Introduction

Satellite borne radar altimetry represents an unlikely meeting point for both solid Earth and ocean sciences. The gravity field associated with the Earth is mainly due to the internal structure of the planet which has a direct bearing on both the dynamics of the ocean and on the precise motion of satellite in orbit about the Earth. A radar altimeter measures the time for a microwave pulse to travel from the instrument to a point immediately below, where it is then reflected back to the satellite. When such an instrument is operated from a satellite over a region of ocean, that measurement contains information relating to the motion of the satellite, the local strength and direction of the Earth's gravity field (both at the sea surface and at the satellite), as well as many components of oceanic and atmospheric variability. Analysis of altimetry requires the separation or correction of all of these effects.

From an oceanographic point of view, the required signal is that of the sea surface height above a theoretical reference surface. This requires that all other contributions to the altimetric range measurement be accurately subtracted out to leave only a precise and accurate sea height measurement. Of the dozen or more signals present in each altimeter measurement the least well modelled is generally the position of the satellite at the time of the observation. As any error in the radial height of the platform at this time will contribute directly to the measured sea height it is crucial to derive satellite ephemerides of a comparable accuracy to that of the instrument. To date this level of accuracy is available only for the US/French satellite TOPEX/Poseidon because of its high altitude and comprehensive tracking systems. For other satellites such as the European ERS-1 and US Navy Geosat at lower altitudes, radial accuracies are several times greater than the

those of the instrument. Since the first satellite altimeter datasets became available, a number of empirical corrections have been suggested and used to remedy this situation. These generally exploit the difference in wavelength scales of orbit error and oceanographic contributions by high pass filtering the data. Unfortunately, these techniques can also have the effect of aliasing long wavelength ocean signals which are important to some studies. The extent of this problem was one of the key objectives of the research undertaken for this thesis resulting in the publication of results in *Jolly and Moore* (1994a). This study also demonstrated the presence in Geosat altimetry of a signal of just 2cm in amplitude at a period similar to the precession period of the satellite relative to the solar angle of 337 days. This contradicts some previous assertions that the effects of solar radiation pressure on the pointing accuracy of Geosat were not observable in the altimetric range measurement.

The superior radial accuracy of TOPEX/Poseidon has been suggested as a useful means of correcting ERS-1 altimetry through minimisation of what are known as dual satellite crossovers. These make use of near coincident observations of the same point on the sea surface by each altimeter to estimate the orbit error contribution associated with the problematic solution of ERS-1 orbits. It was demonstrated in *Jolly and Moore* (1994b) that such a technique provides a significant improvement in the radial accuracy, but at the expense of introducing small aliased long wavelength contributions from TOPEX/Poseidon. As the estimated radial accuracy of TOPEX/Poseidon is around 3-4cm, the contribution from its orbit error is almost negligible; it was shown, however, that certain geographically correlated signals could not only be absorbed in to ERS-1 altimetry, but in some cases amplified in the process.

One of the principal objectives of satellite altimetry is the determination of an accurate map of the mean sea surface elevation (MSS) above a theoretical reference ellipsoid. As altimeter data are generally provided as

10Hz or 1Hz datasets, the MSS analysis of several months or years of data is a computationally demanding task. There are several established techniques which work well for both global and regional analyses. In general the regional analyses are more accurate and better suited to short wavelength recovery, while global studies concentrate on the larger scale features. This distinction is a consequence of the limits of current analysis techniques. Development of a method for the generation of a consistent high resolution global MSS was an important element of the research undertaken. On a regional level, two methods were investigated with a view to development of a global technique by combining one of these with established frequency domain techniques used for the analysis of ground based gravimetry. A statistical Fourier transform filtering technique was developed and has been submitted for publication, *Jolly and Moore* (1995). From 35 days of ERS-1 data corrected by dual crossover minimisation with TOPEX/Poseidon a global MSS was produced and shown to provide a good representation of the marine geoid. This technique will be developed further when Geodetic Phase ERS-1 data become available in order to produce a global high resolution geoid solution of better than 20km resolution and hopefully sub-decimetre accuracy.

To achieve its full potential the analysis of satellite altimetry must provide high accuracy, high resolution and a high degree of consistency and reliability. The traditional restrictions of altimetry relating to orbital coverage and radial accuracy are both being reduced by improved tracking, and a variety of orbital patterns, particularly from ERS-1. The problem of radial accuracy will remain a problem for all low Earth orbiting satellites where precise altimetry is to be analysed and it is in this area that the majority of the work undertaken for this thesis has been directed. From a variety of approaches, contributions have been made to the empirical correction of altimetric measurements as well as to the confidence of such techniques. Novel techniques have produced methods for relating sea level time series at pairs of locations without the need for further orbit

correction. In addition, a globally consistent method has been developed for mean sea surface determination using the millions of altimeter observations of the Earth's oceans.

Chapter 2

Height: the oceans' third dimension

2.0 Introduction

The face of the Earth consists of two very different surfaces: land and sea. The continental crustal areas cover some thirty percent of the globe and in some places rise to several kilometres above the mean radius of the Earth. The ocean's surface, which is the focus of this research, covers the remaining 70% of the Earth and shows more subtle variation. Most people are familiar with the two dimensional outline of the oceans from normal atlases, some of which also contain information on the bathymetry of the ocean floor, which in places is far deeper than the highest mountains to be found on land. What is generally neglected from these maps is the fact that bathymetry, which is generally described relative to a standard ellispoid of revolution, may be significantly different to the ocean depth at a point due to the fact that the surface of the oceans do not conform to a perfect ellipsoid. Under the influences of the Earth's gravity field, the sea surface is raised or depressed by an rms value of 30m, covering a range of -100m south of the Indian sub-continent to +60m in the eastern North Atlantic. In places, this can result in a slope equivalent to a metre or more over tens of kilometres. Other factors such as tides, ocean currents and atmospheric interactions also affect the sea surface to a lesser degree.

This chapter examines influences on the shape of the interface between ocean and atmosphere and what information may be extracted from accurate maps of this topography. Finally, remote sensing is introduced as a means for the recovery of such information on regional and global scales, specifying the requirements for the extraction of various signals.

2.1 Influences on the sea surface

The dominant form of the sea surface is that of the shape of the Earth, that is, an oblate spheroid, described by an ellipsoid with a semi-major axis of around 6378km and flattening 1/298. The precise values of these figures vary from one definition of the reference ellipsoid to another but a few of the more common ones are given in table 2.1

Table 2.1. Various standard reference ellipsoids

Reference model	semi major axis (km)	1/flattening
WGS-84	6378.137	298.25643577
ERS-1 (pre V3.0 data)	6378.1364	298.2564
TOPEX/Poseidon	6378.1363	298.257
Aston standard ellipsoid	6378.1363	298.257

The exact shape of the ellipsoid that best approximates the mean sea level is a consequence of the gravitational attraction of the Earth, trying to pull the Earth into a sphere and the centrifugal effects of inertia at the equator as the planet spins once per day. On smaller scales, the Earth cannot be treated as a perfect gravitational point source located at it's core as irregularities cause distortions of the overall field at a local level. Third body attraction exerts a gravitational influence by way of various tidal perturbations of the oceans while pressure and density variations affect the surface by a number of means. The following sections briefly introduce the cause of some of the major perturbations to the precise shape of the sea

surface and then identify key requirements for their successful identification via remote sensing.

2.1.1 The Geoid

Every massive particle exerts a gravitational force on every other particle of a magnitude given by Newton's law of universal gravitation. This force acts in the direction of a vector joining the geometric centre of each particle, where those particles are both spherical and of uniform density. At great distances from the Earth, it too may be treated as if it were such a particle, for example in the calculation of the Earth's influence on the orbits of the Jovian moons. At it's surface, however, the Earth's gravitational field is not uniform in magnitude and the force is not always directed towards the planets centre of mass. The theoretical surface over which the Earth's gravity field is of equal potential is known as the geoid and is of particular significance to the study of the sea surface which it closely follows .

The largest factor influencing the geoid is that of the oblate shape of the Earth, which has already been mentioned. Significant long wavelength regional variations exist which may be attributed to geological conditions deep below the surface. The Earth has a solid core surrounded by a liquid mantle upon which the continental and oceanic crusts float. Heat generated by the core is transmitted to the surface slowly by immense convection currents of molten rock which help to drive the motion of the continental plates. From the point of view of the gravity field, convection currents produce volumes of hot rising fluid as well as cooler, sinking columns. From thermodynamics there is a difference in the densities of hot and cool currents which result in large scale differences in the local density of the mantle and hence the gravity field experienced at the surface above. These processes are thought to persist for millions of years and can therefore be considered static for the purposes of these investigations.

Floating on the mantle are the continental plates, consisting of two distinct kinds of crust known as the lithosphere: continental crust is 50-100km in depth and constitutes the bulk of the land area; oceanic crust on the other hand is much thinner, of the order 5-10km in depth. In places such as the Mid-Atlantic Ridge, this crustal layer is being continuously added to as the European and North-American plates slowly drift further apart, widening the Atlantic.

Due in part to its greater thickness, the continental crust often contains different rock types, not only within geographical areas but at different depths, for example through sedimentary deposition. Both of these are associated with density variations which in turn cause distortions to the geoid, although at shorter wavelengths than mantle effects.

Of much shorter wavelengths, the actual topography of the solid crustal layers produce density fluctuations which influence the geoid, due principally to the difference in density between the fluid of the ocean or atmosphere and that of the underlying rock. Even when covered by several kilometres of water, topographic distortions on the ocean floor may be found in the geoid and hence in the sea surface, as will be demonstrated in chapter 7.

Finally, there is the influence of the oceans and atmosphere, the fluids of which also have mass and cause distortions to the geoid. Their effect gravitationally is however less pronounced when compared to their other, more direct influences on the sea surface, to be described in the following sections.

2.1.2 Tides

The gravitational forces associated with the Sun and Moon influence the ocean surface through generation of tides. Tides are ocean waves which have a period of several hours and can have amplitudes of several metres.

Constrained by land masses, the ocean is pulled by the gravitational attraction of the Sun and Moon and in addition experiences an inertial force due to the rotation of the Earth about the Earth-Sun and Earth-Moon centres of gravity. The twice daily lunar tide (M_2) is the dominant component of the tidal signal, but up to 390 components can be included in a tidal model. From *Bearman* (1989), table 2.2 summarises some of the principal tidal components, their periods and relative size.

Table 2.2. Some principal ocean tidal components.

Name	symbol	Period (hours)	Coefficient ratio (M_2=100)
Principal lunar	M_2	12.42	100
Principal solar	S_2	12.00	46.6
Larger lunar elliptic	N_2	12.66	19.2
Luni-solar semi-diunrnal	K_2	11.97	12.7
Luni-solar-diurnal	K_1	23.93	58.4
Principal lunar diurnal	O_1	25.82	41.5
Principal solar diurnal	P_1	24.07	19.4
Lunar fortnightly	M_f	327.86	17.2
Lunar Monthly	M_m	661.30	9.1

The magnitude and direction of tidal forces on the ocean can be precisely calculated, but the response of the ocean can be complicated by the local topography and prevailing weather conditions. As the tidal signal is a long period wave travelling round the Earth as it rotates, at the equator, it must travel at a linear velocity of around 450ms^{-1}. This is greater than the theoretical limit for a surface wave in the open ocean [*Bearman*, 1989], resulting in a time lag between passage of the Moon overhead (or 12 hours and 25 minute later) and the tidal maximum. In addition, the presence of land masses prevent tidal bulges from directly circumnavigating the globe and the bathymetry of the oceans constrain the direction of tidal flows. In addition to these, the Coriolis force influences the direction of the tidal

flows, clockwise in the Northern Hemisphere and anti-clockwise in the Southern Hemisphere.

2.1.3 Ocean circulation

Directly and indirectly, the Sun drives ocean currents on a variety of scales, from small eddies a few kilometres across, to currents such as the Gulf Stream which cross the oceans. In addition, the rotation of the Earth on its axis has an influence, both on the strength and particularly the direction of a current.

The most obvious way in which a current can be generated is through coupling of atmospheric winds to the sea surface, thus forcing the movement of a body of water. Secondly, absorption of solar radiation warms the oceans, leading to evaporation which increases the salinity and hence the density of sea water - later precipitation of this atmospheric water vapour results in the opposite. Water density is also changed in polar regions by the formation and melting of sea ice. All of these effects generate differences in temperature or density and in the same way as occurs in the atmosphere, the fluid ocean flows to correct any instability and reduce the potential to a minimum. Once in motion, the Coriolis force acts to perturb the current by changing its direction so that a north flowing current in the Northern Hemisphere is turned slowly to the right into an clockwise circulating current, while the reverse occurs in the Southern Hemisphere. The large scale flow of warm and cold bodies of water through the Earth's oceans has a profound influence on the thermodynamic balance of the planet. As heat is radiated from the Earth in a largely uniform manner, independent of latitude, transport of heat away from the tropics is necessary to ensure that the poles are kept relatively warm and the tropics cooler than they would be if the Earth had no oceans or atmosphere. The precise mechanisms for this heat engine are the focus for investigations such as the World Ocean Circulation Experiment (WOCE)

and will help to predict any global climate changes due, for example, to the greenhouse effect.

2.1.4 Atmospheric effects

The oceans are bounded below by the sea floor and above by the atmosphere - the characteristics of which can vary considerably in a very short period of time. The atmosphere influences the sea surface in two ways, most obviously through wind generation of waves and currents, but also through the weight of the column of air resting on the interface. Of these, it is the effect of air pressure which has the most direct influence on an altimetric range measurement, although significant wave height (SWH) and wind speed do cause changes in the return altimeter pulse which will be described in the following chapter. The inverse barometric effect is the compression or rarefaction of the sea surface by local increases or decreases in atmospheric pressure, leading to an approximate 1cm difference in sea height for each millibar pressure change. Low atmospheric pressure associated with storms can be in excess of 50mbar below standard atmospheric pressure (1000mbar), leading to storm surges of up to 50cm in deep ocean. In shallow seas, the smaller column of water leads to less of a direct inverse barometer effect, but one that can still result in significant changes in sea level, particularly in coastal areas where the flow of water may be restricted.

The flow of water, for a variety of reasons is therefore associated with oceanic pressure differences which result in height differences at the sea surface. In the case of the Gulf Stream, the height difference can be over 1m, but in general according to *Visser et al.* (1993), the static sea surface topography is around 65cm while the dynamic component has an rms height of around 15cm.

2.2 Sea surface analysis

From accurate measurement of sea surface height, there are three major signals which can be extracted. The first of these is the precise shape of the mean surface or marine geoid, secondly there is the dynamic ocean variability signal and finally there is tidal information. This thesis is concerned only with the first two from this list and tides shall not be covered in detail.

2.2.1 Marine geoid

The mean sea surface is composed of two distinct signals, inseparable by sea height analysis alone. Principal of the two is the Earth's geoid, associated with local variations in the gravity field and producing an rms global signal of around 30m and described section 2.1.1. Superimposed on the geoid is the static component of the sea surface topography (SST) which may be attributed to 'permanent' ocean currents. These currents include the major contributions to global ocean circulation and are the subject of intense investigation by the oceanographic community because of the capacity of the oceans to not only transport heat from the tropics to the poles, but their influence as a heat store and therefore their effect on the thermodynamic balance of the planet. Globally, the static SST is thought to have an rms value of around 65cm with certain regions departing from the geoid height by in excess of 1m, *Visser et al.* (1993). Accurate measurement of this signal is a key objective of the World Ocean Circulation Experiment (WOCE) which requires that this signal be separated from the mean sea surface measurement. To distinguish such a measurement, it is necessary to have accurate, independent information about the Earth's geoid at that point. This is where sea surface topography analyses are currently limited. This thesis will deal with precise measurement of the mean sea surface (MSS) at local and global scales in chapters 6 and 7 respectively. For information regarding the independent measurement of the geoid at local

scales, the reader is referred to *Telford* (1976) and on global scales from the precise analysis of satellite orbits. The latter of these two gravity measurements is performed through solution of spherical harmonic gravity field models, for example, as described by *Nerem et al.* (1994) using TOPEX/Poseidon data and *Moore and Ehlers* (1994) with ERS-1 data.

Aston University

Illustration has been removed for copyright restrictions

WAVELENGTH ,km   Rossby radius
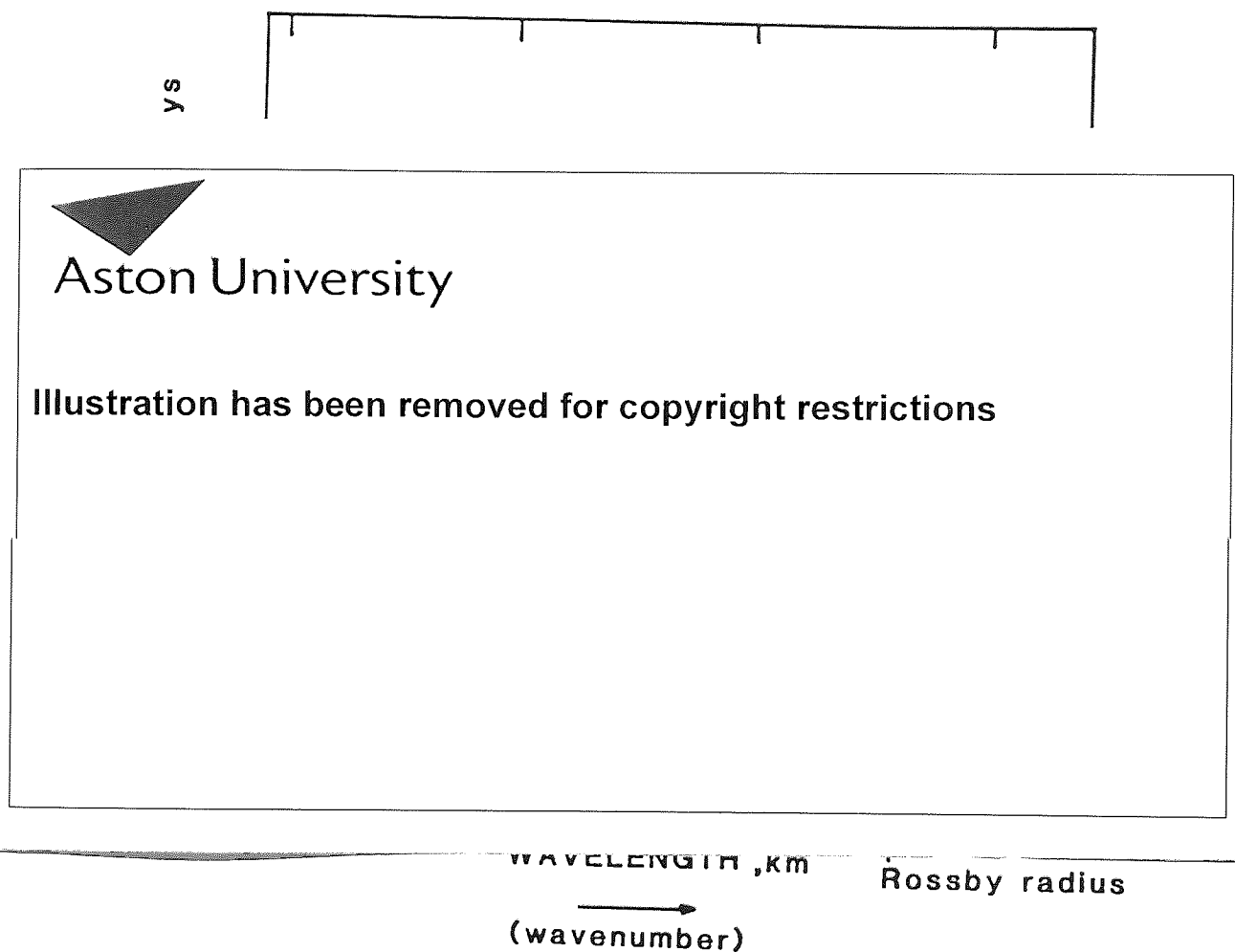
(wavenumber)

Figure 2.1. Schematic representation of the frequency-wavenumber spectrum of circulation energies at middle latitudes, reproduced from Robinson (1985).

2.2.2 Dynamic topography

While it is impossible to separate the static SST from the geoid in the analysis of an altimetric MSS, the variable component of the ocean currents can be determined on various scales. A sketch in figure 2.1 of the frequency-wavenumber spectrum of circulation energy at middle latitudes

from *TOPEX* (1981), reproduced in *Robinson* (1985), plotted in arbitrary units shows that the dominant signals are annual, of wavelengths in excess of around 100km, and meso-scale with wavelengths typically around 100km and of periods greater than approximately 12 days. These signals may be extracted in two ways, either by comparison of 'instantaneous' sea surface maps with the marine geoid introduced in 2.1.1 and described in chapters 6 and 7; or through determination of time series in specific locations or small regions, as will be performed in chapter 5. At specific locations, tide gauge records are commonly used for this purpose, for example *Harangozo et al.* (1993). In open ocean, satellite altimetry provides the coverage and measurement accuracy necessary for extraction of dynamic topography maps, for example, using TOPEX/Poseidon data, *Tapley et al.* (1993). Alternatively, derivation of regional time series were described by *Milbert et al.* (1988) and will be developed as a part of this thesis in chapter 5.

2.3 Remote sensing requirements

Sensors carried on board Earth orbiting satellites can measure four physical characteristics of the sea surface: colour, temperature, sea height or slope, and surface roughness. Of these, colour has not been monitored since the end of the Coastal Zone Colour Scanner (CZCS), carried on board the Nimbus-7 satellite. A replacement for this system will shortly be launched on board the SeaWiFs satellite, thus enabling measurement of the first characteristic - colour. Temperature has been systematically measured by the Advanced Very High Resolution Radiometer (AVHRR), operated by the National Oceans and Atmosphere Administration (NOAA). In addition, the Along-Track Scanning Radiometer (ATSR) on board ERS-1 also measures sea surface temperatures. The final two parameters of height or slope, and surface roughness are best measured by microwaves: principally by radar altimetry, but also in the case of surface roughness by Synthetic Aperture Radar (SAR). This thesis is concerned with the

measurement of sea surface heights; both the static geoidal component and the dynamic ocean variability component.

The quality of any measurement is determined by two factors, the accuracy and the resolution. In the case of estimation of the sea surface height, resolution refers to both spatial and temporal characteristics. Overall, the requirements are dependent on the application to which the surface is of relevance. As the previous sections have described, the precise shape of the sea surface is a consequence of several factors of varying magnitude and wavelength. The principal contributions are listed in table 2.3, all of which contribute to the timing or distortion of the altimeter return pulse. The geoid and static Sea Surface Topography (SST) have no characteristic period to distinguish them and hence cannot be separated without independent information.

Table 2.3 Representative wavelength and periods of ocean surface signals.

Effect	amplitude	Wavelength	Period
Geoid	30m	1-40000km	∞
static SST	1m	1-1000km	∞
dynamic SST	1m	1-100km	1day-1year
steric signal	<10cm	20000km	1year
tides	50cm	10-20000km	1hr-18.6years
gravity waves	1m	1-100m	seconds
wind ripples	<1cm	<10cm	<0.1 sec.

- **Measurement accuracy**

The accuracy of an instrument or technique may be defined in terms of two slightly different quantities, the absolute accuracy and the relative

accuracy. The absolute accuracy determines the ability of an instrument to produce an exact measure, free of systematic biases. The relative accuracy on the other hand defines the effect of random errors on the consistency of measurements of the same signal. For long wavelength and geoid determination, the volume of sampling from a satellite is such that noise may be suppressed by simple averaging techniques. Absolute accuracy is therefore of greatest importance in these cases, hence the efforts made in determining the absolute electronic biases in every altimeter. For medium to short wavelength recovery or in the case of temporal analyses, the consistency of measurements can be more important than the absolute accuracy of a single measurement.

- **Spatial and temporal resolution**

From sampling theory, to completely resolve a signal, is must be accurately measured at double its highest frequency. Wind speed (from surface ripple measurements) and gravity waves on the sea surface would therefore require much higher resolution than is currently possible from a satellite. As will be described in the following chapter a radar altimeter is capable of measuring the statistical characteristics of these effects over a small enough area for a great many applications, see for example *Chelton and McCabe* (1990) and *Carter et al.* (1992) regarding SWH and wind speed determination. Of the other effects, most have a signal which decreases in strength with increasing frequency. From figure 2.1, the dominant oceanographic signals have a wavelength of 100km or more and a period of greater than 12 days. From a single satellite, this is difficult to resolve as it would require 50km sampling every 6 days, or over 60 orbits per day. This is four times faster than can be achieved by the lowest of stable long term Earth orbits - a satellite low enough to complete 16 orbits per day will very quickly be slowed by the Earth's atmosphere and drop out of orbit. Because of the cost of launching and operating satellites, a network of four or more operating simultaneously is unlikely in the foreseeable future. From a

single satellite therefore, a compromise is required which trades off spatial and temporal resolution. Designed for ocean topography research, the TOPEX/Poseidon satellite has a 10 day exactly repeating orbit during which time it completes 127 revolutions, forming a pattern of ground tracks each 2.8° apart (~300km at the equator). It is therefore able to resolve large scale oceanographic features at close to their characteristic period of 17-20 days. In the case of the gravity field, there is considerable signal at wavelengths shorter than this which limits the usefulness of TOPEX/Poseidon for this kind of work. Because the gravity field is essentially static, the repeat sampling period is not important, only the spatial resolution matters. The United States Navy's Geosat Geodetic Mission (GM) did not repeat its ground track during its 18 month mission, instead drifting slowly to produce an unrivalled 5km equatorial spacing. This has been emulated by the ERS-1 Geodetic Phase of two interleaved 168 day missions, resulting in an 8km spacing at the equator.

The ground track is not the only consideration, however, when measuring the sea surface elevation as instrument characteristics are also important. In the next chapter, the performance of the radar altimeter will be described, along with the various corrections necessary to produce each one-second measurement of an area of the sea surface between 2 and 15 km in diameter, separated by around 7km along track.

Chapter 3

Satellite borne radar altimetry

3.0 Introduction

A satellite altimeter is a nadir pointing active microwave sensor designed to measure the time return echoes from the surface of the Earth. The return signals from ocean regions provide information on significant wave height, surface wind speed and, of particular relevance to this thesis, a range measurement from the satellite to the sea surface immediately below. This chapter will introduce the basic principles of altimetry, the necessary corrections and a review of relevant altimeter missions to date.

The operation of all altimeters is broadly similar and unless specifically highlighted the following section describes the operation and characteristics a typical Ku-band satellite radar altimeter (~13GHz).

3.1 Satellite Altimetry

A radar altimeter operates by timing the delay between emission of a short microwave pulse and the subsequent detection of the returned echo, recording the time and distortion of the returned signal. The first requirement for any remote sensing instrument wishing to observe the Earth's surface is that atmospheric attenuation be small enough to detect the electromagnetic signal. From *Rees* (1990), a simple plot of the total zenith attenuation by the atmosphere between ultra-violet and radio frequencies is shown in figure 3.1. Through most of the infra-red region, attenuation is strong due to atmospheric water vapour and gases such as carbon dioxide and oxygen. In the microwave region, however, signal attenuation is small between 10^2 MHz and 10^4 MHz. There is also low attenuation in the

visible region (300-850 nm) where LIDAR was recently tested onboard the Space Shuttle Atlantis - LIDAR is the optical equivalent of RADAR, using a laser to generate the short pulse which gives high spatial resolution but with out the all weather capability of a microwave device. As stated at the end of Chapter 2, for most oceanographic applications, a resolution of a few kilometres is sufficient, which in conjunction with the need for complete coverage without data gaps due to clouds, make microwave instruments the preferred option.
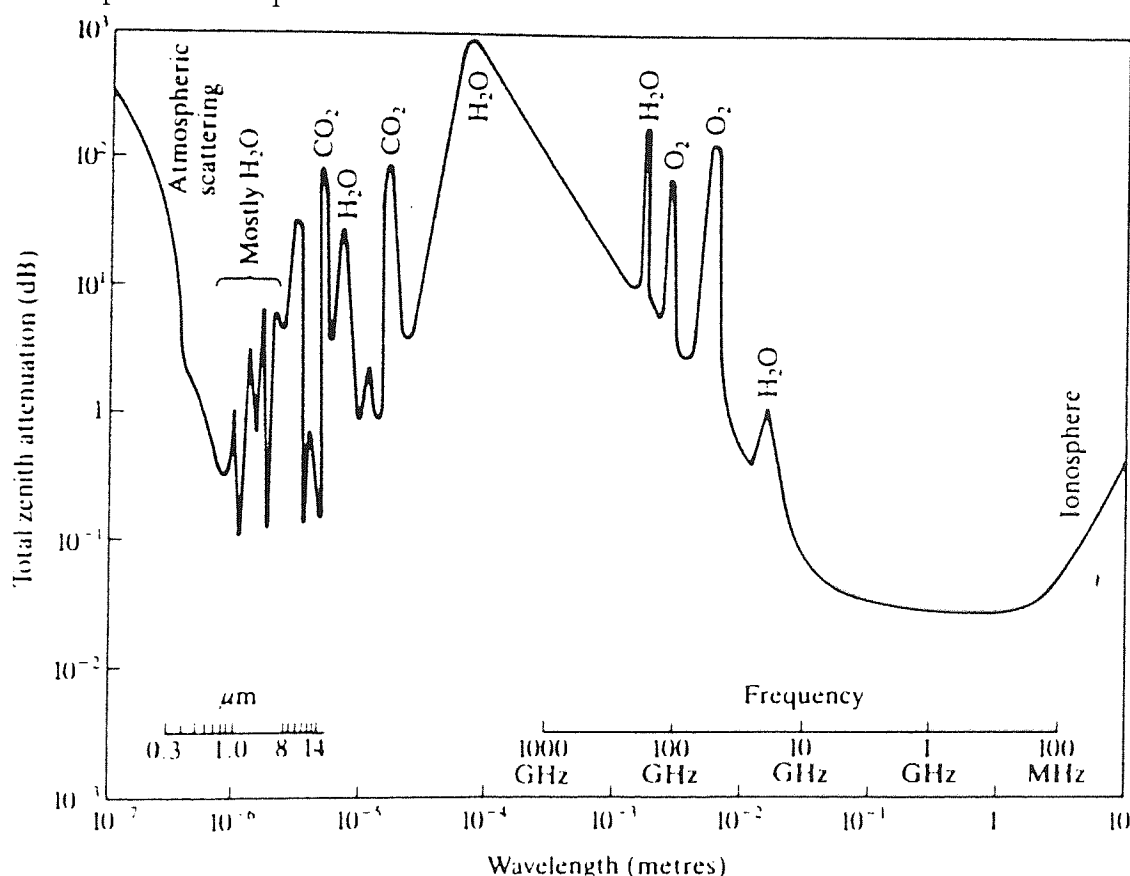
Figure 3.1 Vertical atmospheric water vapour electromagnetic attenuation between ultra-violet and radio frequencies.

3.1.1 Pulse generation and emission

Emission of electromagnetic radiation by an aperture results in diffraction, the width of which, θ, is determined by the wavelength, λ, of the radiation and the size of the aperture, D. From *Hecht* (1987), a uniformly and coherently illuminated circular aperture will produce a diffraction limited beam width θ,

$$\theta = \frac{1.22\lambda}{D}$$

A 13 GHz (λ=2.3cm), Ku-band microwave signal emitted by a 1m antenna will have a beam width of 28 mrad or 1.6°, which from an altitude of 800km covers a disk of diameter 22km. For higher frequencies or large antennas, this figure is correspondingly reduced (the latter being the basis for the resolution of Synthetic Aperture Radar techniques).

There is a second problem associated with the frequency of microwave radiation used: when a surface such as the ocean is illuminated by a coherent single frequency source, differences in optical path length cause constructive and destructive interference which can radically distort the signal received by the instrument. Two points are coherently illuminated if their optical distances from the source differ by less than half the coherence length l_c. This characteristic length is defined by *Rees* (1990) for radiation at a frequency v, of bandwidth Δv as

$$l_c = \frac{c}{\Delta v}$$

where c is the speed of light. Bandwidth is defined as frequency spread of the emitted radiation. For a perfect single frequency transmitter, this produces an infinitely coherent source. In reality, all transmitters have a spectral width, and it is partly to avoid the problems of interference (known as fading) that radar signals are 'chirped', giving them a limited coherence length which is small enough to avoid fading for all but the flattest of reflectors. From *Dumont and Stum* (1992), the bandwidth of the ERS-1 radar altimeter is 330MHz, resulting in a coherence length of around 90cm.

A spot size of 20km is too great for many applications of the data, for example, high resolution geoid determination. There is therefore a need for a larger antenna, a higher frequency or some other refinement to the emitted radiation. A larger antenna on board the satellite is impractical while at higher frequencies, there is greater atmospheric attenuation of the

returned signal, therefore the only alternative is the use of pulse limited altimetry. In this way, the leading edge of the return pulse provides sufficient information for an accurate estimate of the ocean height over a significantly smaller area. This is the second reason for chirping the emitted signal as this can simulate a short but sharply modulated pulse, the distortion of which is analysed on its return.

3.1.2 Pulse reflection

A pulse limited altimeter emits a very short microwave pulse which is reflected from the surface in the following manner, summarised in figure 3.2. In the first instance, the surface is assumed smooth and the radiation incoherent.
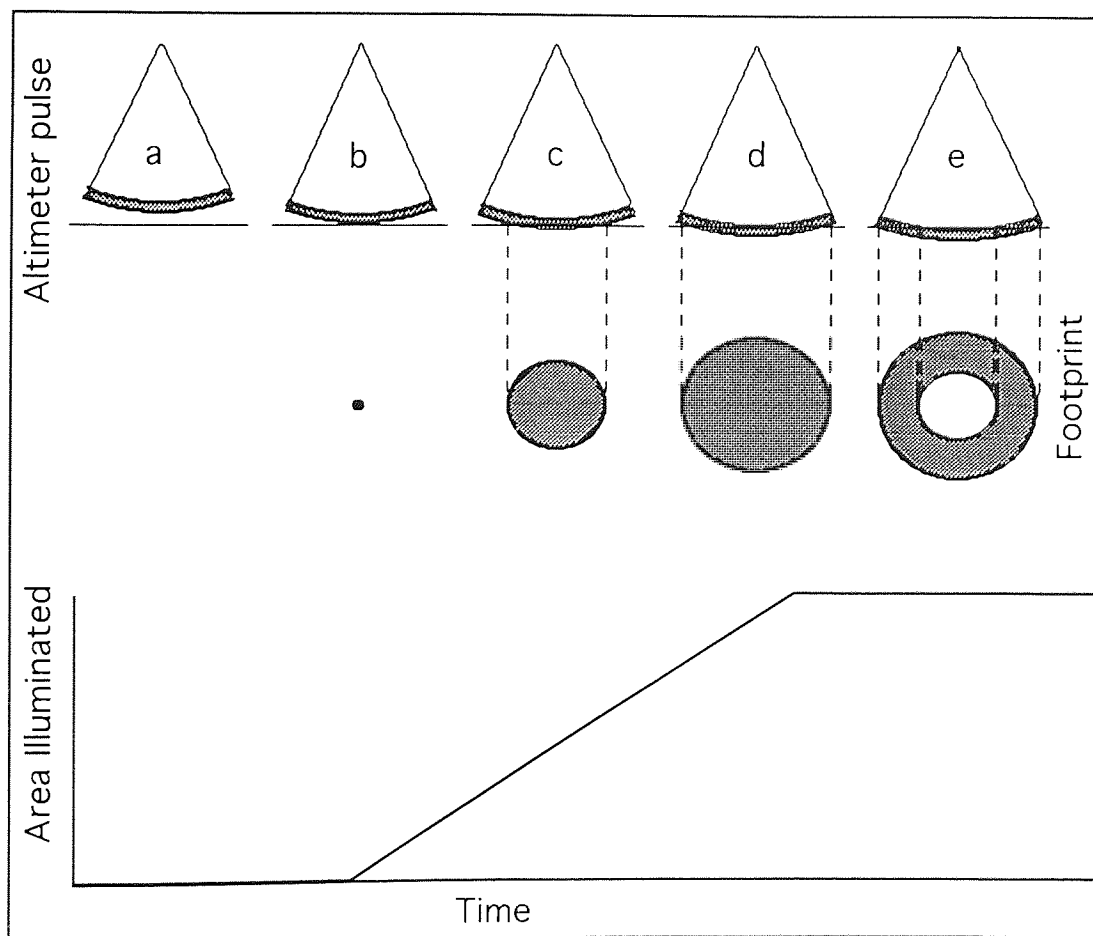
Figure 3.2. Reflection of a pulse limited satellite radar altimeter pulse at a smooth flat surface.

- The leading edge of the pulse strikes the surface and a reflected signal begins to return to the satellite (pulse *b* in figure 3.2).

- The illuminated area grows linearly to form a disk until the trailing edge of the microwave pulse reaches the surface. The returned power is proportional to the illuminated area, thus there is a steady increase in the reflected signal strength (pulses *c-d*, figure 3.2).

- As the emitted pulse continues to propagate, the illuminated area assumes the shape of an annulus which spreads outwards, preserving the area and hence returned power of the disk of maximum area (pulses *c-d*, figure 3.2). Once the annulus size approaches the diffraction limited beam width, the returned signal strength is reduced.

Reflection of the radar pulse at a curved ocean surface is described in more detail by *Chelton et al.* (1989).

A range measurement may therefore be made from the return time of the leading edge of the reflected signal. In addition, where the surface is not smooth, as is usually the case at the sea surface, the ideal return signal is distorted by a number of factors. The presence of waves on the surface cause reflection of the pulse from the wave crests outwith the disks in figure 3.2 at the same time as the illuminated disk is growing. The range measurement spot size is therefore dependent on the average or significant wave height (SWH) within the diffraction limited area, figure 3.3. In the case of incoherent illumination of a smooth sea surface, the spot size of a Ku-band altimeter at 800km altitude is less than 2km. Where waves are present within the diffraction limited footprint of the altimeter, the area of the sea surface which contributes to the return signal is increased. From the time at which the leading edge of the altimeter pulse reaches the wave crests at the satellite nadir, the radius of the footprint increases linearly until the trailing edge reaches the wave troughs at the nadir point. The additional time taken for the trailing edge to reach the wave troughs as
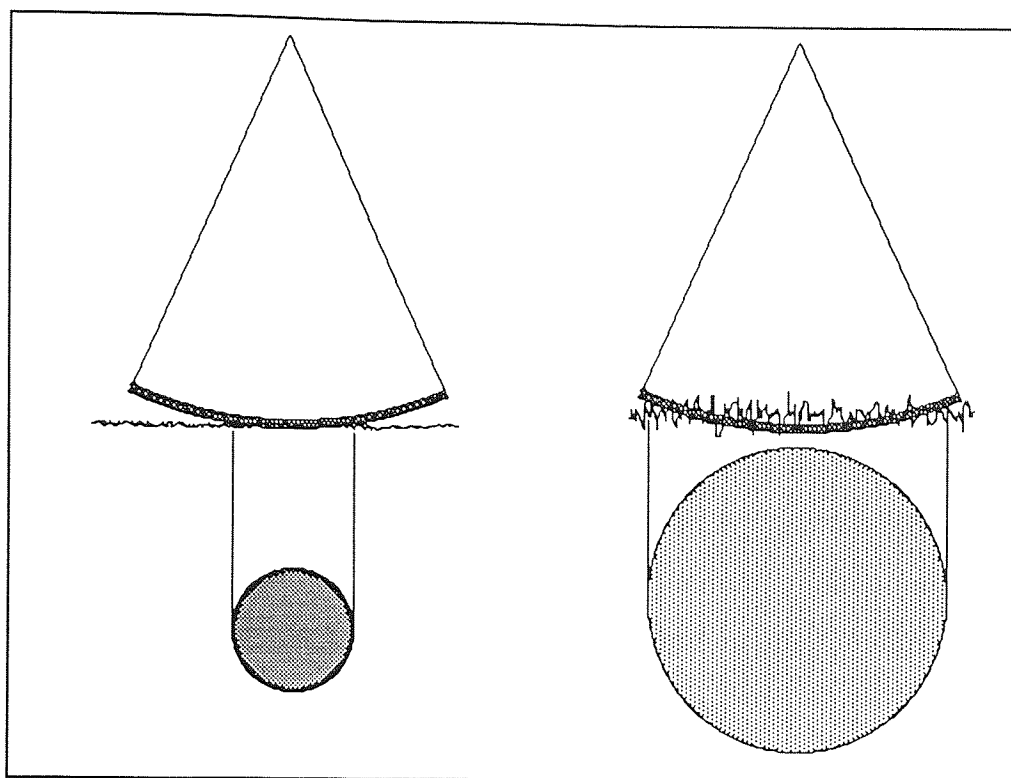
Figure 3.3.   Altimeter pulse spreading by different sea states.

compared to the leading edge reaching the crests results in a larger footprint radius before it becomes an annulus of constant area.   From *Chelton et al.* (1989),   the effective altimeter footprint from an altitude of 800km is summarised in table 3.1.

Table 3.1. Altimeter pulse limited footprint radius as a function of SWH.

Significant Wave Height (SWH)	Effective footprint diameter (@800km altitude)	Effective footprint diameter (@1335 km altitude)
m	km	km
0	1.6	2.0
1	2.9	3.6
3	4.4	5.5
5	5.6	6.9
10	7.7	9.6
15	9.4	11.7
20	10.8	13.4

When the circular footprint is at its maximum extent, the returned signal from the expanding spherical shell is being reflected from wave troughs at the centre and crests at the edge. Overall, the total area of reflectors is not increased by the increased radius of the footprint and hence the returned signal strength is not affected. The distribution of reflecting facets, however, is no longer simple to model and generally results in an error which is dependent upon the SWH and surface roughness caused by wind forcing. This is known as the electromagnetic or sea state bias and will be described in section 3.3.3. By comparison of the rise time and known SWH at the surface, the estimation of SWH may be empirically calibrated, *Carter et al.* (1992). Finally from the returned power, the small scale surface roughness of the ocean may be estimated, and hence the surface wind speed measured.

3.1.3 Pulse detection

The final stage in the propagation of an altimeter radar pulse is detection and analysis by the electronics on board the satellite. Real echoes obtained from the altimeter are the sum of many individual point scatterers within the illuminated surface and hence, the superposition of many amplitudes and phases. The received signal therefore contains the statistical characteristics of the illuminated area convolved with the transmitted pulse. To reduce the uncertainty in these measurements, pulses are transmitted at short intervals and averaged 50 to 100 times to yield 10 or 20Hz average echo profiles which are then stored on board for transmission to the ground station. Processing of the returned altimeter pulse is performed in the frequency domain and involves comparison of the returned signal with an exact copy of the transmitted signal at a slightly reduced frequency, known as the deramping chirp. The time at which the duplicate signal is generated must closely match the return time of the reflected signal in order that enough of the return waveform be recorded for processing. Systematic errors in the timing of the deramping chirp result in

a consistent range bias in the altimetry which cannot be resolved from the waveform alone. This electronic altimeter bias requires careful calibration in order that reliable absolute measurement of the sea surface be made. Further details of the altimeter bias for ERS-1 and its calibration may be found in *Francis* (1993) and *Lam et al.* (1993) for ERS-1 and *Christiansen et al.* (1994) and *Menard et al.* (1994) for TOPEX/Poseidon. Chapter 5 will describe the relative calibration of different altimeters through short arc refinement using dual satellite crossovers.

3.2 Ephemeris modelling

An artificial satellite is acted upon by a number of forces, each of which to a varying degree influence its motion. Precise orbit determination is the analysis of satellite tracking data in order to model the position or ephemeris of the satellite at a given moment. This information is crucial to the interpretation of radar altimeter data as it represents what has been the largest source of uncertainty for all missions except TOPEX/Poseidon which orbits at around 1335km as compared with other altimeter missions which have operated at altitudes of less than 800km. As most remote sensing satellites are placed in circular or near circular orbits, this section will describe the principal forces acting on such a satellite and describe the major sources of uncertainty in the satellite ephemeris produced.

3.2.1 Orbital mechanics

In general, six parameters are required to describe the location of a satellite unambiguously: three of position and three of velocity. These could be expressed in rectangular Cartesian co-ordinates but it is more common to use the Keplerian elements, defined as:

a : Semi-major axis of the elliptic orbit

i : Inclination of the ascending node to the equatorial plane

e :   ellipticity of the orbit

Ω: The angle subtended by the intersection of the orbital and equatorial planes, and the first point of Aries.

ω: The argument of perigee defines the closest point of the orbit to the centre of the Earth (which is at one focus of the orbital ellipse)

M :   The mean anomaly describes the position of the satellite along the ellipse of the orbit (i.e. the angle swept out by a radius vector from perigee assuming constant angular velocity).

Some of these terms are illustrated in figure 3.4 along with θ which defines the position of the Greenwich meridian with respect to the first point of Aries, thereby locating the Earth beneath the satellite.

The Keplerian elements assume that all gravitational attraction acts between the centres of mass of the satellite and the Earth, and that all other forces are negligible. In reality, however, other gravitational and surface forces as well as the anisotropic nature of the Earth's own gravity field at altitudes smaller or comparable with the size of the Earth perturbs the ideal satellite orbit from its nominal, elliptic orbit.

3.2.2 Precise orbit determination

Precise orbit determination of satellites such as ERS-1 and TOPEX/Poseidon involves the numerical integration of the satellite position from an initial state vector using models to estimate the forces experienced by the satellite. The solution of the ephemeris is constrained by any available tracking data, for example Satellite Laser Ranging (SLR) in the case of ERS-1 and SLR, Doppler Orbitography and Radio Positioning Integrated by Satellite (DORIS) and Global Positioning System (GPS) for TOPEX/Poseidon. Altimetry may also be included as tracking data although only in the radial direction [*Shum et al.*, 1990].
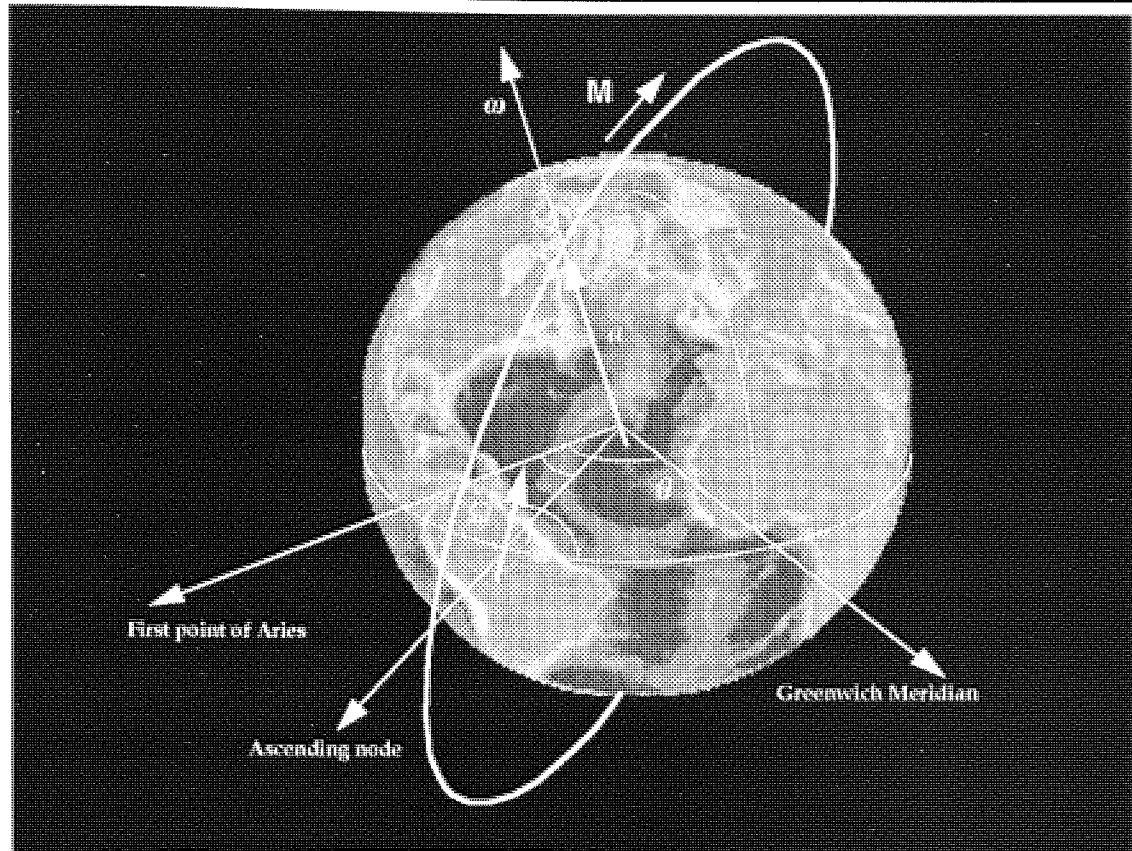
Figure 3.4. Schematic representation of Keplerian orbital elements.

by any available tracking data, for example Satellite Laser Ranging (SLR) in the case of ERS-1 and SLR, Doppler Orbitography and Radio Positioning Integrated by Satellite (DORIS) and Global Positioning System (GPS) for TOPEX/Poseidon. Altimetry may also be included as tracking data although only in the radial direction [*Shum et al.*, 1990].

The principle force, V, of the Earth's gravity field is described by a potential model expressed as a spherical harmonic for a satellite at a radial height, r, above the Earth's centre as,

$$V = \frac{\mu}{r}\left[\sum_{l=2}^{\infty}\left(\frac{R_e}{r}\right)^l \sum_{m=0}^{l}\left(C_{lm}\cos m\lambda + S_{lm}\sin m\lambda\right)P_{lm}(\sin\phi)\right)$$

$$(3.1)$$

- 3rd body gravitational attraction from the Sun, Moon, Jupiter, Mars, etc.

- Luni-Solar tidal forces - both direct and indirect through distortion of the Earth and hence, the Earth's gravitational influence.

- Atmospheric drag.

- Direct solar radiation pressure.

- Albedo radiation reflected from the Earth's surface.

- Earth's infra-red radiation pressure.

Further details of these forces may be obtained from *Roy* (1978), *Wakker et al.* (1983), and *Ehlers* (1993).

3.2.3 Ephemeris modelling errors

The final accuracy of an ephemeris solution is dependent on several factors: the availability and quality of tracking data; the duration of the orbital solution; the size and temporal characteristics of the various force models; and finally, on the software used to generate the ephemeris. Errors therefore can be attributed to mismodelling and noise present in the raw data, about which little can practically be done. Poor modelling can arise from measurement errors in for example, satellite tracking station co-ordinates, atmospheric effects and measurement biases. Alternatively, mismodelling may be due to dynamic factors of gravitational or non-gravitational origin. Gravitational errors are introduced through errors in third body attraction, the gravitational coefficient, μ, and the non-spherical potential coefficients, C_{lm} and S_{lm}. Non-gravitational errors are usually due to atmospheric drag, solar radiation pressure and mass expulsion. This final point generally makes it impossible to determine accurate satellite ephemerides during an orbital manoeuvre. The effect on the accuracy of an orbit of each of these gravitational and non-gravitational contributions are

discussed in detail by *Engelis* (1988) and *Moore and Rothwell* (1990) and will be summarised in the Chapter 4.

3.3 Environmental corrections

An altimeter radar pulse transmitted from the satellite must travel through the Earth's atmosphere before being reflected at the surface to return through the atmosphere to the satellite. The time elapsed during this journey multiplied by the speed of light in a vacuum produces a measure of the optical distance travelled by the pulse. This measure differs from the true distance due to the refractive influence of the atmosphere and reflection characteristics of the surface. In addition, the range measurement may contain oceanographic signal not of relevance to a particular study. Correcting for these effects is therefore necessary if useful information is to be derived from the data.

3.3.1 Atmospheric propagation

The tropospheric and ionospheric layers of the atmosphere make the greatest contributions to retardation of the altimeter pulse. These effects are not constant, however, and because the pulse travels through the same region twice before it is received, the effects of small changes in the optical path length are amplified by a factor of two.

• **Wet and dry troposphere**

The troposphere is the lowest part of the atmosphere, extending from ground level to a height of approximately 11km and contains the greatest mass of air (90% of the atmospheric mass is below a height of 16km (*Rees*, [1990]), presenting the largest optical retardation of a microwave signal. There are two ways in which the troposphere interacts with the altimeter pulse, firstly through the mass and density of dry air through which the pulse must travel, and secondly, the pulse is retarded by atmospheric water

vapour.   Respectively,   these effects are known as the dry and the wet tropospheric corrections.

The dry correction may be derived from an estimate of the mass and density of the column of air through which the pulse must travel.   This may be inferred from the ground level atmospheric pressure, P, in millibar and is expressed by *Cheney et al.* (1991) as

$$\Delta r_{dry} = -2.277 \, P \, (1+0.0026\cos 2\phi)$$

where ϕ is the latitude of the observation point.   This produces a dry tropospheric correction in the range 2.2-2.3 metres,   with an accuracy estimated to less than a few millimetres.

The wet tropospheric correction is not as accurate as it depends upon the amount of water vapour present along the optical path.   There are two means by which this correction may be estimated: direct simultaneous measurement using passive microwave radiometers at two or more frequencies;   alternatively,   from meteorological predictions of the water vapour content at the time and location of the altimeter measurement.   Of these,   the latter can generally resolve only the long wavelength variations in water vapour concentrations without the real time information provided by the former approach which is therefore more reliable.   The primary mission of US Navy's Geosat was to map the marine geoid which has a signal significantly greater than oceanographic signals at comparable wavelengths;   a water vapour radiometer was,   therefore,   not provided. From *Cheney et al.* (1987),   Geosat data must be corrected either by using radiometry from other satellites or by using the Fleet National Oceanographic Centre (FNOC) model,

$$\Delta r_{FNOC} = -2.277(0.05+(1255/T))(e)$$

where T is the surface atmospheric temperature in °K and e is the surface water vapour pressure in millibar.   The FNOC correction was used for all Geosat altimetry in this thesis.

Later satellites do make use of simultaneous radiometry and produce corrections of up to 35cm, Water vapour retardation data are provided for ERS-1 by way of the Microwave Sounder instrument, a passive radiometer which operates at the dual frequencies of 23.8GHz and 36.5 GHz.

• **Ionosphere**

The ionosphere consists of several ionised layers within the Earth's atmosphere, extending from an altitude of 70km up to a several hundred kilometres. Extreme Ultra-Violet (EUV) and X-radiation from the Sun impinge on this layer and have sufficient energy to liberate electrons from the atoms found at this height. Microwaves interact strongly with these free electrons, thereby retarding the signal. The refractive index, n, of this layer is a function of the number density of free electrons, N_e, given by *Kingsley and Quegan* (1992) as

$$n^2 \approx 1 \pm 80.6 \frac{N_e}{v^2}$$

for a microwave frequency of v. The minus sign applies when considering the phase velocity, while for signal such as an altimetric radar pulse, the plus sign is used, i.e. for the group velocity. The range correction is therefore expressed as

$$\Delta r_{iono} = \int_S n \, \mathrm{d}s \approx \frac{80.6}{v^2} \int_S N_e \, \mathrm{d}s$$

where s is the distance along the path S from the altimeter to the sea surface.

As with the wet tropospheric correction, the ionospheric range correction is highly variable and is best obtained by direct measurement using a dual frequency microwave instrument. With the exception of the NASA Radar Altimeter (NRA), all other altimeters to date have been single frequency. Comparison of the return times of the NRA's Ku and C-band radars, provides a reliable estimate of the total electron count (TEC) and hence the optical path length correction. In the case of other altimeters, the TEC can be derived from measurements of solar activity and models describing the

interaction of the received solar radiation with the ionosphere. The TEC may also be estimated from dual frequency satellite tracking systems such as DORIS, PRARE or GPS.

The net result of ionospheric retardation of the radar pulse is a correction which depends upon the level of solar activity at the time of the observation and the time of day, i.e. the incidence angle and subsequent intensity of solar radiation on the upper atmosphere. For Geosat, the TEC models have been shown to have deficiencies during periods of high solar activity (which occurred towards the end of the mission) as well as poorly modelling the fall in TEC on the nightside of the planet. The French Solid State ALTimeter (SSALT) altimeter is single frequency, therefore, ionospheric corrections are determined from TEC models and DORIS tracking data. For ERS-1, ionospheric range delay is calculated from the Bent model, based on the analysis of 12 month running averages of the CCIR sun spot numbers.

3.3.2 Tidal corrections

The largest dynamic variation of the sea surface is attributable to third body tidal attraction, the effects of which must be removed from altimetry in order to investigate the static and non-tidal signals. There are two slightly different tidal motions at the sea surface: the fluid tides of the ocean and the vertical displacements of the solid Earth tides. From *Pugh* (1987), tides in the open ocean have mean amplitudes of up to 50cm, with considerable geographic variation. In contrast, solid Earth tides are generally less than 10cm.

• Ocean tides

As described in chapter 2, the major solar and lunar ocean tides are caused by third body gravitational attraction constrained by local bathymetry and the presence of coastal boundaries. In deep water areas far from any

land mass and of even bathymetry, the vertical component of ocean tides is straightforward to model. The transverse components, however, are more complicated, particularly where it is necessary to include boundary constraints. To compound this problem in deep ocean it is difficult to accurately measure tidal heights, thus making validation of large scale tidal models extremely difficult. In shallow areas, validation is simpler due to the proximity of coastal regions where tide gauges may be fixed and monitored. Finally, for consistent correction of altimetry the models must have global coverage although this does not preclude the use of local tidal models for regional studies.

Tidal models used to correct the altimetry used in this thesis have been derived in part from the analysis of the temporal variability of ocean signal from previous altimeter missions. For Geosat data, *Schwiderski* (1980), has provided the tidal model which has a global rms accuracy of around 4cm, *Woodworth and Hughes* (1995). In shallow water and enclosed seas, such as the North Sea used in chapter 6, this accuracy can be much reduced. The ERS-1 dataset supplied by Eurimage contains an improved version of the Schwiderski tidal model while TOPEX/Poseidon data supplied by Aviso make use of a different tidal model which covers all latitudes to 72°, based on analysis of Geosat altimetry, *Cartwright and Ray* (1990), which is accurate to around 3.25cm globally, *Woodworth and Hughes* (1995). For the dual satellite crossover work in chapters 4 and 6, this model was also used to correct ERS-1 altimetry, thus providing consistent tidal corrections for the two datasets. This does mean, however, that all ERS-1 altimetry above 72° were ignored during these investigations.

Finally, there is an interaction between the solid Earth and the fluid ocean tides through the loading and unloading of the ocean tide on the crust. These result in an additional non-equilibrium tidal signal and correction of the altimeter range is generally performed by incorporation of this signal with the ocean tide correction.

• **Solid Earth tides**

The 'solid' Earth is in fact pliable enough to respond to the same gravitational forces which generate the ocean tides. From *Pugh* (1987), the natural modes of oscillation of the Earth are around 50 minutes (considerably shorter that their oceanic equivalents), resulting in a faster response time to the gravitational forces. The elastic Earth therefore yields quickly to conform to an equilibrium tidal signal without the locally dependent lag times associated with ocean tides. Earth tides can therefore be determined from a function of the second order spherical harmonic which describes the equipotential surface of the varying gravitational field associated with the Moon, Sun and other planetary influences. The tidal amplitude is a proportion of the equilibrium value Q, described by the Love numbers L_1 and L_2 which are elastic constants, quoted by *Pugh* (1987) as being in the range $L_1 = 0.604 \rightarrow 0.630$ and $L_2 = 0.299 \rightarrow 0.310$. The semi-diurnal Solid Earth tidal displacement, Δh_{SE}, is then

$$\Delta h_{SE} = \left(1 + L_1 - L_2\right)\frac{Q}{g}$$

where g is the gravitational acceleration constant ($\sim 9.81 \text{ms}^{-2}$). Love numbers for the other tidal frequencies are not well known due to the difficulty of accurately measuring the crustal displacements and have not been included in altimetric range corrections.

3.3.3 Ocean-atmosphere interaction

The fluid boundary between air and water is susceptible to atmospheric currents and pressures as well as those of the ocean. The influences of the inverse barometer effect and sea state bias must, therefore, be accounted for in the interpretation of altimeter range measurements.

● **Inverse barometer effect**

Section 2.1.4 introduced the inverse barometer effect as the depression of the sea surface by the weight of the column of air resting on it. The instantaneous atmospheric pressure, P, in mBar at the observation point can be obtained from the modelled dry tropospheric estimate, Δr_{dry}, (section 3.3.1) by using the equation quoted by *Cheney et al.* (1991) as

$$P = \frac{\Delta r_{dry}}{-2.277(1+0.0026\cos 2\phi)}$$

where ϕ is the latitude of the observation. Also from *Cheney et al.* (1991), the inverse barometer correction, Δr_{invB}, can then be derived in metres from,

$$\Delta r_{invB} = -0.009948(P-1013.3)$$

● **Sea state bias**

A radar altimeter measures the range from the satellite to the ocean surface. The value obtained is an estimate of the distance to the mean reflecting surface which is generally different to the mean sea surface. From *Chelton et al.* (1989), there are two contributory factors which influence the sea-state bias correction to be applied to altimetric range measurements, electromagnetic bias (EM-bias) and modelling errors introduced by on-board processing. The former of these is associated with the significant wave height at the sea surface which is assumed to present a Gaussian distribution of reflecting facet angles. As ocean wave troughs are generally flat as compared to the crests which are peaked, the true distribution is non-Gaussian and results in stronger reflection from the wave troughs than from the crests. This has the net result of biasing the range measurement by a proportion of the SWH. Wind speed is also thought to influence the EM-bias through differences in surface roughness between crest and trough, as described in theoretical studies by *Barrick and Lipa* (1985), *Srokosz* (1986) and *Rodriguez et al.* (1992) and through

empirical analysis by *Douglas and Agreen* (1983), *Walsh et al.*(1991) and *Carnochan* (1994).

Corrections based on the theoretical studies to date have not yet proven to be as reliable as empirical estimates, with the result that all sea state biases applied throughout this thesis have been derived from empirical corrections. These corrections, Δr_{ssb}, are generally based on a simple linear proportion of the SWH, h_{SWH}, i.e.

$$\Delta r_{ssb} = k_{ssb} h_{SWH}$$

where the constant k_{ssb} is generally expressed as a percentage. For Geosat, the value obtained by *Douglas and Agreen* (1983) of (2.0±0.6)% was implemented, based on analyses of GEOS-3 and SEASAT; for ERS-1 a correction was derived by *Carnochan* (1994) at (5.951±0.057)%; for the NRA and SSALT, the correction is based on the four parameter model,

$$\Delta r_{ssb} = h_{SWH}\left(a + bU + cU^2 + dh_{SWH}\right)$$

where U is wind speed as estimated from the surface roughness measured by the altimeter. The parameters reported by *Gaspar et al.* (1994) are for the NRA: 1.93, 0.368, -0.014, -0.268; for the SSALT: 5.12, 0.233, -0.011, -0.176; for terms *a,b,c* and *d* respectively.

3.4 Mission Profiles

In 1973, a radar altimeter made the first range measurement to the sea surface from space. Flown on board NASA's Skylab, the experiment demonstrated the feasibility of satellite borne radar altimetry to oceanographers. With an instrument accuracy of 1m and a radial orbit accuracy of 10's of metres, the measurements were of little practical use. The launch three years later of GEOS-3 brought routine altimetry one step closer, but with radial accuracies still measured in metres, the data were capable of resolving only the largest of features. August 1978 saw the

launch of SEASAT, the first dedicated oceanographic remote sensing satellite, carrying amongst other instruments, a radar altimeter and a laser retro-reflector for precise orbit determination. Oceanographers finally had decimetre accurate range measurements from a satellite the radial accuracy of which was estimated to be approximately 50cm. Unfortunately, SEASAT suffered a catastrophic failure 100 days later and no further data were obtained.

Over the past ten years, several missions have gathered many millions of altimetry measurements and have permitted many characteristics of the sea surface to be investigated. The following sections introduce the three satellites from which data have been used for the research described in this thesis.

3.4.1 Geosat

The United States Navy's GEOdetic SATellite, Geosat, was launched on March 12, 1985 to an altitude of 785km at an inclination of 108°. The satellite completed two missions in slightly different orbits. The first of these was the 18-month Geodetic Mission (GM), during which time the satellite was allowed to drift in an orbit which never repeated, but had near repeats at 3, 23 and 78 day periods. This orbit produced a dataset which is still unequalled in terms of spatial resolution as the equatorial spacing of the ground tracks was typically less than 5km, and reduced to 2-3km at latitudes of 60°. As a result of this resolution, these data were classified by the US government and only a subset south of 30°S have so far been released into the public domain. Tracked by the 46 stations of the U.S. Defence Mapping Agency, the satellite ephemerides were calculated by the Naval Surface Weapons Centre to an accuracy of around 60cm rms (*Cheney et al.*, 1991). This radial accuracy was on occasions degraded by the passive gravity gradient attitude control system which allowed off nadir excursions of 1 degree or more, as discussed by *Kilgus* (1987). *Holdridge* (1988) states

that the underlying cause of Geosat's attitude problem appears to be related to solar radiation pressure (with small magnetic, atmospheric and gravity field contributions). *Cheney et al.* (1991) state that the attitude excursions of the platform had no effect on the altimetric sea height measurement when the satellite was in its precise tracking mode, but as will be shown in Chapter 5, there is a distinct correlation between the radial measurement and the long term orientation of the satellite orbit to the solar angle.

Beginning on October 1, 1986, Geosat was manoeuvred into it's Exact Repeat Mission (ERM) where it remained until it's failure in January 1990. The nominal orbit for Geosat gave identical ground track coverage to the near 17-day repeat orbit of the prior SEASAT mission which allowed data to be released into the public domain. During this time, the NSWC continued to determine precise ephemerides for the satellite, these however, remain classified and the public domain data made use of orbits determined by the Naval Astronautics Group (NAG) using a reduced tracking dataset from just four stations, three of which were located on in North America, and the fourth in Hawaii. This produced a 3m rms radial accuracy from 2-day discontinuous ephemeris solutions with a one day overlap. Finally, all the orbits during the ERM were recalculated by NASA using an improved geopotential model (Goddard Earth Model T2, GEM-T2) in place of the GEM-10 previously used for the NAG orbits) and data from an additional six tracking stations. The radial accuracy of these orbits was closely linked to levels of solar activity which rose steadily throughout the mission, and as a result, the orbit precision at the start of the ERM was in the region of 10-25cm, rising to 40-60cm by the end (*Cheney et al.,* 1991). Due to the availability of data at the time Geosat data were being used for this research, the inferior NAG orbits were used to obtain sea height information from the altimeter range measurements.

3.4.2 ERS-1

The first European Remote Sensing Satellite (ERS-1) was launched in July 17, 1991 from Kourou, French Guiana into a near polar sun-synchronous orbit, inclined at 98°, at an altitude of around 780km. The satellite, based on the Multi-Mission Platform developed for the French SPOT series of imaging satellites carries a payload of active and passive remote sensing instruments. These include the Active Microwave Instrument which combines the operation of a Synthetic Aperture Radar and a Wind Scatterometer; a passive Along Track Scanning Radiometer which combines an infra red radiometer with a microwave sounder for measurement of sea surface temperature; two tracking systems, the laser retroreflector for use with ground based Satellite Laser Ranging (SLR) systems and the experimental Precise Range And Range-rate Equipment (PRARE) which failed shortly after launch. As a consequence of this failure, ERS-1 orbits are not as accurate as had been hoped. Radial accuracies are reported however which are considerably better than were achieved for Geosat, for example, *Scharoo et al.* (1994) and *Massemann et al.* (1994) report values of 10-20cm. Finally, the other active microwave device is the radar altimeter which is of particular relevance to this thesis.

There are several key stages within the ERS-1 mission, summarised in table 3.3. The initial commissioning phase during which the satellite was placed in a three day repeating orbit, followed by another three day repeat orbit for the first ice-phase, during which the ground tracks were shifted slightly. The main, multi-disciplinary phase then followed during which ERS-1 was placed in a 35 day repeating orbit, thus reducing the equatorial track spacing to 80km. The penultimate scheduled part of the mission is the geodetic phase which will consist of two interlaced cycles of a 168-day repeat pattern. The geodetic phase provides a cross-track sampling of 8.3km at the equator and around 4.2km at 60° of latitude. For the first time in the public domain, altimetry will be available close to the pulse limited

resolution of the altimeter in both along track and cross track directions (Geosat during its GM had a 5km equatorial cross track resolution, at the time of writing, however, these data are still classified but are expected to be released in the coming months). On April 20, 1995, ERS-2, a nearly identical satellite to ERS-1, was launched into a 35-day repeating orbit, 20 minutes ahead of ERS-1, thus giving it a one day ground track separation. During the commissioning of ERS-2, ERS-1 will be operated simultaneously for calibration and validation exercises. There will then follow a 9 month Tandem mission where both ERS-1 and ERS-2 will operate concurrently. After the Tandem mission, it is planned to switch ERS-1 off until such time as ERS-2 fails, at which time an attempt may be made to re-activate ERS-1.

Table 3.2. Operational phases of ERS-1.

	Repeat period	orbits/period	Equatorial track spacing
Commissioning phase	3 days	43	932km
First ice phase	3 days	43	932
Multi-disciplinary phase	35 days	501	80
second ice phase	3 days	43	932
geodetic phase	2x168 days	4822	8.3
Tandem phase (with ERS-2)	35 days	501	80

3.4.3 TOPEX/Poseidon

The first dedicated altimetric satellite, TOPEX/Poseidon, was launched on August 10, 1992 into a 1335km, near 10-day repeating orbit at an inclination of around 66°. It is a joint venture between the Unites States of America's National Aeronautic and Space Administration (NASA) and France's Centre National de Etudes Spatial (CNES), and consists of two different altimeters. The NASA Radar Altimeter (NRA) is a dual

frequency (Ku/C-band) instrument of a conventional design, which, because of the dual frequencies used is able to provide more accurate atmospheric propagation corrections to the range measurement. The CNES Solid State ALTimeter (SSALT) is, as the name implies, a solid state device which requires less power to operate and is considerably lighter and smaller thus making it cheaper to launch. The platform has three different tracking systems: Satellite Laser Ranging (SLR), Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) and GPS (Global Positioning System). SLR operates in the same way as for other satellites, making use of the same global network of laser ranging stations. GPS is a space based positioning system which relies on triangulation to four or more of a network of 24 satellites in 12-hour orbits. DORIS is based on measuring the Doppler shift of an omnidirectional microwave signal to obtain range rate information (PRARE is primarily a measure of the two way range to a ground transponder), from a transponder on board the satellite (DORIS), or to a network of transponders on the ground (PRARE). The DORIS network consists of a large number of permanent and mobile tracking stations which give TOPEX/Poseidon far greater tracking coverage than the current or planned PRARE network. In addition, it has been shown by *Lerch et al.* (1994) that the TDRSS network of communications satellites are capable of coarse satellite tracking (to accuracies of around 1km). Because of TOPEX/Poseidon's altitude, precise orbit determination is less influenced by atmospheric drag and gravity field modelling errors than lower satellites such as ERS-1 and Geosat, orbits from a combination of the three main systems therefore result in rms radial accuracies of around 3-4cm, *Tapley et al.* (1994); as compared to 10-20cm for ERS-1 from SLR tracking and crossover differences.

TOPEX/Poseidon gives oceanographers excellent temporal resolution when observing meso-scale variations and the dynamic component of the sea surface topography, but, the cross track spacing as a result is coarser than most altimetric missions to date. In the repeat period of 9.97 days,

TOPEX/Poseidon completes just 127 orbits and hence the equatorial spacing of each nominal track is over 300km. This spatial resolution is, however, sufficient to achieve the objectives of that part of the mission which gives TOPEX its name, i.e. the ocean TOPography EXperiment.

Chapter 4

Empirical orbit error correction using crossovers

4.0 Introduction

For a variety of reasons, altimetry from the NASA/CNES satellite TOPEX/Poseidon is unusual in that it has an estimated radial accuracy of under 5cm. For all other altimeter satellites to date, radial position of the satellite has been the single largest source of uncertainty in the dataset. To extract the full oceanographic potential from the data, therefore, it has been necessary to develop a number of empirical correction methods to remove this error. This chapter will start by describing the spatial and temporal characteristics of these errors and summarise some of the methods commonly used to reduce their influence on the data. This is followed by a detailed description of two of the most reliable, if computationally intensive techniques: those of single (SXO) and dual (DXO) satellite crossover minimisation. In the process, the validity and consequences of such methods shall be investigated.

4.1 Orbit Errors

Section 3.2 introduced the motion of a low Earth orbiting satellite and the principal forces acting upon it, all of which are imprecisely known. The indeterminacy of the forces acting on the platform results in uncertainties in the estimated location of the satellite some time later. As with most mechanical systems, orbit errors have spatial and temporal characteristics: some errors for example, are correlated to the orientation of the Earth as the satellite passes overhead, while others can be linked to the satellite's or the

orbit's orientation to the Sun, Moon or other planets. These errors all have static or constant, systematic elements in addition to harmonic, time and spatially varying components as well as anharmonic variations due to mismodelled and unmodelled contributions. It is these final, anharmonic errors which limit the length of arc for which a continuous ephemeris may be accurately determined through modelling of the satellite forces. As force modelling and the quality and quantity of satellite tracking data have improved, so the length of arcs have increased and confidence in the precise location of the satellite improved

The mathematical description of radial orbit errors (of particular importance to altimetry), is generally expressed as a sum of constant, harmonic and secular-harmonic terms. For all satellites there exist orbit errors which are strongly correlated to the spatial orientation of the Earth and its gravity field below. For an exactly repeating orbit, this strong correlation may be simply expressed in terms of time due to the periodic nature of the orbit and its repeat pattern. Therefore, all the relevant orbit errors can be effectively described without the direct use of geographic variables such as latitude and longitude. Detailed descriptions of time dependent orbit errors have been presented by a number of authors, including *Engelis* (1988) and *Moore and Rothwell* (1990), as the summation of several components attributable to errors in, the gravity field (Δr_{grav}); the initial position vector of the satellite(Δr_{ipv}); and surface forces such as solar radiation pressure, Δr_{srp}, and atmospheric drag, Δr_{drag}. This may be written to first order as,

$$\Delta r = \Delta r_{grav} + \Delta r_{ipv} + \Delta r_{srp} + \Delta r_{drag} \tag{4.1}$$

Each component has a different dependence upon the orbital elements of mean anomaly, M, and the orientation of the satellite to the Earth, $\psi_{k,m}$, which, for integers k and m, may be written in terms of the Keplerian elements defined in section 3.2, and the Earth orientation, θ, relative to the first point of Aries, as,

$$\psi_{km} = k(M+\omega) + m(\Omega - \theta) \tag{4.2}$$

The precise form of the components of 4.1 are given by *Moore and Rothwell* (1990) for terms up to and including sinusoidal contributions of $2M$ which are ignored in the following summary as this component is generally significantly less than the dominant bias and one cycle per revolution terms. Expressed in terms of the mean anomaly and the time t relative to some fixed time in the orbit t^*, the radial orbit error components are

$$\Delta r_{grav} = A_{g1} + A_{g2}\cos M + A_{g3}\sin M + \left(t - t^*\right)\left(A_{g4}\sin M + A_{g5}\cos M\right)$$
$$+ \sum_{k,m}\left[C_{km}\cos\psi_{km} + S_{km}\sin\psi_{km}\right] \tag{4.3}$$

$$\Delta r_{ipv} = A_{i1} + A_{i2}\cos M + A_{i3}\sin M + A_{i4}\left(t - t^*\right)\sin M \tag{4.4}$$

$$\Delta r_{srp} = \left(t - t^*\right)\left(A_{s1} + A_{s2}\sin M + A_{s3}\cos M\right) \tag{4.5}$$

$$\Delta r_{drag} = \left(t - t^*\right)\left(A_{d1} + A_{d2}\cos M\right) \tag{4.6}$$

where the parameters A_{g1}, etc. are related to errors in the coefficients of the spherical harmonic expansion of the gravity field, C_{km} and S_{km}; and are functions of the orbital elements, a, e and i (the orbital elements may be assumed constant for a particular satellite orbit). The coefficients $A_{\alpha j}$ (α='i','s','d','g') are constant throughout a single orbital solution but may have little or no correlation between different solutions, *Chelton and Schlax* (1993).

• **Orbit error approximation**

With the exception of bias terms, the radial orbit errors can be assumed to have sinusoidal components which are of constant or time dependent amplitude. The secular bias terms in Δr_{srp} and Δr_{drag} on the other hand, change slowly over a period of days and will therefore not be of concern in the short arc orbit correction to follow in this chapter, although they are of crucial importance in long arc determination. The final summation term in Δr_{grav}, however, consists of a great many individual frequencies (some

of which can be shown to be in resonance) where different elements of the error share common frequencies or integer multiples thereof. Spectral analysis of these error terms by *Chelton and Schlax* (1993), *Carnochan* (1994) have shown that the majority of the power is concentrated at or near one cycle per revolution, i.e. of the same period as the harmonic and secular-harmonic errors present due to the other error sources. From *Tai* (1991), the total time dependent radial orbit error of a low Earth orbiting satellite may therefore be approximated over distances of a few thousand kilometres up to several revolutions by a simple sinusoid and bias expression, i.e.

$$\Delta r = A \sin \dot{M}(t - t_o) + B \cos \dot{M}(t - t_o) + C \qquad (4.7)$$

For shorter arc segments, this may be further simplified to a bias and tilt or even a bias. Most arcs described in this thesis are between a quarter and half a circumnavigation (10000-20000km) in length and are therefore best described by the sinusoid and bias approach. In this chapter, equation 4.7 is sufficient to describe the radial orbit error contamination of satellite altimetry, but in the following chapter, the full expansion of equation 4.1 will be required.

4.2 Single satellite crossovers

The intersection of ascending and descending altimeter ground tracks provides multiple measurements of the same location within which the errors present in each measurement may be significantly different and therefore distinguishable. A crossover difference is defined simply by the subtraction of two height measurements of the ocean surface from each arc at the intersection of an ascending and descending altimeter track; explicitly,

$$X = h_i - h_j \qquad (4.8)$$

where the h_i and h_j represent the ocean heights as measured at the crossover location on the ascending and descending arc respectively.

From a simple model of the orbit error consisting of only a bias and sinusoid (equation 4.7) to represent the dominant components, it can be seen from equation 4.8 that the orbit error contamination of a crossover difference measurement is of the form

$$\Delta X = \left(A\cos M_a + B\sin M_a + C\right) - \left(A\cos M_d + B\sin M_d + C\right) \qquad (4.9)$$

From symmetry, the mean anomaly of the ascending observation, M_a, is approximately the negative of that for the descending arc when measured from the northern or southern latitudinal extreme, i.e.

$$M_a \approx -M_d$$

Therefore the orbit error contamination described by equation (4.9) becomes

$$\begin{aligned}\Delta X &\approx A\cos M + B\sin M + C - A\cos(-M) - B\sin(-M) - C\\ &\approx 2B\sin M\end{aligned} \qquad (4.10)$$

Cancellation of the bias and cosine terms has the result that these components of the orbit error cannot be resolved in a crossover difference. The absence of any bias information leads to an ambiguity in the absolute height measurements that can be inferred from crossover data while the absence of the cosine terms has the effect of masking a potentially large error component, as will be demonstrated in section 4.4. Despite this, a crossover difference does contain a significant proportion of the orbit mismodelling error present in altimeter data.

The above description assumes that the only difference between two altimeter observations of the same geographical location is due to orbit error, that is both observations are made simultaneously. This is obviously not the case as any change in sea height at a particular location during the time it takes for the satellite to make its second pass over the area will be absorbed directly into the crossover difference. Sea height fluctuations include tidal variations as well as longer term, mesoscale variability due to changes in ocean current and atmospheric loading effects. In order to act as an accurate estimate of orbit error the effects of these

variations must be kept to a minimum which may be achieved through the imposition of a time limit between ascending and descending passes. The time limit is determined by the decorrelation time of contaminating signals.

Where data from ascending and descending arcs are obtained from different orbital solutions the terms in equation 4.9 no longer cancel exactly to produce the simple expression of equation 4.10. The cosine and bias terms are likely to be correlated from one arc to another resulting in partial cancellation of these terms. As no two orbital solutions determined in a consistent manner from similar tracking data will ever be completely uncorrelated some ambiguity will always remain in these terms. This is particularly the case for consecutive orbits where some smoothing of the boundary is common, *West* (1986). Therefore, when approximating the sinusoidal radial orbit error through minimisation of single satellite crossover differences, it is not sufficient to solve only for the sine term of equation 4.10 when the time interval between ascending and descending terms exceeds that of the orbital solution.

Having established the nature of a crossover difference measurement, and its relation to the radial orbit error, it may be exploited to improve the relative accuracy of satellite altimetry and thus the quality of oceanographic results obtained from it. This requires two stages: the first being the precise location and determination of the crossover followed by minimisation of the error component resolved by the crossover difference.

4.2.1 Crossover location

Analytic description of a satellite orbit and its ground track is a complex and difficult process and in many cases is unnecessary if an empirical approach can be used with sufficient accuracy. The short term motion of a satellite is smooth and easily predicted thereby permitting the iterative
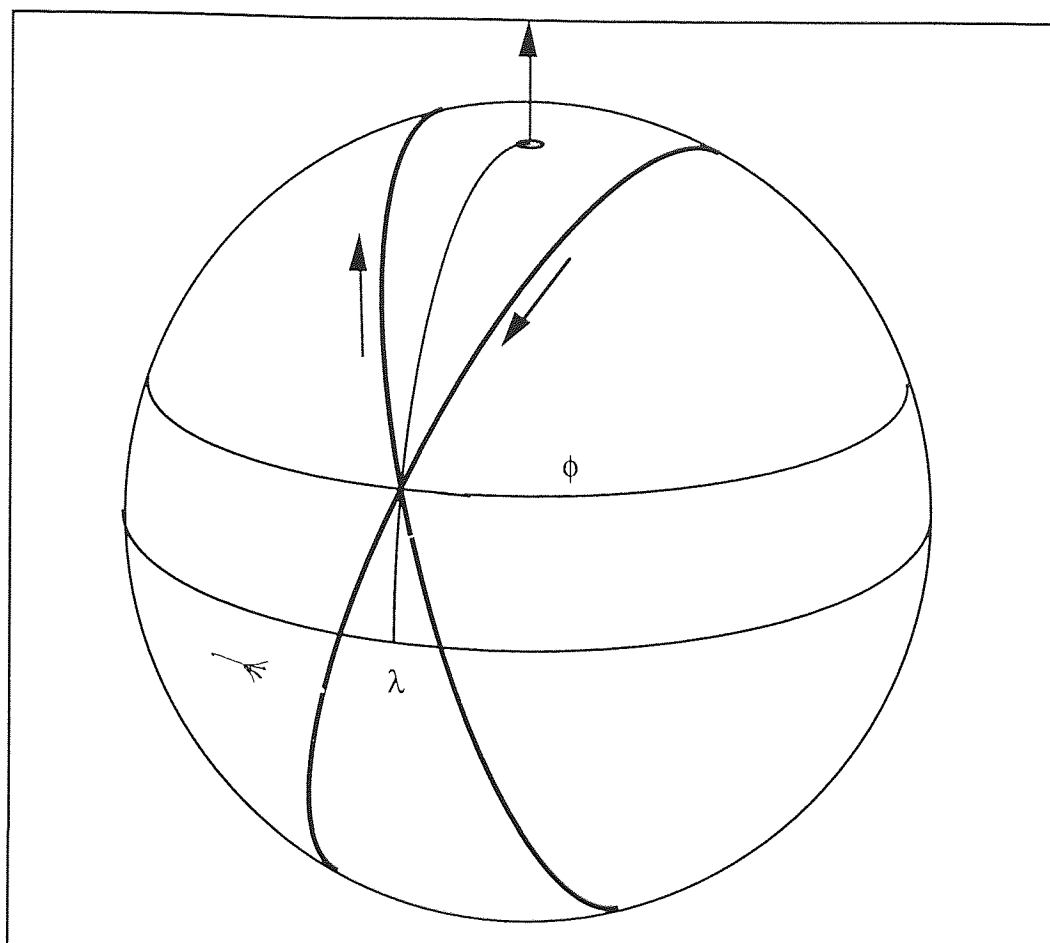
Figure 4.1.   Geometric location of a single satellite (Geosat) crossover.

determination of a crossover location by a process of refinement.   An approximate location for the crossover can be found from the application of symmetry to the shape of a near circular orbit and the knowledge of whether the crossover between two arcs is likely to be in the northern or southern hemisphere.   From figure 4.1, it is clear that the longitude of the crossover occurs at approximately the midpoint of the equator crossing locations appropriate to the ascending or descending nature of the particular arcs,   where the equator crossing longitudes are obtained through interpolation of the orbital ephemeris dataset.   The ephemeris dataset is then searched to find the times and a more precise estimate of the crossover location.   Using this information the altimeter dataset then provides the necessary refinement to these estimates to generate an accurate crossover

location and time for each arc by iterating until the result falls within predefined bounds. Where there is sufficient altimetry, the height at this time is determined by a linear or cubic interpolation of adjacent points, using the rejection criterion that there must be at least six valid altimeter observations in the ten seconds surrounding the crossover time. In the same manner, the latitudes and longitudes of the crossover are refined until, finally, the interpolated heights from each arc are subtracted to yield a single satellite crossover difference.

4.2.2 Sinusoidal minimisation of crossover differences

Empirical orbit error correction in this thesis is based on approximation of the radial error over distances of between a quarter and half a revolution with a sinusoid of the form

$$\Delta r = A \cos M + B \sin M + C \tag{4.11}$$

This is not the only representation that has been used in the past, there being numerous examples of polynomial approximation of the radial error using bias, bias and tilt and quadratic formulae, as described for example by *Marsh et al.* (1982), *Fu and Chelton* (1985) and *Cheney et al.* (1986). Even in the case of a quadratic fit, these have been shown by *Tai* (1988) to seriously misrepresent the error over distances greater than a few thousand kilometres. The first attempts by *Rummel and Rapp* (1977) to use a sinusoid to approximate the radial error were abandoned as unpromising, but *Tai* (1988) introduced the method of constrained sinusoidal crossover adjustment upon which crossover minimisation in this thesis is based.

Crossover minimisation is performed as a least squares problem, where the single satellite crossovers can be defined in terms of sinusoidal approximations to the radial orbit error plus some oceanographic and other signals, to be treated as noise, η, i.e.

$$X_{ij} = \left[A_i \cos\left(\frac{2\pi t_i}{T}\right) + B_i \sin\left(\frac{2\pi t_i}{T}\right) + C_i \right]$$

$$- \left[A_j \cos\left(\frac{2\pi t_j}{T}\right) + B_j \sin\left(\frac{2\pi t_j}{T}\right) + C_j \right] + \eta$$

where A_i, B_i and C_i are the sinusoidal coefficients of the radial orbit error on the ascending arc i and A_j, B_j and C_j are the same for the descending arc j. The period of the sinusoid is set by the orbital period, T, and the precise phase is defined by the times, t_i and t_j, since some fixed reference for the ascending and descending arcs respectively.   The crossover difference, X_{ij}, is therefore defined by the difference of the two sinusoids pertinent to the two crossing arcs, plus the ocean signal/noise term, η.   A system of equations can therefore be formed from the data, consisting of one equation per crossover and one triplet of terms $(A,\ B$ and $C)$ per arc.   This may be more conveniently expressed in matrix notation as

$$\mathbf{H}\mathbf{x} = \mathbf{z}$$

where the sum of the sin and cosine terms are arranged in $\mathbf{H}$, the crossover differences from the data are placed in $\mathbf{z}$ and the coefficients in $\mathbf{x}$.   By denoting the transpose as ^{T} and the inverse as $^{-1}$ *Moritz* (1980) describes how terms in the vector $\mathbf{x}$ may be therefore estimated as the vector through

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{z} \qquad\qquad (4.12)$$

Equation 4.12 is not, however, sufficient to produce a unique solution because of the way in which the crossover is defined: for a single crossover, there are an infinite number of possible values for the individual heights of the ascending and descending measurements, each of which may yield exactly the same crossover difference.   For example, the addition of a bias to both measurements would not change the value of the crossover difference, thus giving crossover adjustment both its strength and its weakness.   That is, a crossover is insensitive to the geoid height and any errors in the geoid model above which the ocean height might be calculated because any error

is constant for a given location and cancels exactly, leaving only a relative measure of the difference in height at that point.

The relative measures of ocean height and orbit error result in a singular solution to equation 4.12. To produce a unique solution one of two approaches must be used: either by fixing one arc through assigning it a zero or known orbit error; or by imposition of a constraint on the whole solution. As the orbit error of the dataset is unknown, the only way of fixing the orbit error of one arc is to assume it to be precise where the error coefficients A and B equal zero. The choice of arc is arbitrary and therefore open to the vagaries of poorly modelled arcs which could have the effect of contaminating the entire dataset. This potential unreliability makes the arbitrary choice of a fixed arc unattractive in comparison the use of a constraint based on the estimated accuracy of the data.

From *Lawson and Hanson* (1974), the non-uniqueness may be removed by applying the following constraint,

$$\sigma^T I x = 0$$

where I is the identity matrix and σ is the vector of weights relating to the estimated accuracy of individual terms in $\mathbf{x}$. The solution of equation 4.12 therefore takes the form,

$$\hat{x} = (H^T H + \sigma^2 I)^{-1} H^T z \tag{4.13}$$

In the case of single satellite crossovers using GEOSAT data, all arcs were assumed to be of the same accuracy and no favour was given to the sine or cosine terms for each arc. The weights in σ were therefore identical and estimated from the overall accuracy estimate of the NSWC orbits as described by *Cheney et al.* (1991) as of the order 60cm. This results in the constraint that the total radial orbit error in the crossover should not exceed approximately one metre; this does not mean, however, that individual terms in the empirical equation 4.11 cannot exceed 1m. As these terms are frequently highly correlated, the values of A and B may both exceed 1m

without producing an rms correction to the data in violation of the estimated accuracy. A constraint was therefore chosen of 3m for each term, corresponding to a weight of 0.1 (i.e. $\sim 3^{-2}$). The value of this constraint was then reduced on an experimental basis until the solution became unreliable. By reducing the constraint, the corrections became slightly larger at first, reducing the rms crossover after fit and then stabilised, showing little improvement at lower constraint values. The improvement of fit was at the expense of the correlation between terms and a compromise constraint of 0.01 was then selected which was small enough to constrain the solution without over suppressing the corrections, thus producing a lower final rms crossover difference, but without significantly raising the mean correlation. A value of 0.01 corresponds to a loose constraint of 10m of the orbit error terms and produced corrections in line with the 60 cm expected accuracy of the data of 57cm rms.

4.3 SXO methodology

A subset of Geosat crossovers south of 35°S were extracted from the crossover data record (XDR) released by NOAA and described by *Cheney et al.* (1991) for use in validation of empirical single satellite crossover minimisation. The methods used are based on those of constrained sinusoidal correction described by *Tai* (1988) and summarised in the previous section, and the generation of a reference dataset to allow sequential correction of individual arcs, described by *Milbert et al.*(1988). This section will deal with the exact methods used to improve the radial accuracy of Geosat XDR arcs of up to 11500km in length in order that they may be used for sea level variability studies to be described in the following chapter.

4.3.1 Generation of a reference SXO dataset

The Geosat XDR dataset consists of almost 65 million crossovers, of which over 44 million are supported by valid altimetry to allow calculation of a crossover difference. Simultaneous solution using 13000 arcs from the 2.5 year dataset would obviously be the best way of producing a consistent, high quality correction. The volume of data is however prohibitive and an alternative means of determining the 26000 orbital correction terms must be sought. Piecewise correction involves the use of a single reference dataset upon which a simultaneous correction may be made using the resources available, followed by sequential correction of any remaining data via crossovers with this reference set. This implies that most crossovers will be determined with positioning data derived from different orbital solutions, as discussed earlier in this chapter. Fortunately, as the dataset contains crossovers from observations up to two and a half years apart, it may be assumed that the effect of using different orbit solutions as well as the presence of ocean signal in the data will be randomly distributed about the error signal being sought. Using crossover difference differences in the next chapter, this assumption will be validated.

From initial trials using equation 4.11, the bias term was found to be too highly correlated to the cosine term to be of any independent significance. A simple sinusoid with no discrete bias component was therefore used to model the orbit error, which due to the use of short, 11500km, arcs should also absorb any orbit error bias to be found in the crossovers.

Following an approach set out by *Milbert et al.* (1988), a reference dataset was formed from the first 25 days of the Geosat GM, corresponding to slightly in excess of the 23-day near repeat of the GM. Application of a 3σ rejection level on the dataset of 358 arcs produced 19367 crossovers, from which a solution was obtained. Table 4.1 summarises the results from

simultaneous solution of the 716 orbit error terms using a constraint of $\sigma^2=0.01$.

Table 4.1. Simultaneous constrained sinusoidal correction of 25-days of Geosat XDR data, solving for 716 orbit error terms on 358 arcs up to 11500km long.

Dataset	No. of SXO's	Mean	RMS
Raw crossovers	19367	-14.8cm	77.7cm
corrected ($\sigma^2=0.01$)	19125	-1.3cm	12.1cm

After rejection of 1.2% of the original crossovers greater than 2m, the improvement in crossover differences following removal of a sinusoid is evident. Additionally, from the reduction in mean crossover differences, it is clear that the cosine term is sufficient in this case to remove any bias contamination without solution of additional terms. From examination of the sinusoidal correction terms, some of the deficiencies of the long arc solutions used to generate the XDR dataset become apparent. Plotting sine and cosine coefficients against descending equator crossing longitude of the relevant arc, a clear geographically correlated signal can be seen in figure 4.2 for the cosine coefficients which appears to be absent from the sine terms. Also clear from the cosine terms in particular, is a systematic difference in the coefficients on alternate days.

4.3.2 Sequential orbit error correction

Between the 25 day reference set of 358 arcs, there are an average of 60 crossovers with each of the remaining arcs of the XDR dataset, thus providing sufficient information to correct each arc on an individual basis. This was achieved by least squares minimisation similar to that used in section 4.3.1: in this case, however, the reference arcs are fixed and

assumed free of orbit error. The solution of each is therefore limited to the inversion of a simple 2x2 matrix. From an rms crossover difference before correction of typically 80-100cm on each arc, sequential correction of the data produces a dataset of rms value in the range 15-20cm rms. This
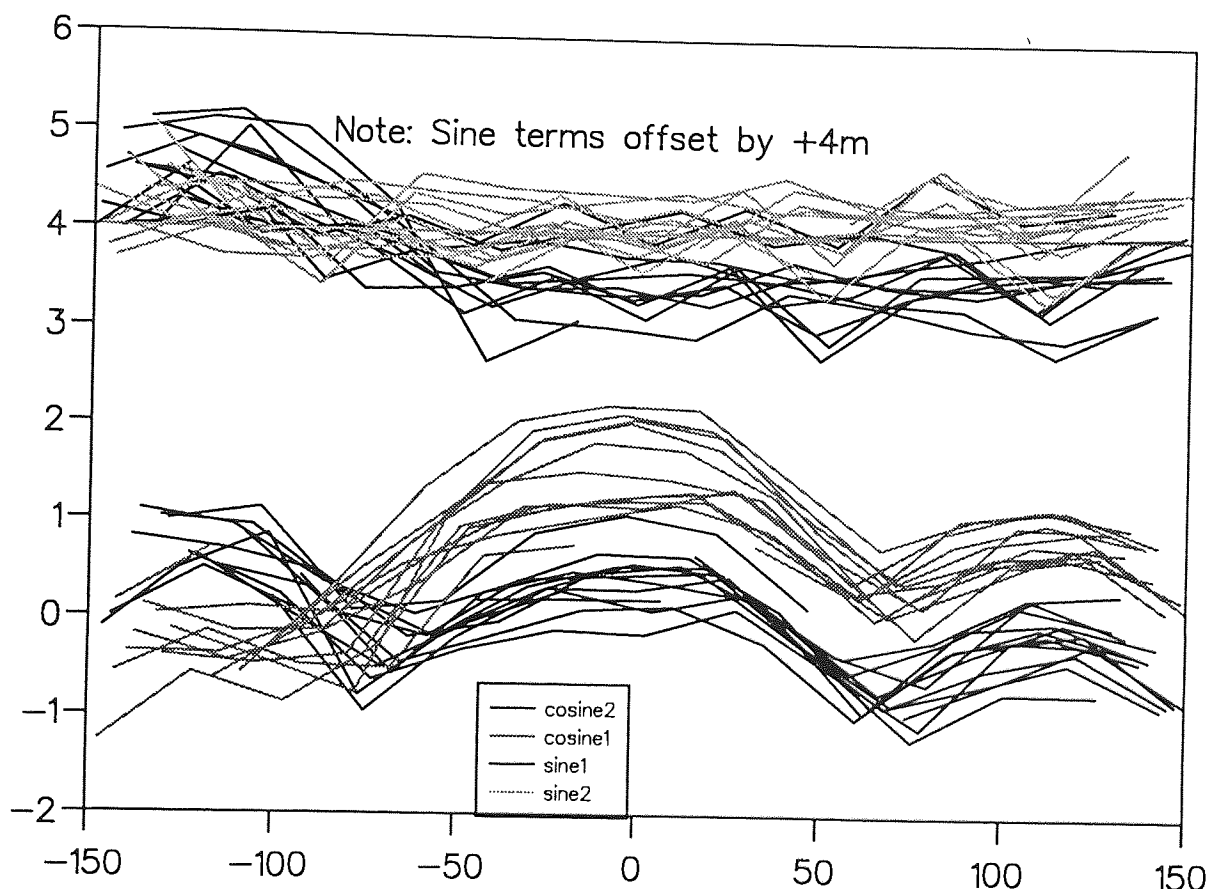
Figure 4.2. Sine and cosine orbit error correction coefficients plotted against equator crossing longitude of the descending arc for alternating odd and even days, illustating the geographical dependence of these terms.

improvement was found to be consistent in time and shows no particular degradation as the interval between the reference period and the epoch of a particular arc is increased. In space however a significant geographically correlated orbit error was identified in the raw data which was found to persist after correction, albeit at a much reduced level. These findings are illustrated in figures 4.3 and 4.4, where raw and corrected crossover differences have been averaged into $10°$ latitude by $20°$ longitude bins. Before correction an overall mean of -14.5cm was present, indicating a consistent difference in heights measured on ascending and descending arcs

at identical locations. On a local level, figure 4.3 clearly shows that in a number of locations the mean crossover difference approached one metre in magnitude, with a particularly distinct longitude dependence. Such a signal was found both in the full dataset illustrated as well as in the reference dataset, thus suggesting a static geographically correlated orbit error associated with the direction of motion of the satellite. Following
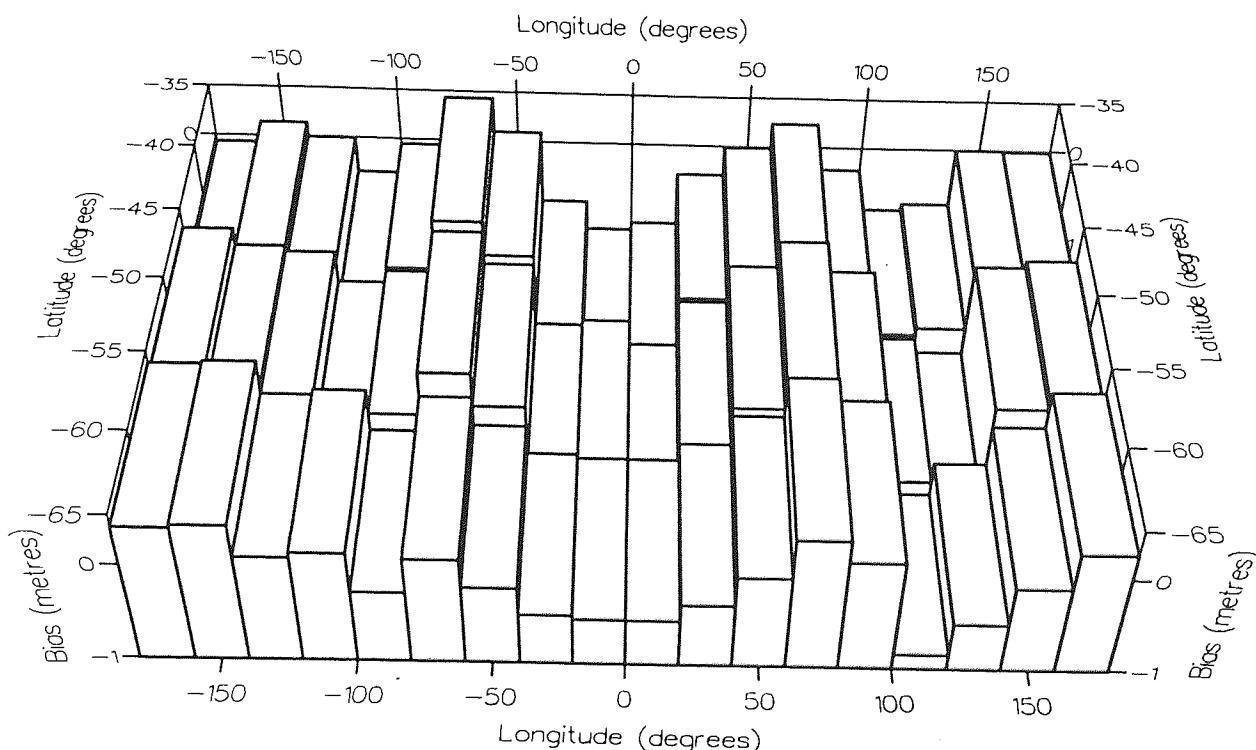
Figure 4.3. Uncorrected crossover residual biases, averaged into 20° longitude by 10° latitude bins. Mean calculated over the whole area = -14.5 cm.
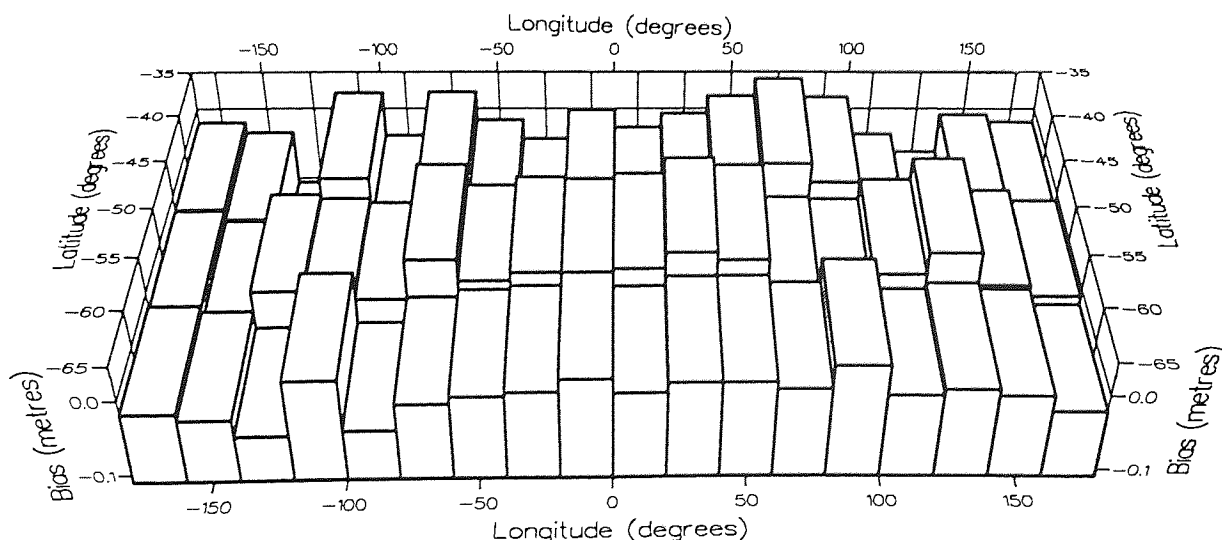
Figure 4.4. Empirically corrected crossover biases, averaging into the same bin as figure 4.3, with an overall mean of 0.1 cm.

sinusoidal correction (figure 4.4) the overall mean was reduced to 0.1cm between latitudes 35°S and 65°S, but still showed significant local discrepancies of 5-10cm in places. This equates well with the observation of a geographical trend in the cosine correction terms (which also absorb any bias in the crossovers) found in the reference set solution.

By plotting the sine and cosine terms produced in these solutions against equator crossing time, figure 4.5a illustrates the presence of a two day signal. This corresponds to the duration of each orbit solution, *Cheney et al.* (1991), of two days and demonstrates that while the errors from different solutions are not equal, there is a correlation between the radial orbit error in arcs separated in time by more than the period of each solution. The correlation plot of sine terms in figure 4.5b, however demonstrates a daily correlation pattern, thus illustrating different sensitivities of the two terms to the orbit errors present. These findings are thought to be an artefact of the orbital solution used to generate the XDR dataset as there is little physical basis for such a strong signal and no theoretical description of a 48 hour correlation of orbital terms has previously been described. In line with previous findings, these results clearly indicate that a significant improvement is possible in Geosat orbital accuracy through the minimisation of single satellite crossovers. The presence of a geographically correlated orbit error which is not cancelled by the determination of crossover differences, reported by *Jolly and Moore* (1994a) has not been confirmed by any independent source, but it is thought likely that the error is attributable to mismodelling of the gravity field and its effect on the calculated height of the satellite. This is different from a geoid error which cancels exactly except were the geographical co-ordinates of the satellite nadir are seriously in error for ascending or descending measurements. The theme of geographically correlated errors will be developed further in the following section with regard to dual satellite crossovers and in the next chapter through crossover difference differences (CDD's).
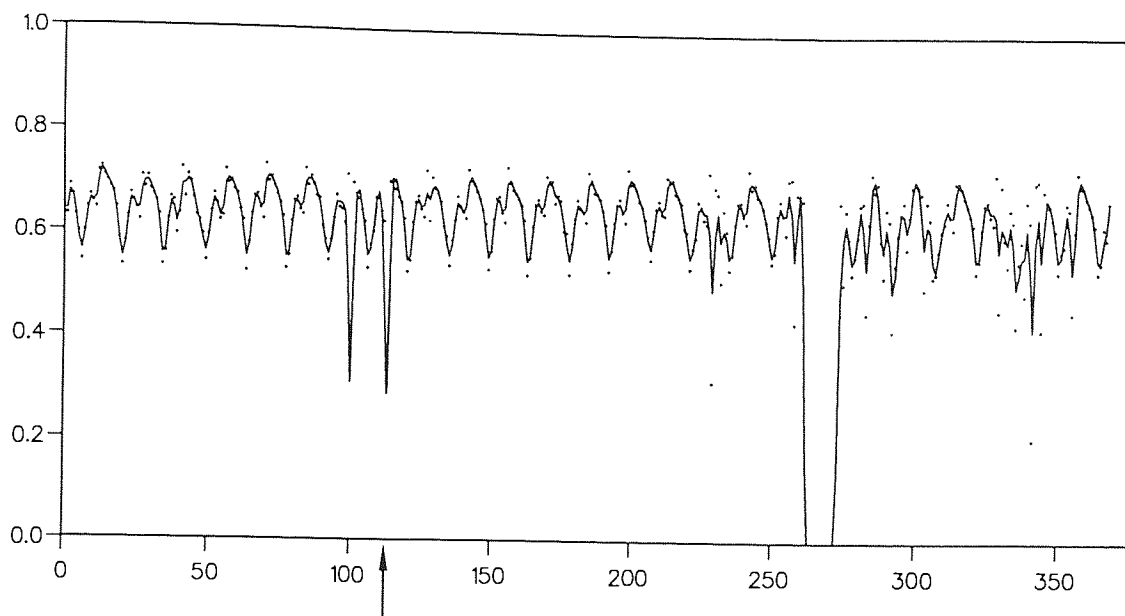
Figure 4.5a. The correlation of cosine empirical sinusoid correction terms for Geosat
XDR data in the Southern Ocean.   Correlation evaluated with arc number 113
shown as an arrow, from inverrsion of the normal matrix.
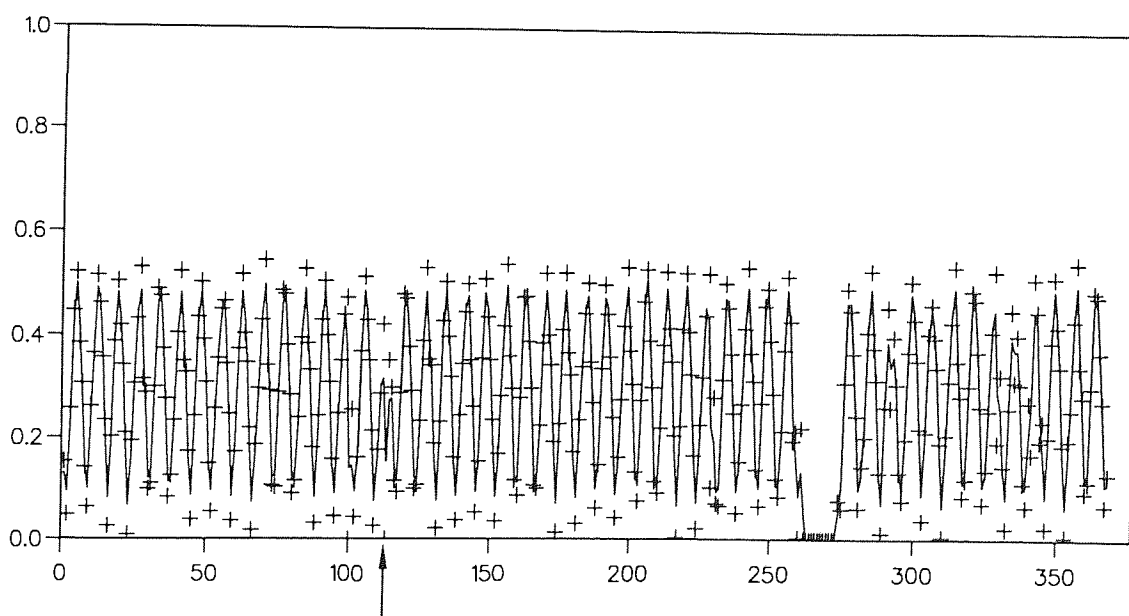
Figure 4.5b. As for figure 4.5a, for the sine coefficient.

4.4 Dual satellite crossovers

With the launch in 1991 of ERS-1, followed in 1992 by TOPEX/Poseidon,
there have been two concurrent altimetric missions gathering data on the
ocean surface for the first time.   Of these satellites, TOPEX/Poseidon is at a

greater altitude that ERS-1 (1335km compared with 785km) which makes modelling of atmospheric and gravitational forces considerably more reliable. Combined with the three tracking systems used on the US/French mission as opposed to ERS-1's SLR only tracking, the radial accuracy of TOPEX/Poseidon is 3-4cm rms [*Tapley et al.* (1994)], considerably less that ERS-1's 10-20cm rms [*Scharoo et al.* (1994)]. Firstly by *Born et al.* (1986), and then with specific reference to ERS-1 and TOPEX/Poseidon by *Shum et el.* (1990), dual satellite crossovers have been suggested as a possible means of improving ERS-1's radial accuracy to a similar level as TOPEX/Poseidon. Defined in a similar way to single satellite crossovers, a dual crossover, D, may be written as

$$D = h_T - h_E$$

where h_T and h_E are the altimeter measured ocean heights at a crossover location from each satellite. In common with single satellite crossovers, any static signal or error common to both observations, such as the geoid, cancels exactly during dual crossover determination. However, the cancellation of some orbit error terms demonstrated in equation 4.10 does not now occur. In the case of the dominant one cycle per revolution radial orbit errors for each satellite, the resultant error in the dual crossover can be expressed as

$$\Delta D = \left(A_T \cos M_T + B_T \sin M_T\right) - \left(A_E \cos M_E + B_E \sin M_E\right)$$

As stated in chapter 3, there is a bias in any altimetric measurement which may be attributed to the processing electronics on the spacecraft, specific to each altimeter. In calculating a dual crossover difference, the biases of the two altimeters used cancel incompletely, leaving a relative bias in all dual crossovers; the value of this bias is dependent upon whether an NRA - ERS-1 or an SSALT - ERS-1 dual crossover difference (DXO) has been determined and its value provides a valuable independent check on the calibrated estimate for each altimeter. Bias estimates are approximately -45cm for ERS-1 (*Francis* [1993] and *Lam et al.* [1993]), -15cm for the NRA

and 2cm for the SSALT (*Christiansen et al.* [1994] and *Menard et al.* [1994]), thus providing a priori bias estimates of 30cm and 47cm between ERS-1 and each of the NRA and the SSALT respectively.

Explicitly therefore, to approximate the ERS-1 radial orbit error (Δr_{ERS-1}) as a one cycle per revolution sinusoid and assuming TOPEX/Poseidon orbits to be precise and accurate, it is necessary to solve for the following terms,

$$\Delta r_{ERS-1} = A_i \cos M_{ERS-1} + B_i \sin M_{ERS-1} + b_{T(P)}$$

where A_i and B_i are arc dependent sinusoidal orbit error terms for an ascending or descending arc i, while $b_{T(P)}$ describes the two relative altimeter biases, common to all DXO's between ERS-1 and either the NRA or the SSALT.

Solution of these terms is obtained in a similar manner to that described for Geosat in section 4.3.2 where TOPEX/Poseidon data become the fixed reference set; solution may then be performed sequentially for the sinusoidal ERS-1 components. To estimate the relative biases, however, a simultaneous solution of as many arcs as possible was undertaken, where the normal matrix was of an augmented, block diagonal form.

Before solution, however, the crossover must first be located, a task more complicated than in the case of single crossovers due to the fact that dual crossovers do not occur at the longitudinal midpoint of the equator crossings.

Dual crossover location

A full discussion of the process of dual crossover location may be found in *Ehlers* (1993), a summary of which follows.

ERS-1 is in a near polar retro-grade orbit of 98° inclination, while TOPEX/Poseidon follows a pro-grade 66° inclination orbit. Consider the case of a descending ERS-1 arc crossing the path of that part of a TOPEX/Poseidon arc north of the Equator. That is, where the descending

equator crossing of ERS-1 is between the ascending and descending Equator crossing longitudes of TOPEX/Poseidon - this case is illustrated in figure 4.6.
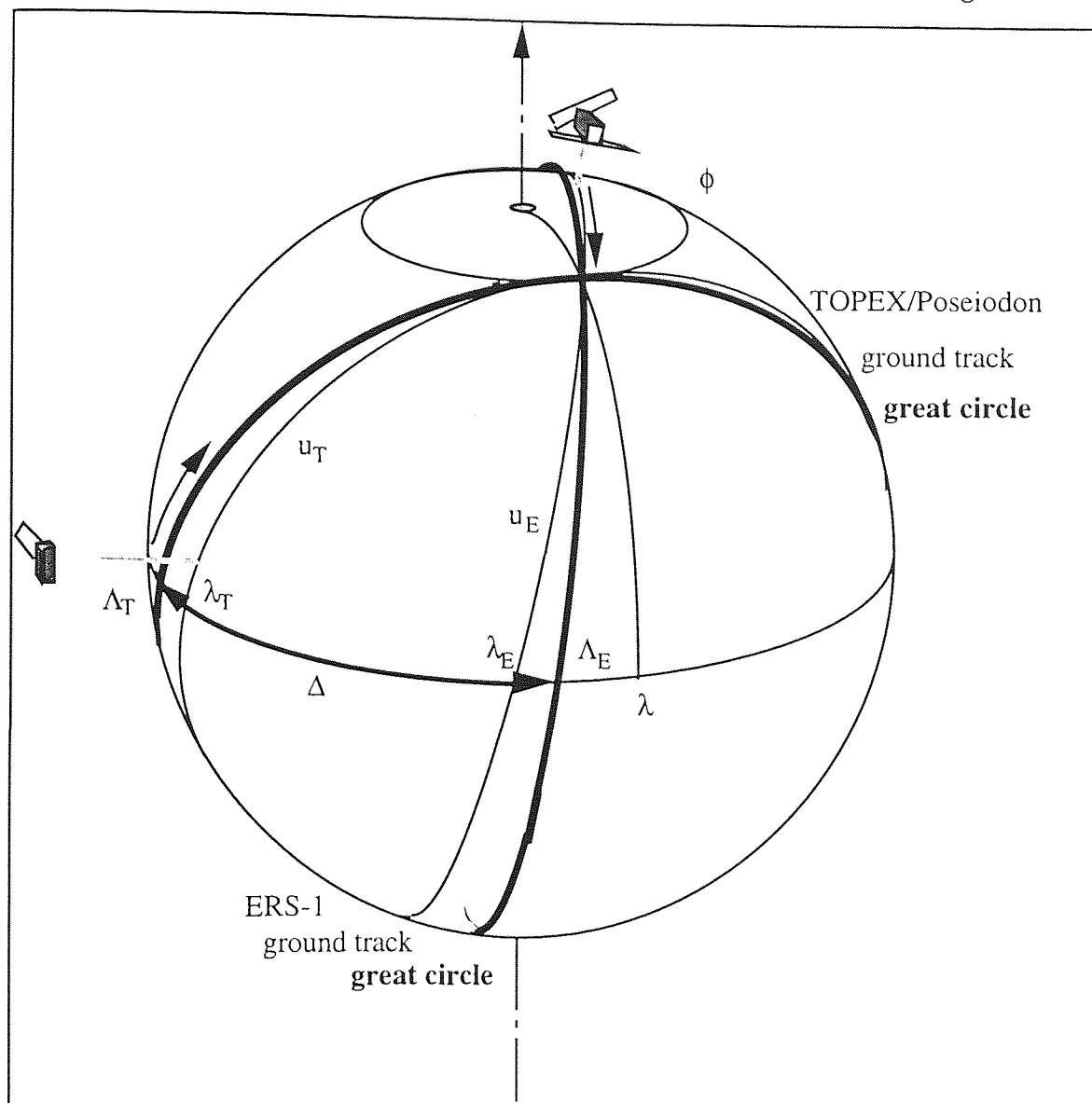
Figure 4.6. Geometric location of a particular dual crossover between a descending ERS-1 track and a TOPEX/Poseidon track in the Northern hemisphere.

Also shown on figure 4.6 are the great circles formed at the intersection on a spherical Earth of the orbital planes of the satellites at the exact times, for each satellite, they pass the dual crossover location. The longitudinal separation, Δ, of the intersections of the great circles with the equator, denoted as Λ_T and Λ_E, may be found from the equator crossing longitudes of each arc, λ_T and λ_E, and the time interval, τ_T and τ_E, between the dual

crossover point and the equator. The subscripts T and E denote ERS-1 and TOPEX/Poseidon respectively, and

$$\Lambda_E = \lambda_E + \tau_E \dot{\omega} \tag{4.14a}$$

$$\Lambda_T = \lambda_T - \tau_T \dot{\omega} \tag{4.14b}$$

where $\dot{\omega}$ is the mean angular rotation rate of the Earth and the difference in sign of equations 4.14a and b is a consequence of the pro-grade orbit of TOPEX/Poseidon and the retro-grade orbit of ERS-1. The separation of these longitudes is

$$\begin{aligned}\Delta = \Lambda_E - \Lambda_T &= \left(\lambda_E + \tau_E \dot{\omega}\right) - \left(\lambda_T - \tau_T \dot{\omega}\right) \\ &= \lambda_E - \lambda_T + \left(\tau_E - \tau_T\right)\dot{\omega}\end{aligned} \tag{4.15}$$

From the spherical cotangent formula,

$$\cot a \sin b = \cot A \sin C + \cos b \cos C$$

the value of u_E may be determined from

$$\cot u_E = \frac{\cot i_T \sin i_E + \cos \Delta \cos i_E}{\sin \Delta} \tag{4.16}$$

A similar expression may also be derived for u_T through the identity

$$\sin i_E \sin u_E = \sin \phi = \sin i_T \sin u_T$$

which gives a value of

$$u_T = \sin^{-1}\left\{\frac{\sin i_E \sin u_E}{\sin i_T}\right\}$$

and hence the first estimate of the dual crossover times on each arc may be derived from

$$\tau_E = \frac{u_E}{2\pi} T_E$$

$$\tau_T = \frac{u_T}{2\pi} T_T$$

the process from equation 4.16 is then iterated until adjustments to the times τ_E and τ_T are smaller than some chosen limit.

The other three combinations, covering ascending ERS-1 tracks in the northern hemisphere and all dual crossovers in the southern hemisphere can be derived in a similar manner to produce the times necessary to find the relevant altimeter observations and interpolate to an accurate dual crossover location. As with single crossover differences, rejection criteria are applied in order to ignore dual crossovers with insufficient altimetry to produce an accurate difference measurement, i.e. where there are insufficient points to fit a cubic spline accross the time of the dual crosover.

Dual crossover results

Dual crossover data were obtained from cycle 9 of the ERS-1 35 day repeat (MJD 49006-49041) and from cycles 12 through 16 inclusive for the 10 day TOPEX/Poseidon dataset (MJD 48997-49046). For ERS-1 the required heights above the TOPEX reference ellipsoid were derived from the altimeter data using 7 near 5 day arcs computed using the SATAN-A software. The orbits were determined from laser range data and ERS-1 crossovers by *Moore and Ehlers* (1994). TOPEX/Poseidon heights were taken directly from the NASA orbit provided with the CD-ROM data records. Following rejection of approximately 1% of dual crossovers larger than ±60cm, remaining DXO's were used with a Gaussian weighting function of decorrelation length of 5 days. This contrasts with the method used in SXO minimisation due mainly to the fact that the two datasets are coincident in time, whereas a sequential approach was necessary for the Geosat XDR data. The dataset comprised approximately 48000 dual crossovers with rms value of 34.8 cm prior to adjustment (using zero *a priori* relative biases) and 9.3 cm after the sinusoidal adjustment procedure. Values of b_T=33.0cm and b_P=48.7 cm and $A^{(i)}$, $B^{(i)}$ which gave a 12.7cm rms adjustment were recovered. On modifying for non-zero mean of the sinusoidal component given the geographical distribution of the data over each arc relative biases of (29.1±3.3)cm and (43.8±3.6)cm were established for the NRA and the SSALT respectively. The sine and cosine coefficients for

ascending and descending arcs are plotted as Figures 4.7 and 4.8. Clearly evident are large adjustments for the long arc MJD 49031-49036 which were comparatively poorly determined from 216 laser ranges and 594 crossovers. The non-zero mean of the cosine terms during this period implies a geographically correlated error which is unobservable in the ERS-1 single satellite crossovers used during precise orbit determination of these ephemerides.
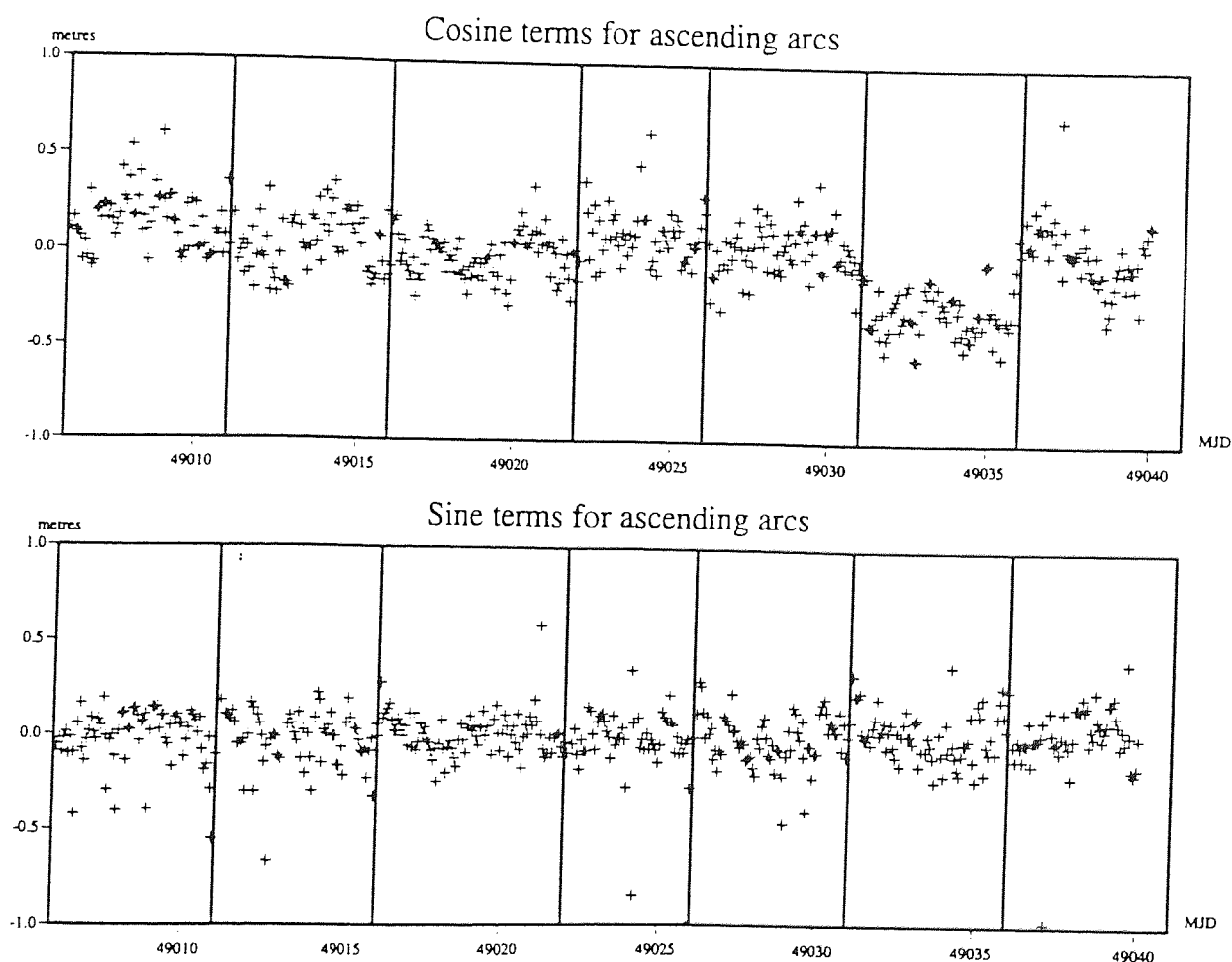
Figure 4.7. Sine and cosine coefficients of ascending ERS-1 arcs (MJD 49006-49041).
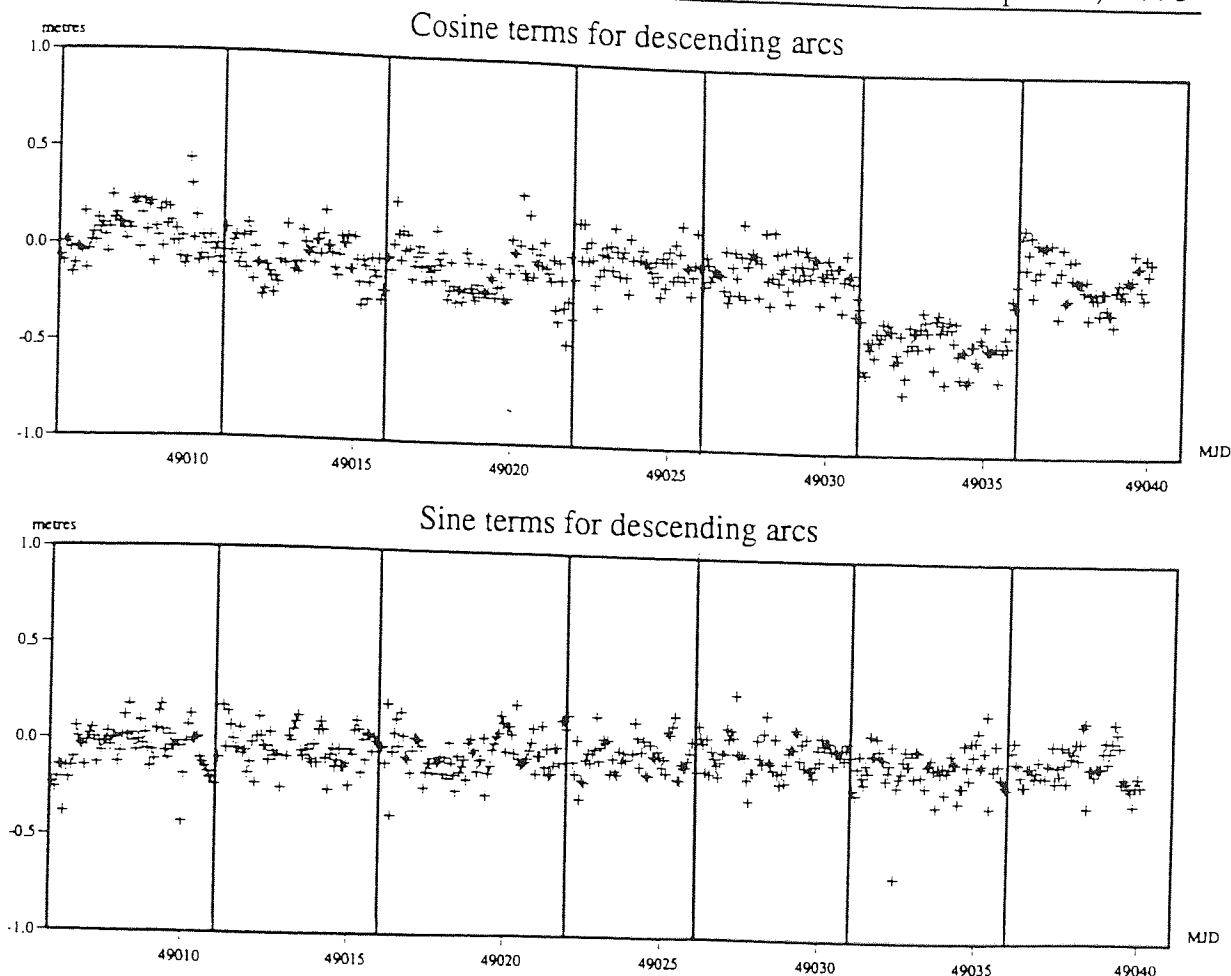
Figure 4.8. Sine and cosine coefficients of descending ERS-1 arcs (MJD 49006-49041).

4.5 Orbit error aliasing through DXO minimisation

With an estimated radial accuracy of 3-4cm, TOPEX/Poseidon has the best accuracy of any altimetric mission flown to date. For this reason, there is a great deal of interest in using the accuracy of TOPEX/Poseidon to refine the radial position of the much lower altitude ERS-1. As with any empirical correction of one dataset with another, this carries with it the risk of cross-contamination of errors from one satellite to the other - this includes not just orbit error, but all effects which may act to downgrade an altimetric observation.

In this section, through simulation of TOPEX/Poseidon orbit errors, the exchange of signal from one dataset to the other is investigated. Radial errors were grouped into two classes, those which are static and

geographically correlated, Δr_G, and those which have a more time dependent, temporally correlated nature, Δr_T. Explicitly, these have been modelled in this study by

$$\Delta r_T = \sum_k \sum_m \left(A_{km} \cos \psi_{km} + B_{km} \sin \psi_{km} \right) \tag{4.17}$$

$$\Delta r_G = \sum_n \sum_m P_{nm}(\sin \phi)\{ C_{nm} \cos m\lambda + S_{nm} \sin m\lambda \} \tag{4.18}$$

where P_{nm} are the unnormalised Legendre polynomials; ϕ and λ are respectively geocentric latitude and longitude of the satellite and the term ψ_{km} is defined in terms of the mean anomaly M and the relative angle between satellite, Ω, and Greenwich, θ, right ascensions $\Omega - \theta$, i.e. $\psi_{km} = kM + m(\Omega - \theta)$. Dual crossover minimisation is performed in a least squares sense from the measured dual crossover, D, and the radial orbit error components of ERS-1 and TOPEX/Poseidon , Δr_{ERS-1} and $\Delta r_{T/P}$ respectively, i.e.

$$\min \sum_i \left(D - \Delta r_{T/P} + \Delta r_{ERS-1} \right)^2 \tag{4.19}$$

where the summation is over all data points.

Firstly, however, it is important to recognise some of the mechanisms by which error signal may be transferred and even amplified through the sampling patterns imposed by dual crossover locations.

4.5.1 Aliasing due to DXO sampling patterns

The major cause of absorption of errors in the empirical correction of satellite altimetry is the distribution of data and how those data are fitted and corrected. Conventional orbit determination is performed through numerical integration of the satellite ephemeris to produce a position and velocity vector, usually at regular intervals in time. In empirical correction of altimetry using single or dual crossovers, the along track sampling of the dataset is not uniformly spaced as can be see in figure 4.9 for

one cycle of TOPEX/Poseidon with 6 days of ERS-1. As this figure shows, the vast majority of DXO data are at the southern extremes of each arc, with a similar, yet smaller spatial density at the northern extremes. Least squares solutions do not perform well at the boundary of a dataset, *Lawson and Hanson* (1974), therefore the distribution of data in figure 4.9 can only exacerbate this problem.
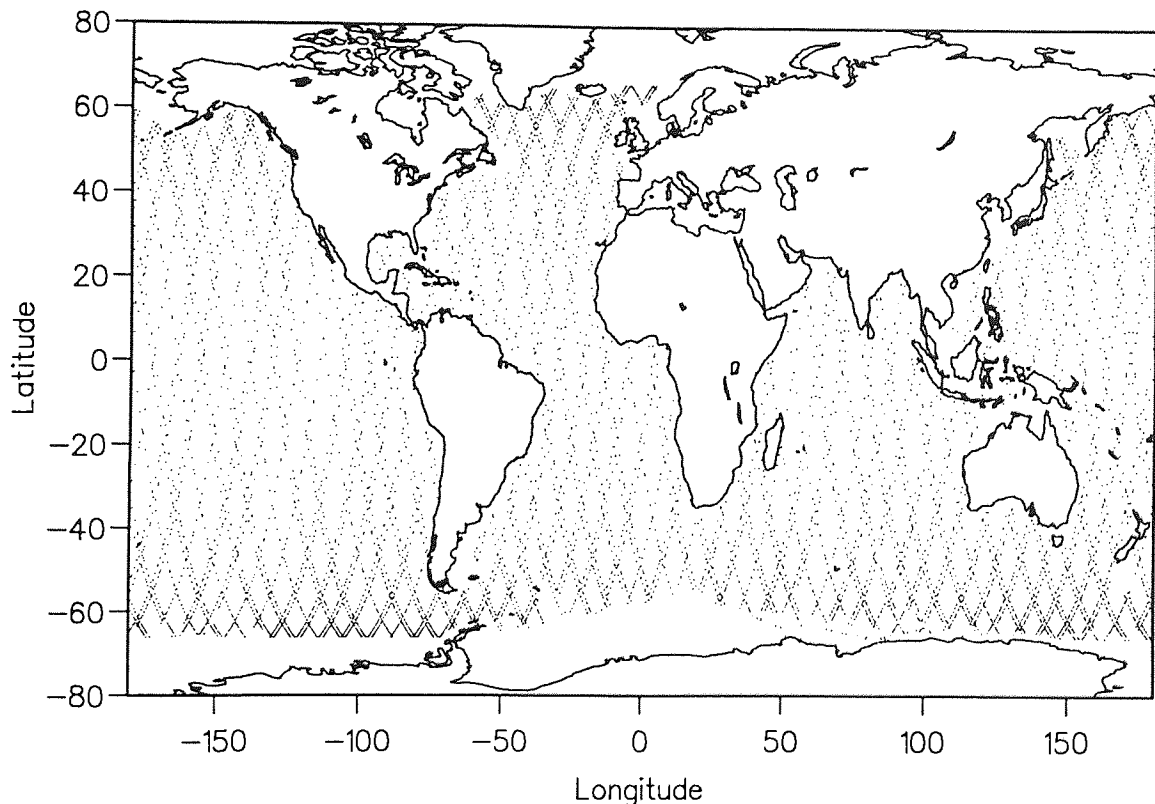
Figure 4.9. Dual crossover locations from 6 days of ERS-1 (MJD 49006-49011) with cycle 12 of TOPEX/Poseidon (MJD 48997-49007).

Regularly sampled data

The simplest mechanism for absorption through dual crossovers is a direct consequence of trying to fit a signal of one frequency with a sinusoid of another frequency. As figure 4.10 shows, the geometry of this example is such that evenly weighted data from regular samples along the length of a 2 cycle per revolution function cause the best fit sinusoid to be non-zero. This would obviously be a maximum were both signal and approximation

of the same period, leading to complete absorption of the signal. Where the periods of signal and approximation differ, however, the absorption will be reduced by an extent dependent upon the relative values of the two periods, becoming negligible when they are sufficiently different.
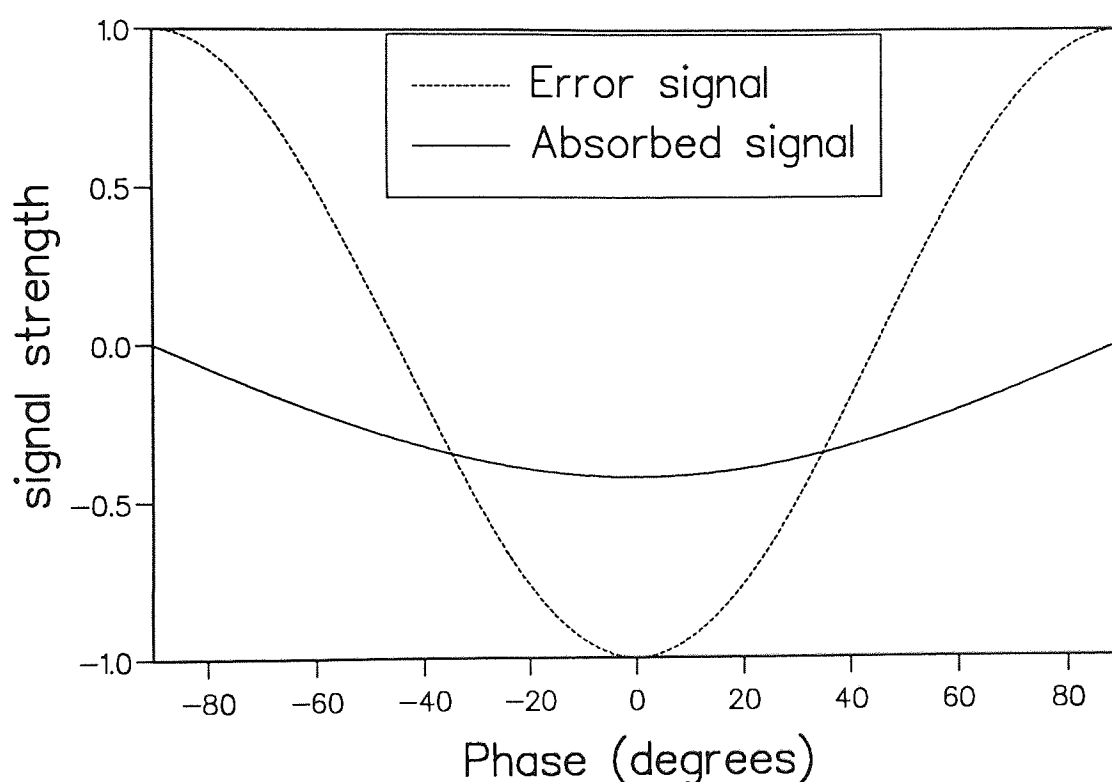
Figure 4.10. Absorption of a two cycle per revolution orbit error by a one cycle per revolution empirical correction.

Irregular sampling of data

The sine function has a mean which, if evaluated along its entire length, would be exactly zero. If however it were sampled more often around values of $+\pi/2\pm 2n\pi$ (n is an integer), i.e. around its maxima, and there was little or no sampling at $-\pi/2\pm2n\pi$, the result would be a positive mean. Such is the problem encountered with dual-crossovers, due to the fact that they occur irregularly in both time and space. Geographically, dual-crossovers are more densely located at high latitudes than around the equator (see figure 4.9) thus influencing the spatial and temporal sampling patterns of orbit errors in the correction of ERS-1 altimetry.

Figure 4.11 illustrates this case for a simulated three cycle per revolution error to which has been fitted two one cycle per revolution approximations. The first of these (the dotted line) was fitted to data sampled regularly at one degree intervals - leading to no absorption. The second function (the solid line) was fitted to the samples indicated by the vertical bars on the broken line. This irregular sampling has introduced a substantial alias signal due to the greater numbers of observations approaching ±90°. Where genuine satellite ground tracks are interrupted by land masses, the symmetry of sampling may be further disrupted, making aliasing of this sort more severe for some arcs.
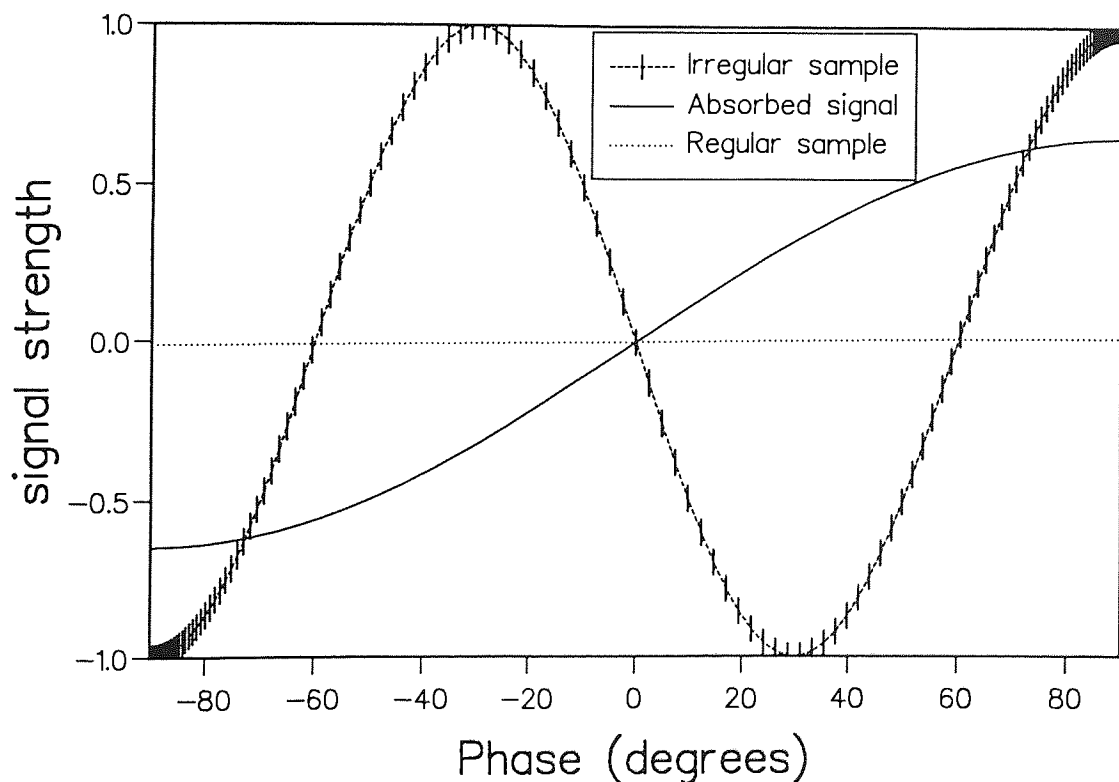
Figure 4.11. Absorption of orbit error signal through irregular sampling of the signal.

Asymmetric sampling of 2-D surfaces

A regular two-dimensional surface may be sampled along a line traversing it in such a way as to produce an irregular one-dimensional profile. Such is the problem with spatially sampling components of the Earth's gravity field error along a satellite ground tracks. Expressing the geographically correlated orbit error in terms of legendre polynomials, equation 4.17, generates a surface which, for a particular n and m, have a total of m changes of sign along a circle of latitude (except where $P_{nm}(\sin \psi)$ is zero), and n-m maxima or minima along a line of longitude from one pole to the other (with the same exception, i.e. where $\cos m\lambda$ or $\sin m\lambda$ is zero - whichever is relevant).

The inclination, altitude and equator crossing point of the satellite determine the location and shape of a particular ground track. Gravity field error terms are longitude independent when $m = 0$, therefore the sampled profile will be determined simply from the availability of data over ocean regions and will otherwise be the same for all arcs. Where $m \neq 0$, individual profiles can be very different. Figure 4.12 illustrates two typical ERS-1 ground tracks across the surface $n = 8$, $m = 4$, while figure 4.13 shows their associated profiles and the absorbed orbit error signal for a sinusoid fitted to these datasets. Profile A is nearly symmetric about the equator and of approximately constant amplitude. When irregularly sampled (see section 4.2), there is a small absorbed signal which has a strength of 29% that of the TOPEX/Poseidon error signal. Track B, however, shows a modulated amplitude along its length and is anti-symmetric about the equator. In conjunction with the irregular sampling pattern shown, an absorption strength of 89% is observed. This is extremely high for what might otherwise be considered a high frequency error, not susceptible to such aliasing.
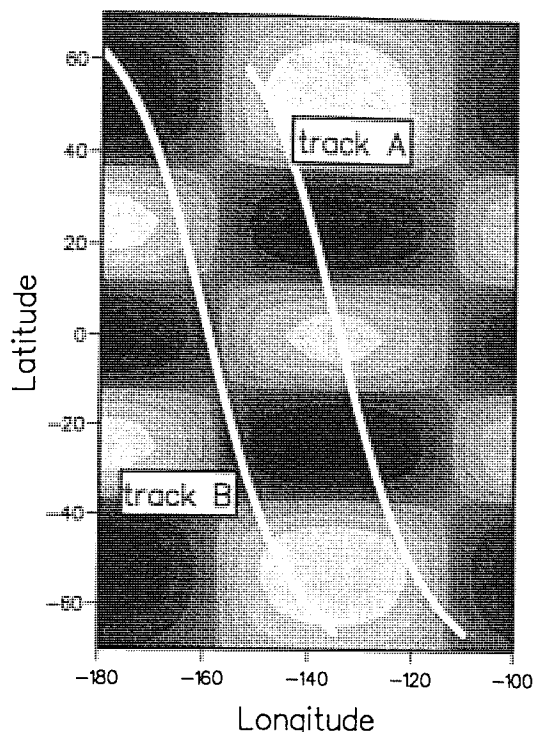
Figure 4.12. ERS-1 ground tracks over an n=8, m=4 geographically correlated error signal.

4.5.2 Aliasing of geographically correlated errors

Orbital solutions obtained for TOPEX/Poseidon and ERS-1 using the same gravity field model will generally result in different geographically correlated errors for each satellite. Therefore unlike single crossovers, geographically correlated orbit errors will not cancel in the determination of dual-crossover differences. Absorption of TOPEX/Poseidon orbit errors were investigated numerically using the
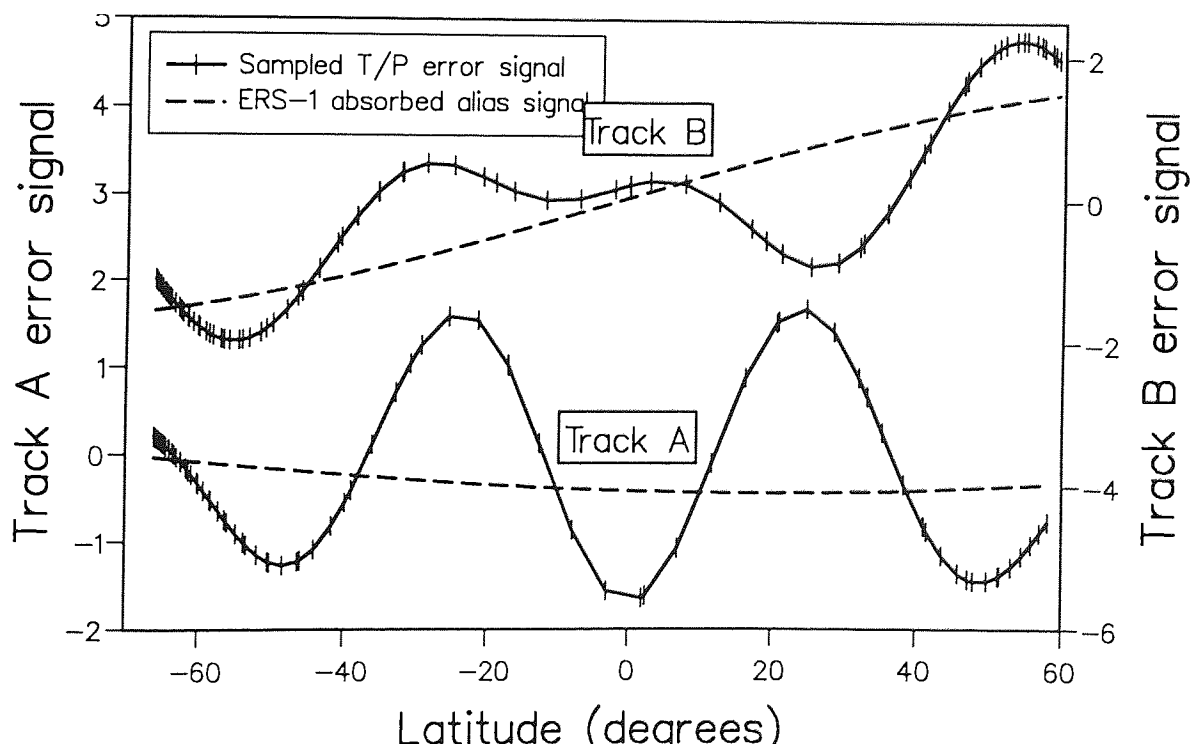
Figure 4.13. Profiles derived from sample of tracks shown in figure 4.12.

model in equation 4.18. Absorption of an error of this form may be illustrated for the case k=1, m=1, including both sine and cosine coefficients of equation 4.18, i.e. C_{11}^T and S_{11}^T

$$\Delta r_{11} = \left(C_{11}^T \cos \lambda + S_{11}^T \sin \lambda\right)\cos \phi \,, \tag{4.20}$$

From spherical trigonometry, it may be shown that

$$\cos \lambda = -\frac{\tilde{C}\sin M - \tilde{S}\cos i \cos M}{\cos \phi} \tag{4.21}$$

where $\tilde{C} = \cos(\Omega - \theta)$ and $\tilde{S} = \sin(\Omega - \theta)$ are slowly varying. Substituting equation 4.21 into equation 4.20 reveals that

$$\Delta r_{11} = \left(S_{11}^T \tilde{C}_{ERS-1} - C_{11}^T \tilde{S}_{ERS-1}\right)\cos i_{ERS-1} \cos M_{ERS-1}$$
$$-\left(C_{11}^T \tilde{C}_{ERS-1} + S_{11}^T \tilde{S}_{ERS-1}\right)\sin M_{ERS-1}$$

The coefficients of ERS-1's sinusoidal orbit error correction therefore absorb this TOPEX/Poseidon orbit error as

$$A = \left(S_{11}^T \tilde{C}_{ERS-1} - C_{11}^T \tilde{S}_{ERS-1}\right)\cos i_{ERS-1}$$

and

$$B = -\left(C_{11}^T \tilde{C}_{ERS-1} + S_{11}^T \tilde{S}_{ERS-1}\right)$$

The term $\cos i_{ERS-1}$ is small for ERS-1 ($i_E \approx 98°$), therefore the sine coefficient B will absorb most of this error. For example, for $\Omega = \theta$, it follows that $\tilde{S}_{ERS-1} = 0$ and $\tilde{C}_{ERS-1} = 1$, leading to complete absorption of the C_{11}^T orbit error into B.

Numerical analyses of terms up to degree and order 50 were performed separately for each ERS-1 arc and for each term in the simulated orbit error. This involved solving with a sine or cosine TOPEX/Poseidon coefficient individually for each value of n and m of interest.

Figures 4.14 and 4.15 show the results of a sine or cosine coefficient for each term in the TOPEX/Poseidon orbit error, while tables 4.2 and 4.3 summarise absorption of the low degree and order terms. At high degree
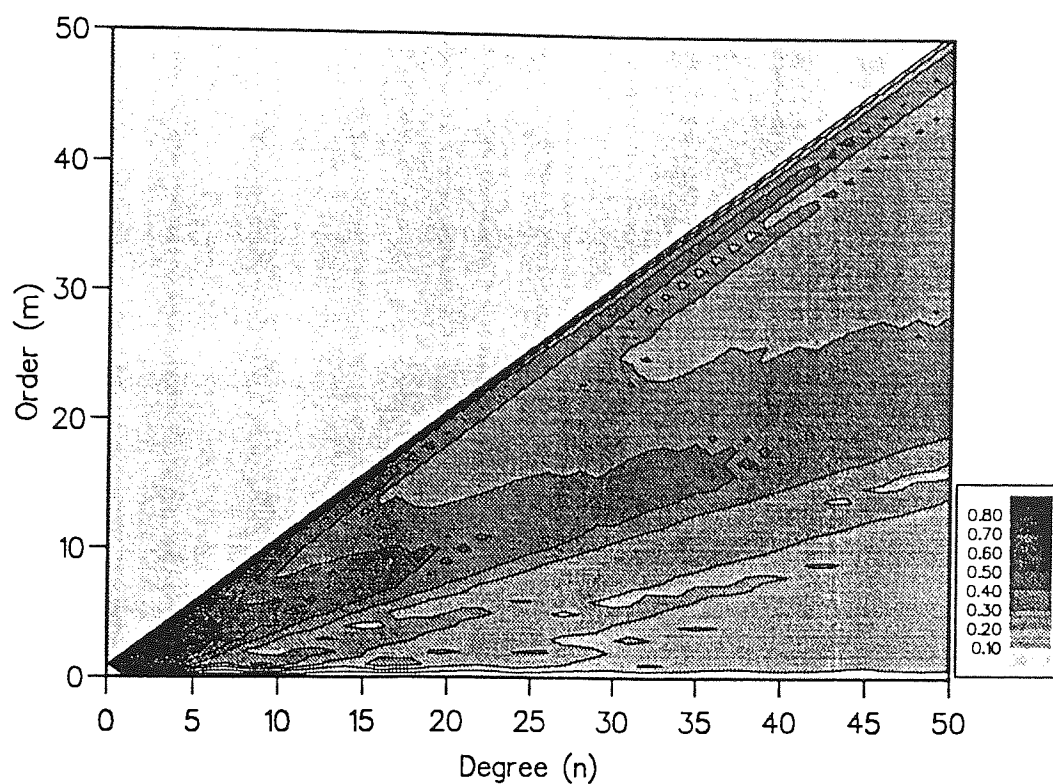
Figure 4.14. ERS-1 absorption of simulated TOPEX/Poseidon geographically correlated orbit errors - sine terms.
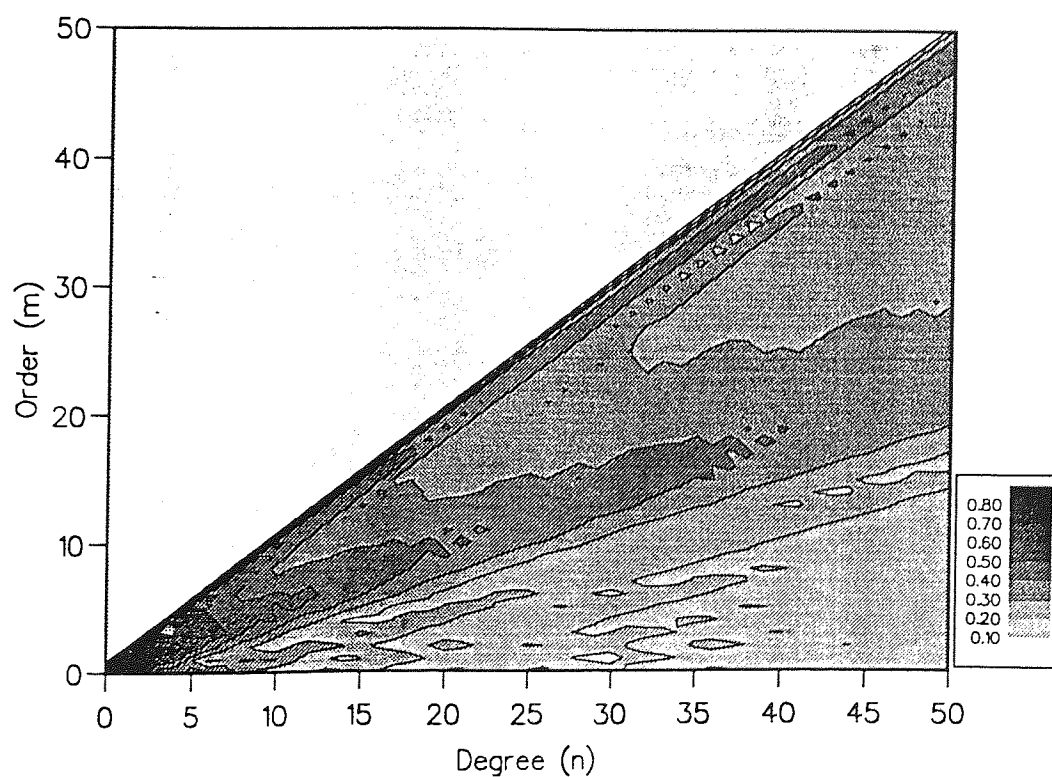
Figure 4.15 As for figure 4.14 - cosine terms.

and order, both terms show very similar absorption patterns. Mean absorption of the cosine terms is highest for low degree and order, reaching a maximum of complete absorption when $n=1$, and over 90% absorption where $n \leq 3$ and n≈m. The plot of sine term absorption shows a similar pattern, except where m=0, where no absorption takes place.

The most interesting feature of both plots is the illustration that above $n \approx 6$, maximum absorption occurs for $n \approx 2m$. This is a direct consequence of the irregular sampling along track which results when determining DXO's. Other smaller maxima are visible and are a result of higher order aliasing of the same type as the principal maximum. The relationship n≈2m is attributable to the geometry of the ground track of ERS-1 relative to TOPEX/Poseidon and for satellites of different altitudes and inclinations this approximate expression would be different.

Table 4.2. Absorption of $P_{nm}(\sin\phi)\cos m\lambda$ terms x1000.

m	n 0	1	2	3	4	5
0	984	1000	765	513	461	678
1	0	1000	937	776	732	517
2	0	0	978	858	828	768
3	0	0	0	922	789	783
4	0	0	0	0	863	730
5	0	0	0	0	0	802

Table 4.3. Absorption of $P_{nm}(\sin\phi)\sin m\lambda$ terms x1000.

m	n 0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1000	936	844	706	515
2	0	0	978	867	810	750
3	0	0	0	926	806	782
4	0	0	0	0	858	715
5	0	0	0	0	0	808

The cumulative effects of these absorptions was investigated through use of the JGM-2 gravity field covariances, *Nerem et al.* (1993). The unit values of C_{11}^T and S_{11}^T used above were replaced with the equivalent terms from the covariance matrix, thus giving an estimate of their relative strength. It should be noted, however, that the covariances are not equivalent to the error terms, simply that the energy at each frequency may be representative of what the radial orbit error might be. By calculating the total rms orbit error for TOPEX/Poseidon and ERS-1 heights corrected via minimisation of equation 4, it was found that approximately 50% of the total TOPEX/Poseidon error was directly absorbed into ERS-1.

4.5.3 Aliasing of temporally correlated errors

From equations 4.17 and 4.19, DXO's were minimised to solve for A and B using $\Delta r_{T/P} = \Delta r_{km}$. This was accomplished by selecting individual combinations of k and m for each solution and ascribing a unit value to either the sine or cosine coefficient in the simulated TOPEX/Poseidon orbit error (B_{km} or A_{km} respectively) and setting all other coefficients in equation 4.17 equal to zero. Analytical solution of this expression for an arbitrary k and m is complicated, but the consequences can be illustrated for the case $k=1$, $m=0$ and $B_{km}=0$, i.e. a dominant one cycle per revolution error, assumed in this case to be zero at the equator crossings; written as $\Delta r_{T/P} = C_{10} \cos M_{T/P}$. Enforcing equation 4.19 and writing in full, this becomes

$$A \cos M_{ERS-1} + B \sin M_{ERS-1} = C_{10} \cos M_{T/P} \qquad (4.22)$$

From spherical trigonometry, it may be shown that

$$\sin i_{ERS-1} \sin M_{ERS-1} = \sin \phi = \sin i_{T/P} \sin M_{T/P} \qquad (4.23)$$

where i_{ERS-1} and $i_{T/P}$ are the orbital inclinations to the equatorial plane of ERS-1 and TOPEX/Poseidon respectively. Combining 4.22 and 4.23 and

treating sine and cosine terms separately, it can be seen that $B=0$ and therefore

$$A = \frac{\sin i_{ERS-1}}{\sin i_{T/P}} C_{10}$$

i.e. such a TOPEX/Poseidon orbit error is amplified by the ratio of the sines of the inclinations of the two satellites, in this case by 8.2%, and then absorbed into the term A and hence directly into 'corrected' ERS-1 altimetry.

Investigation of other terms was undertaken numerically with the results being illustrated in Figures 4.16 and 4.17 for cosine and sine terms respectively. Plotted in both figures are the mean absorption and its variability determined from the solution of all individual arcs. In both cases, absorption is significant only at low degree and order and markedly different for sine and cosine terms. For $|k| \geq 1$, the variability of the sine terms is negligible, but for $k=0$, $m<6$, there can be significant absorption in a single arc despite low mean absorptions (<0.25) for all values of k and m: This implies a strong longitude or arc dependence for the absorption. The behaviour of the TOPEX/Poseidon cosine terms are slightly different, in this case for $k=0$, the mean and variability are low, indicating little absorption for all arcs. At $k=\pm1$, the means have significant values (1.082 for $m=0$, as predicted). At higher degrees and orders, the means and variabilities are close to zero. The variabilities of absorption of TOPEX/Poseidon cosine terms demonstrate some distinction between odd and even values of k. At $k=0$, the means at low order are small but significantly greater than for $k=\pm1$ where variability is very close to zero. The greatest variabilities are seen at $k=\pm2$, followed by $k=\pm4$, with $k=\pm3$ possessing variabilities similar to those of $k=\pm6$.

Mean of sine term absorptions
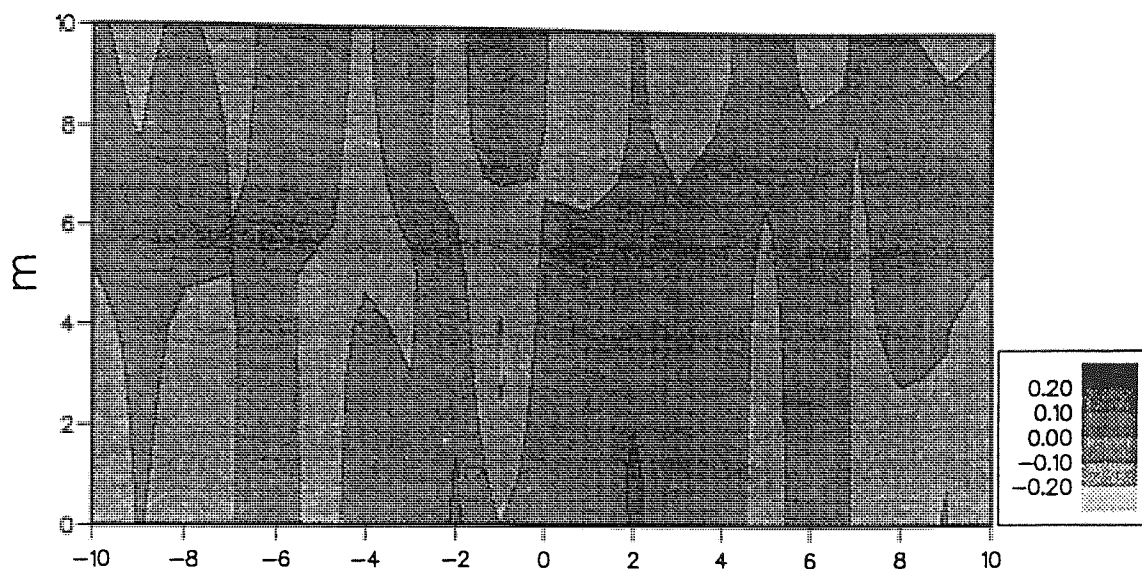
Variability of sine term absorptions
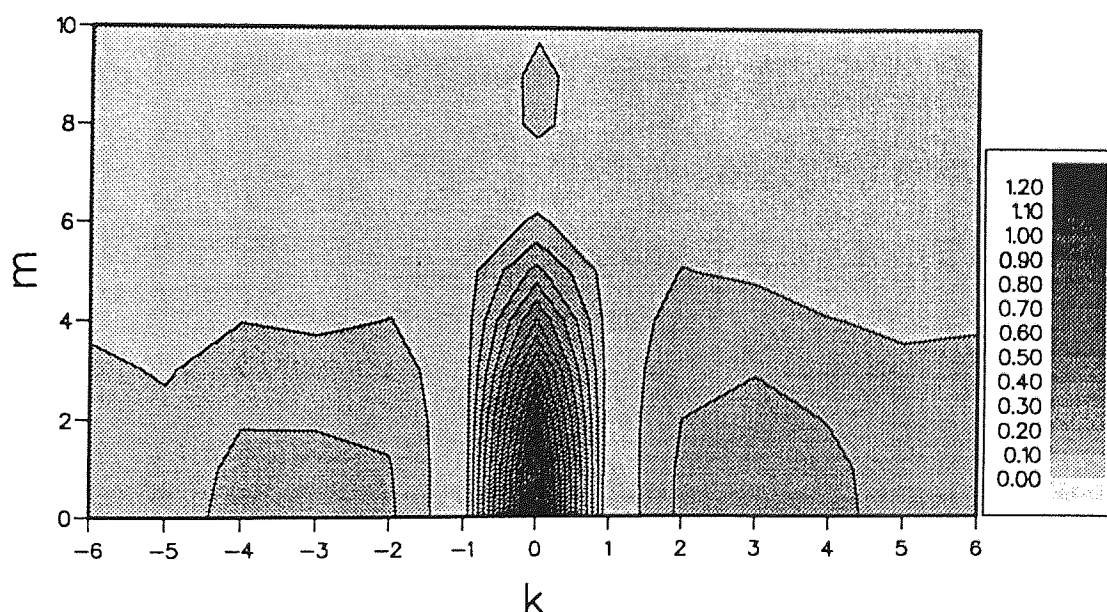
Figure 4.16. Absorption of ERS-1 simulated TOPEX/Poseidon temporally correlated orbit errors - sine terms.

Overall these results indicate that at low degree and order, significant and in some cases amplified TOPEX/Poseidon time dependent orbit errors are aliased directly into ERS-1 orbit error corrections. The effect of this is to introduce long wavelength errors into any sea level profiles derived from data corrected by such empirical techniques.

Mean of cosine term asbsorptions
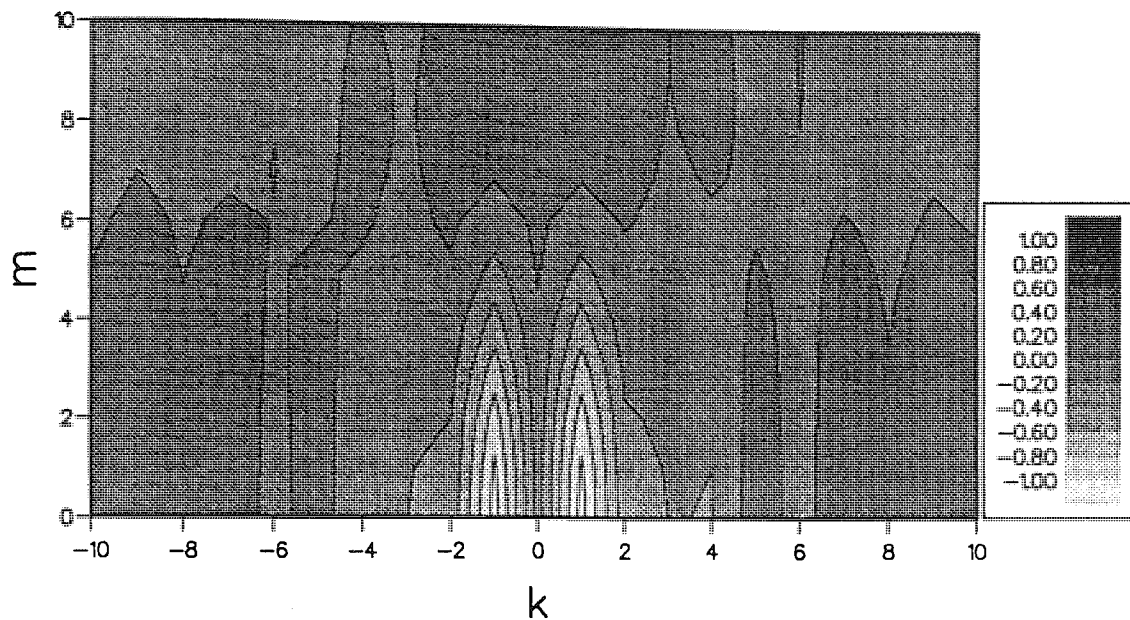
Varibility of cosine term absorptions
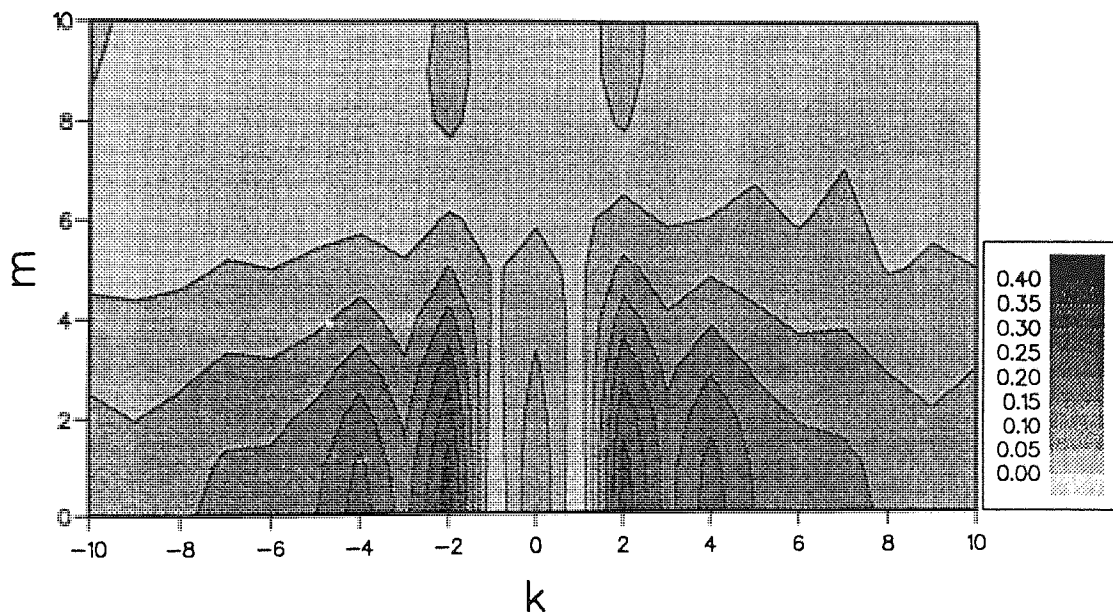
Figure 4.17. As for figure 4.16 - cosine terms.

4.6 Conclusion

Empirical correction of satellite altimetry to reduce the influence of radial orbit errors is a well established technique. A number of approaches have been suggested and used over recent years to improve the quality of oceanographic results from analysis of these data. All empirical methods have exploited the long wavelength nature of the dominant one cycle per revolution error either by spectral analysis or by least squares fit and subtraction of an appropriate function to the data. This chapter has introduced the sinusoidal correction of altimetry using an accurate dataset as a reference. In the case of Geosat, this involved the selection of a subset of the available XDR data and a simultaneous constrained solution of 358 arcs over 25 days, thus giving the necessary reference arcs with which the remaining two and a half years of data could be corrected. Significant improvements in the rms crossover differences after fit as well as some improvement in the geographically correlated discrepancies between ascending and descending data were observed. Because of the use of two day orbital solutions during the generation of the XDR dataset, a two day signal was also detected in the values of radial correction terms.

For ERS-1, the reference set of arcs came from the TOPEX/Poseidon mission whose superior radial accuracy permitted the direct usage of altimetry from long-arc solutions supplied with the data without any further enhancement. This again demonstrated an improvement in the rms crossover differences between the two satellites and allowed values of (29.1 ± 3.3)cm and (43.8 ± 3.6)cm to be determined for the relative biases between ERS-1, and the NRA and the SSALT respectively.

From simulation of various spatial and temporal errors which may be present in TOPEX/Poseidon data, the absorption into ERS-1 altimetry via dual crossover minimisation was investigated. It was found that whilst a significant proportion of some components was absorbed, this constituted

an error signal in the ERS-1 dataset that was significantly smaller than other sources. That is, for 50% absorption of a pessimistic 3cm TOPEX/Poseidon geographically correlated orbit error, of the say 10.0cm accuracy of dual crossover corrected ERS-1 altimetry, 9.9cm is attributable to sources other than absorption of the TOPEX/Poseidon error.

TOPEX/Poseidon is unusual as an altimetric satellite in that it has a radial accuracy of under 5cm; dual crossover minimisation makes use of the accuracy of TOPEX/Poseidon data to refine the radial accuracy of ERS-1, but there are other geographically and temporally correlated errors which may influence the validity of dual crossover minimisation. For example where different tidal or atmospheric propagation models have been used for each dataset, consistent geographically correlated differences will be produced in the range measurements used to obtain estimates of the sea surface heights. These inconsistencies are to a certain extent unavoidable due to, for example, the use of different radiometer measurements to estimate wet troposphere corrections and they will be transferred in exactly the same manner as would a radial orbit error. Therefore, while dual crossover techniques will improve the radial accuracy of ERS-1 sea height measurements, they will always be inferior to that of TOPEX/Poseidon.

Chapter 5

Single satellite crossover analyses

5.0 Introduction

An impediment to the accurate analysis of satellite altimeter data is uncertainty in the exact position of the platform in space. This orbit error is slowly being eliminated by improved gravity field and surface force modelling but is still a problem and is likely to remain so for some time to come, especially for lower orbit missions such as Geosat and ERS-1. Reduction of this error by empirical means is a well-established method which removes a sizeable proportion of the error by approximation of the dominant one cycle per revolution component. A common factor in such techniques is that the arc length over which the error is estimated is longer than the wavelength of the ocean signal being investigated. Using the methods described in section 4.3 to correct the Geosat XDR dataset, two analyses are described in this chapter: the first to extract the relative ocean variability signal at a given location, while the second attempts to assess the extent of long wavelength aliasing of the ocean signal during sinusoidal orbit error correction.

Verifying that there is no significant aliasing or removal of long wavelength ocean signals has typically involved one of two approaches: the comparison of tide gauge records with altimetric time series [e.g. *Harangozo et al.*, 1993], thereby providing independent measures of the sea surface by two different techniques, or statistical comparison of different approximation models [e.g., *Tai*, 1989]. Unfortunately for the first of these validation methods, tide gauges tend to be located in harbours at the coast or on islands, while altimeters best measure heights in the open ocean. In conjunction with other errors and uncertainties, extrapolation of results to

a common location further degrades any comparison of the independent time series, thus making definitive validation impossible. Statistical interpretation of altimetry corrected by different empirical methods has so far provided the best validations for this style of orbit error correction. Spectral and variability studies [e.g., *Denker and Rapp,* 1990; *Bhaskaran and Rosborough,* 1993] show that certain models filter the data to a greater or lesser degree and have varying characteristics for the duration of the solutions (i.e., usually being worse at the ends of the arcs used). The weakness of these approaches lies in the comparison of like with like, that is of one least squares approximation with another. To provide conclusive evidence that empirical correction of the once per revolution error is safe, it is necessary to eliminate this error from the data entirely and to compare empirically corrected results with error free results. Such a validation exercise is presented here using crossover difference differences (CDD's) to cancel errors common to four ocean height observations. CDD's were first suggested by *Moore and Rothwell* (1990). The result is a data set that is free of orbit error contamination yet retains oceanographic information suitable for comparison with empirically corrected results from conventional crossover difference analysis.

With the results of the ocean signal recovery methods from the first half of the chapter, consistent validation of sinusoidal orbit error removal is obtained by comparison with CDD analysis developed in the second half of the chapter. The strength of this validation lies in the fact that it verifies the effectiveness of the method rather than the accuracy of the final data. That is to say, it has demonstrated that empirical correction, as described, does not alias large-scale ($\lambda \approx 2500$km), long-period ($T \approx 1$ year) ocean signal to an unacceptable degree.

5.1 Ocean height determination from crossover differences

As defined in section 4.2, a crossover difference is defined by changes in the measured ocean height between ascending and descending observations of the same location. The crossover difference therefore measures the change in ocean height, atmospheric propagation, orbit error and all the other constituents and errors in each observation. Provided error contamination can be reduced to an acceptable level, it is possible to extract an estimate of the change in ocean height over the period between two observations. This estimate is a relative measure, and without further information, no absolute height information can be inferred - it is for this reason that Geosat crossover data are available from the classified GM phase but the actual altimetry from which it is derived is not. Given a number of crossovers at a particular location, therefore, it is a simple task to estimate the changes over time of the ocean height as compared with say the first (unknown) height. Such a time series, however, would contain very little data if actual altimetry were used as very few crossovers occur at identical locations, with the result that relative heights can be deduced only at discrete locations.

Milbert et al. (1988), showed that meso-scale signals could be extracted from crossovers by accepting data from a region of ocean and performing a least squares fit of a vertical piston term to the data, thus providing a single height estimate for the region as a whole. That is, by taking an area large enough to give the required temporal resolution, signal up to a high spatial frequency cut off could be extracted from the data. During the Geosat ERM, the satellite ground track was maintained to within ±1km of a nominal path (towards the end of the mission, fewer corrections were made in order to maintain fuel and operation as long as possible and this limit was progressively increased). The XDR dataset contains only the first twelve months of this phase during which most crossovers around a nominal

crossover location were located within a diamond shape of sides 1km in length. Crossovers during the GM phase are more uniformly distributed over the sea surface and are not grouped around nominal locations. Finally, crossovers between the two missions are located in a locally near uniform manner along ERM tracks. To exploit the full two and a half years of Geosat XDR data, therefore, a significant area must be chosen in order to include all three crossover types (GM/GM, GM/ERM and ERM/ERM) and locations covering the period. Because of the geometry of satellite ground tracks, the exact size and shape of the area chosen is dependent on its the latitude; for example *Milbert et al.* (1988) used a region measuring 2° in latitude by 10° in longitude near the equator which gives very high temporal resolution. Despite this resolution, Geosat data are too noisy to translate directly to a meaningful time series without some filtering of the data. The method to be used for this is known as optimal interpolation and exploits the interdependence within different observations of common signals and errors.

5.2 Optimal interpolation

Filtering altimeter data by optimal interpolation was proposed by *Wunsch and Zlotnicki* (1984), based on techniques described by *Moritz* (1980) as least squares collocation for the extraction of geoid height estimates from various data sources. The method is also known by several other names, for example, least squares prediction, optimal mapping, the inverse method/mapping, the Gauss-Markov method, Kolmogorov-Wiener prediction and Krigging. As these names essentially describe the same basic method this text shall use 'optimal interpolation' to describe all of these. The method will be used for this purpose in the next chapter, but it may also be applied to time series analyses, provided realistic descriptions of the signals and their errors are available.

• **Mathematical background**

The interdependence of the signal and errors present in an altimetric dataset may be described in terms of covariance functions, where the auto-covariance of a property f is given by the expectation value of its square, i.e.

$$C_{ff} = \langle ff^* \rangle$$

where f is a function and f^* its complex conjugate, or in matrix notation with real valued data,

$$C_{gg} = \langle gg^T \rangle \tag{5.1}$$

where g is a vector of data points with a mean value of zero and T denotes the transpose of that vector. The cross-covariance is described in a similar way for two vectors g and h as

$$C_{gh} = \langle gh^T \rangle \tag{5.2}$$

Given a set of observations l of an unknown signal s, it is possible to estimate the values of s for each observation. An estimate $\hat{s}$ may, therefore, be defined via a matrix transfer function H, written,

$$\hat{s} = Hl \tag{5.3}$$

The error (ε) in the signal estimate is the difference between that estimate and the true signal,

$$\varepsilon = \hat{s} - s$$

The error covariance of this can be expressed as the expectation value of the square of the function, i.e.,

$$\text{cov}(\varepsilon) = \langle \varepsilon \varepsilon^T \rangle \tag{5.4}$$

The product $\varepsilon \varepsilon^T$ can be expanded as follows,

$$\varepsilon\varepsilon^T = (\hat{s} - s)(\hat{s} - s)^T$$

$$= (Hl - s)(Hl - s)^T$$

$$= (Hl - s)(l^T H^T - s^T)$$

$$= Hll^T H^T - sl^T H^T - Hls^T + ss^T$$

From the definition of equation 5.4, the error covariance becomes,

$$\langle \varepsilon\varepsilon^T \rangle = \langle Hll^T H^T \rangle - \langle sl^T H^T \rangle - \langle Hls^T \rangle + \langle ss^T \rangle$$

As the transfer function H is independent of the values of s and l, it follows that

$$= H\langle ll^T \rangle H^T - \langle sl^T \rangle H^T - H\langle ls^T \rangle + \langle ss^T \rangle$$

Therefore from definitions in equations 5.1 and 5.2,

$$= HC_{ll}H^T - C_{sl}H^T - HC_{ls} + C_{ss} \tag{5.5}$$

The definition of the cross-covariance can also be shown to have the following property,

$$C_{sl} = \langle sl^T \rangle = \langle ls^T \rangle^T = C_{ls}^T$$

equation 5.5 can, therefore, be rewritten,

$$\langle \varepsilon\varepsilon^T \rangle = C_{ss} + HC_{ll}H^T - C_{sl}H^T - HC_{sl}^T$$

$$= C_{ss} + HC_{ll}H^T - C_{sl}H^T - (C_{sl}H^T)^T$$

which may be re-expressed as

$$= C_{ss} + (H - C_{sl}C_{ll}^{-1})C_{ll}(H - C_{sl}C_{ll}^{-1})^T - C_{sl}C_{ll}^{-1}C_{ll}C_{ll}^{-1^T}C_{sl}^T$$

$$= C_{ss} - C_{sl}C_{ll}^{-1}C_{sl}^T + (H - C_{sl}C_{ll}^{-1})C_{ll}(H - C_{sl}C_{ll}^{-1})^T$$

cancelling the identity matrix where $I = C_{ll}^{-1}C_{ll}$ gives the error covariance as the summation of two functions,

$$\langle \varepsilon\varepsilon^T \rangle = A + B$$

where

$$A = C_{ss} - C_{sl}C_{ll}^{-1}C_{sl}^T$$

$$B = \left(H - C_{sl}C_{ll}^{-1}\right)C_{ll}\left(H - C_{sl}C_{ll}^{-1}\right)^T$$

Optimal interpolation is based on the premise of generating a minimum variance solution which corresponds to the sum of A and B being a minimum. The matrices C_{ss} and C_{ll} are positive definite, therefore, for all possible transfer functions H, it follows that B≥0. As A is independent of H, the minimum variance occurs for B=0, or

$$H = C_{sl}C_{ll}^{-1}$$

Therefore from equation 5.3, the optimal estimate of a signal s can be found from

$$\hat{s} = Hl = C_{sl}C_{ll}^{-1}l \qquad (5.6)$$

where the error covariance of the estimate is

$$C_{\varepsilon\varepsilon} = A = C_{ss} - C_{sl}C_{ll}^{-1}C_{sl}^T \qquad (5.7)$$

The crux therefore of this approach is whether the cross-covariance C_{sl} and the auto-covariance C_{ll} can be determined with sufficient confidence. Additionally, as the matrix C_{ll} is of rank equal to the number of observations, for a large dataset (more than a few thousand points) it becomes impractical to store and invert the matrix. Using available technology, this method is therefore restricted to analyses of small well defined datasets, for example, the mean sea surface analysis of Geosat data as performed by *Blanc et al.* (1991)

5.3 Ocean height time series interpolation

Time series generation from satellite crossover differences is dependent on two factors: choice of data from an appropriate small area or pixel and selection of appropriate covariance models for the ocean signal and observation errors.

• Pixel selection

Altimetric satellites are generally placed in near-circular orbits in order to simplify acquisition of range measurements by the altimeter. As a consequence of such an orbit, all crossovers of a particular time difference can be found at approximately the same latitude, with variations due mainly to the slow drift and decay of the orbit. The same applies to crossovers from exactly repeating missions where the time differences are equal for all crossovers at a particular latitude, plus or minus integer multiples of the orbital repeat period. A plot of crossover points from the first 25 days of the Geosat GM in figure 5.1 clearly shows crossover locations to be at consistent latitudes within the first 23-day near repeat period. The following two days are visible as 'double' tracks of parallel dots. Crossovers within each near repeat occur at common latitudes, while crossovers between near-repeats occur at separate but also consistent latitudes. To maximise the volume of data selected to generate a time series and hence give the highest temporal resolution possible, the pixel should be centred on one ERM crossover latitude. Further, because of the 17 day repeat period, each nominal crossover location is sampled only twice (one ascending and one descending) in each period. To improve the temporal sampling, two or more nominal crossovers locations surrounded by suitable margins may be selected. The exact size depends on the minimum acceptable spatial resolution - which is downgraded by the use of larger pixels. Geosat tracks are separated by around 80km at 60° of latitude, and therefore to resolve mesoscale ocean signals of spatial size of the order of 100km, pixels must be limited to individual crossover locations. The pixels chosen were therefore 2.0° (~110km at 60°N/S) in longitude, centred on nominal ERM crossover locations and 1.1° (~120km) in latitude in order to provide a reasonable number of GM/ERM and GM/GM crossovers.
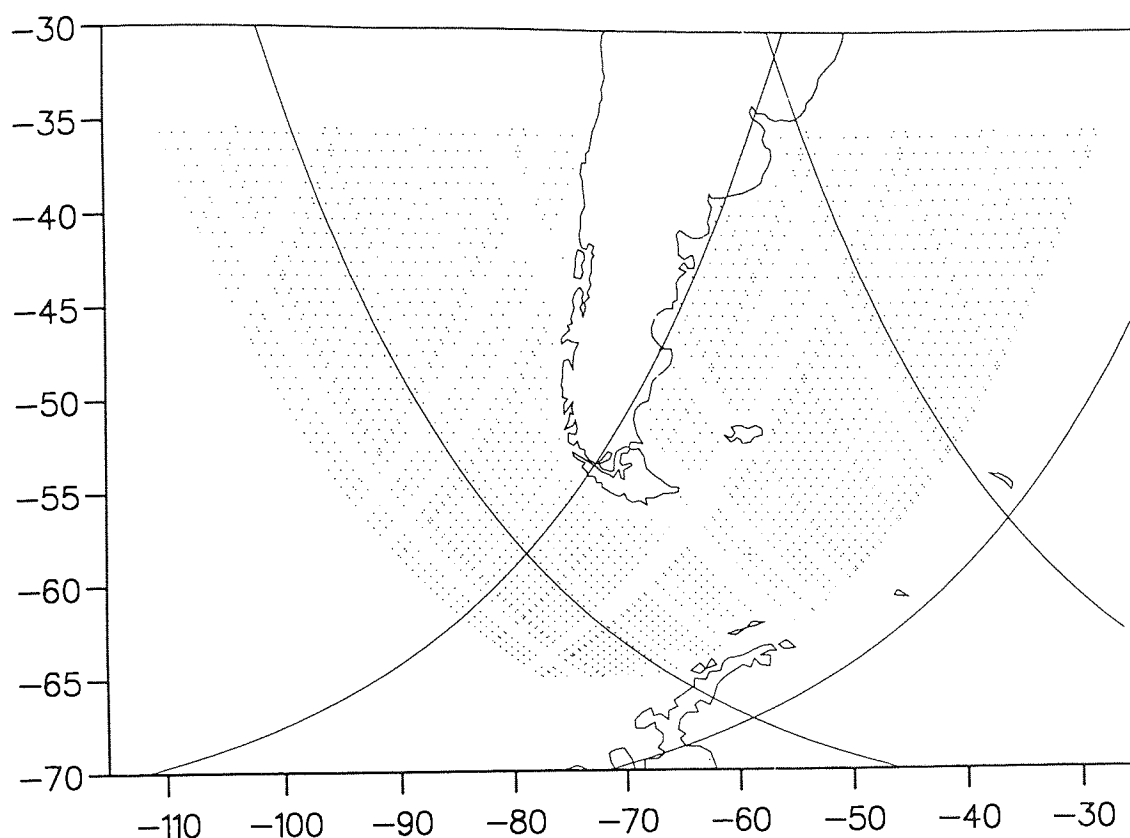
Figure 5.1. Single satellite crossover locations from the Geosat GM - due to the exact repeat nature of this mission, all crossovers occur repeatedly over the same locations (±1km).

• **Covariance models**

The covariance of mesoscale ocean signal has been given by a number of authors, for example, *Mazzega and Houry* (1989) and *Wunsch and Zlotnicki* (1984), who proffer a Gaussian temporal covariance of decorrelation time of the order 17-20 days. For a pixel of the size used by *Milbert et al.* (1988), this certainly produces a distinct signal. however, part of the reason for producing a time series in this study was for comparison with CDD results produced later in this chapter. As will be shown, the significant effects of geographically correlated errors mean that results are best produced from small pixels of width no more that 2° in longitude. The chosen covariance function was of the form,

$$C_{pq} = \sigma^2 \exp\left(-\frac{|t_p - t_q|}{b} \right)$$

(5.8)

where σ^2=10 cm and b=430 orbital periods (~30 days). This expression is different from the Gaussian distribution used by *Milbert et al.* (1988), (who additionally used σ^2=5 cm and b=15 days). The reason for this difference is mainly due to the fact that the pixels used to fit the piston term were one quarter the size, with therefore one quarter the data, of those used by *Milbert et al.* (1988). To compensate for this, it was necessary to interpolate with a function that was less localised, but which would retain a decorrelation time similar to that of ocean mesoscale signals, that is, 17-20 days [*Milbert et al.*, 1988]. For this reason, the explicit, non-Gaussian, form of equation 5.8 was used with $b \approx 30$ days. The correlation length of this function (time difference where C_{pq} is half the maximum value, at $t_p=t_q$) is marginally greater than that of the ocean at 20.8 days, but it has the advantage over the Gaussian of attributing more weight farther from the interpolation point, thus giving a smoother estimate of the ocean signal. The different value of σ^2 chosen can be attributed to the fact that Milbert produced time series in the equatorial belt where ocean variabilities are lower than, in particular, the Southern Ocean where data for this study were selected. The value of 10cm is, however, consistent with that used by *Wunsch and Zlotnicki et al.* (1984) for the North Atlantic Ocean.

Finally, the covariance of the error present in relative height estimates must be modelled. In the absence of any detailed information of the exact nature of tidal, atmospheric and other modelling errors, the covariance of the differences in these errors was assumed to have a white noise spectrum, and were therefore modelled simply as a random noise. The optimal predictor, equation 5.6, therefore, may be written as

$$\mathbf{h}_s = C_{st}\left(C_{tt} + \sigma^2_{noise}I\right)^{-1}\mathbf{h}_t$$

where $\mathbf{h}_s$ and $\mathbf{h}_t$ are the predicted and measured heights respectively and σ_{noise} is the noise term. The amplitude of σ_{noise} was selected on the grounds of root sum square of each of the errors present in the data, multiplied by a factor of 2 because the derivation of each relative height estimate is obtained
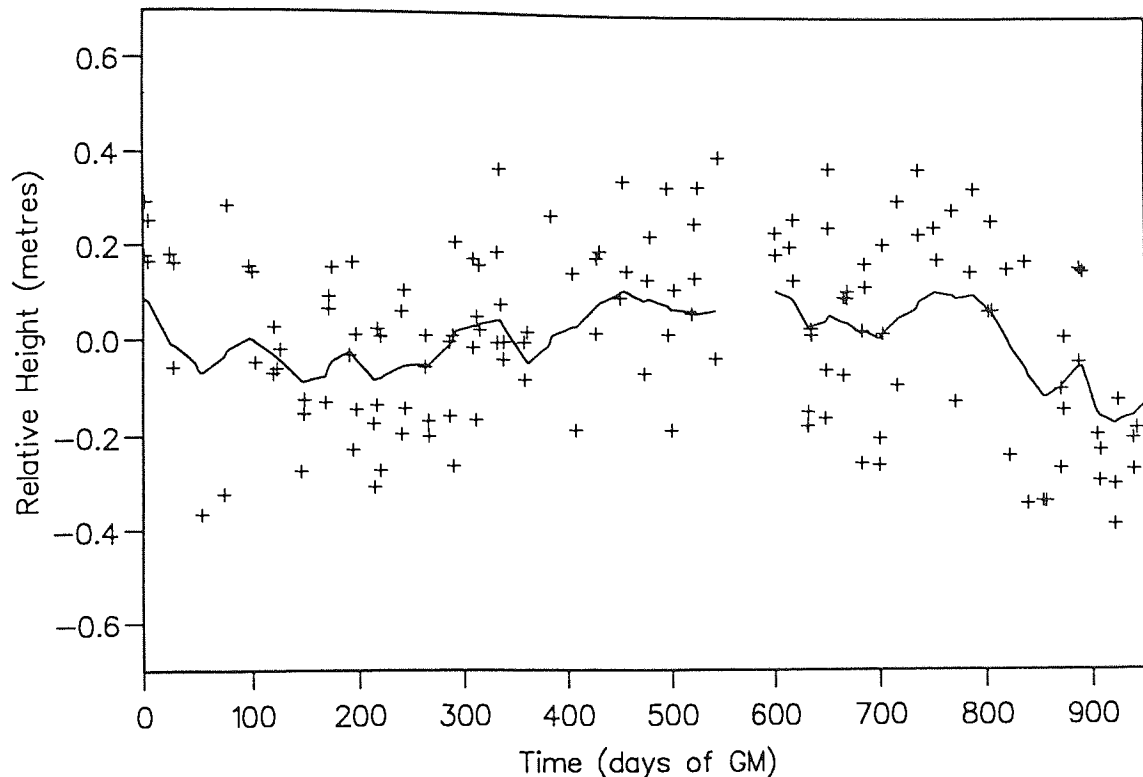
Figure 5.2. Relative ocean heights derived from crossover difference data around 43.5°S, 45°W. (+: piston heights; ___: Optimal estimate of ocean height.)

from two measurements (i.e. from crossover data). The estimated error constituents were: residual orbit error (20cm); atmospheric propagation (10cm); tidal mismodelling (10cm); instrument errors (5cm); and other minor errors such as electromagnetic bias errors and inverse barometric modelling errors (5cm). Thus giving a value of σ_{noise} =36cm.

The time series produced for a typical area, 1.1° in latitude and 2.0° in longitude, centred on 43.5°S and 45.0°W is shown in figure 5.2. The data gap corresponds to the transition from GM to ERM in November 1986, but with this exception, the method appears to have extracted a distinct signal from extremely noisy data.

5.4 Crossover difference differences

Any unmodelled variation in the amplitudes $A_{\alpha j}$ of equations 4.3-4.6 in section 4.1 will be slow over a single orbital period T_0 to the extent that the

amplitudes may be assumed constant. Hence the change in orbit error from t to $t+T_o$ will be

$$\Delta r(t + T_o) - \Delta r(t) = T_o A_{i4} \sin M$$
$$+ T_o \left(A_{s1} + A_{s2} \sin M + A_{s3} \cos M \right)$$
$$+ T_o \left(A_{d1} + A_{d2} \cos M \right)$$
$$+ T_o \left(A_{g4} \sin M + A_{g5} \cos M \right) \qquad (5.9)$$
$$+ \sum_{k,m} \left(C_{km} \cos \psi_{km}^{\lambda+\chi} + S_{km} \sin \psi_{km}^{\lambda+\chi} \right)$$
$$- \sum_{k,m} \left(C_{km} \cos \psi_{km}^{\lambda} + S_{km} \sin \psi_{km}^{\lambda} \right)$$

where $\psi_{km}^{\lambda+\chi}$ and ψ_{km}^{λ} are the angular arguments described in equation 4.2, evaluated with the relevant values of $(\Omega-\theta)$ for observations at $(t+T_o,\ \lambda+\chi)$ and (t,λ) respectively. The distance χ is defined as the separation of consecutive equator crossing longitudes, that is, the angle through which the Earth has rotated relative to the orbital plane during the period T_o,
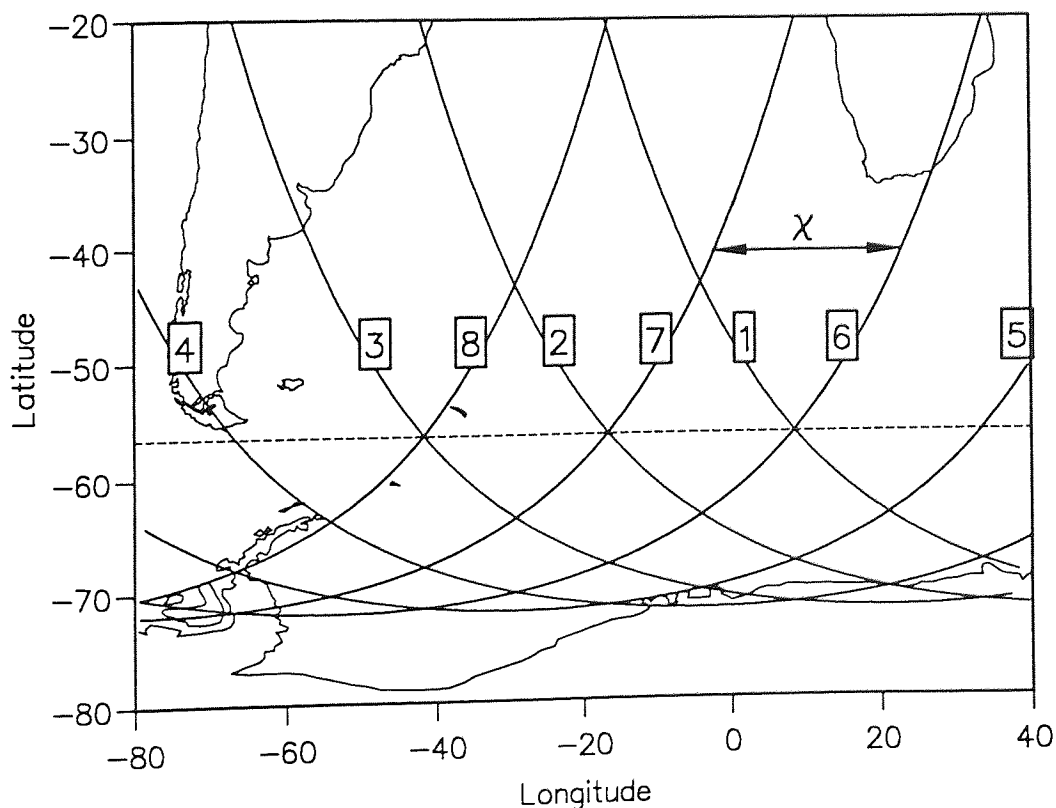
Figure 5.3. Pattern of consecutive Geosat tracks over the South Atlantic, each track separated by χ in longitude.

illustrated in figure 5.3. If T_o denotes the orbital period in seconds, then

$$\chi = \frac{360°T_o}{86400}$$

For Geosat, $\chi \approx 25.1°$. Substituting ξ to denote the difference in the spherical harmonic gravity field mismodelling errors,

$$\xi(\lambda + \chi, \lambda) = \sum_{k,m} \left(C_{km} \cos \psi_{km}^{\lambda+\chi} + S_{km} \sin \psi_{km}^{\lambda+\chi} \right)$$
$$- \sum_{k,m} \left(C_{km} \cos \psi_{km}^{\lambda} + S_{km} \sin \psi_{km}^{\lambda} \right)$$

Then equation 5.9 may be simplified to

$$\Delta r(t + T_o) - \Delta r(t) = T_o \left\{ A_{s1} + A_{d1} + \left(A_{i4} + A_{s2} + A_{g4} \right) \sin M \right.$$
$$\left. + \left(A_{s3} + A_{d2} + A_{g5} \right) \cos M \right\} + \xi(\lambda + \chi, \lambda)$$

From *Moore and Rothwell* [1990], terms A_{i4}, A_{g4}, and A_{g5} have the orbital eccentricity as a factor; Geosat was in a near-circular orbit (that is, with a very low eccentricity), therefore, in comparison to the other coefficients, these terms are negligible and may be ignored, thus leaving

$$\Delta r(t + T_o) - \Delta r(t) = T_o \left\{ A_{s1} + A_{d1} + A_{s2} \sin M \right.$$
$$\left. + \left(A_{s3} + A_{d2} \right) \cos M \right\} + \xi(\lambda + \chi, \lambda) \tag{5.10}$$

Equation 5.10 describes the difference in orbit error of two observations made exactly one orbital period apart. In this time, the satellite will have completed a single orbital revolution and the two observations will be at identical geographical latitudes but displaced χ in longitude due to the revolution of the Earth. Furthermore, assuming that the pair of observations were obtained from consecutive ascending arcs, it would follow that if a descending arc crossed the location of the first observation, then one orbital period later, a second descending arc would similarly cross at the location of the second observation, that is, on the second of the ascending arcs. This is illustrated by the broken line in figure 5.3 for three ascending and three descending arcs. For a crossover $X_{i,j}$ occurring at t_i and

t_j for the ascending and descending arcs i and j, respectively, the second crossover at a distance χ away on the same circle of latitude corresponds to observations at time t_{i+1} and t_{j+1}, for arcs one orbit later.

Equation 5.10 can be determined for both ascending and descending pairs of arcs. At the crossover point of an ascending and a descending arc, the magnitudes of the mean anomalies for each arc are identical but of opposite sign. On using equation 5.10 and the relationships between the mean anomalies, most of the errors can be seen to cancel, leaving only

$$\Delta C_{i,j} = 2T_0 A_{s2} \sin M + \Delta\xi(\lambda + \chi, \lambda) \tag{5.11}$$

This relates to a pair of crossovers at the same latitude, where M is constant. Variations in the error, $\Delta C_{i,j}$, are due entirely to the difference $\Delta\xi$ in ξ between the two crossovers and to changes in the solar radiation pressure (SRP) term A_{s2} (section 5.4.1 gives more detail about the SRP component).

For an ascending pair of arcs i and $i+1$, and a descending pair of arcs j and $j+1$, the error term $\Delta C_{i,j}$, in equation 5.11 will be present in the altimeter derived value,

$$C_{i,j} = \left[h(\lambda + \chi, t_i + T_0) - h(\lambda, t_i) \right] \\ - \left[h(\lambda + \chi, t_j + T_0) - h(\lambda, t_j) \right] \tag{5.12}$$

where $h(\lambda, t)$ denotes the altimeter height measurement at time t and longitude λ, corrected for environmental effects. Alternatively, equation 5.12 may be written to yield the difference in the crossover differences, that is,

$$C_{i,j} = X_{i+1,j+1} - X_{i,j}$$

hence the name crossover difference difference (CDD).

Being free of the static geographically correlated, constant and once per revolution orbit errors, the CDD measurement contains only the difference in the change at the two locations between the time of the ascending and descending pairs of arcs of the following constituents: ocean heights, the

spherical harmonic gravity field induced orbit error, systematic ephemeris modelling errors (due to SRP), tidal modelling errors, atmospheric propagation modelling errors, ocean loading uncertainties and instrument noise.

In common with empirical correction, this procedure leads to removal of the steric signal due to seasonal heating of the oceans. This is a well-known consequence of most methods of radial orbit error correction using altimetry, *Chelton and Schlax* (1993), and is not important for this study so long as the signal is absorbed by both methods, as is likely to be the case. All but the first three of the aforementioned constituents may be considered to possess a near-random distribution about zero and may be apportioned to a random term δ. The CDD may then be written as

$$C = \Delta h + \Delta \xi + \Delta s + \delta \tag{5.13}$$

The systematic error Δs is due to SRP mismodelling, the removal of which is described in section 5.4.1. The spherical harmonic gravity field induced error difference $\Delta \xi$ is independent of time and will be identical for all CDD's obtained from the same pair of locations. Ocean signal Δh on the other hand is obviously strongly spatially and temporally dependent but mesoscale and other longer period components may be considered virtually constant over brief time periods, for example, less than a day. The change in the ocean height at any point will be small when the time difference between observation epochs is also small. As the CDD measures the difference between two of these changes, the ocean signal difference Δh will be negligible for $t_i \approx t_j$ in comparison to the other terms. Further, by averaging over many such CDD's, the noise component δ will average to zero, thereby leaving only the spherical harmonic gravity field induced orbit error component, that is,

$$\overline{C} \approx \Delta \xi$$

Calculation of this mean CDD when the time between ascending and descending arcs is small with respect to significant ocean mesoscale periods gives the gravitational orbit error component, thus allowing correction of all other CDD's for which the time difference is greater, making them virtually free of all orbit error contamination. In addition, the use of crossover residuals $X_{i,j}$ and $X_{i+1,j+1}$ will eliminate any geoid errors from the altimeter data.

Application of these methods to data from the areas to be described in the validation exercise in section 5.5 (north of Tristan da Cunha and over the Cape Basin) improved the 88 cm (rms) crossover residuals in two stages. Calculation of CDD's produced a data set free of most once per revolution orbit error; the rms CDD prior to correction was 72 cm, equivalent to the removal of a 50.6cm rms orbit error signal. The large residual was mainly due to remaining gravity field orbit error which, when removed, brought the CDD rms down to 30 cm. This is approximately $\sqrt{2}$ times the value for empirically corrected crossover residuals, as might be expected from the fact that CDD's measure the difference in the change in sea level between a pair of well separated locations (~1500km apart).

The limiting factor in this process is variation in the coefficients of the orbit error terms over the time for one orbital period T_0. If ephemerides for both observations have been determined from the same orbital solution, the error due to the initial position vector will cancel in repeat pass analysis and hence in CDD's. From *Chelton and Schlax* (1993), where epochs pertain to different long-arc solutions, there will be little or no correlation between the orbit errors at these times resulting in potentially large differences in the coefficients. In the case of Geosat, the orbits for the crossover difference data released by NOAA were determined from overlapping two-day arcs. The probability that two consecutive orbits, T_0 apart, lie across the solution boundary is therefore

$$p = \frac{T_o}{T_{soln}}$$

where T_{soln} is the duration of the long-arc solution. For Geosat, this gives around a 3.5% chance that the orbit errors from two consecutive arcs are poorly correlated (from figure 4.5, the correlation of cosine terms is no greater than 0.7). The correlation of errors toward either end of the solution are also reduced by the fact that consecutive two-day arcs are overlapped and smoothed by a cosine function [*West*, 1986], thus making the accurate elimination of these errors more hazardous. Fortunately, these circumstances apply only to a minority of consecutive arc pairs, and this method may be used to remove the once per revolution error from most differenced observations.

5.4.1 Systematic Solar Radiation Pressure Error

For a non-Sun-synchronous satellite such as Geosat, the position of the Sun relative to the orbital plane varies with a frequency of

$$v = \dot{\lambda}_\Theta - \dot{\Omega}$$

where λ_Θ is the right ascension of the Sun and Ω the right ascension of the ascending node of the satellite; for Geosat, the corresponding period is approximately 337 days. This variation in the incidence angles of solar radiation produces a perturbation which has the effect of slowly modulating the once per revolution component of the orbit error. If modelled as a sinusoid, this systematic error may be expressed as

$$\Delta h_{srp}(t) = F \exp i(vt + \alpha) \exp i(M + \beta)$$

where $v = \frac{1}{337} cycle / day$, α and β are simple phase terms, F is the amplitude of the error and $i = \sqrt{-1}$ This can be simplified by introducing

$$\gamma = \alpha + \beta$$

and rewriting as

$$\Delta h_{srp}(t) = Fe^{i(vt+M+\gamma)}$$

A CDD measurement is made by differencing four height observations; therefore the precise form of this error component in equation 5.13 may be found from

$$\Delta s = \left(\Delta h_{srp}^a - \Delta h_{srp}^d\right)_\lambda - \left(\Delta h_{srp}^a - \Delta h_{srp}^d\right)_{\lambda+\chi}$$

$$= F\left\{\left[e^{i(vt_a+M+\gamma)} - e^{i(vt_d-M+\gamma)}\right]\right.$$

$$\left. -\left[e^{i(v(t_a+T_o)+M+\gamma)} - e^{i(v(t_d+T_o)-M+\gamma)}\right]\right\}$$

$$= Fe^{i\gamma}\left[1 - e^{ivT_o}\right]\left(e^{iM}e^{ivt_a} - e^{-iM}e^{ivt_d}\right) \qquad (5.14)$$

where subscripts a and d denote ascending and descending tracks respectively. Equation 5.14 consists of constant and time dependent elements. The constant terms may be simplified to

$$K = Fe^{i\gamma}\left[1 - e^{ivT_o}\right]$$

thus reducing 5.14 to

$$\Delta s = K\left(e^{iM}e^{ivt_a} - e^{-iM}e^{ivt_d}\right) \qquad (5.15)$$

From equation 5.15, it can clearly be seen that some of the 1 cycle/revolution error remains in the CDD measurement, modulated with a period of 337 days. This characteristic timescale permits identification and removal of the error in a time series analysis.

To validate the nature of this error signal, time series were determined around a complete circle of latitude using the method to be described in the next section. A total of 244 series were generated for 1.6° longitude by 0.3° latitude pixels, each centred on ERM crossover locations at 56.5°S, with a mean separation of 1.47° in longitude. All ocean signal and higher-frequency errors were eliminated by calculating the mean of the 244 time series at regular time intervals. A clear signature near 337 days is evident

in figure 5.4. Also shown in figure 5.4 are sinusoids of period 337 days, fitted separately to the error for the GM and ERM periods of the mission, where amplitudes of around 2 cm were observed. Also evident is a small bias, particularly during the GM which may be associated with residual gravity field modelling errors.
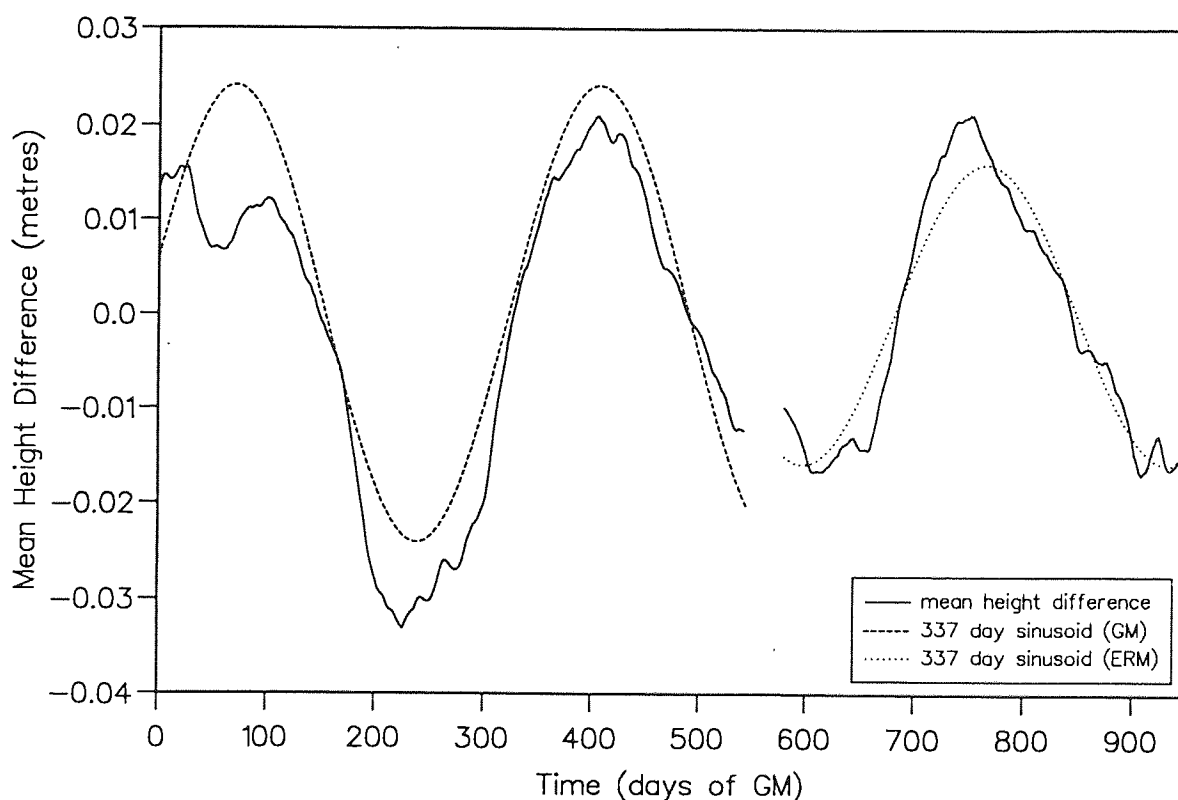
Figure 5.4. Orbit error due to systematic solar radiation pressure modelling errors, derived from the mean of 244 time series 1.47° apart along the 56.5°S circle of latitude.

5.4.2 Height Difference Time Series solution

The starting point for the determination of height difference time series was Geosat crossover difference data from two regions of the South Atlantic, the first being that used in section 5.3 for the crossover difference time series solution, and the second, 25° east, just off the South African coast, over the Cape Basin. These data were corrected as described in the previous section for gravitational orbit error, using CDD's corresponding to a time gap between ascending and descending arcs of around 0.4 days, during

which time, any mesoscale ocean signal could be assumed negligible. A gap significantly greater than the 0.4 days used would result in a larger ocean signal in the data, hence masking the spherical harmonic orbit error $\Delta\xi$ of equation 5.13. Approximately 15% of the data were discarded because of pairs of arcs which were derived from separate orbital solutions or because of degradation of the correlation by the two-arc smoothing of the orbit across the solution boundary. The remaining data provided an extremely consistent estimate of the spherical harmonic orbit error (this is illustrated in figure 5.5 for the same circle of latitude used in section 5.4.1, i.e. at 56.5°S). Close inspection demonstrated a slight difference (less than 5 cm) between the values obtained from the GM and ERM phases of the mission at similar locations. This small difference is not surprising. Although the orbits during the two missions were very similar, they were not identical; therefore under slightly different gravity field influences a small difference in this property was inevitable. Of primary importance is the consistency with time for each individual mission. For the one and a half years of the GM, and the one year of the ERM, no discernible change in this component of the orbit error was found. The spherical harmonic error correction was therefore calculated separately for each mission, and then applied to all the other CDD's, where the time between the two pairs of arcs was greater and, consequently, the ocean signal non-zero. In is interesting to note that comparison of Geosat CDD's with those reported by Moore and Rothwell (1990) for SEASAT in figure 5.6 show an excellent correlation, differing significantly only in terms of a single bias value of around 80cm (indicated by the shifted, lower line).

Solution of the CDD data set to produce time series was performed in exactly the same manner as for the crossover difference solution of section 5.3, but the results were height differences with time between the pixel at λ and its partner at $\lambda+\chi$ instead of the relative heights at a single location. Figure 5.7 shows a typical height difference time series between two regions or pixels, centred on 43.5°S, at longitudes of 20°W and 45°W.
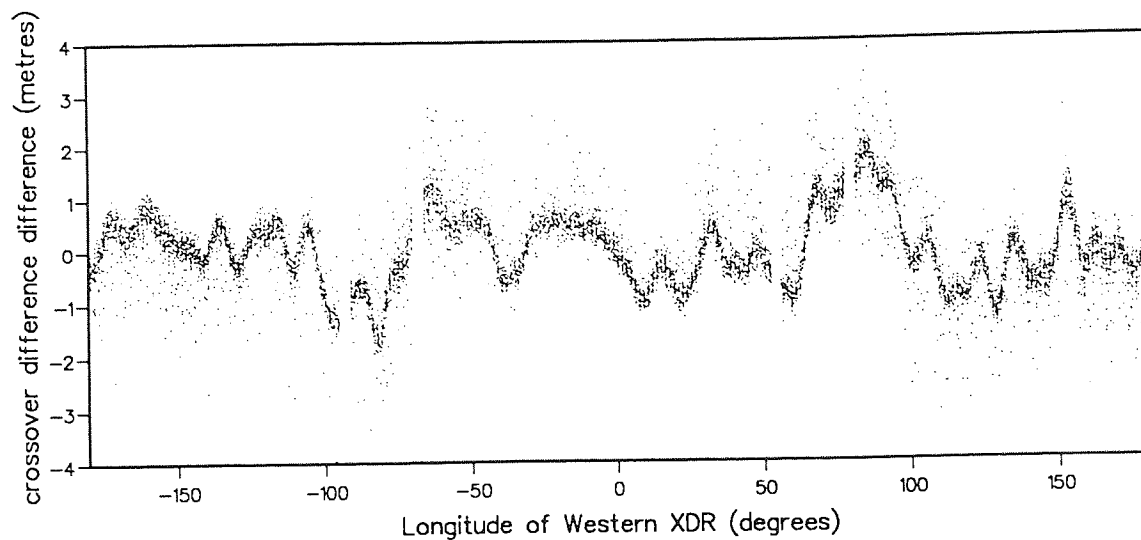
Figure 5.5. Difference in spherical harmonic gravity field induced orbit error at 56.5°S between points χ apart in longitude.
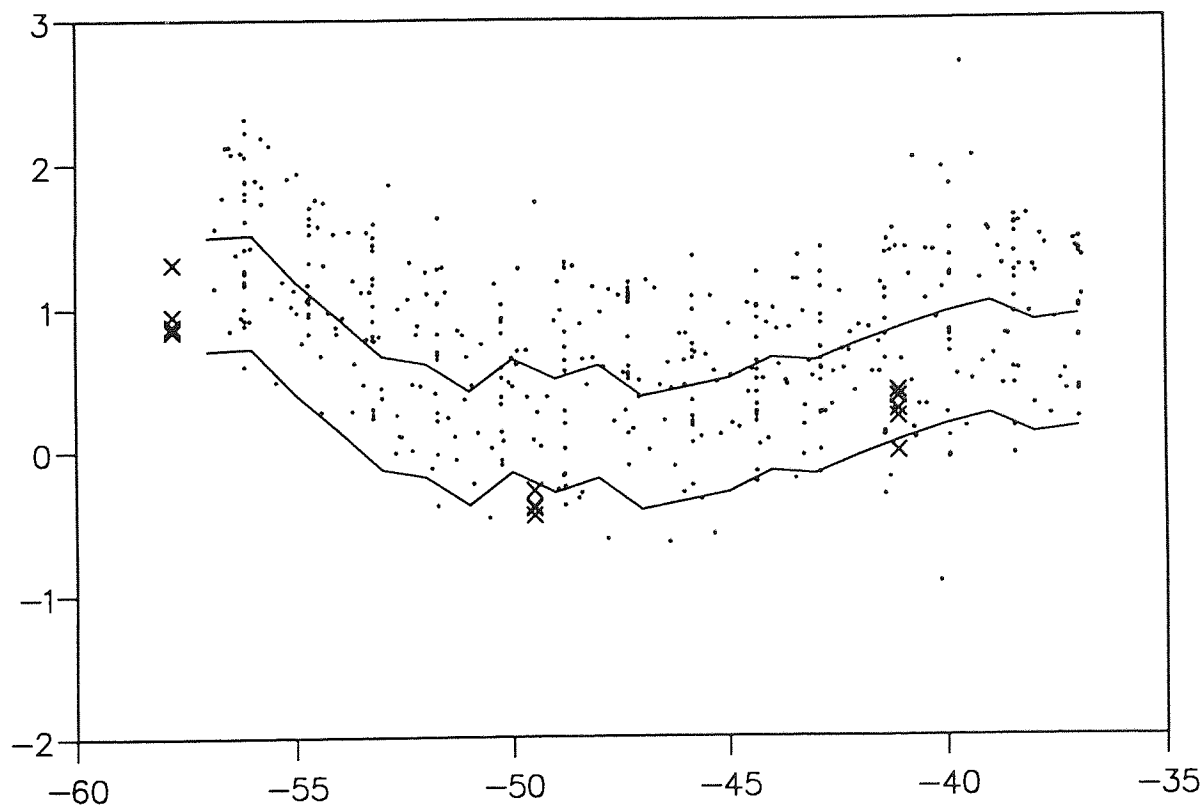
Figure 5.6. Comparison of Geosat and SEASAT CDD's as reported in *Moore and Rothwell* (1990) - Geosat denoted by dots, SEASAT by crosses.
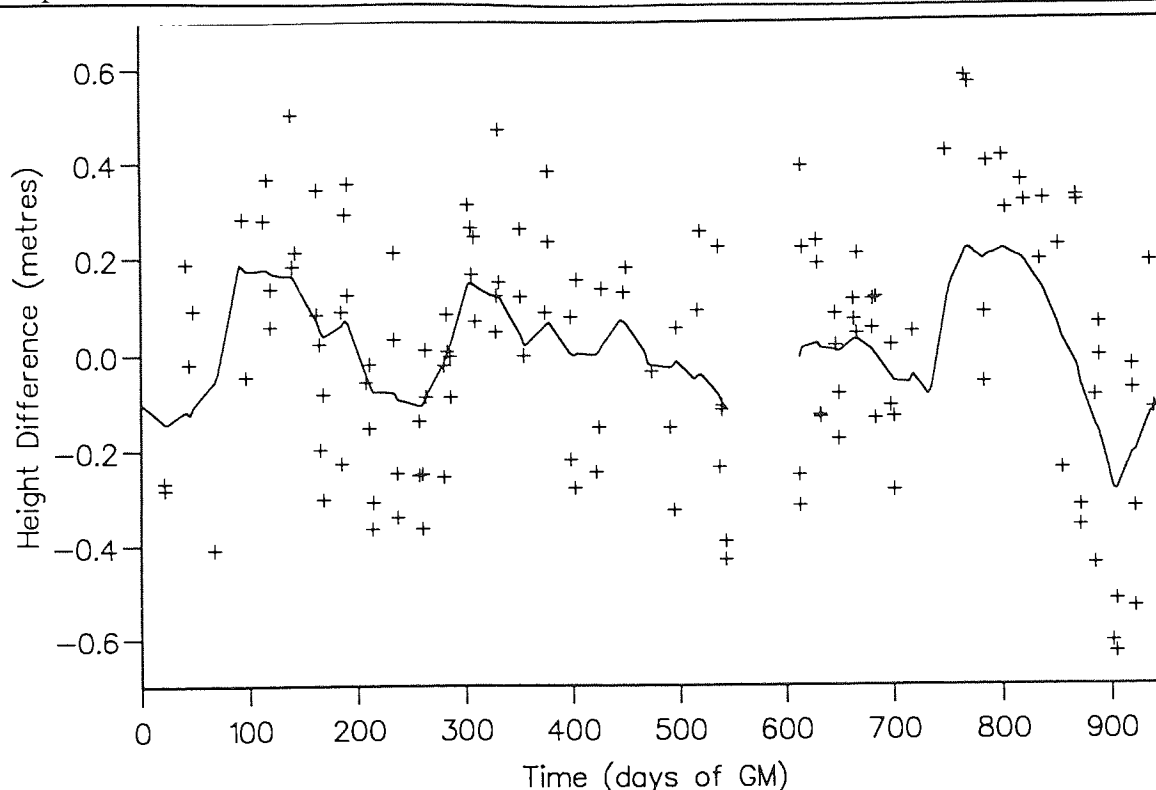
Figure 5.7. Height difference time series derived from crossover difference difference data, using crossover around 43.5°S, 45°W and 20°W (+: piston heights; ___: Optimal estimate of ocean height.).

5.5 Comparison of Results

Comparison of results obtained from the methods in sections 5.3 and 5.4 was undertaken by deriving time series from the common datasets. Calculating CDD's produced a time series of ocean height differences, while using corrected crossover differences produced time series of relative ocean heights. To compare the two, crossover difference time series were produced in pairs where the pixels were ~25° apart. These two series were then subtracted to give height differences between the pixels - exactly what was produced by the CDD time series analysis, where $\chi \approx 25°$.

Figure 5.8 displays the resultant time series from four pairs of such pixels centred on 36.6°S, 17°-9°W for the western pixels, and 8°-16°E for the eastern, each of which is 1.1° in latitude by 2.0° in longitude. The correlations in table 5.1 show the agreement of the two series to be extremely good, especially for pixels A and D. For pixels B and C, the

match is still good but may have been corrupted slightly by the proximity of the island of Tristan da Cunha in the western set of pixels. This may have increased the tidal mismodelling errors for these areas, thus reducing the signal to noise ratio of the data, making the fit of both methods to the three signals (two time series and one CDD series) less reliable.
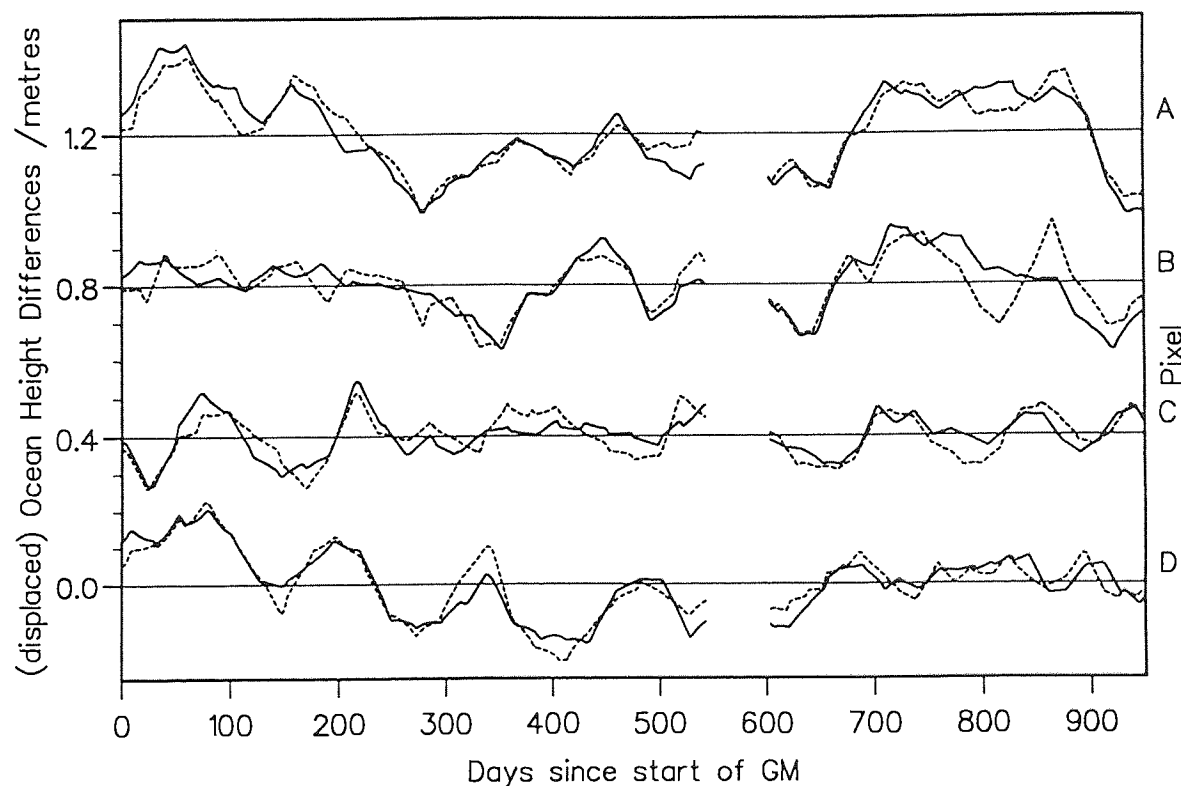
Figure 5.8. Comparison of height difference time series: from differences in crossover height time series (broken line, section 5.3); and crossover difference difference time series (solid line, section 5.4). Pairs of series are displaced by multiples of 40 cm for ease of presentation.

As both CDD's and empirical sinusoidal correction remove the annual steric ocean signal, the residuals demonstrate only the differences in ocean signals which have some degree of longitude dependence. Analysis of the residuals in figure 5.9 reveals how much of this ocean signal has been absorbed. A number of conclusions may be drawn from table 5.1. Fitting a bias and tilt to the residuals showed that there was a mean tilt of -0.4 cm per year from an initial bias of 0.5 cm on day 1 of the mission. Over the 2.5 year data set this indicated a maximum drift of less than 1.5cm, well within what may be considered an acceptable error margin to vindicate the use of a

25 day reference set of arcs. Independently, a sinusoid of period 1 year was fitted to the residuals, from which the maximum annual signal was found to be less than 2 cm with a mean value under 1.5 cm for the four pixels. This sets a limit on the annual non-steric long-wavelength ocean signal absorbed erroneously by the empirical correction of the Geosat altimeter data.
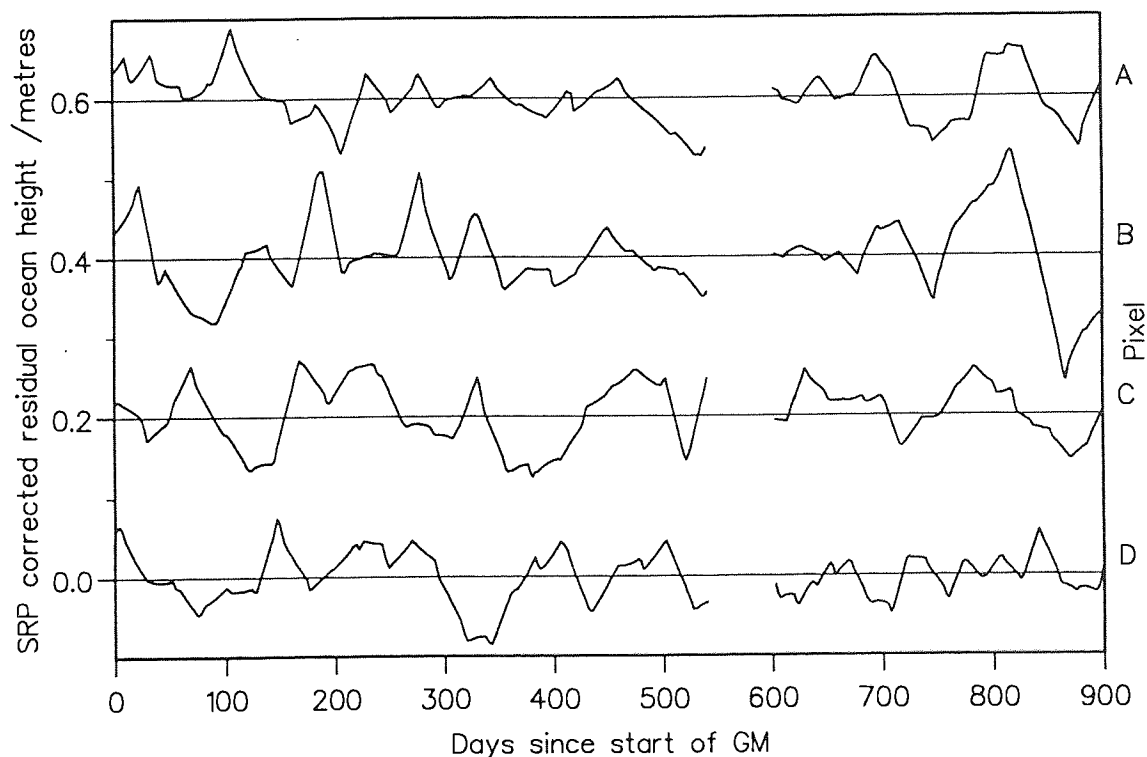
Figure 5.9. Solar radiation pressure corrected residual between series in figure 5.8, displaced by multiples of 20cm.

Table 5.1. Analysis of crossover difference difference residuals in figure 5.9.

	Pixel			
	A	B	C	D
Correlation	0.96	0.76	0.73	0.94
RMS residual	3.2cm	4.3	3.2	2.9
Bias at t=0	1.1cm	0.5	0.3	0.1
Tilt	-0.9cm/yr.	-0.4	-0.3	-0.1
Cosine coefficient	1.3cm	1.5	-1.1	-1.1
Sine coefficient	0.3cm	-1.0	-0.6	0.2

5.6 Secondary time series generation

Section 5.4 described the production of a time series from crossover difference difference data: a time series that is virtually free of any orbit error contamination and as such is purely ocean signal (plus residual ocean and atmospheric modelling errors). For this reason, a CDD time series may also be used for oceanographic analysis. This section introduces one such application.

The simplest application of a CDD time series is a rearrangement of the formula for the ocean height difference at two points,

$$C(\lambda,\lambda+\chi) = h(\lambda) - h(\lambda+\chi) \, ,$$

which becomes,

$$h(\lambda+\chi) = h(\lambda) - C(\lambda,\lambda+\chi) \, , \tag{5.16}$$

allowing determination of the relative ocean height at a point χ to the east or west of a point at longitude λ, where the ocean height is already known without the need to carry out further orbit error correction analyses. This may be performed using an existing altimeter-derived time series or tide gauge data. As an illustration of this figure 5.10 shows the altimeter-derived relative ocean height time series at 36.6°S, 9.0°E off the coast of South Africa, over the Cape Basin (broken line). Shown along side as the solid line is the time series produced from the combination of an altimeter time series near Tristan da Cunha and the CDD time series performed as described in equation (5.16). For the spherical harmonic gravity field error, only data corresponding to these two areas were required to derive a correction. To remove the systematic SRP error present in this Geosat crossover difference data set, a sinusoid of period 337 days was fitted to several such time series on the same circle of latitude. This correction used eight time series derived for the validation exercise in section 5 and produced a small improvement in the rms residuals of around 5 mm when compared with a time series generated without this correction.
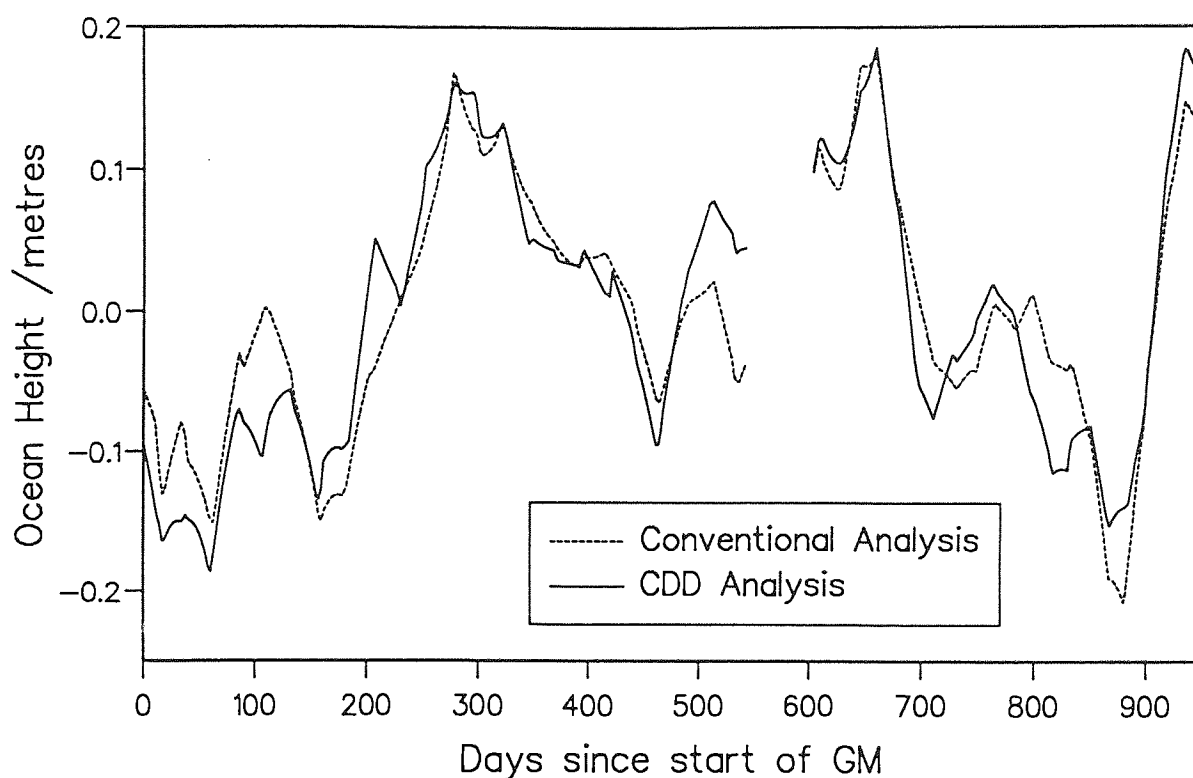
Figure 5.10. Comparison of conventional ocean height time series generation and secondary crossover difference difference generation of ocean height time series.

In principal, this process may be repeated by making consecutive 25° "steps" around the circle of latitude to obtain a time series at any longitude. In practice, however, residual errors and uncertainties reduce the correlation between the genuine and the secondary time series at each step until no useful result can be obtained. Ultimately, however, good tidal and atmospheric modelling in conjunction with an accurate mean sea surface may permit the swift and above all accurate time series evaluation at any point on a circle of latitude from any altimeter mission, whatever the quality of the orbits.

5.7 Conclusion

Oceanographers require an accuracy of 10cm or better from altimetry to produce worthwhile results. The contamination of the radial range measurement from most low Earth orbiting satellites is significantly greater

than this and therefore requires further analysis to remove the characteristically long wavelength orbit error signals present in the data. Single satellite crossover minimisation described in Chapter 4 suggested that a considerable improvement could be made to the estimated radial accuracy of Geosat data by using constrained sinusoidal analysis to form a reference set of arcs with which the remaining 2.5 years of data could be corrected.   It has been shown that a relative ocean height signal can be extracted from a dataset corrected in such a way - despite the obvious remaining noise contamination illustrated in figure 5.2.   In the absence of any independent measure of the ocean height during this period and at this particular location, the validity of the observed signal cannot be confirmed. Using crossover difference differences, however, the methods of correction can be rigorously checked and validated, thus improving the confidence of the relative height results.

The sensitivity of CDD's to small, yet systematic, errors was shown by the isolation of a 2cm solar radiation pressure modelling error (figure 5.4), removal of which left only the raw change in the sea height difference between the times of ascending and descending observations.   Using an identical analysis regime, CDD's were used in place of crossover differences to obtain ocean height difference time series which were then used to validate the results obtained from sinusoidally corrected crossover differences.

Comparison of these time series conclusively demonstrated that sinusoidal correction is a reliable method for the reduction of orbit error contamination of satellite altimetry.   There was found to be no significant drift of ocean signal due to the use of a reference set of arcs at the beginning of the GM phase, and no definite long period ocean signal absorption. Finally, the rms agreement of the comparison is such as to suggest that optimal interpolation has been successful in accurately filtering the ocean

signal from the noise relative height estimates obtained from solution of a small area piston term.

Having established the confidence of sinusoidal correction of single satellite crossovers, it may reasonably be assumed that the same confidence may be applied to dual satellite crossovers, in particular given the use of full ascending or descending ERS-1 tracks as compared with 11500km Geosat tracks. The next chapter will therefore look at mean sea surface (MSS) recovery from altimetry, leading ultimately to the generation of a precise, high resolution MSS map from sinusoidally corrected ERS-1 data.

Chapter 6

Regional mean sea surface determination

6.0 Introduction

With a spacing of around 7km, one-per-second radar altimetry from satellites such as ERS-1 and Geosat provide dense coverage of the Earth's oceans along track. Across track, however, the sampling interval can be several hundreds of kilometres, presenting a problem in the oceanographic analysis of data. Additionally, even for 'exactly repeating' missions, there is usually a drift of ±1km over time in the cross-track location of the satellite on a particular arc. Combined with the fact that along track sampling rarely occurs at identical points in the orbit for each pass, it is virtually impossible to find two altimeter observations of exactly the same point. Fortunately, the spatial spectrum of the MSS, SST and meso-scale variability are such that the majority of their signal is at wavelengths greater than the spot size of the altimeter (tables 2.2 and 3.1), thus permitting interpolation of the sampled dataset to common locations for subsequent analysis.

This chapter looks at interpolation of altimetry by two different methods: the first makes use of localised base functions, the heights of which are evaluated by least squares minimisation; the second is a statistical technique similar to that described in section 5.2 - optimal interpolation. The techniques are used to derive mean sea surfaces for an area of the North Sea from Geosat ERM data.

6.1 Base function interpolation

Simple averaging techniques for the interpolation of satellite altimetry are reliant on the initial reduction of orbit errors before sea surface heights

can be interpolated.   Because of the arbitrary nature of the choice of weighting function and the piecemeal way in which the heights are calculated at the interpolation grid points, statistical interpretation of the results may be unreliable.   A better means of interpolating the surface is therefore a simultaneous solution of both orbit correction terms and sea surface heights.   The following method is based on those described by *Lelgemann* (1981) and *Müller and Groten* (1991).

In order to estimate the ocean height at an arbitrary point within a regular grid of surface heights requires some form of interpolation, for example bilinear interpolation.   As an alternative, the values of the interpolation points may be thought of as coefficients of individual two dimensional base functions, whereby the height at any point is determined by the sum of contributions from all the base functions.   Each base function β_i has a shape determined in advance, after which, a scaling factor α_i can be derived through least squares minimisation of the data.   Choice of the functions β_i and their distribution is flexible but should meet a number of criteria in order to best represent the surface.   Firstly, the whole region of interest should lie within the correlation length of at least one base function, where that correlation length should be chosen to reflect the data sampling pattern, the base function distribution pattern and the physical characteristics of the signal being interpolated.   The base functions should also reflect something of the nature of the covariance function of the MSS, which implies that the base function should depart from zero only over some small part of the globe.   *Müller and Groten* (1991) suggest the use of a function $\beta_i^{M\&G}$ of the form

$$\beta_i^{M\&G} = \frac{c_i}{\left(\psi^2 + c_i^2\right)^{3/2}}$$
(6.1)

where the correlation length is determined by the constant c_i, and ψ is the separation of the base function point, k, and the altimeter observation, j. From spherical trigonometry this can be written,

$$\cos\psi = \sin\phi_j \sin\phi_k + \cos\phi_j \cos\phi_k \cos(\lambda_j - \lambda_k) \tag{6.2}$$

where ϕ represents the latitude of the point, λ its longitude and R_e is the mean radius of the Earth. Over a small area, *Müller and Groten* (1991) state that c_i may be considered a constant when a regular grid of base functions are used. In equation 6.1, c_i has dimensions of LENGTH, as does the spherical distance, ψ; as a result, the dimensions of $\beta_i^{M\&G}$ are *LENGTH*$^{-2}$. Therefore to result in a product of the correct dimension, the units of α_i must be *LENGTH*^{3}.

To ensure that α_i has dimensions of LENGTH, a general base function was chosen to be of the dimensionless form

$$\beta_i = \frac{c_i^3}{\left(\psi^2 + c_i^2\right)^{3/2}} \tag{6.3}$$

The correlation length $\psi_{correlation}^i$ for a particular base function β_i may be found by substitution of β_i with the value equal to half its maximum (when $\psi = 0$, from which it follows that $\beta_{maximum} = 1$), i.e. rearranging equation 6.3

$$\psi_{correlation}^i = \sqrt{\left(2^{2/3} - 1\right)c_i^2} \approx 0.766c_i$$

Reconstruction of the height at any location is possible though the summation of the individual contributions from each base function, once the α_i coefficients are known

$$h' = \sum_i \alpha_i \beta_i \tag{6.4}$$

Conversely, this expression can be used to solve for the individual α_i terms. Perhaps the simplest means of achieving this is by least squares minimisation. Taking the squared residuals of equation 6.4, the following expression is obtained,

$$I = \sum_k \left(h_k - \sum_i \alpha_i \beta_i\right)^2 \tag{6.5}$$

The minimum of I is found when the partial derivatives with respect to each of the unknown parameters are zero, i.e.

$$\frac{\partial I}{\partial \alpha_j} = 2\sum_k \left(h_k - \sum_i \alpha_i \beta_i \right)(-\beta_j)$$

$$= 2\sum_k \left(\sum_i \alpha_i \beta_i \beta_j - h_k \beta_j \right) = 0 \qquad (6.6)$$

Using matrix notation, this may be re-expressed in the form Ax=b (as in section 4.2.2), where x is a vector containing the unknowns and the elements of matrix A and vector b are

$$A_{ij} = \sum_k \beta_i^{(k)} \beta_j^{(k)} \text{ and } b_i = \sum_k h_k \beta_i^{(k)}$$

One advantage of least squares is the possibility of solving for many different terms simultaneously and in a consistent manner. It is a simple task to extend equation 6.5 to include a description of the orbit error. For an area of this size, orbit error can reasonably be approximated by a straight line and hence, the description added to equation 6.5 was in the form of a bias and tilt model, i.e.

$$I = \sum_k \left(h_k - \sum_i \alpha_i \beta_i - \sum_{m,n} (a_m \Delta t_n + b_m) \right)^2$$

where the subscripts m and n denote, respectively, the arc number and the observation number on that arc. The term Δt_n is the time of the observation n since some known reference point on the arc, while a_m and b_m are the arc dependent unknowns to be solved for. Due to the requirement of least squares that the rank of the normal matrix A be equal to the number of unknowns, there is a practical limit to the number of terms that can be solved for and hence a limit to either the resolution or the size of the area covered by the MSS produced. To maintain spatial resolution, for this section, 9599 Geosat altimeter observations from 6 months of the ERM were selected over an area of the North Sea.

The sea height measurements were obtained from the GDR orbits produced by the Naval Astronautics Group (NAG) which had an accuracy of around 70cm rms [*Cheney et al.*, 1991]. Base functions were located in order to best represent the true, non-uniform sampling of the sea surface by distributing them on a diamond shaped grid similar to the local satellite ground track pattern, illustrated in figure 6.1, with a mean separation equal to one quarter the inter-track spacing and a correlation length which varied with latitude due to the shape of the ground tracks which are closer together at higher latitudes. This represented a compromise between the along and cross-track spacing of altimeter observations and gave 961 (31x31) base functions and 294 (2x147) orbital parameters to solve for. A correlation length equal to half the separation of the nearest base functions at the same latitude was used to determine individual values of c_i, thus ensuring complete coverage of the area with out excessive low pass filtering of the ocean signal.
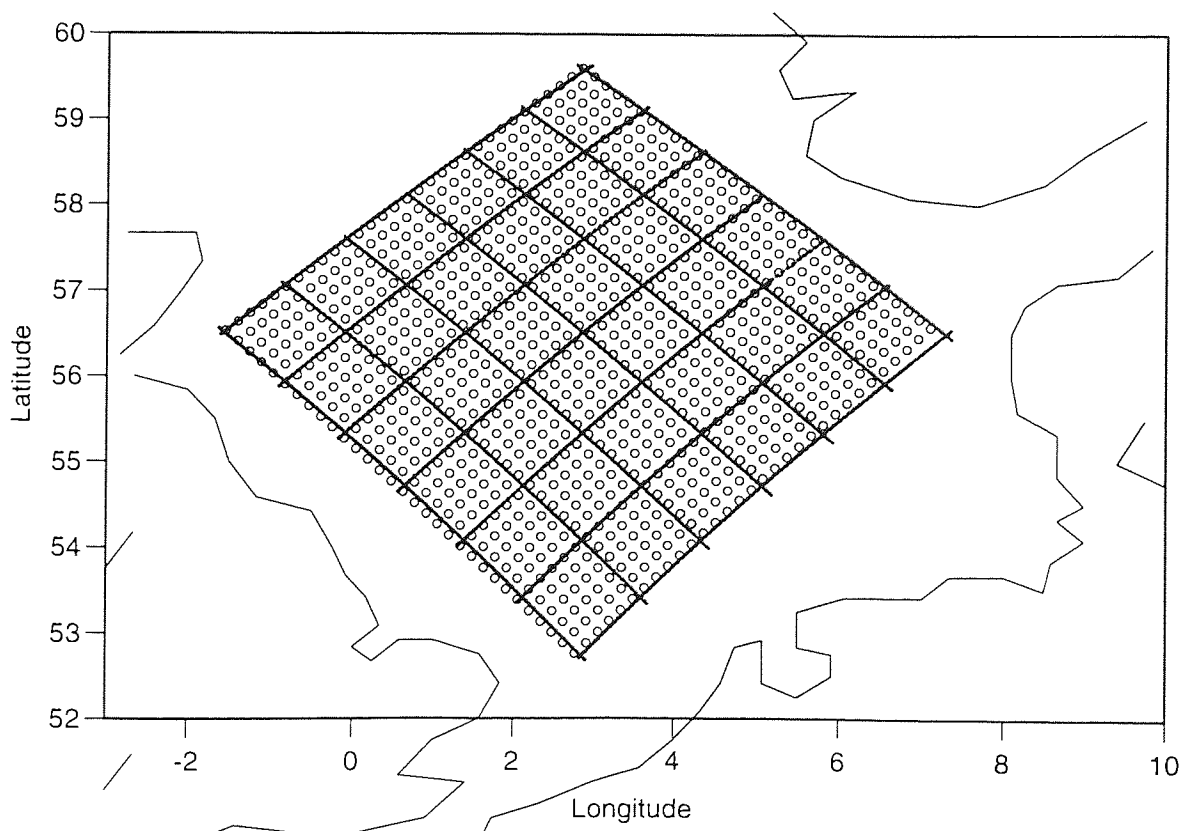
Figure 6.1. Base function distribution pattern of 961 locations, overlaid by altimeter data points.
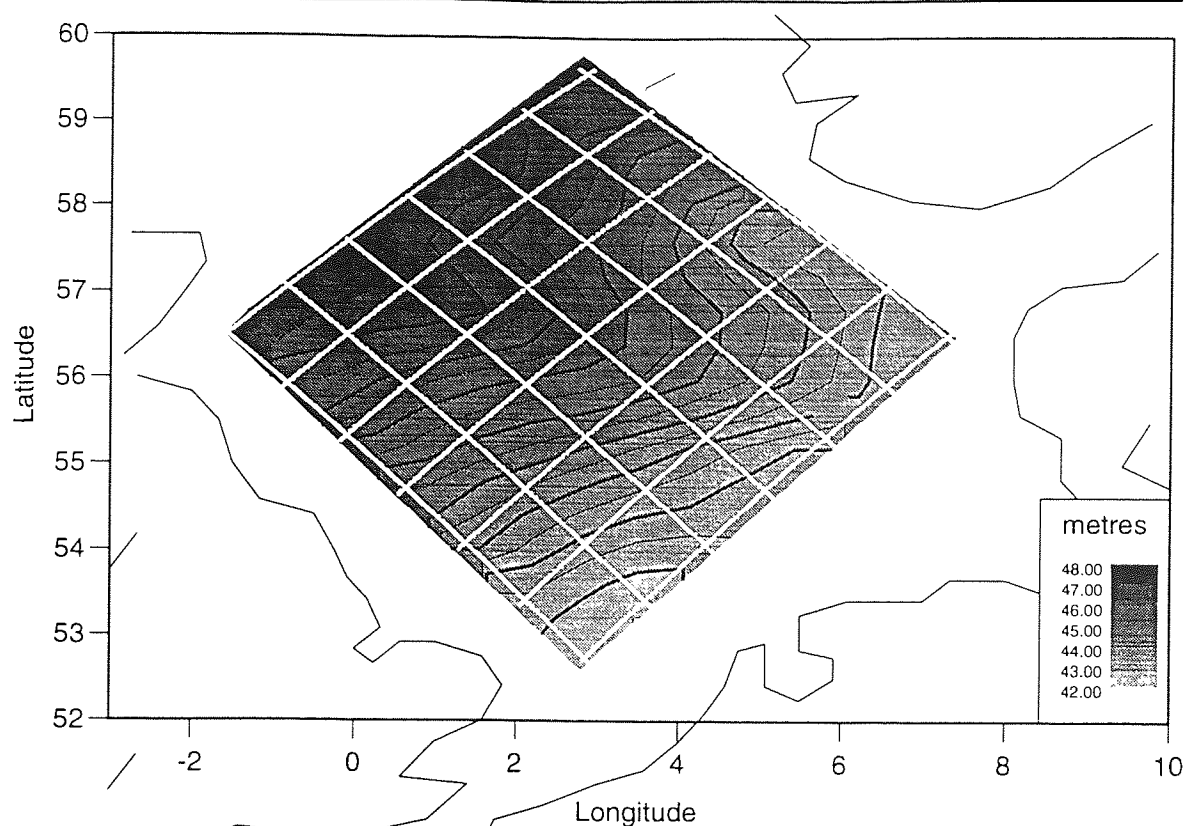
Figure 6.2.   Interpolation of 6 months of Geosat ERM data in the North Sea using 961
              base functions.

Using the full 9599 point dataset, solution took 2 hours and 5 minutes on a VAX-8650 mainframe computer for the total of 1255 terms. When compared to a local high resolution geoid, *Stewart and Hipkin* (1989), there was found to be a significant slope due to aliasing of long wavelength geoid signal into the bias and tilt orbit error correction. By fitting and removing this slope with a bias and two orthogonal tilts by least squares there was found to be an improvement from 84.1cm for altimetry before fit to an agreement of 18.1cm between the altimetric MSS and the geoid model afterwards. Following removal of this slope from the MSS to replace the long wavelength signal, a corrected MSS with the associated comparison with the geoid is displayed in figure 6.2.

6.2 Optimal interpolation

Most mean sea surface interpolation methods require approximation of the orbit error by some arbitrary long wavelength function, either before

solution or simultaneously in the case of the base function approach. There is however a great deal of indeterminacy in such approximations and a potentially better method called optimal interpolation was proposed by *Wunsch and Zlotnicki* (1984), based on techniques described by *Moritz* (1980) as least squares collocation, the mathematical background of which has already been described in section 5.2. The method does not explicitly remove orbit error, but instead exploits the spatial and temporal characteristics of the signals and errors present in the data, to filter and identify those which are important to the analysis. To implement this method effectively requires realistic descriptions of the signals and their errors.

The covariance models used in this analysis were as follows, and relate to characteristics of the Geosat ERM dataset with NAG orbits.

• Mean sea surface covariance function

A reasonable first guess of the MSS heights can be obtained from an accurate geoid model derived from analysis of orbit perturbations, e.g. GEM-T2 [*Marsh et al.* ,1989] or including altimetry data, e.g. JGM-2, *Nerem et al.* (1994). The formal errors (δC_{lm} and δS_{lm}) in the spherical harmonic coefficients can give the power spectrum of the long wavelength residual geoid from the sum of these errors as

$$\sigma_l^2 = R^2 \sum_{m=0}^{l} \left(\delta C_{lm}^2 + \delta S_{lm}^2 \right)$$

where R is the mean radius of the Earth and σ_l is referred to as the degree variance, defined for l up to degree and order L_1. To extract medium and short wavelength features, the power spectrum can be extended using a modification of Kaula's rule of thumb [*Kaula*, 1966] as described by *Blanc et al.* (1991), i.e.

$$\sigma_l^2 = \frac{1}{2} 10^{-7} \frac{1}{l^{4.5}}$$

The a priori covariance function can then be determined for a spherical angle ψ, from the degree variances by

$$C_{MSS}(\psi) = \sum_{l=2}^{\infty} \sigma_l^2 P_l\{\cos(\psi)\} \tag{6.7}$$

where P_l are the Legendre polynomials. Practically, a summation over infinite terms is impossible therefore a cut-off must be introduced which will produce a high frequency limit in the solution, corresponding to wavelengths shorter than

$$\lambda_{\text{cut-off}} = \frac{2\pi R}{L_{\text{max}}}$$

This summation is computationally intensive and was therefore simplified by calculation of values of $C_{MSS}(\psi)$ at regular intervals of $\cos(\psi)$. Using the GEM-T2 degree variances, the results of this were then linearly interpolated to produce a good estimate of the actual covariance for any given spherical angle.

• Orbit error covariance function

As described in Section 4.1, the dominant orbit error component has a frequency of one cycle per revolution which may be simply approximated by a sinusoid. Assuming terms at other frequencies to be negligible in comparison, the covariance function of such an error can be approximated by an exponentially decaying cosine [*Wunsch and Zlotnicki,* 1984]. For two observations Δt apart in time, their covariance C_{orb} may be expressed as

$$C_{orb}(\Delta t) = \sigma_{orb}^2 \cos(\omega_o \Delta t) \exp\left(-\frac{\Delta t^2}{T_{orb}^2}\right) \tag{6.8}$$

That is, the variance of the error signal multiplied by a cosine of frequency ω_o equal to one cycle per revolution, the amplitude of which decays exponentially at a rate determined by T_{orb}. Geosat's nominal radial accuracy is approximately 70cm (σ_{orb}) for the ERM with NAG orbits and a decorrelation time of 5 days is assumed.

• Mesoscale ocean signal covariance

Mesoscale variability is highly geographically dependent; for example regions of the Southern Ocean have variabilities several times greater than those found in the Mediterranean Sea. According to *Mazzega and Houry* (1989), the covariance however is more consistent globally and can be approximated by a Gaussian function of space and time. With decorrelation lengths of 130km and 17 days, the covariance function may be written

$$C_{meso}(\psi, \Delta t) = \sigma^2_{meso} \exp\left(-\frac{\psi^2}{R^2_{meso}}\right) \exp\left(-\frac{\Delta t^2}{T^2_{meso}}\right)$$

From *Zlotnicki et al.* (1989), the Northeast Atlantic may be considered to have a homogeneous variability of magnitude 10cm rms (σ_{meso}); for the North Sea, this is probably an overestimate but due to the problems of tidal modelling in a shallow enclosed sea, the excess was assumed to account for any short wavelength tidal mismodelling which is likely to have a similar covariance.

• Atmospheric modelling covariance

The second of the environmental corrections, due to atmospheric propagation effects, has an uncertain covariance but has been assumed to have a form

$$C_{atm}(\psi, \Delta t) = \sigma^2_{atm} \cos\left(\frac{2\pi\psi}{\psi_o}\right) \exp\left(-\frac{\psi^2}{R^2_{atm}}\right) \exp\left(-\frac{\Delta t^2}{T^2_{atm}}\right)$$

Blanc et al. (1989) quote values for the variance of 10cm rms (σ_{atm}) and a cosine period of 3200km (ψ_o), and spatial and temporal decorrelation of 1700km (R_{atm}) and 5days (T_{atm}) respectively.

• Method and results

To form the auto-covariance matrix, all of the above contributions were included in addition to the covariance of the altimeter instrument,

assumed to be white noise, σ_{inst}=5cm and modelled with a Dirac-delta covariance function, δ_{inst}. The final auto-covariance is therefore

$$C_{ll}(\psi, \Delta t) = C_{MSS}(\psi) + C_{orb}(\Delta t) + C_{meso}(\psi, \Delta t) + C_{atm}(\psi, \Delta t) + \sigma_{inst}^2 \delta_{inst}$$

The cross-covariance function however, contains only the relation between the data and the signal at the points of interest, i.e. the MSS covariance, therefore

$$C_{sl}(\psi) = C_{MSS}(\psi).$$

In order to provide a comparison, least squares collocation was used on a subset of that dataset analysed in the last section using base functions. To reduce the rank of C_{ll} to under 2000 for inversion on a VAX-8650 mainframe, each fifth point in the dataset was arbitrarily selected, and C_{ll}
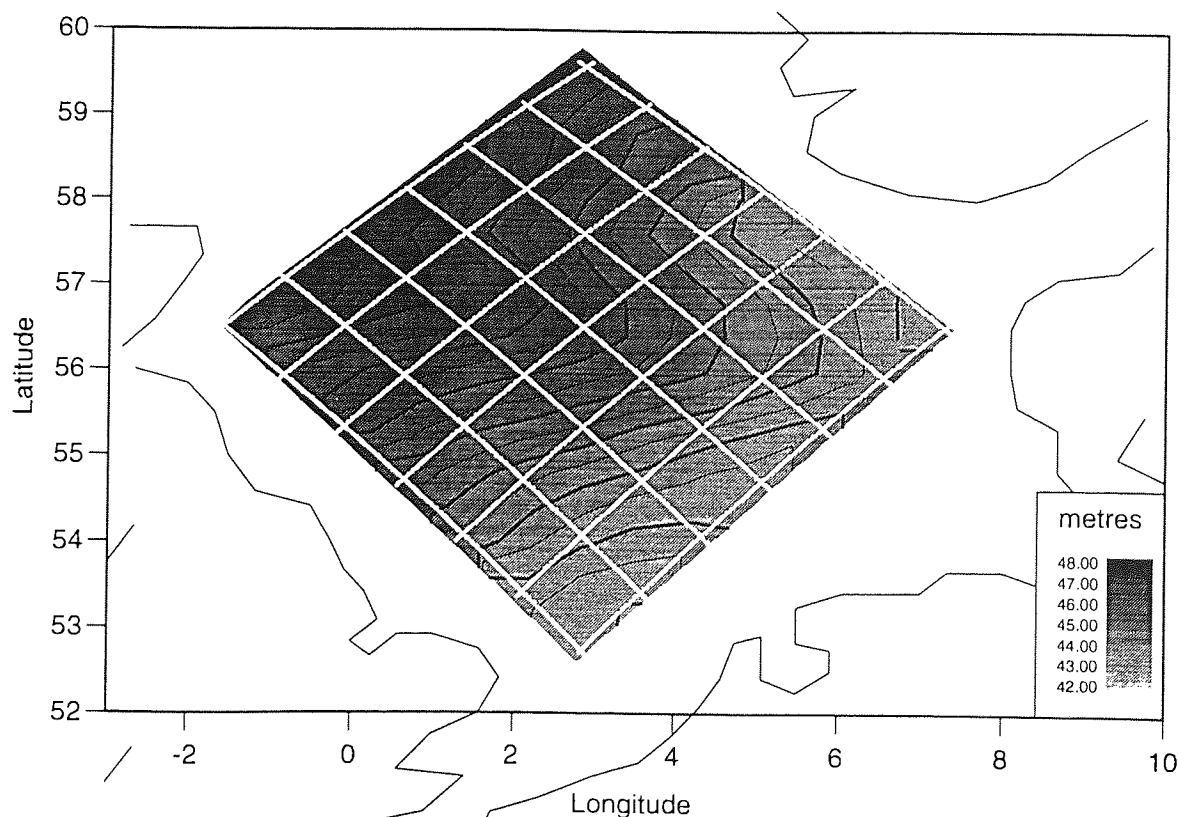
Figure 6.3. Optimally interpolated mean sea surface in the North Sea from 6 months of Geosat ERM data.

formed from the separation in time and space of each observation. The same 0.25°x0.25° grid from the previous section was used to interpolate the surface for comparison. C_{sl} was therefore formed between each grid point

and each data point using only the MSS covariance function. Inversion of C_{ll} took approximately six hours of processor time to complete, as compared to around two hours for the base function inversion. The result of this solution is shown in figure 6.3. When compared to a local geoid model [*Stewart and Hipkin,* 1989], for the area, the short arc lengths used resulted in loss of absolute accuracy and a small bias and tilt correction was required. Afterwards, the comparison of the Hipkin model with the optimally interpolated Geosat data gave an rms difference of only 12cm.

6.3 Intercomparison and analysis of results

The two methods described in this chapter, namely, base functions and optimal interpolation, are limited by the size of matrix to be inverted at some stage in the analysis. To investigate the use of these techniques, six months of Geosat ERM data were selected covering an area of the North Sea, corresponding to that shown in figure 6.1. This comprised 9599 observations from 147 arcs, spread over 7 ascending and 7 descending repeat ground tracks - approximately 8-10 passes per track. Using the methods described in sections 6.1 and 6.2, mean sea surfaces were produced for comparison with an independent model: this was chosen to be a local North Sea geoid model produced by *Stewart and Hipkin* (1989), from corrections to the OSU86 [*Rapp and Cruz*(1986)] geoid using gravimetric data, illustrated in figure 6.4. Comparison of the MSS solutions was performed by subtraction of the Hipkin Geoid, thus producing the residual surfaces shown in figures 6.5 and 6.6 for the base function and optimal interpolation methods respectively. In the actual surfaces of figures 6.2 and 6.3, it was evident that both showed similar features to the geoid; the
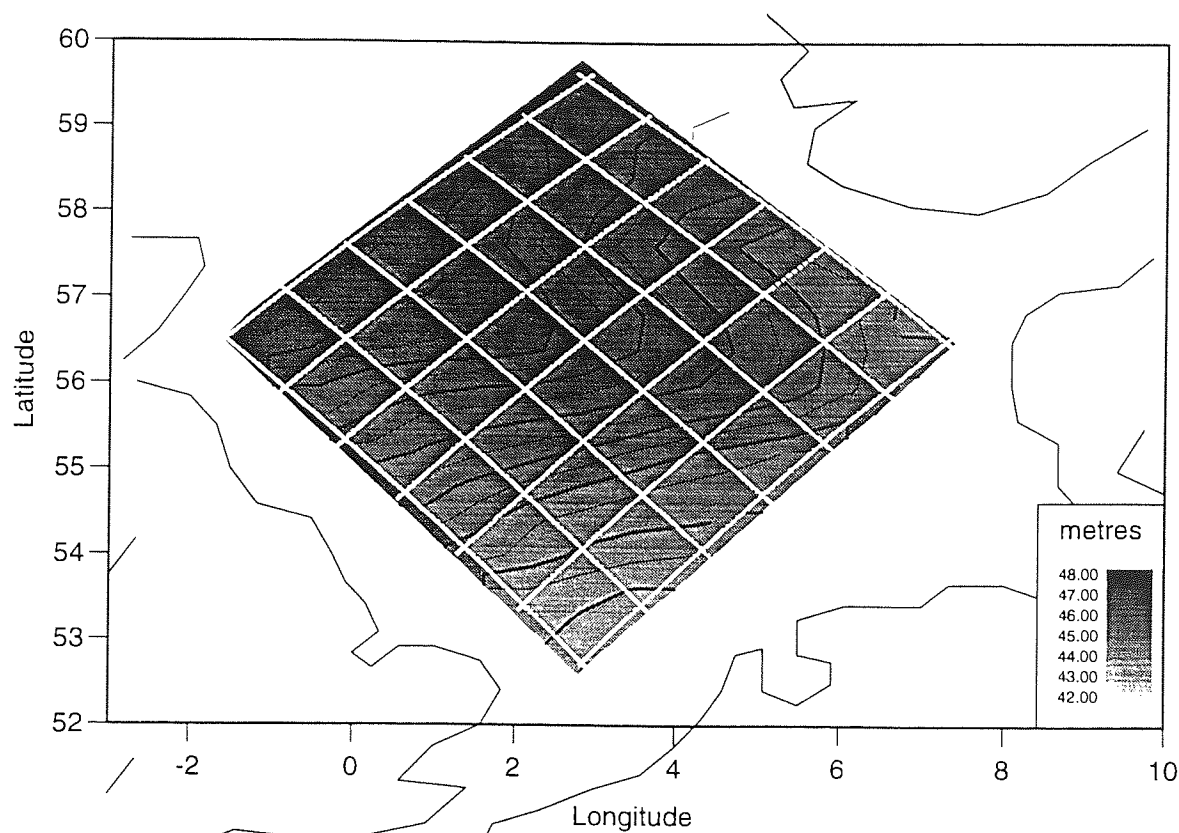
Figure 6.4. The Hipkin gravimetric geoid of the North Sea.
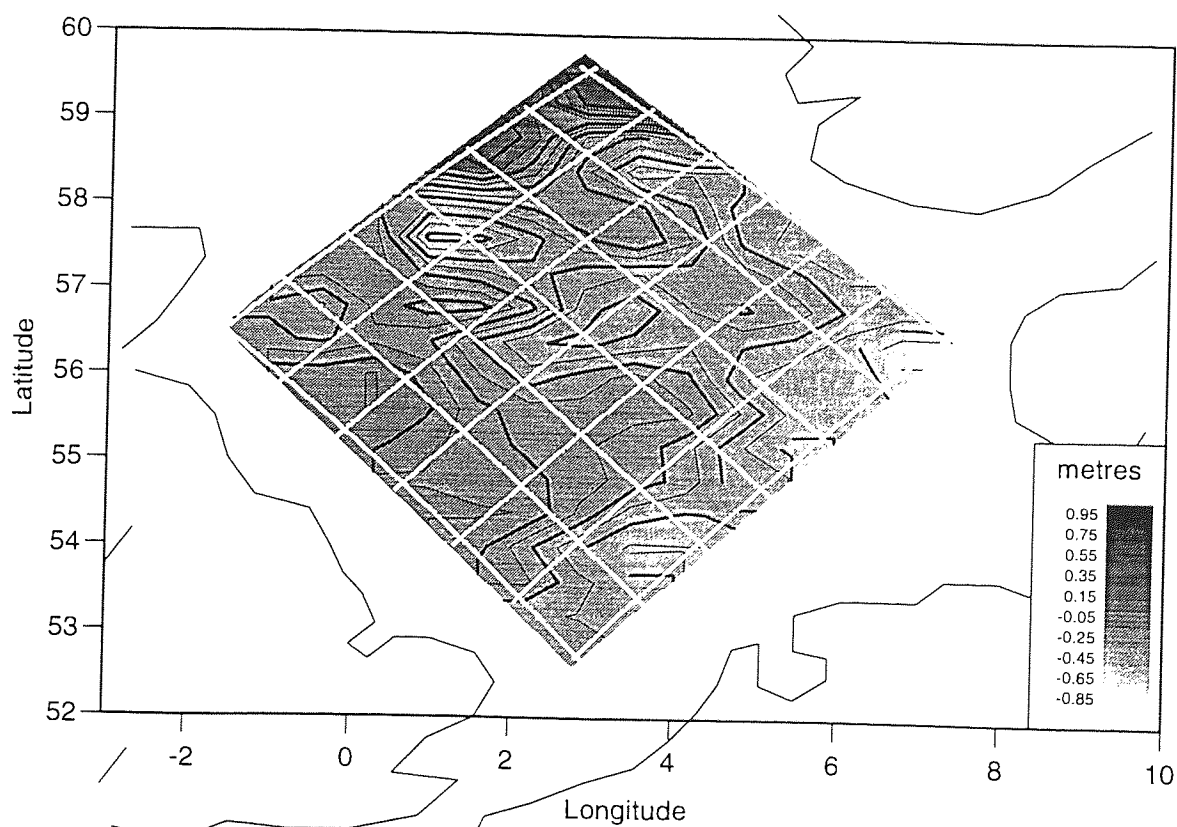
Figure 6.5. Residual base function MSS following subtraction of the Hipkin geoid.
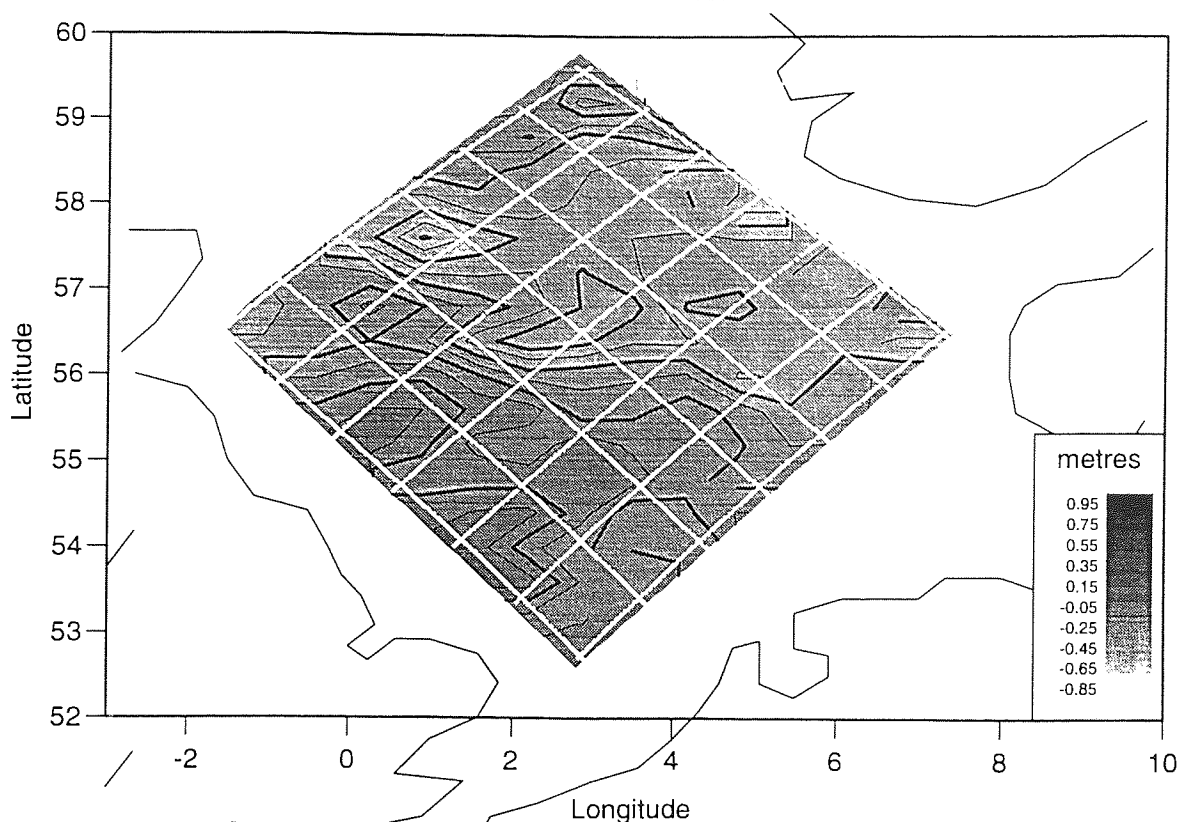
Figure 6.6. Residual optimally interpolated MSS following subtraction of the Hipkin geoid.

residual surface, however, clearly show that there were also significant differences. In both cases, a large bias is the most significant error. Therefore, it is evident that removal of long wavelength orbit errors has corrupted the surface by also removing some of the long wavelength geoid signal from the altimeter data, reducing the absolute accuracy of the results.

Fitting by least squares a bias and orthogonal tilts, the errors of figures 6.5 and 6.6 were reduced from an rms of 54.7cm (including the a priori bias) for the base function surface to 18.5cm. Similarly for the optimally interpolated surface the rms's were reduced from 42.0cm to 11.8cm. These corrected residuals are displayed in figures 6.7 and 6.8, where similarities in the pattern of residuals become distinguishable. In addition to some small scale features there is a significant latitude dependent difference at a wavelength of around 4°, present in both residual surfaces. This may be due to one of a number of factors, for example, errors in the geoid model due to gravimetric sampling or other causes, or as a result of quasi-static sea surface topography not included in the geoid model. Upon differencing the

base function and optimal interpolation surfaces, the consistency of these differences to the geoid may be investigated. Figure 6.9, shows the raw differences between the two solutions, showing a distinct bias and a significant tilt. Upon removing these effects, it becomes evident that the 4° latitudinal signal is common to both surfaces as it is not visible in the difference surface. It may be concluded, therefore, that it is a an artefact of the analysis procedure, most probably due to the limited extent of the dataset and insufficient low frequency sampling. The same cannot be said for short wavelength discrepancies as these show a distinct and systematic difference dependent upon the analysis procedure used.

Following removal of the bias and tilts from the base function/optimal interpolation residual surface, the rms differences were reduced from 26.8cm to 13.3cm: this is greater than the 11.8cm residual from comparison of the optimal interpolation and geoid model surfaces, indicating that the base functions used in this analysis do not best represent the altimeter data. Examination of figure 6.10 clearly shows a definite signal on a similar scale to the satellite ground track spacing. This error signal is more pronounced at higher latitudes, indicating a difference in the way in which data are treated in these areas. This would perhaps suggest that simply relating the correlation length of the base functions to their separation is not sufficient and it may be necessary to include some dependence on the density of data in the region. Poor choice of the parameter c_i in the base function equation 6.3 was shown by *Lelgemann* (1981) to result a surface containing 'spikes' where c_i was chosen to be too small. This typifies the limitation of the base function approach, as it requires careful choice of correlation length to accurately interpolate that data and where those data are not evenly distributed, as in the case of satellite altimetry, the choice of base function becomes even more difficult.

Interpolation of altimetry over a small region of ocean is, therefore, best performed by optimal interpolation, where the number of observations
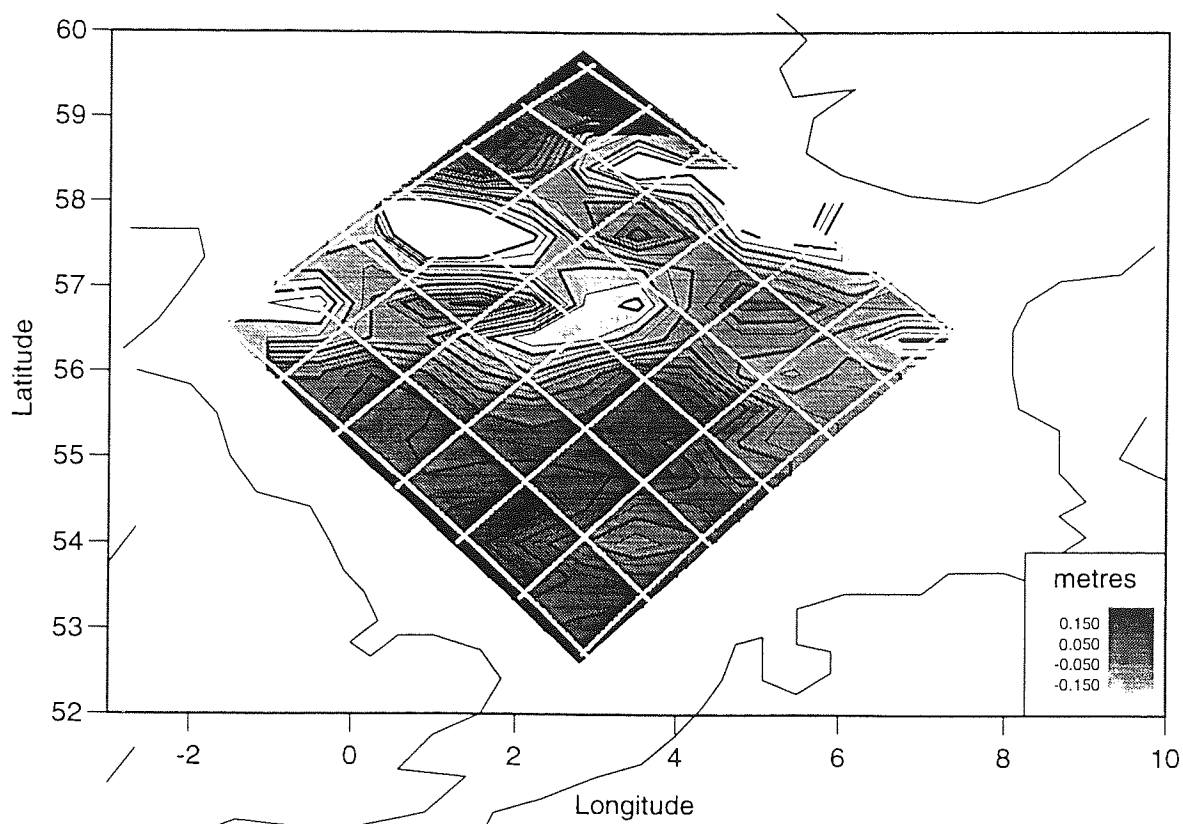
Figure 6.7. Base function MSS residual following subtraction of a plane approximation, fitted by least squares.
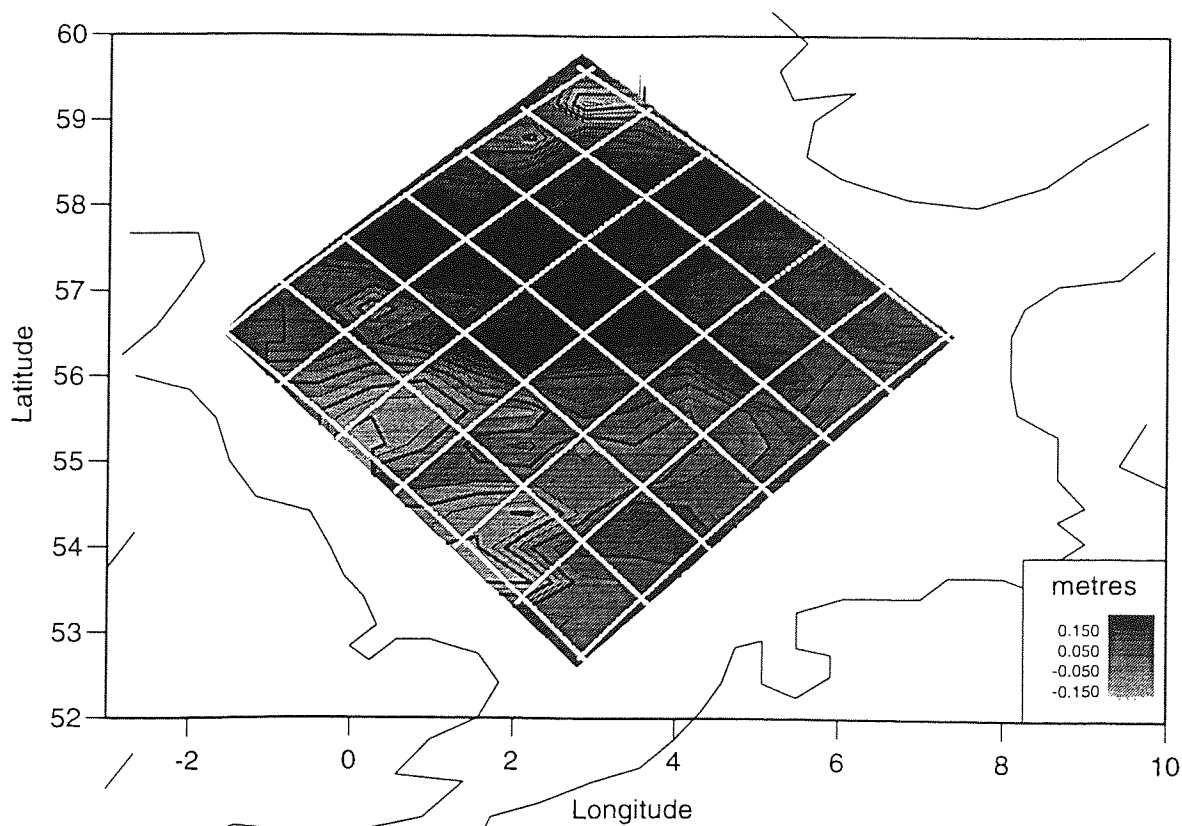
Figure 6.8. As in figure 6.7, from the optimally interpolated MSS residual.
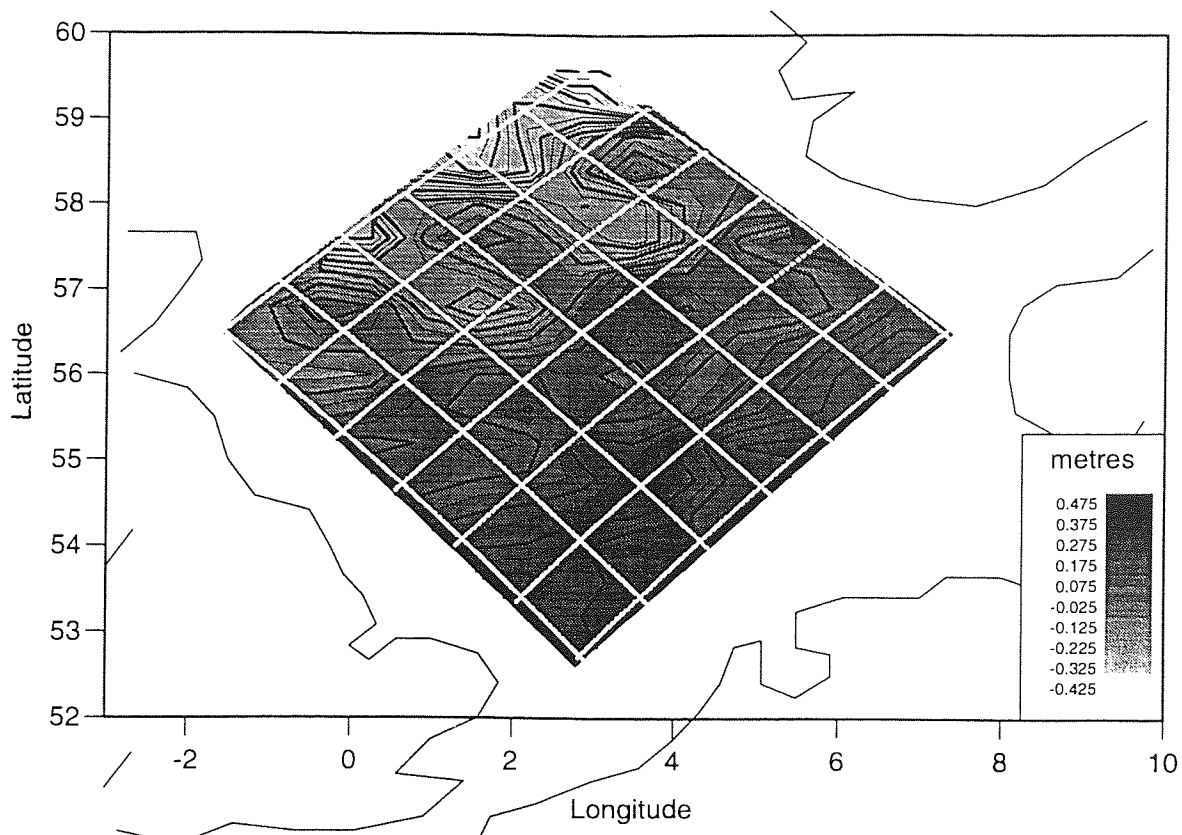
Figure 6.9. Residual MSS generated by subtraction of base function and optimally interpolated mean sea surfaces.
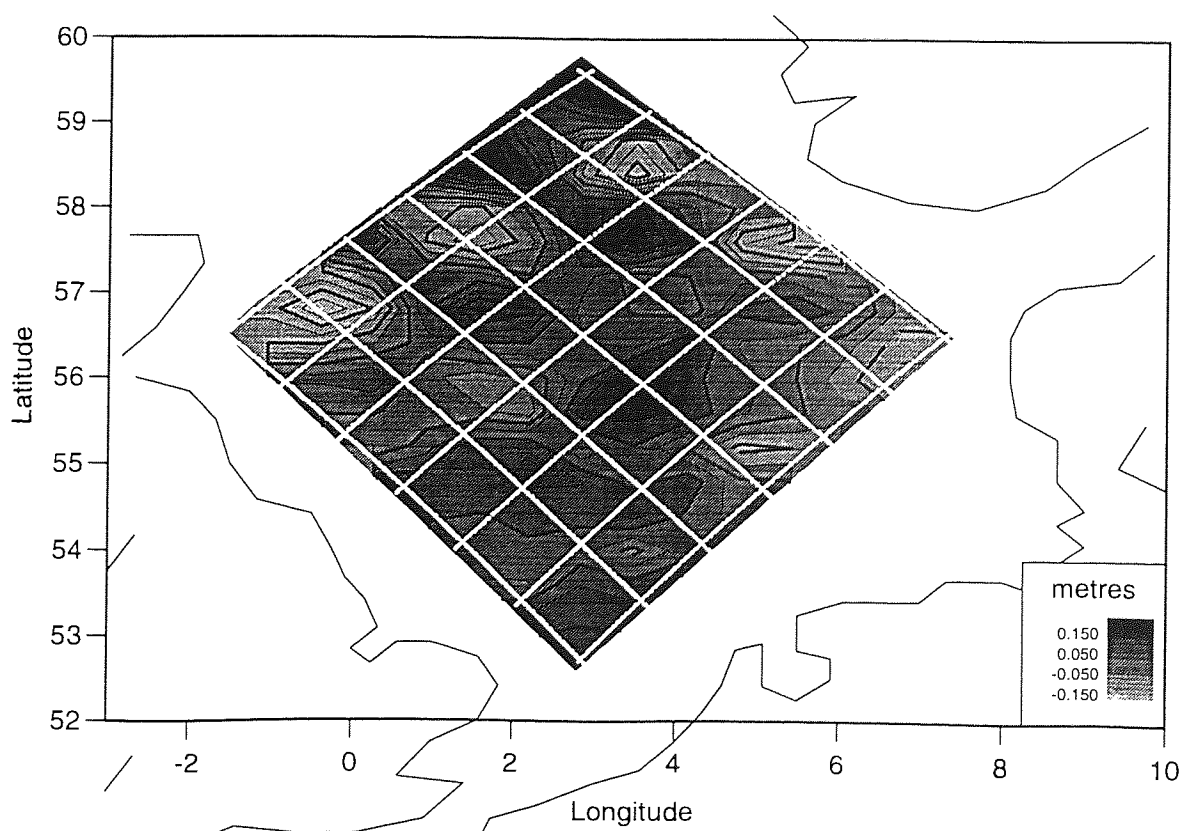
Figure 6.10. As for figure 6.9, following subtraction of a least squares fitted plane approximation.

required is of the order of a few thousand points. The accuracy, therefore, of the interpolated estimate of the ocean height is dependent only on the accuracy of the data and modelled corrections and covariances, and is not influenced by the choice of any crucial parameters, as is the case for the base function method. Where there are too many observations to consider using optimal interpolation, base functions do represent a worthwhile compromise. This can be seen in figure 6.10, below 56°N, where the distinct pattern of the northern half is not as clear. Therefore, careful choice of correlation parameters and base function locations could undoubtedly improve the accuracy of this surface, thus reducing the influence of latitude and data density.

6.4 Conclusion

Analysis of satellite tracking data is capable of accurately resolving the Earth's gravity field down to wavelengths which depend on the height of the satellite orbit. On a local level, ship borne gravity surveys can produce highly accurate and detailed maps at a resolution of as little as a kilometre. Between these two extremes lies analysis of satellite radar altimetry.

Geosat ERM altimetry is derived from tracks separated by 164km at the equator, this sets the lower limit on the resolution of geoid maps derived from this dataset. This chapter has demonstrated that the upper limit for high resolution geoid studies is determined by the computing power available. In the case of base function interpolation, a matrix of rank 1255 was inverted in a couple of hours to yield a surface which compared with the ground truth gravimetric geoid to around 18cm once a long wavelength correction had been applied. In the case of optimal interpolation, a 12 cm comparison was achieved after the rank 2000 matrix had taken over six hours to invert using 20% of the full dataset. These two methods each produced geoid maps over a comparatively small area of the North Sea from a raw altimeter dataset severely contaminated by orbit errors estimated

at approximately 60-80cm rms.   The consistency of these maps demonstrated that there was some erroneous signal introduced by the base function approach, perhaps due to the choice of a decorrelation function which did not fully represent the data. This hypothesis was supported by the fact that the optimally interpolated geoid had a smaller rms difference when compared to the ground truth gravimetric geoid than it did to the base function surface.

From these results, it may be concluded that the greater the accuracy of the covariance modelling used during interpolation of the sea surface from satellite altimetry, the more accurate the results should be.   This is in addition to the influence to the accuracy of the raw altimeter data   Optimal interpolation is therefore the better suited of these two approaches for accurate determination of the marine geoid from satellite altimetry. The following chapter will extend this method to the analysis of global datasets through implementation of optimal interpolation in the frequency domain.

Chapter 7

Global mean sea surface analyses

7.0 Introduction

The great advantage of satellite borne radar altimetry is its global coverage, to restrict analyses to purely regional surfaces therefore is to waste some of the instruments potential. This chapter describes two techniques for dealing with the immense volume of data generated by a satellite altimeter in the course of its mission. The first method to be described is based on simple and computationally straightforward weighted averaging of altimetry in order to generate global geoid maps of acceptable quality from any altimeter dataset. The second method seeks to apply optimal interpolation, as described in the previous chapter, in the frequency domain. This is based on methods already used for ground based gravimetry collocation where measurements can more easily be made on a regular grid. Results for both methods are based on the analysis of ERS-1 data from its 35-day repeat cycle, with some use also made of TOPEX/Poseidon data for the former of the two techniques; and in order to refine the orbital accuracy by dual crossover minimisation (section 4.4) for the latter.

7.1 Gaussian weighted average interpolation

Interpolation of a dataset may be performed in a variety of ways, depending on the required accuracy or resolution sought. The simplest methods are those devoid of any dependence on the physical characteristics of the signal being interpolated. The fastest and least accurate regime is nearest neighbour interpolation which involves the arbitrary selection of

the closest observation to the point of interest. This provides a fast and efficient means of producing a surface the accuracy of which is no better than that of any individual observation and may often be worse. An improvement on this is, for example, bilinear interpolation, whereby data bounded at say four corners by the grid points are fitted by a bias and orthogonal tilts to produce a square tile from which the mean height and slope of the tile can be determined. This improves the local accuracy of a surface as each grid point may now be generated from more than one observation. In this approach, all data are generally equal within the boundary but data outwith this area are ignored, leading to the possibility of inconsistencies at the boundaries.

Because both methods described in chapter 6 deal only with local data, they are efficient of both time and computer memory and can be used to generate a surface of any required resolution. A second set of interpolation regimes, based on the arithmetic mean of the data can also be used to generate surfaces. This was done for the first global altimetric mean sea surfaces, produced from Geos-3 by *Marsh et al.* (1981) and from SEASAT data by *Marsh and Martin* (1982). These spatial averaging methods require that the weighted mean be determined for each grid point using all the available data. In general, this is prohibitively slow but where the weighting function is localised this method can be used efficiently by considering data only within a given radius or cap size of individual grid points. With a weighting function derived from the spherical separation of observation and grid points, this method was investigated in order to provide 'quick look' surfaces for raw ERS-1 and TOPEX/Poseidon altimetry.

For a satellite in a circular or near circular orbit, the longitudinal separation (in degrees) of ascending or of descending ground tracks is constant; this does not imply that the true distances (in kilometres) are constant - only that for N equator crossings, at a latitude ϕ there are always N crossings of that circle around it's complete 360° circumference (where ϕ is

within the latitudinal extreme of the orbit). The density of altimeter observations per unit area therefore increases with latitude as the circumference of the circles of latitude diminish. A weighting function and grid were chosen to reflect this fact by reducing the mean separation of grid points with increasing latitude - this is the only way in which the spatial resolution differences can be incorporated into this approach.

Because of the differences in spatial resolution at different latitudes, ϕ, a locally homogeneous grid was chosen where grid spacing was latitude dependent. On a regional scale, the grid was approximately square, with sides of length l_ϕ,

$$l_\phi \propto l_0 \cos\phi$$

where l_0 is the chosen equatorial grid spacing. In order to ensure full Nyquist recovery of all cross-track signal, the equatorial grid spacing was chosen to be half the cross-track separation of the particular satellite - in the case of ERS-1 this corresponded to 1002 grid points along the equator with a separation of 0.359° or around 40km. At a latitude of 60° North or South, this was reduced to a grid size of 0.359° in longitude and 0.180° in latitude, or a rough square of side 20km.

As the grid is inhomogeneous, the weighting function should reflect this in order that the cumulative weights at each grid point be similar (assuming a complete altimetric dataset free of any gaps). For this reason, the weights were calculated from a Gaussian function with latitude dependent decorrelation,

$$G(\psi) = \exp-\left(\frac{\psi}{\psi_\phi}\right)^2$$

where ψ is the spherical separation of the grid point, i, and the altimeter observation, j. From spherical trigonometry this can be written,

$$\psi = \sin\phi_i \sin\phi_j + \cos\phi_i \cos\phi_j \cos(\lambda_i - \lambda_j) \tag{7.1}$$

The radius at which the Gaussian weighting function has a value of half its maximum is defined to be the decorrelation length. The decorrelation coefficient ψ_ϕ determines this length and may be calculated from

$$\psi_\phi = \frac{\psi_{\text{decor}}}{\sqrt{\ln 2}}$$

where ψ_{decor} denotes the desired decorrelation length. Using a cap-size $\psi_{decor} = 3l_\phi$, data at a spherical distance greater than ψ_{decor} were ignored (this corresponds to a negligible $G(\psi)$) and the height of each grid point was calculated from

$$h_{int} = \frac{\sum h_{obs} G(\psi)}{\sum G(\psi)} \tag{7.2}$$

To allow for the latitudinal variation of sampling density of the satellite altimeter, a cosine factor was introduced in determination of the decorrelation length. Refinement of the equatorial decorrelation length was based on an iterative process, best illustrated with cycle 2 of TOPEX/Poseidon data. Starting from a value for the decorrelation length equal to the mean separation of adjacent ground tracks, this value was progressively reduced until a balance was obtained between excessive smoothing across two or more adjacent tracks, and failure to generated a reliable MSS height estimate between tracks. Plotting the cumulative weights calculated for the TOPEX/Poseidon dataset (the denominator of equation 7.2), figure 7.1 illustrates the first case of excessive smoothing (ψ_{decor} =track spacing); figure 7.2 shows what was considered to be the 'best' case (ψ_{decor} =one quarter track spacing); figure 7.3 demonstrates the effect of too small a decorrelation length (ψ_{decor} =one sixteenth track spacing), resulting in clear minima in the cumulative weights which correspond to TOPEX/Poseidon track patterns. Figure 7.2 also clearly shows a disparity in the weights with latitude, this may be partially associated with the low inclination of TOPEX/Poseidon and hence the greater range of sampling densities within the region of study than was modelled.
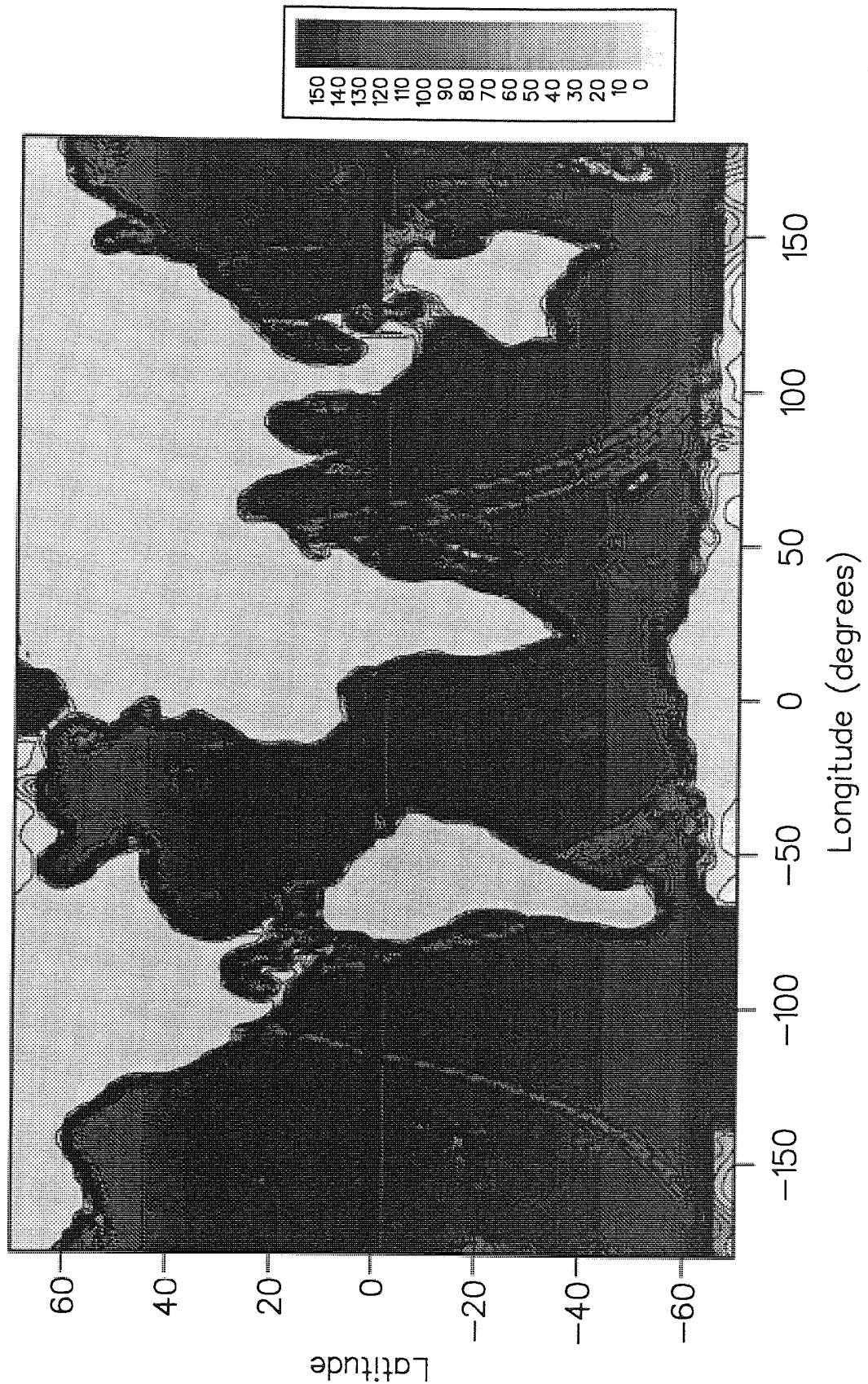
Figure 7.1. Cumulative weights associated with the Gaussian weighted averaging of TOPEX/Poseidon cycle 2 altimetry. In this figure, the decorrelation coefficient, ψ_{decor}, is equal to the cross-track spacing (~300km at the equator).
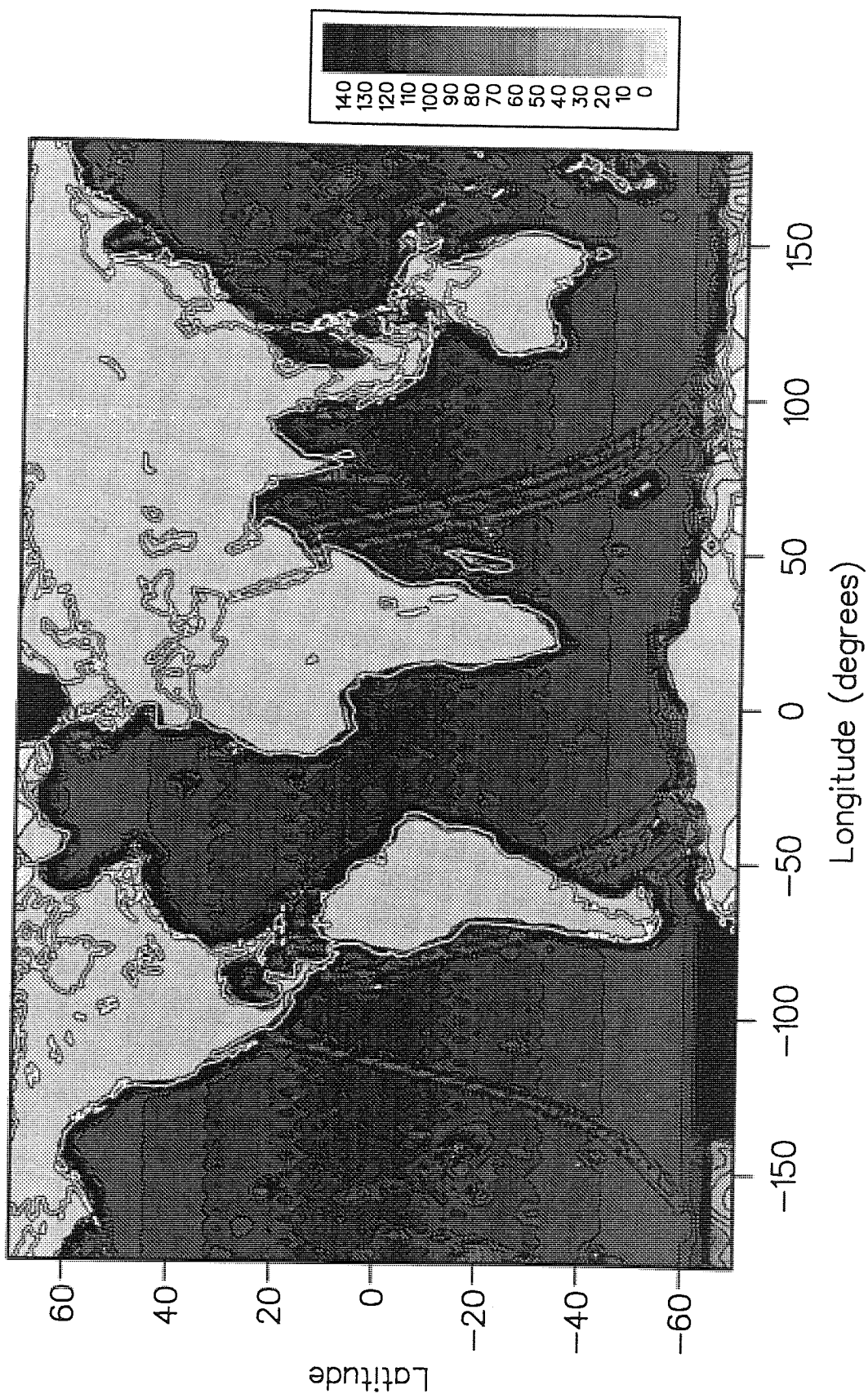
Figure 7.2. As for figure 7.1, but with ψ_{decor} equal to one quarter the inter track spacing.
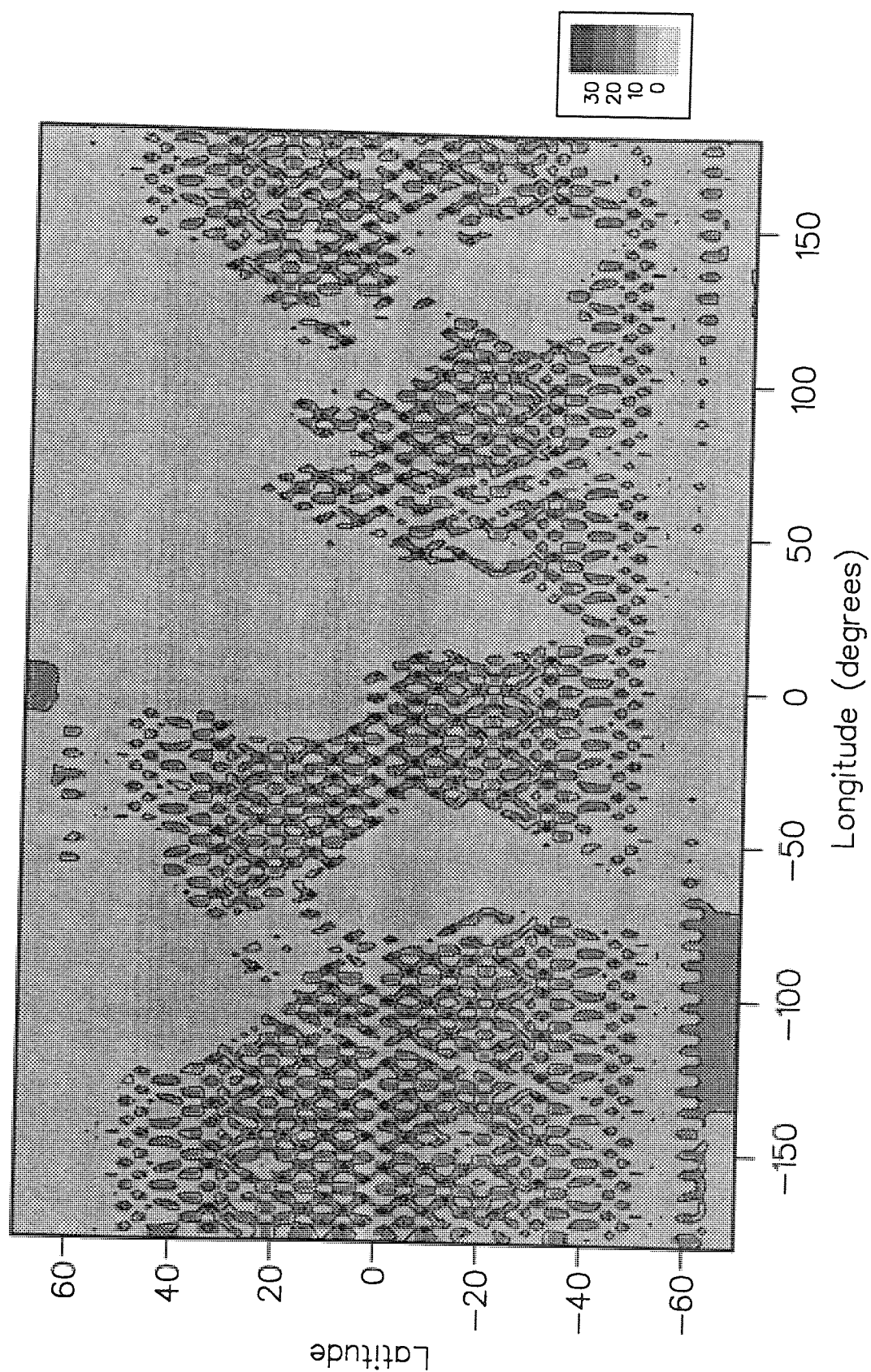
Figure 7.3. As for figure 7.1, but with ψ_{decor} equal to one sixteenth the inter track spacing.

For TOPEX/Poseidon the decorrelation coefficient was selected to have a value of one quarter the inter-track spacing or half the grid spacing, equivalent to approximately 79km at the equator.

Using data from cycles 2 and 3 of TOPEX/Poseidon, this interpolation regime was tested by producing 2° longitude by 2° latitude MSS maps between the latitudes of 66° North and South. TOPEX/Poseidon results for cycle 2 are shown in figure 7.4. Because of the 300km cross track separation, when the cumulative weights are plotted on a 2°x2° grid in figure 7.2, each individual missing track can be resolved - two descending in the Indian Ocean and another three over South America, along with one missing ascending track over the Pacific Ocean. A similar MSS map was also produced from cycle 3 of the TOPEX/Poseidon dataset; as it closely resembles figure 7.4, it will not be shown here. It will however be used at the end of this chapter for intercomparison of results.

7.2 Optimal Fourier transform interpolation

Over the past decade or so, many millions of altimeter measurements have been made of the sea surface, from various satellites in different orbits, with different accuracies and with different error characteristics. The only rigorous method of combining these data into a single global dataset is through optimal interpolation but because of the constraints of storing and inverting the auto-covariance matrix this is, and will probably always remain, wholly impractical. The strength of optimal interpolation lies in its use of the interdependence of observations to filter signal, errors and noise through accurate modelling of the covariances of the data. This allows for irregular sampling with different instruments, each of which may require different corrections, e.g. altimetry and tide gauge data.
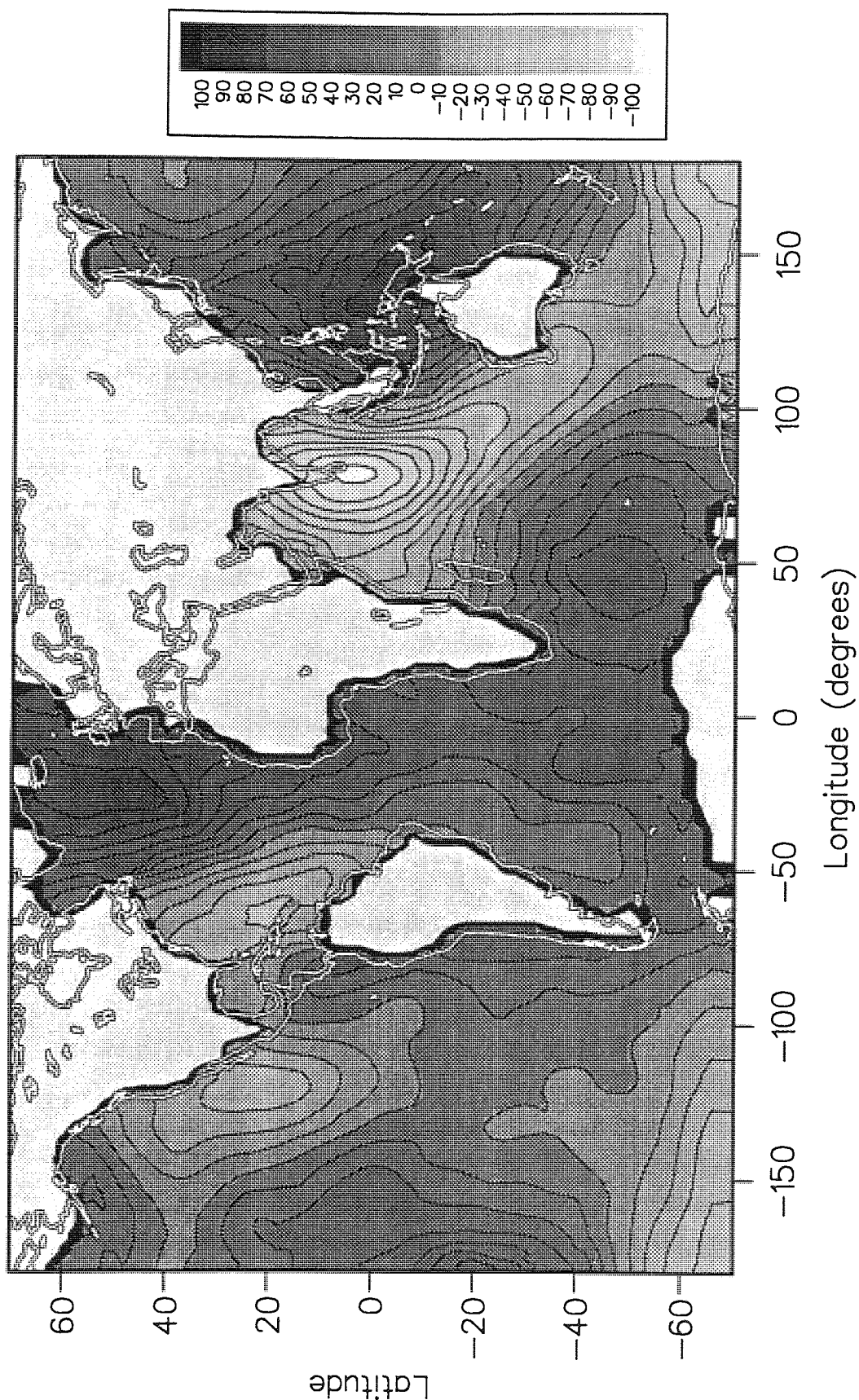
Figure 7.4. Global mean sea surface from cycle 2 of TOPEX/Poseidon data, interpolated by Gaussian weighted averaging onto a rectangular 2°x2° grid using 254 equatorial grid points.

Altimetric sampling of the ocean surface is, however, regular but not of uniform density - because of the shape of an orbital ground track, the number of observations per unit area is much greater at high latitudes than at the equator. The regular nature of a satellite orbit implies that optimal interpolation may be generalised to the point where the covariance may safely be described in terms of the relation of data to an arbitrary point. That is, instead of calculating unique covariances between every individual point, regularly spaced data are described by their relative position on an even grid. This is not the case for a general, complete altimeter dataset, but, for a subset containing only ascending or descending arcs it is possible to specify any location using only the equator crossing longitude of an arc and the time elapsed since the satellite traversed the equator on that arc. This results in a regular two dimensional grid of evenly spaced data. Generalisation of least squares collocation has been done previously with regularly spaced gravimetric and airborne gradiometry data through the use of Fourier transform theory to replace the covariances of optimal interpolation with Power Spectral Densities (PSD's), reviewed by *Schwarz et al.* (1990). The methods to be described in this section deal with optimal interpolation of satellite altimetry in the frequency domain.

7.2.1 Fourier transform theory

Any continuous function may be characterised with equal accuracy in either of two equivalent descriptions: conventionally by representation in space or time; or alternatively, by representation in frequency, be it spatial or temporal. These different representations form what is known as a Fourier transform pair, the operators of which are given in the case of a two dimensional function $h(x,y)$ by the integral

$$H(u,v) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} h(x,y) e^{-2\pi i(ux+vy)} dx\, dy = F\{h(x,y)\} \qquad (7.3)$$

where the spectrum $H(u,v)$ is the Fourier transform, $F\{.\}$ of the function at frequencies u and v, defined in terms of the respective wavenumbers k_x and k_y as

$$u = 2\pi k_x, \quad v = 2\pi k_y$$

and i is the imaginary unit ($i = \sqrt{-1}$). The equivalent inverse operation is defined in the same manner as,

$$h(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(u,v) e^{2\pi i(ux+vy)} du\ dv = F^{-1}\{H(x,y)\} \qquad (7.4)$$

The inverse Fourier transform operator $F^{-1}\{.\}$, therefore differs from equation 7.3 only in the variables over which the integration is evaluated and in the absence of a negative sign from the exponential. The function $h(x,y)$ and the spectrum $H(u,v)$, form a Fourier transform pair, represented by the symbol $\leftrightarrow$, i.e.

$$h(x,y) \leftrightarrow H(u,v) \quad \text{or} \quad H(u,v) \leftrightarrow h(x,y)$$

These definitions present a powerful tool in the analysis of data and functions through the many properties and identities associated with the transforms. A complete description may be found in many texts, for example, *Bracewell* (1978), or any good optics or electrical engineering text book. Of relevance to this thesis are the correlation functions of data and the result of their convolution. The correlation operator, $\otimes$, was defined by *Bendat and Piersol* (1980) as

$$R_{ab} = a(z) \otimes b(z) = \int_{-\infty}^{\infty} a(z_0) b(z + z_0) dz_0$$

for which the spectral equivalent is

$$a(z) \otimes b(z) \leftrightarrow A^*(w)B(w)$$

where $A(w) \leftrightarrow a(z)$ and $B(w) \leftrightarrow b(z)$, and A^* represents the complex conjugate of A. Convolution of two functions is defined, *Bendat and Piersol* (1980), by the operator, $\times$, as

$$a(z) \times b(z) = \int\limits_{-\infty}^{\infty} a(z_0)b(z - z_0)\mathrm{d}z_0 \tag{7.5}$$

which has the spectral counterpart

$$a(z) \times b(z) \leftrightarrow A(w)B(w)$$

• Discrete Fourier transform

In practice, most real data are finite in extent and defined only at discrete points, making evaluation of equations 7.3 and 7.4 impossible unless it is assumed that the data constitute an estimate of the continuous function, which is zero except at the observations points. Where these observation points exist on a regular $M \times N$ grid separated by Δx and Δy, however, equations 7.3 and 7.4 may be re-expressed as the summation of series, i.e.

$$H(m\Delta u, n\Delta v) = \sum_{k=0}^{M-1}\sum_{l=0}^{N-1} h(k\Delta x, l\Delta y) e^{-2\pi i\left(\frac{mk}{M} + \frac{nl}{N}\right)}, \quad m = 0, 1, \dots, M-1, \quad n = 0, 1, \dots, N-1$$

$$h(k\Delta x, l\Delta y) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} H(m\Delta u, n\Delta v) e^{2\pi i\left(\frac{mk}{M} + \frac{nl}{N}\right)}, \quad k = 0, 1, \dots, M-1, \quad l = 0, 1, \dots, N-1$$

thus generating a spectrum evaluated at regular intervals of frequency, Δu and Δv. By convention, this discrete Fourier transform pair (DFT) are represented by the integer indices and wavenumbers only, i.e.

$$h(k, l) \leftrightarrow H(m, n).$$

Sampling a continuous function puts limits on the upper and lower extent of the frequencies, v, that may be resolved in the spectrum. For M samples at intervals of Δx, these limits are given by

$$v_{\min} = \Delta v = \frac{1}{M\Delta x}$$

$$v_{\max} = \pm\frac{M}{2}\Delta v = \pm\frac{1}{2\Delta x}$$

From this it follows that the spectrum is defined at M equal intervals of frequency, separated by Δv.

In the case of a large dataset, evaluation of the DFT is a slow and laborious summation requiring a great deal of computer time. The Fast Fourier Transform (FFT), *Cooley and Tukey* (1965), however, permits evaluation of the transform in a considerably shorter time if data are located on a regular grid, reducing the complexity for a two dimensional transform from an $(MN)^2$ dependence for the DFT to $MN \log(MN)$. The FFT algorithm is most efficient when M and N can be reduced to the products of the prime factor 2. Part of the time saving is achieved through the imposition of certain assumptions about the data: principally that over the interval, the signal is harmonic with a period equal to the total sample length, i.e. that the Mth observation of a series from *0* to *M-1* would be identical to the zeroth observation. It is therefore necessary to window the data appropriately in order to reduce contamination from boundary effects. The FFT is most efficient when M and N are composed only of the products of the prime factor, 2.

7.2.2 Optimal interpolation in the frequency domain

Section 5.2 described the derivation of the optimal estimator equation 5.6 in terms of the cross-covariance and auto-covariance matrices (C_{sl} and C_{ll} respectively),

$$\hat{s} = C_{sl} C_{ll}^{-1} l \qquad (7.6)$$

where $\hat{s}$ is the estimated signal as derived from the vector of observations l. The estimator, a_{sl} may be described as,

$$a_{sl} = C_{sl} C_{ll}^{-1}$$

or, rearranging

$$C_{sl} = a_{sl} C_{ll}$$

For data on a regular two dimensional grid, this can be rewritten as a convolution integral,

$$C_{sl}(x_s - x_l, y_s - y_l) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_{sl}(x_s - x, y_s - y)C_{ll}(x - x_l, y - y_l)\, dx\, dy \qquad (7.7)$$

In one dimension an expression of the form of equation 7.7 may be rewritten,

$$\int a(z_s - z)b(z - z_l)dz = \int a(k_o)b(k - k_o)dk_o \qquad (7.8)$$

where the following substitutions have been made,

$$k_o = z_s - z$$
$$k = z_s - z_l$$

The right hand side of equation 7.8 is equivalent to the definition of the convolution operator, equation 7.5. Thus, equation 7.8 can be written,

$$\int a(k_o)b(k - k_o)dk_o = a(k) \times b(k) \leftrightarrow A(w)B(w)$$

By analogy with the one dimensional case, equation 7.7 can also be written as a convolution integral and hence in the frequency domain as

$$P_{sl}(u,v) = A_{sl}(u,v)P_{ll}(u,v) \qquad (7.9)$$

where P_{sl} and P_{ll} are the Fourier transforms of the cross-correlation and auto-correlation functions respectively, and A_{sl} is the Fourier transform of the least squares estimator a_{sl}. Rearranging equation 7.9, the estimator A_{sl} may be written as

$$A_{sl} = \frac{P_{sl}}{P_{ll}} \qquad (7.10)$$

and hence, the vector of height estimates, $\mathbf{h}_s$, from the altimeter dataset $\mathbf{h}_l$ is

$$\mathbf{h}_s = F^{-1}\{A_{sl}\}\mathbf{h}_l$$

Alternatively, a more efficient means of evaluating $\mathbf{h}_s$ utilises the property (equation 7.11) that convolution and multiplication form a Fourier transform pair of operations, i.e.

$$H_s = A_{sl}H_l \leftrightarrow \mathbf{h}_s = F^{-1}\{A_{sl}H_l\} \qquad (7.11)$$

7.2.3 Power spectral density evaluation

From equation 7.11, optimal interpolation in the frequency domain is dependent entirely on the production of Fourier transforms of the auto and cross-correlation functions, as well as of the altimeter dataset. The Fourier transform (F{.}) of a correlation function is also known as the Power Spectral Density (PSD), P, which from *Schwarz et al.* (1991) is related to the correlation function (R_{ab}) and the covariance function (C_{ab}) by

$$P_{ab} = F\{R_{ab}\} = F\{C_{ab} - \mu_a \mu_b\} \qquad (7.12)$$

where μ denotes the mean value of the signal. For μ equal to zero for one or both of the signals, the correlation function is equivalent to the covariance; in determining the mean sea surface over a large area, this is a justifiable assumption - hence, to ensure this, the mean value for each surface was subtracted before the PSD was derived from the Fourier transform of the covariance function.

Equation 7.12 states that the Power Spectral Density (PSD) may be determined from the Fourier transform of the correlation function of a dataset; from the definition, that PSD will be real valued and positive. Using continuous, integral, Fourier transform theory does yield such a result, but because of anisotropic sampling over a finite area, the use of discrete Fourier transform methods such as the FFT can lead to problems. In general, for a correlation function, the result is real, with a virtually negligible imaginary component due to round off errors in the process. Unfortunately, the requirement of producing a positive result is more problematic as incomplete sampling of the correlation function can lead to negative energies at some frequencies. This can be found especially at sharp boundaries in a spectrum due to the effect of ringing. In this application the consequences of even small negative energies at only a few out of several thousand frequencies can be disastrous because of the way in which the final weighting function (A_{sl}) is formed. In the ideal case of error free data, a_{sl} is equal to the Dirac delta function, i.e. giving exact duplication of

the original dataset. A delta function is formed by the combination of equal amounts of all frequencies present in the spectrum, this follows from the fact that the Fourier transform of the delta function is a function of constant amplitude. As each frequency may be thought of as representing a sine wave in the spatial domain, the delta function is formed through the superposition of all sine waves in the spectrum, leading to constructive interference at a single point and exact cancellation at all other points. If some of the values in the spectrum are made negative, however, then this will have the effect of maintaining their magnitude in the spatial domain but shifting their phase by 180°, i.e. swapping crest and trough of the sine wave and leading to incomplete destructive interference around the focus of the delta function. Because of these effects, all attempts at using the same covariance function as were used in section 6.2 proved unsuccessful as phase errors in the determination of both P_{sl} and P_{ll} proved extremely difficult to eradicate completely. Determination of these power spectra was therefore performed via alternative means.

It is well recognised that the power spectral density can be obtained by two complementary methods: the first using the correlation functions of the data proved in this instance, to be too problematic; the alternative is to use the actual datasets to derive a power spectrum and it is this approach which yielded real valued positive spectra for both the auto and cross PSD's.

The auto-PSD of a dataset is a measure of the total energy at a given frequency present in that data and from *Bendat and Piersol* (1980), the auto-PSD may be defined as

$$P_{ll} = | F\{l\}^{*} F\{l\} | = | F\{l\} |^{2} \qquad (7.13)$$

where * denotes the complex conjugate of the spectrum. Similarly, the cross-correlation function describes the dependence of one dataset on another and as such, the cross-PSD is defined as

$$P_{sl} = | F\{s\}^{*} F\{l\} | \qquad (7.14)$$

The cross-PSD can therefore be determined from the altimeter dataset and the estimated, filtered dataset. In the case under investigation, s is the unknown being sought and as such, an alternative must be found that will produce a good approximation to the energies of the correlation function at each frequency. To achieve this, s is replace with g, a vector containing the heights of an accurate high resolution geoid model at points coincident with the altimetric data. The geoid model used was a combination of the JGM2 geoid up to degree and order 70, the OSU91 geoid from degree and order 71 to 360 and a 20x20 sea surface topography (SST) model derived from analysis of ERS-1 altimetry for the dates 48481-48532 (3-day repeat data) and 48848-48917 (35-day repeat data); respectively, *Marsh et al.* (1993), *Rapp et al.* (1991) and *Moore and Ehlers* (1994). Finally, for both auto and cross-PSD evaluation, the window functions must be identical. Because altimetry is distributed within a belt centred around the equator, the dataset is harmonic in longitude, but discontinuous in latitude, where the FFT assumption fails. Therefore it is necessary to apply a latitude dependent function such as a Hanning window in ϕ, defined by

$$f_H(\phi) = 0.5 \left[1.0 - \cos\left(2\pi \frac{\phi - \phi_{min}}{\phi_{max} - \phi_{min}} \right) \right]$$

By multiplying the data by a window function, there is an effective convolution of data and window to produce a smoother spectrum; choice of window may also be made on the grounds that convolution of this type will reduce noise in the ultimate spectrum. The auto-PSD of ascending ERS-1 data is shown in figure 7.5. Signal is clearly evident in the centre of the figure, extending out as far as the cross track sampling permits, visible as a blue ring around the central peak. Beyond this radius, noise dominates and in the final solution, these frequencies are arbitrarily set to zero. The axes in figure 7.5 are labelled in arbitrary units of wavenumber. In the x-direction, the unit corresponds to (501 x track spacing)[-1]; in the y-direction, the unit is (550 seconds)[-1]
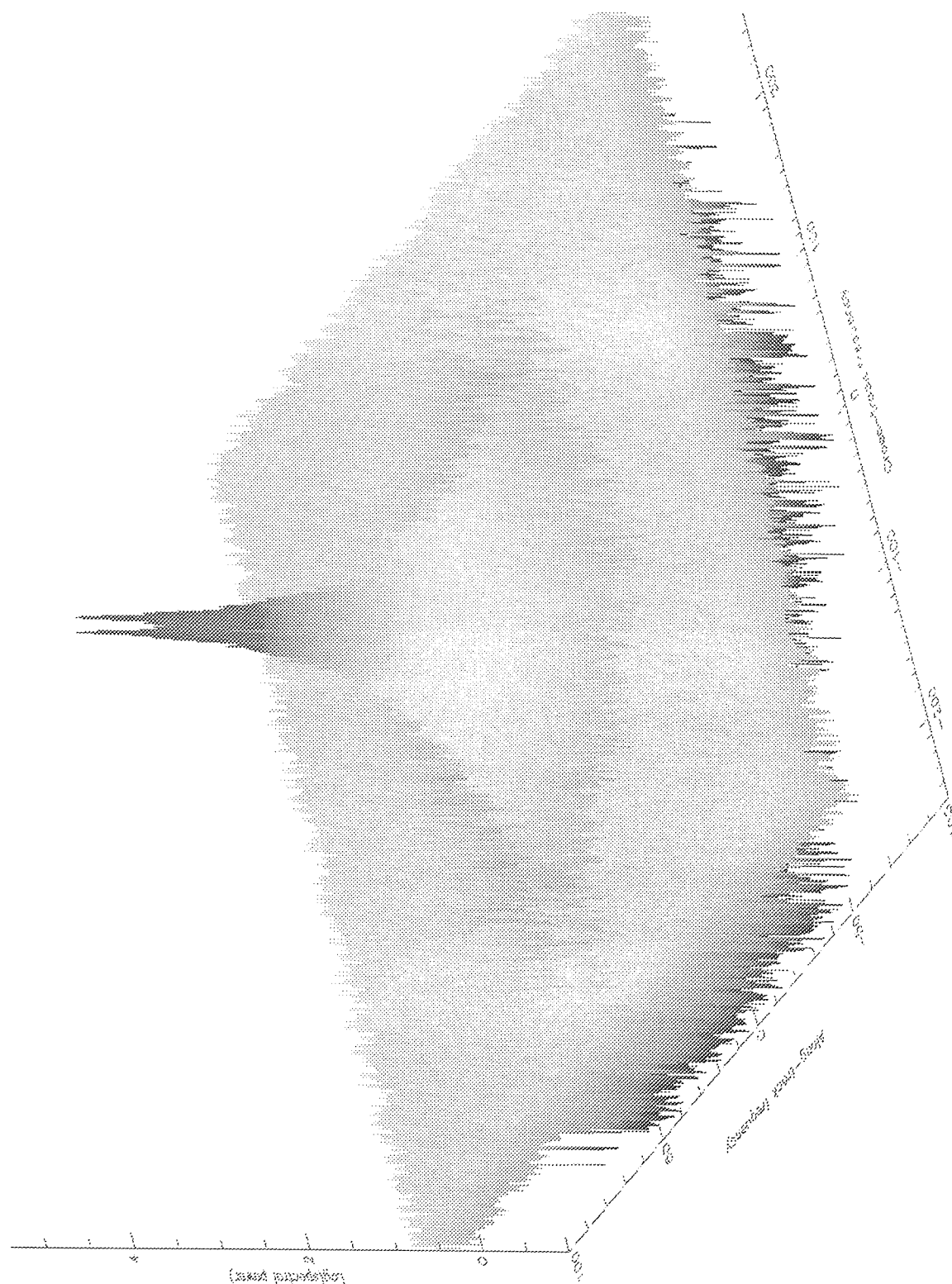
7.5.　　The auto power spectral density of ascending ERS-1 altimetry from cycle 9, labelled in arbitrary units of wavenumber. In the x-direction, the unit corresponds to (501 x track spacing)$^{-1}$; in the y-direction, the unit is (550 seconds)$^{-1}$

Final evaluation of the estimated height was then obtained from

$$h_s = F^{-1}\left\{\frac{P_{gl}}{P_{ll}} H_l\right\}$$
(7.15)

Although equation 7.15 is true for any function, whether it be discrete or continuous, in practical terms for a large dataset the fast Fourier transform (FFT) algorithm must be used to derive all of the spectra and this requires that discrete data be evenly distributed on a regular grid. Altimeter data are not, however, evenly distributed when projected onto any map of the globe. To obtain a regular grid, the dataset must be split into ascending and descending sub-sets; in this way, any observation can be described in terms of the equator crossing longitude of the short arc on which it sits, and by the time since that equator crossing or some other fixed reference on each arc. That is, for an observation i, it may be assigned two orthogonal indices x and y, as illustrated in figure 7.6, for an ascending grid. Similarly for a descending grid, there is an exclusive set of co-ordinates to describe any point on that grid. In this way, using equation 7.15, two intermediate MSS's can be formed, the combination of which is dealt with in the next sub-section.

The second requirement of the FFT method is that the dataset not only be on a regular grid, but that the grid be complete - i.e. with no datagaps. This is not a problem for the model geoid surface, but for altimetric data, gaps exist due to the presence of land as well as due to any problems with the altimeter where no data are available. For example, in the ascending dataset there is a consistent gap in the North Atlantic where data stored by on-board magnetic tapes are downloaded to the ground station, during which time the altimeter cannot be used. This is clearly visible as a vacant swathe in figure 7.6, showing the available altimetry from ERS-1 on separate, regular ascending and descending arcs for cycle 9. Note the slight distortion of the continental shapes, particularly at the southern tip of South America. Following methods described by *Kim et al.* (1994), to
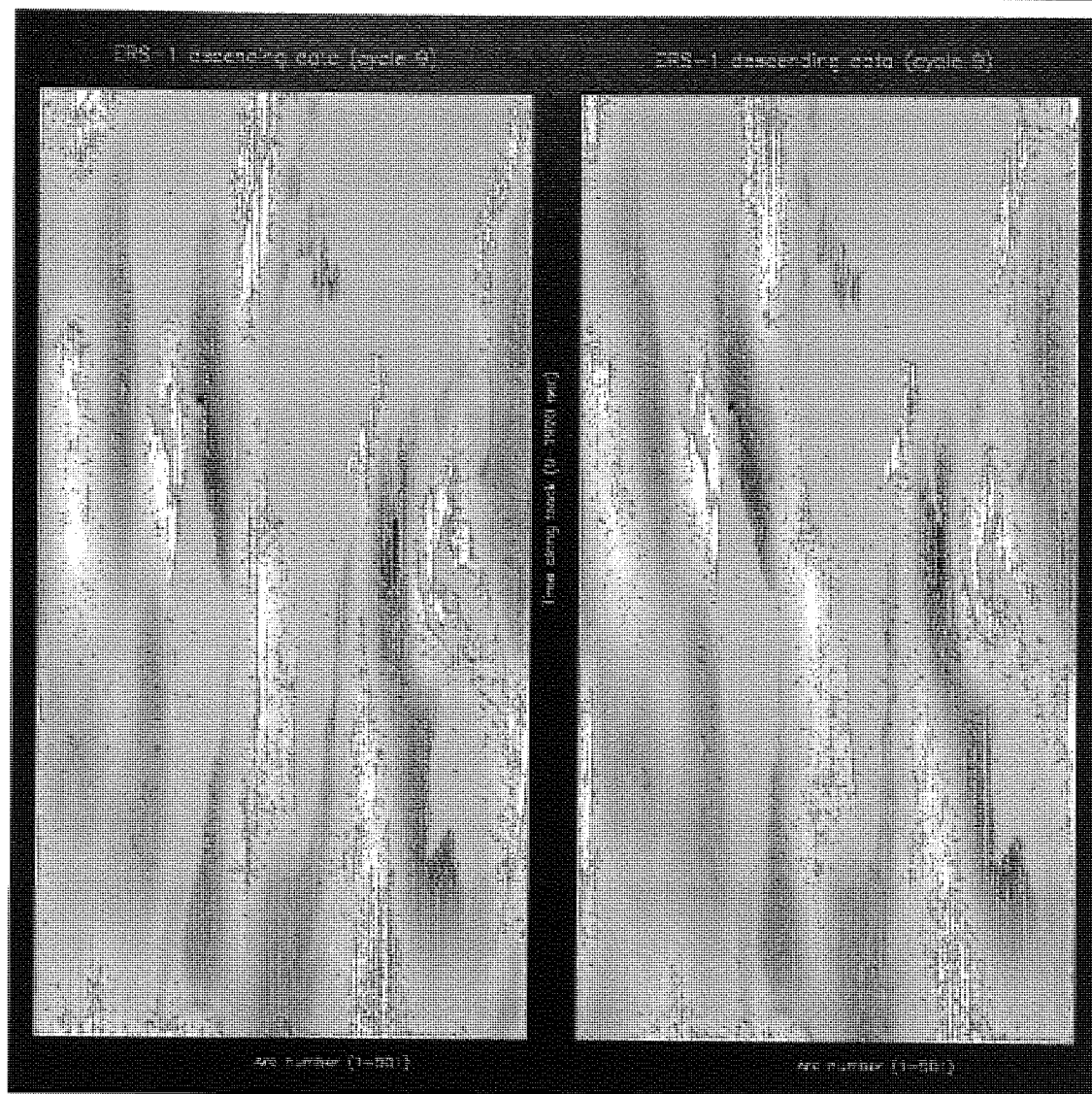
Figure 7.6. Ascending and descending ERS-1 altimetry grids before substitution of modelled surface heights to replace absent data.

compensate for the data gaps, heights from the geoid model were substituted into the data gaps in order that a smooth, well behaved surface was available for determination of the auto and cross-PSD functions. Within these surfaces, there are two problems which relate to the determination of the PSD's which arise due to the extent of data coverage combined with long wavelength and bias differences between geoid and altimeter datasets. The first of these arise as a consequence of the limited latitudinal extent of altimeter coverage, thus restricting the spectral resolution of the data at low frequencies. To reduce this problem, the surface model complete to degree and order 4 was removed from both

datasets, this consisted of the JGM2 gravity field to 4x4 and the SST model to 4x4. This reduced the rms signal being sought from around 31m to just under 12m, in the process, removing some of the potential ambiguities from the results. As mentioned in chapter 3, most altimetry contains a number of bias and long wavelength uncertainties which corrupt the datasets. To counteract this problem, a bias term was fitted to the differences between the altimetry and the sea surface model where there was sufficient altimetry to derive such a figure, producing a value of 62cm for ERS-1 (it should be noted that this figure is not exclusively attributable to altimeter electronics biases but undoubtedly aliases other long wavelength signatures). In these ways, when altimetric datagaps were filled using the geoid model, the dangers associated with sharp boundaries were reduced. Between the geoid model and each of the altimeter surfaces, the rms differences were around 26cm, which includes shallow water data of inherently poorer quality due to tidal modelling errors and the failure of certain ocean/atmosphere assumptions in shallow or enclosed seas. From this point on, all surfaces relate to signals above degree and order 4 unless otherwise stated.

With a geoid model complete between degree and order 4 and 360, and a complete altimetric dataset, evaluation of the PSD's may be undertaken in accordance with equations 7.13 and 7.14 using a latitudinal Hanning window.

7.2.4 Interpolation to a standard grid

Following derivation of the PSD's, equation 7.11 may be applied using a FFT to filter the raw data and produce a refined dataset of the heights at each of the original grid locations. This still results in separate ascending and descending surfaces which are of little practical use on their own. The next step in the process must therefore be to interpolate these data onto a common grid, consistent with that to be used for all other data. If, as has

just been mentioned, a FFT algorithm is used in equation 7.11, many of the advantages of this optimal method are lost as the data must be treated as the raw input to the first iteration of any interpolation regime. This may be efficient of time in particular, but will not produce the best possible results. Greater consistency may be achieved by using a slower discrete Fourier transform (DFT) method as follows,

$$h_s = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} A_{sl}H_l e^{2\pi i\left(\frac{mx^*}{M}+\frac{ny^*}{N}\right)} \tag{7.16}$$

where x^* and y^* are not integers as in the case of FFT but represent the real number co-ordinate of the required interpolation point, located somewhere between the grid points of the regularly spaced data. The above summation is considerably slower than FFT generation of a surface, but does allow for the surface to be transferred from the altimeter grid, unique to each ascending or descending phase of an orbit, onto a common grid, regular in latitude and longitude which may then be used for all altimeter datasets and further, in their subsequent combination.

The exact values of x^* and y^* are found through the following process, the first stage of which is to establish the shape of a standard ascending arc: in determining the PSD's, the relative positions in time, latitude and longitude of each altimeter observation relative to the equator crossing is noted; from this, an estimate can be made of the latitude and the relative longitude of each point on the ascending grid. This information is used with a cubic spline interpolation routine to reverse the process and produce firstly for any arbitrary latitude within the extremes of the data, an estimate of the time since the equator crossing on this standard orbit. By the same approach, once the latitude is known, an equator crossing longitude relative to the given location may be interpolated, after which it is a simple process to identify the nearest equator crossing East or West of this point. The precise fractional values of x^* and y^* are thus available for any point and may be used in equation 7.16. This must be performed for each point

on the new standard grid and as such, the total time for this procedure is directly proportional to the number of grid points. For example, a to interpolate a 501x192 ERS-1 ascending or descending grid onto a global 1°x1° grid between 55° N/S, requires around 10 hours of processor time on a Sun Sparc 10; if this were reduced to a 0.5°x0.5° grid the interpolation would take approximately 40 hours. This obviously compares very poorly with a single inverse FFT which is completed in under five minutes and is, therefore, a practical bottleneck in any MSS determination.

A suitable compromise can be found by exploiting one of the features of the Fourier transform, namely, the frequency limit imposed by the discrete dataset. For a one dimensional function sampled N times at intervals of Δ, the unit frequency of its FFT would by $\frac{2}{N\Delta}$ cycles per unit length, thus any signal at a frequency greater than $\frac{1}{2\Delta}$ cannot be resolved. Conversely, for a given spectrum of maximum frequency u, no useful information is available on a wavelength less that $\frac{1}{u}$. To interpolate to say M evenly spaced points between each sample point in the first example, zeros may arbitrarily be added at frequencies greater than that determined by the sampling interval - up to M times that value. As the values added are zero, they do not contribute any energy to the spectrum and hence contribute nothing to the interpolated signal. However, inverse FFT of the new augmented spectrum produces a signal with MN points over the same interval where only N were originally sampled at a spacing, therefore, of $\frac{\Delta}{M}$. The principal applies in two dimensions and for a modest value of M, a high resolution MSS can be produced from which bilinear interpolation is sufficient to refine the precise heights to locations x^* and y^*. Using a value of $M=4$, the augmented inverse FFT allows an interpolated surface to be generated in approximated 5 minutes of CPU time to accuracies of better than 1cm when compare with the full DFT mean sea surface.

7.2.5 Optimal data collocation

From *Moritz* (1980), the basis of optimal interpolation is the least squares estimator of equation 7.6/5.6,

$$\hat{s} = C_{sl}C_{ll}^{-1}1 \tag{7.17}$$

which describes the optimal estimation of a signal s from the dataset l. This may also be performed for multiple datasets, either simultaneously by increasing the size of C_{sl}, C_{ll} and l; or by stepwise evaluation of s. For the example of collocation of two datasets, the above vectors and matrices may be partitioned thus:

$$1 = \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} \qquad C_{sl} = \begin{pmatrix} C_1 & C_2 \end{pmatrix} \qquad C_{ll} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \tag{7.18}$$

where the inverse of C_{ll} is found from the matrix

$$C_{ll}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{7.19}$$

Combining equations 7.17, 7.18 and 7.19, the optimal, collocated estimate, $\hat{t}$, of the signal, s, may be written

$$\hat{t} = \begin{pmatrix} C_1 & C_2 \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} \tag{7.20}$$

The individual sub-matrices in equation 7.19 are as follows,

$$\overline{B}_{22} = \left(C_{22} - C_{21}C_{11}^{-1}C_{12} \right)^{-1}$$
$$B_{12} = -C_{11}^{-1}C_{12}\overline{B}_{22}$$
$$B_{21} = -\overline{B}_{22}C_{21}C_{11}^{-1}$$
$$B_{11} = C_{11}^{-1} - C_{11}^{-1}C_{12}B_{21}$$
$$\phantom{B_{11}} = C_{11}^{-1} + C_{11}^{-1}C_{12}\overline{B}_{22}C_{21}C_{11}^{-1}$$

Substituting these into equation 7.20, the signal estimate becomes

$$\hat{t} = \left(C_1 B_{11}^{-1} - C_2 \overline{B}_{22} B_{21} B_{11}^{-1} + C_1 B_{11}^{-1} B_{12} \overline{B}_{22} B_{21} B_{11}^{-1} \right) l_1 + \left(C_2 \overline{B}_{22} - C_1 B_{11}^{-1} B_{12} \overline{B}_{22} \right) l_2$$

This may be simplified through definition of the operators

$$f_1 = C_1 B_{11}^{-1} - C_2 \overline{B}_{22} B_{21} B_{11}^{-1} + C_1 B_{11}^{-1} B_{12} \overline{B}_{22} B_{21} B_{11}^{-1}$$
$$f_2 = C_2 \overline{B}_{22} - C_1 B_{11}^{-1} B_{12} \overline{B}_{22}$$

from which, the estimated signal can be written

$$\hat{t} = f_1 l_1 + f_2 l_2 \tag{7.21}$$

Once again, as the right hand side of equation 7.21 approximates an integral operator, from which it may be shown that the products are the result of discrete convolutions. Therefore, in the frequency domain the equivalent expression of equation 7.21 is

$$\hat{T} = F_1 H_1 + F_2 H_2$$

similarly, the operators F_1 and F_2 are

$$F_2 = P_{2g} \overline{P}_{22_1}^{-1} - P_{1g} P_{11}^{-1} P_{12} \overline{P}_{22}^{-1}$$
$$= \frac{P_{2g}}{\overline{P}_{22}} - \frac{P_{1g} P_{12}}{P_{11} \overline{P}_{22}}$$
$$= \frac{P_{2g} P_{11} - P_{1g} P_{12}}{P_{11} \overline{P}_{22}}$$
$$= \frac{P_{2g} P_{11} - P_{1g} P_{12}}{P_{11} P_{22} - |P_{12}|^2}$$

where by analogy with $\overline{B}_{22}$,

$$\overline{P}_{22} = \frac{P_{11} P_{22} - P_{21} P_{12}}{P_{11}}$$

Similarly,

$$F_1 = P_{1g} P_{11}^{-1} - P_{2g} \overline{P}_{22}^{-1} P_{21} P_{11}^{-1} + P_{1g} P_{11}^{-1} P_{12} \overline{P}_{22}^{-1} P_{21} P_{11}^{-1}$$
$$= \frac{P_{1g}}{P_{11}} - \frac{P_{2g} P_{21}}{\overline{P}_{22} P_{11}} + \frac{P_{1g} P_{12} P_{21}}{P_{11} \overline{P}_{22} P_{11}}$$
$$= \frac{P_{1g} P_{22} - P_{2g} P_{21}}{P_{11} P_{22} - |P_{12}|^2}$$

In general terms therefore, for two input signals and a single output, the frequency domain operator at a particular frequency (u,v) can be written

$$F_j(u,v) = \frac{P_{jg}(u,v)P_{jj}(u,v) - P_{ig}(u,v)P_{ij}(u,v)}{P_{ii}(u,v)P_{jj}(u,v) - |P_{ij}(u,v)|^2}, \quad \begin{array}{l} j=1, \ i=2 \\ j=2, \ i=1 \end{array} \quad (7.22)$$

Underlying this expression are several assumption, namely that the input signal and noise are uncorrelated, as are the input and output noise. Additionally, equation 7.22 is undefined where the correlation between datasets 1 and 2 approaches unity for a particular frequency, i.e. when

$$P_{ii}P_{jj} \rightarrow |P_{ij}|^2$$

This implies a linear correlation between datasets 1 and 2,

$$H_2(u,v) = D(u,v)H_1(u,v)$$

and hence,

$$P_{11} \cong P_{22} \cong P_{12} \cong P_{21}$$

Therefore, for F_j at a frequency where this singularity occurs, it is sufficient to simply sum the auto-PSD of the two datasets and hence use the simpler expression from *Schwarz et al.* (1990),

$$F_j(u,v) = \frac{P_{jg}(u,v)}{P_{11}(u,v) + P_{22}(u,v)} \qquad (7.23)$$

Due to the high correlation of the ascending and descending intermediate results, equation 7.23 was used in preference to equation 7.22 in order to avoid singularities in the operator. In the case of higher frequency elements, the lower signal to noise ratio reduces the correlation of the two datasets, whereupon equation 7.22 may be used. Because of time constraints, this point was not examined in detail and for consistency, equation 7.23 was implemented for all results in this section.

7.2.6 Methodology

Processing of data is performed in several stages, starting with the pre-processing of raw altimetry and ending with a global MSS map between latitudes 55° North and South.   The limited extent in latitude was implemented for two reasons, the first was due to uncertainty in the effect excess curvature of the ground tracks at high latitudes might have and secondly as a consequence of the first stage of processing.   This involves the use of TOPEX/Poseidon altimetry and precise orbits to refine ERS-1 range measurements through dual crossover minimisation, as described in section 4.4.   Dual crossover corections may therefore be unreliable outwith the latitude limits of the TOPEX/Poseidon orbit, i.e. 66°N/S.   As the principal objective at this stage was proof that the method could successfully recover geoid signal, the latitudinal boundaries were chosen to minimise any possible boundary effects.

Separate ascending and descending grids are then formed from all the available altimetry, sub-sampled to 0.5Hz along track using a Gaussian weighted filter of correlation length 1 second.   This produces a grid of 501 columns and 960 rows, as illustrated in figure 7.6. This clearly shows the presence of datagaps over the continents as well as demonstrating the geographical distortion at high latitudes - particularly evident at the southern tip of South America.   During this process, the geographical locations of datagaps were noted to allow substitution of model values from the combined JGM2, OSU91 and SST ocean surface model.   A constant bias offset was solved for where valid altimetry were available which was then deducted from the ocean surface model for missing grid points.   The geographical co-ordinates of each of the grid points were then used to generate a second grid of identical size containing only the modelled ocean surface heights.   Both of these grids were then windowed using identical latitudinal Hanning functions.   Auto and cross PSD's were then generated

as described by equations 7.13 and 7.14 respectively, displayed for ascending ERS-1 data in figure 7.5.

Evaluation of the spectral optimal estimator A_{sl} was then undertaken by calculating the ratio at each frequency of the two power spectra, equation 7.11. Due to the small sample size of one dataset for each PSD, some noise was present at each frequency. To limit the influence of this contamination, values of A_{sl} outwith a predetermined range for each frequency were capped to the closest boundary of that range. Figure 7.7 shows the values of the optimal estimator in terms of isotropic spherical harmonic degree. The ideal value for each estimator is 1.0, shown as a horizontal line and the maximum and minimum limits are shown above and below the majority of the data as solid black lines. Low pass filtering of the geoid signal is evident from this graph where the value of the estimator drops close to zero above degree 360 as this is the maximum frequency provided by the sea surface model. For this reason, no capping was implemented above this frequency, but instead the estimator was assigned a value of zero.

Generation of intermediate mean sea surfaces was then expedited by the methods described in the previous section using an augmented grid with $M=4$ to inverse FFT the solution onto the predefined common, rectangular latitude/longitude grid. Collocation of these separate ascending and descending surfaces was performed by FFT of these grids and imposition of a filter which described the limiting along and cross track sampling characteristic of the altimeter and the ground tracks in terms of the spatial frequencies represented in the spectrum of the regular, common grid. Using filters appropriate to each ascending or descending altimeter dataset, therefore, marine geoid signals were combined using equation 7.23 on an individual frequency basis before FFT back into the spatial domain to form the final MSS. The combination filters for each solution were in the form of stepped elliptical masks applied to the intermediate solutions in the

frequency domain. Four parameters were required to determine each filter: semi-major axis from the along track resolution, limited by the maximum resolution of the sea surface model of degree and order 360; semi-minor axis from the cross track resolution, equivalent to degree and order 250; inclination from the satellite inclination, thus defining the direction in the frequency domain of the highest resolvable frequency; finally the weight of the filter was determined from the volume of data. As the expected accuracies and volumes of data for the ascending and descending datasets are comparable, each binary filter was equally weighted.
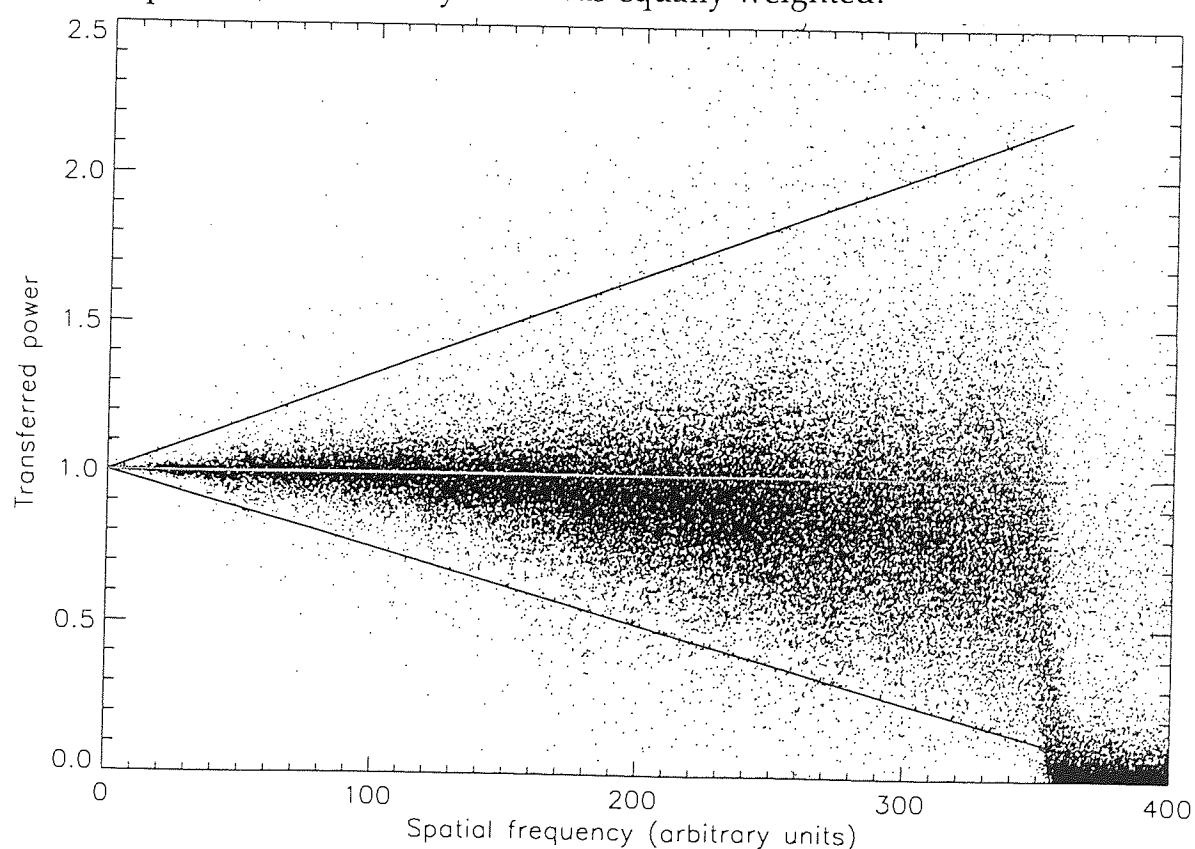

Figure 7.7. The optimal estimator A_{sl} in the frequency domain, plotted against equivalent spherical harmonic degree frequency.

7.3 Comparison of global results

Global surfaces in this chapter were determined by two different techniques, by simple Gaussian weighted averaging and by frequency domain optimal Fourier transform interpolation (OFTI). The averaging technique described in section 7.1 is by far the fastest and least demanding of computer memory. Surfaces generated by OFTI on the other hand demand more of the computer hardware and are extremely sensitive to aliasing errors. Because of the need to remove the long wavelength geoid before analysis of the altimeter datasets, solutions obtained by OFTI were not directly compared with those from the simple interpolation scheme. In addition, as will be shown, there are serious flaws in the simple Gaussian averaged surfaces which would make any comparison of limited value.

7.3.1 Gaussian interpolation results

Section 7.1 described the generation of MSS maps from altimetry by means of simple Gaussian weighted averaging, producing results which illustrated the large scale features of the geoid and perhaps illustrate some of the difference in resolution of the TOPEX/Poseidon and ERS-1 satellites. This section is concerned with the accuracy and consistency of these results.

Comparing first the MSS maps obtained from consecutive TOPEX/Poseidon cycles 2 and 3, figure 7.8 shows the scale of the problems associated with this interpolation regime. With a contour spacing of 10cm, the radial accuracy of the satellite data is totally obscured by the effect of missing arcs, particularly in the Indian Ocean where figure 7.1 clearly showed there to be missing arcs for the cycle 2 dataset. In other regions, isolated pockets of poor or absent data can be seen to seriously contaminate the picture, for example in the North Atlantic and many areas of the Pacific.
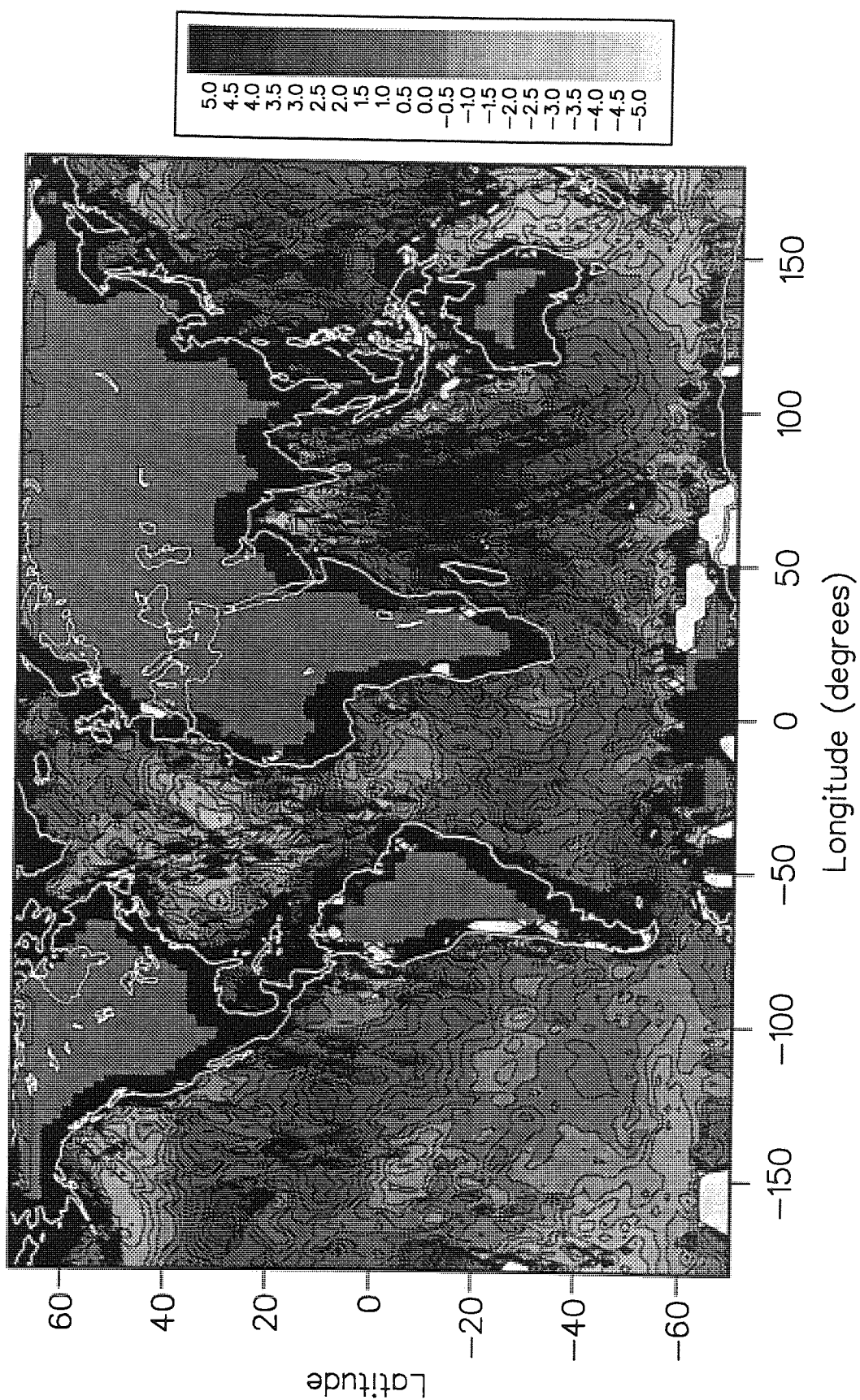
Figure 7.8. Difference in cycle 2 and cycle 3 TOPEX/Poseidon Gaussian weighted averaging MSS.

As a consequence, it is impossible to distinguish what may be dynamic topography from erroneous signal introduced by the interpolation method.

The photograph in figure 7.9 contains the detailed MSS obtained from the ERS-1 dataset (this shows a global map formed from the complete ERS-1 dataset for cycle 9, in practise only the values between 66°N/S were used further). As no orbit correction was applied, the radial accuracy of this dataset is around 25cm, as can be distinguished by the presence of ripples in some regions. These stripes are thought to be related to the poorly determined arcs mentioned in section 4.4, illustrated in figures 4.7 and 4.8.

In conclusion, this method can produce a MSS map from satellite data only where there is excellent coverage. Missing data have been shown to introduce differences of several metres between solutions, thus making the extraction of any oceanographic signal from the results extremely difficult.

7.3.2 Optimal Fourier transform interpolation results

In order to demonstrate the effect that orbit error can have on an altimetric MSS, two solutions were obtained from the ERS-1 cycle 9 dataset. The first of these contained altimetry derived from orbital solutions performed at Aston using the SATAN-A software package, described by *Rothwell* (1989) and *Ehlers* (1993) without any empirical correction. The second dataset was formed by empirical dual crossover minimisation of the long arc dataset, as described in section 4.4. All environmental and instrument corrections for the two datasets were identical in order that the results be consistent.

Numerically, these surfaces show agreement of 11.5cm and 10.2cm with their associated altimeter datasets, i.e. from the long arc solution and DXO corrected altimetry respectively. Through a combination of the fact that OFTI is essentially a filtering technique, combined with deficiencies in the short wavelength sea surface model (shown in figure 7.7 and described in section 7.2.6), comparison of the raw DXO corrected data with the sea
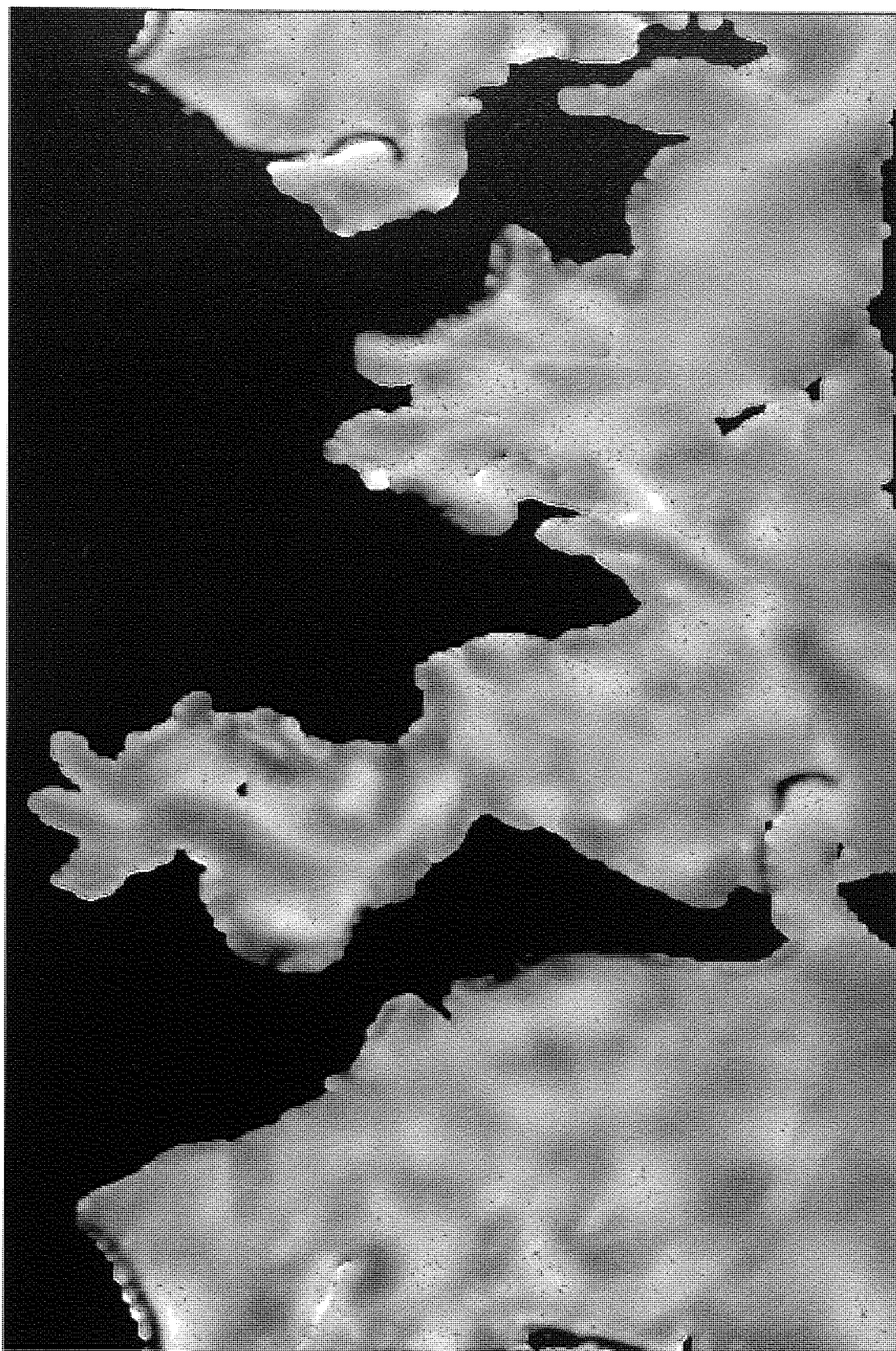
Figure 7.9.   Global mean sea surface from cycle 9 of ERS-1 altimetry, interpolated by
         Gaussian weighted averaging onto a rectangular 1°x1° grid using 254 equatorial
         grid points.

surface model showed an rms difference of around 26cm while comparison of the gridded solution also revealed a disparity of around 26cm rms. It should be pointed out that the former of these comparisons was obtained from the observations points, i.e. where there was altimetry, while the latter was evaluated at all grid points not over land where altimeter coverage was occasionally incomplete. A comparison of the two surfaces is given in figure 7.10 and clearly shows the influence of orbit error on the MSS, resulting in an rms difference of around 13cm. This is larger than the rms correction applied (section 4.4) and may be due to a number of factors.

- Differences are calculated at gridded locations and not at the altimeter ground points. This may influence the reliability of estimates where data coverage is poor.

- Through rejection of poorly resolved arcs during the DXO processing, there are areas in the uncorrected map supported by poor altimetry which have no data in the corrected solution

- Use of the FFT may have introduced aliasing at or near the cross track sampling interval, thus introducing an erroneous signal.

From analysis of a single cycle, it is difficult to identify which of these is the principal cause of errors in the final surface. Comparison of the best available dataset with a corrupted dataset in the form of the original long arc solution altimetry has shown that the method is at least self consistent while comparison with the geoid model shows that the gridded solution is of a similar accuracy to the along track data.

As stated in Chapter 2, the short wavelength geoid is strongly related to the local bathymetry; by comparison with the topographic image in figure 7.11, the MSS image of figure 7.12 can be seen to resolve all of the major ocean trenches and ridges. In particular, the Hawaiian chain of volcanic islands are clearly visible stretching westwards - beyond the extent of the
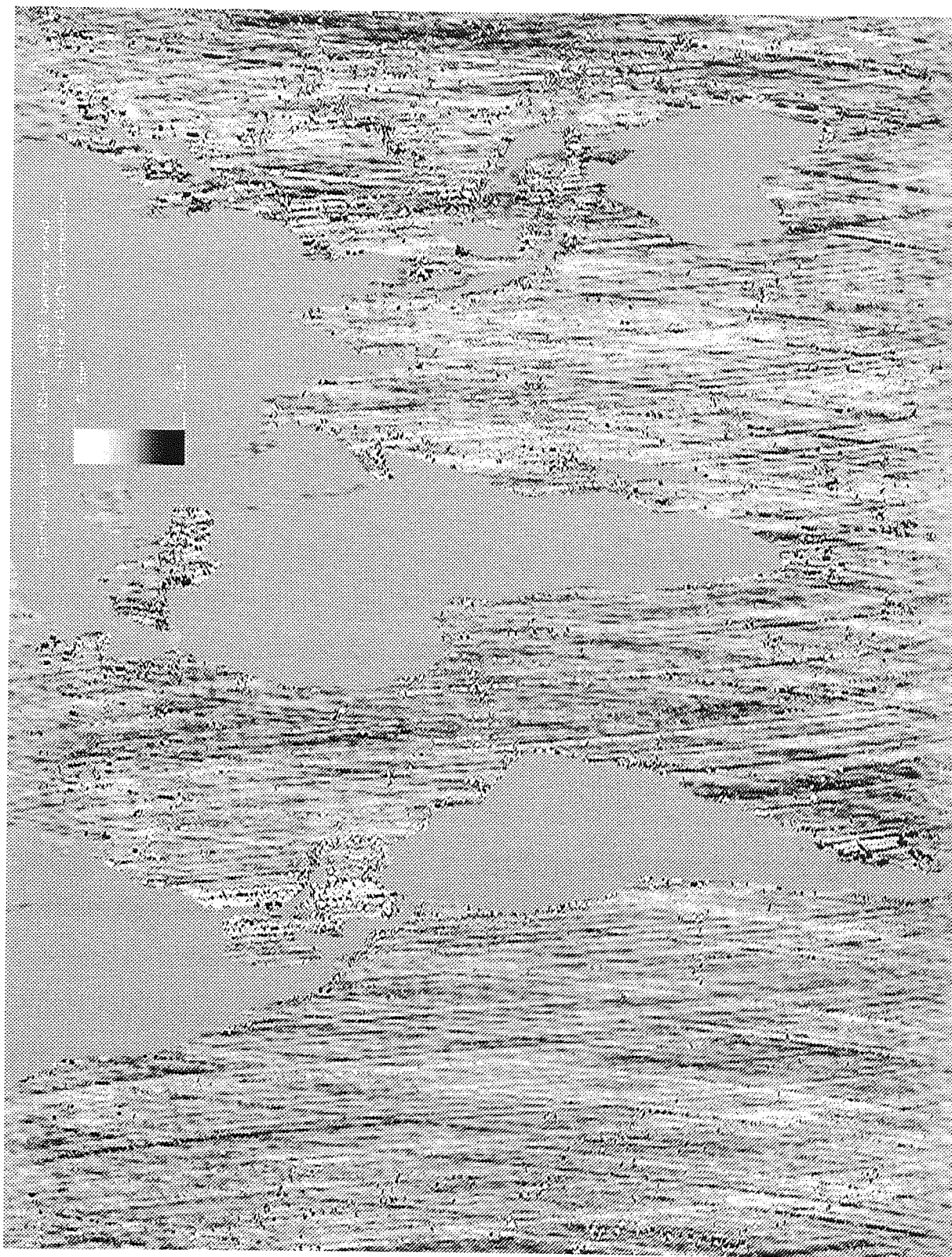
Figure 7.10. Greyscale difference plot of long arc and dual crossover corrected mean sea surfaces.

surface islands. Particularly prominent around the north-western edges of the pacific plate are the series of subduction trenches, clockwise from south of Alaska at the Aleutian trench, round to the south of Japan and the Marianas trench. As this represents a single, globally consistent solution, examination of local detail may be performed by simple enlarging the MSS image rather than by additional piecewise solutions.

Closer examination of the North Atlantic in figures 7.13 and 7.14 clearly shows the central spreading ridge crossed by the numerous fracture zones. The edge of the continental North American plate is clearly visible extending progressively further out from the coast from 35°N to the northern limit of the image. The shallow region of the Gulf of Mexico shows a great deal of small variations, typical of old sedimentary continental shelves extending eastwards to the crescent shaped Puerto Rico trench. Off the south-west coast of Spain, a range of sea mountains is also visible, as far south as the Canary Islands. Prominent off the western tip of Africa are the Cape Verdi Islands as two projections from the otherwise smooth long wavelength signal of the thin oceanic crustal geoid.

The MSS around the Australia shows several significant features, in particular, the Karmadec and Tonga Trenches running in a generally northerly direction from New Zealand's North Island. The Tonga trench has been formed by subduction of the Pacific Plate by the Tonga Micro-plate, which is moving at a relative speed of 24cm/yr; from *Bevis et al.* (1995), this represent the fastest tectonic motion to be seen at any plate boundary. The Island of Tasmania can also be clearly seen to be a part of the greater continental mass of Australia. To the north of Australia, the shallow nature of the South China Sea can be seen in the plan view, bounded to the south by Java and to west by Sumatra. The broad Chago-St Paul Ridge which runs from the Horn of Africa south-east before turning east at the Amsterdam-St Paul Plateau to pass south of Australia as the Indian Antarctic Ridge is visible through the south-western corner of figures 7.15
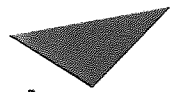
Figure 7.11. Solid surface topography illustrating in particular the bathymetry of the oceans. Image courtesy of the National Geophysical DataCente.

and 7.16, between latitudes 20°S and 30°S. This ridge is crossed by a series of small ripples, generally running at an angle of around 45° to the ridge axis. It should be noted that due to the use of the 10% Hanning window function, the features towards the south of the image are greatly suppressed.

7.4 Conclusion

On March 15 1995, ERS-1 completed its second interleaving geodetic phase, forming what is essentially a single 336-day cycle of 4822 orbits at an equatorial spacing of 8km. This dataset consists of approximately 20 million 1Hz altimeter observations over the entire globe - more than can possibly be analysed by either optimal interpolation or base function interpolation. To produce a globally consistent MSS solution requires either simple weighted average interpolation or some form of spectral domain regime. This chapter has demonstrated that while the broad structure of the marine geoid can be resolved by the weighted averaging method, it is too unreliable to produce useful results.

Figure 7.12. Global optimally interpolated MSS from cycle 9 ERS-1 dual crossover corrected altimetry, presented on a 0.25° longitude by 0.1° latitude grid.

Figure 7.13. Close view of the North Atlantic geoid derived from figure 7.12, viewed from directly above, illuminated from above.
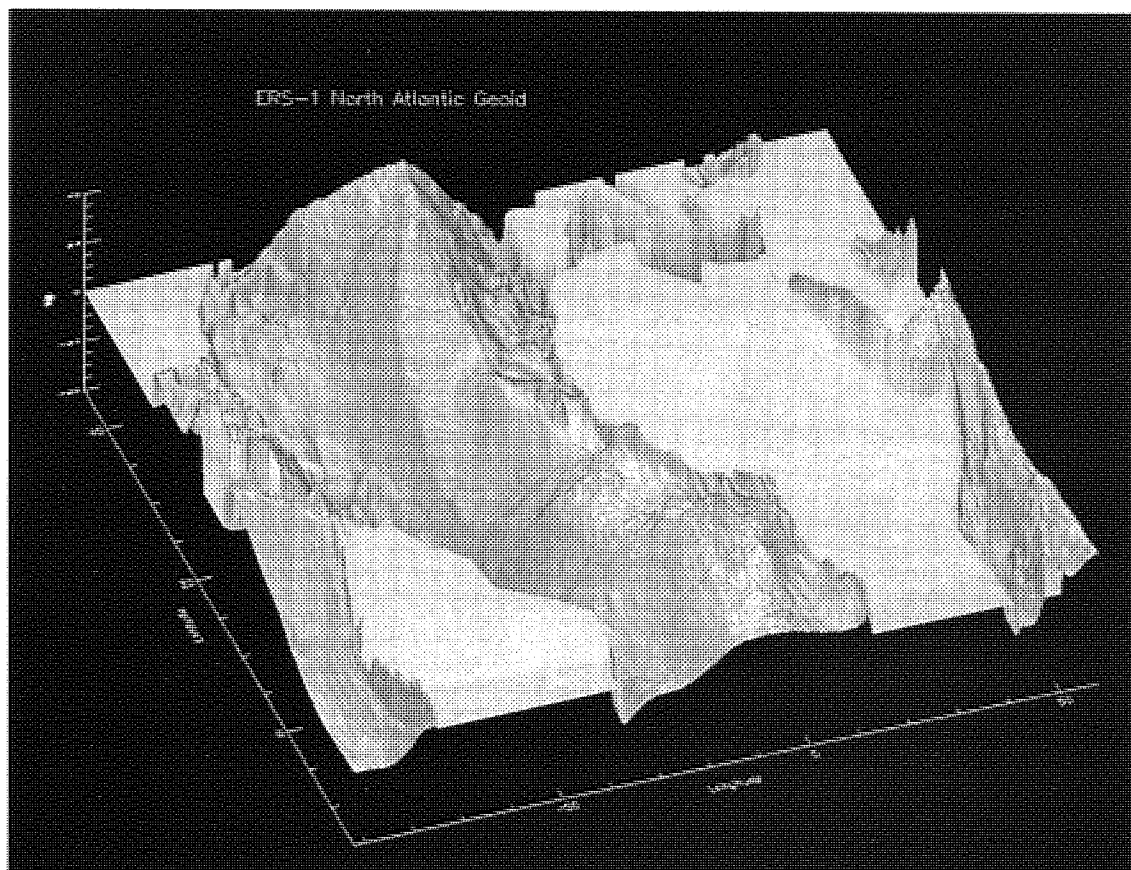

Figure 7.14. As for figure 7.13, viewed from the South West, illuminated from above.
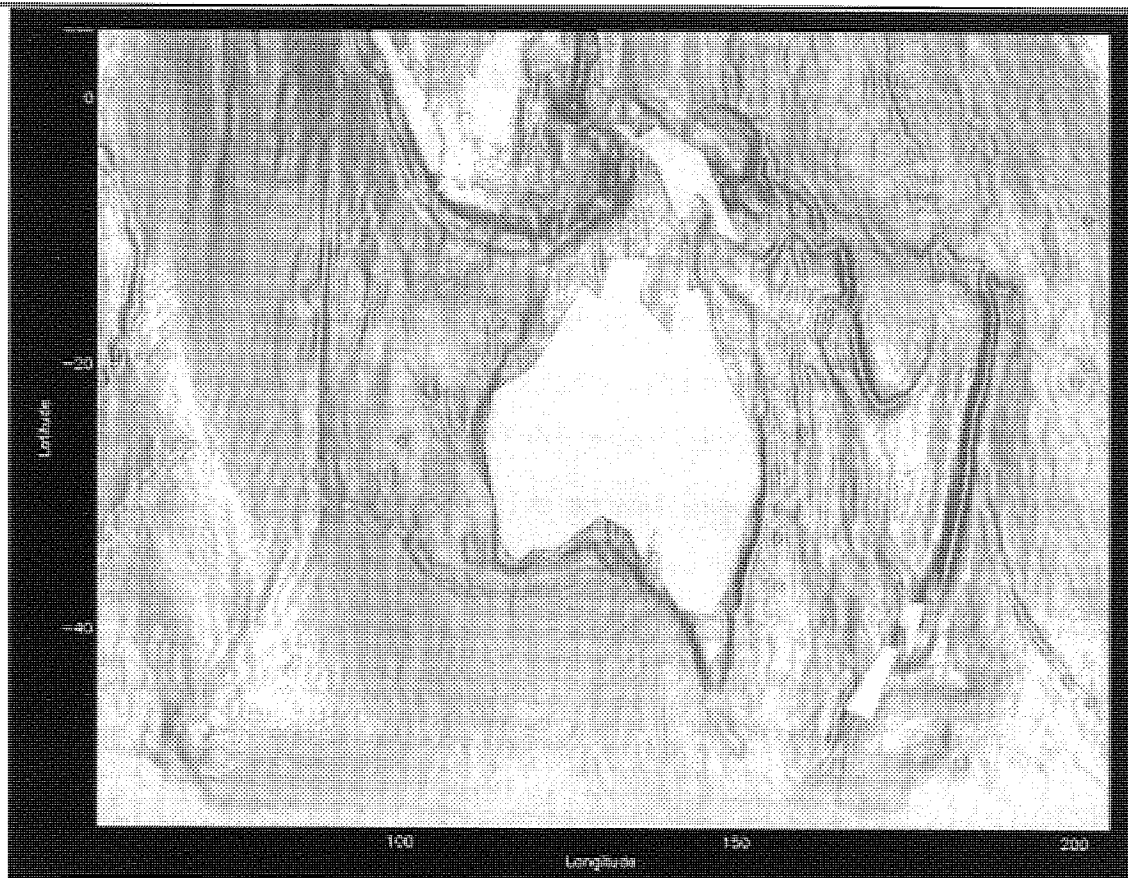
Figure 7.15. Close view of the geoid around Australia derived from figure 7.12, viewed from directly above.
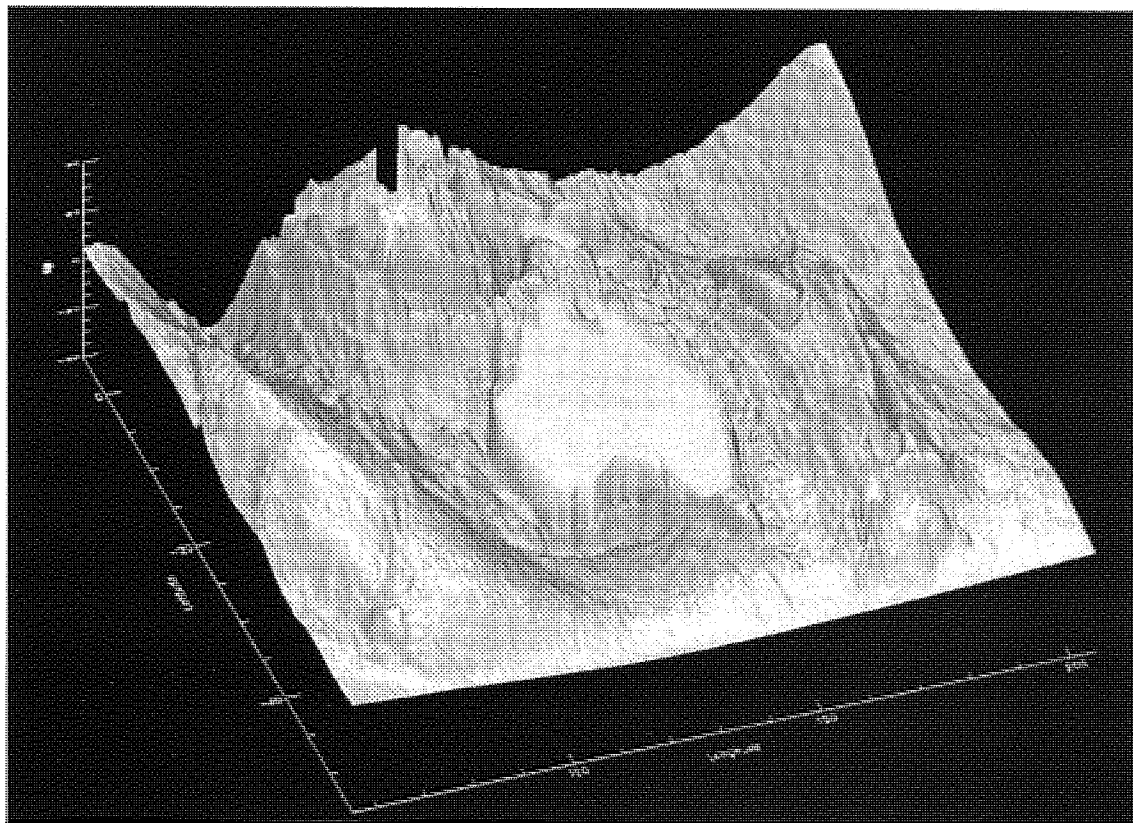
Figure 7.16. As for figure 7.15, viewed from the South-West, illuminated from above.

From techniques described for the interpolation of regularly spaced gravimeter data, spectral analysis techniques have been developed and tested in this chapter using 35-day dual crossover corrected ERS-1 altimetry. It has been shown that this technique is capable of producing geologically significant results from this comparatively small dataset. To do the same for the Geodetic Phase data will require several remaining problems to be addressed, namely, modelling of the short wavelength geoid above degree and order 100 where the OSU91A model has been shown to be under valued as well as above degree and order 360 where no model information is available. This method also has computational limitations due to the size of the dataset involved, i.e. requiring the Fourier transform of a 4822x3000 matrix - an alternative to the augmented FFT transformation may therefore have to be considered. This does not detract, however, from the strength of optimal Fourier transform interpolation as a basic technique. It is clearly superior to spatial domain techniques of chapter 6 because of the ability to generalise analysis to the point where the full global dataset may incorporated into a single two stage solution.

Chapter 8

Conclusion

8.0 Summary

The majority of the original work undertaken for this thesis has been described in chapters 4 to 7, and has dealt with the correction and analysis of satellite borne radar altimetry. The main emphasis of the work has been directed towards the empirical correction of errors in the modelled radial height of the satellite at the time of each altimeter range measurement. This work was then developed by an investigation of both dynamic and static ocean signal recovery from these corrected measurements.

Radial orbit correction techniques exploit the long wavelength nature of ephemeris modelling errors in order to distinguish the comparatively shorter ocean signals from the error signatures. Two methods were examined: single satellite crossover minimisation using 2.5 years of Geosat crossover data and dual satellite crossover minimisation using altimetry from the first pair of altimetric satellites to operate concurrently - ERS-1 and TOPEX/Poseidon. The Geosat study demonstrated through examination of crossover difference differences (CDD's) that from 11,500km track lengths, sinusoidal approximation of the one cycle per revolution orbit error produces a significant improvement in the accuracy of altimetric measurements without significant aliasing of long wavelength ocean signals. During the analysis of the CDD's from a complete circle of latitude in the Southern Ocean, several significant features of the radial orbit error were observed. These included a large geographically correlated error which exceed 1m in places and a small, 2cm, sinusoidal error of similar period to the rotation rate of the solar angle with respect to the satellite's orbital plane. The latter of these two error signals demonstrates the subtlety

of some of the forces which influence satellite orbits as well as the sensitivity of the altimeter instrument which recorded the error in modelling that force.

Dual satellite crossover minimisation was based on the same techniques as used and tested with the Geosat dataset, with the exception that the reference dataset of assumed negligible radial error was taken to be that of TOPEX/Poseidon. ERS-1 altimetry were then adjusted using a bias and one cycle per revolution sinusoid. This resulted in ERS-1 data of better than 10cm rms accuracy and an estimate of the relative altimeter biases of (29.1±3.3)cm and (43.8±3.6)cm from ERS-1/NRA and ERS-1/SSALT respectively (NRA: NASA Radar Altimeter; SSALT: Solid State ALTimeter). From theoretical and numerical modelling of the interaction of TOPEX/Poseidon orbit errors with sinusoidal correction of ERS-1, it was shown that significant and in some cases amplified error signals could be directly aliased into ERS-1 altimetry though the use of dual satellite crossovers. In particular this showed that in the case of geographically correlated errors from TOPEX/Poseidon, comparatively short wavelength features could be aliased into the long wavelength ERS-1 orbit error approximation. This may be attributed to the non-uniform sampling of orbital ground tracks; data shortages due to land in the altimeter footprint and the geometry of the two satellite ground tracks

The static component of altimetric measurements is a consequence of the local strength and direction of the Earth's gravity field at the sea surface. Recovery of the marine geoid signal was examined on both regional and global scales using Geosat, ERS-1 and TOPEX/Poseidon data. For Geosat altimetry in the North Sea it was shown that despite a 70cm radial orbit error, the marine geoid could be resolved to an accuracy of better than 12cm when compared to a regional high resolution gravity model. This accuracy was achieved using Optimal Interpolation, which although able to extract accurate signals from otherwise noisy data, is limited by the availability of

computing power. For this reason, and in order to perform global analyses using ERS-1 data, the technique was modified to operate in the frequency domain, where the relative correlations of regularly gridded altimetry were exploited. While sensitive to both long and short wavelength aliasing, this technique was shown to produce significant results from a single cycle of ERS-1 35-day data.

8.1 Future development

The value of data from satellite altimeter missions has historically been seriously limited by the contamination of data with uncertainty from the satellite's exact radial position. The advent of near continuous tracking by networks such as the Global Positioning System (GPS) will help to improve the orbital accuracy of future missions. This has been demonstrated by TOPEX/Poseidon which has achieved an accuracy of around 1-2cm. Missions such as ERS-2 and the Geosat Follow On (GFO) will be at lower altitudes than the US/French platform and hence atmospheric drag and higher order gravity field uncertainties will still make orbital modelling difficult. Empirical techniques will therefore remain an important tool, not only in the analysis historical altimeter data, but for future low altitude missions. In particular, dual satellite crossovers (DXO) may be available for ERS-2/GFO, which respectively will have PRARE and GPS tracking in addition to satellite laser ranging. This will introduce problems associated with dual crossover minimisation and subsequent aliasing when cross contamination of orbit errors and other geographically and temporally correlated errors will be significant.

As was stated at the end of the last chapter, ERS-1 has now completed its Geodetic Phase of 4822 revolutions in an interlaced 168-day pattern. This represents the highest resolution altimeter dataset publicly available and is likely to lead to the release of the remainder of the Geosat Geodetic Mission (GM) dataset north of 30°S in the near future. It is extremely unlikely that

any altimetric mission in the near future will again include a geodetic orbit pattern, therefore once the GM data are released, it will not be added to. This will present a finite dataset for analysis of the marine geoid. There are several established techniques for the analysis of regional, high resolution geoid and gravity anomalies from altimetry. The method described in chapter 7 represents one of the few which may be capable of producing a globally consistent geoid map from a complete dataset. It is acknowledged that there still exist certain obstacles to be overcome before the optimal Fourier transform interpolation can be considered an operational technique for geoid recovery at all wavelengths. Principal amongst these is the need to generate an accurate model surface for evaluation of the cross power spectral density - this is, however, a problem for all such inverse techniques.

8.3 Concluding remarks

Satellites in orbit about the Earth pass over all of the planets surface water and since the first experimental use of the Skylab altimeter in 1973, many millions of measurements have been made of the sea surface height. These data have come from a succession of satellites, starting with GEOS-3, briefly on SEASAT, followed by the first long term mission to produce measurements of sufficient accuracy, Geosat. More recently two satellites currently in operation are ERS-1 and TOPEX/Poseidon which will be followed by ERS-2 and the Geosat Follow on mission. All of these altimetric missions have required a major effort to evaluate the precise radial height of the satellite at the exact time of a particular observation. TOPEX/Poseidon has undoubtedly benefited from its range of tracking systems and almost continuous tracking coverage, which combined with its high altitude have produced orbital ephemerides of 1-2cm accuracy. For lower satellites, radial orbit error has always been the largest single source of uncertainty in the data and for the foreseeable future it will remain so.

There has therefore always been a demand for reliable empirical techniques for the refinement of satellite altimetry, leading to the development of various filters and approximations. This thesis has described an investigation of the accuracy and reliability of one of these techniques, whereby the dominant one cycle per revolution component of the radial error is approximated by a sinusoid. By constrained least squares solution, it was shown that such a method not only improves the accuracy of altimetric sea height measurements, but that it does not in the process remove excessive meso-scale ocean signal. By adapting this technique to examine dual satellite crossovers, it was again found to produce a distinct improvement in the measurement accuracy of the altimeter. Close examination of the theoretical absorption of specific TOPEX/Poseidon orbit error components revealed a slight degradation of ERS-1 accuracy, but at a level much below other contaminating influences.

Following the investigation of various mean sea surface recovery techniques, a significant improvement was made in the use of optimal interpolation for the analysis of global altimetric datasets. By moving from spatial domain to frequency domain analysis, it was shown that inverse methods could be applied to satellite data by exploiting the regular nature of separate ascending or descending datasets. Following completion of ERS-1's Geodetic Phase, optimal Fourier transform interpolation will provide a means of incorporating the complete 336-day data record into a single, consistent geoid of the whole marine environment.

References

Barrick, D.E. and B.J. Lipa, Analysis and interpretation of altimeter sea echo. *Adv. Geophys.*, **27**, 61-99, 1985.

Bearman, G., (Ed.), *Wave, tides and shallow water processes*, Open University, Walton Hill, Milton Keynes, UK, 1989.

Bendat J. G., and A. G. Piersol, *Engineering application of correaltion and spectral analysis*, Wiley-Interscience Publication, John Wiley and sons Inc., New York, 1980.

Bevis, M, F.W. Taylor, B.E. Schutz, J. Recy, B.L. Isacks, S. Helu, R. Singh, E. Kendrick, J. Stowell, B. Taylor and S. Calmanat, Geodetic Observations of very rapid convergence and back-arc extension at the Tonga arc, *Nature*, **374**, No. 6519 1995.

Bhaskaran, S., and G.W. Rosborough, Computation of regional mean sea surfaces from altimetry data, *Manuscr. Geod.*, **18**, 147-157, 1993.

Blanc, F., S. Houry, P. Mazzega, and J.-F. Minster, High resolution, high accuracy altimeter derived mean sea surface in the Norwegian Sea, *Mar. Geod.*, **14**, 57-76, 1991.

Born, G.H., B.D. Tapley, M.L. Santee, Orbit determination using dual crossing arc altimetry, *Acta Astronautica*, **13**, 157-163

Bracewell, R., *The Fourier transform and its applications*, 2nd edition, McGraw Hill, New York, 1978.

Carnochan, S., *Sea-state bias for ERS-1*, private communication, 1995.

Carter, D. J. T., P. G. Challenor and M. A. Srokosz, An assessment of Geosat wave height and wind speed measurement, *J. Geophys. Res.*, **97**, 11383-11392, 1992.

Cartwright, D., and R. Ray, Ocean tides from Geosat altimetry, *J. Geophys. Res.*, **95**, 3069-3090, 1990.

Chelton, D. B., and P. J. McCabe, A review of satellite altimeter measurement of sea surface wind speed; with a proposed new algorithm, *J. Geophys. Res.*, **90**, 4707-4720, 1990.

Chelton, D.B., and M. G. Schlax, Spectral characteristics of time-dependent orbit errors in altimeter height measurements, *J. Geophys. Res.*, **98**, 12,579-12,600, 1993.

Chelton, D.B., M. G. Schlax, D. L. Witter and J. G. Richman, Geosat altimeter observations of the surface circulation of the Southern Ocean, *J. Geophys. Res.*, **95**, 17,887-17,904, 1990.

Chelton, D.B., E.J Walsh and J.L. MacArthur, Pulse compression and sea level tracking in satellite altimetry, *J. Atmos. Oceanic Technol.*, **6**, 407-438, 1989

Cheney, R. E., B. C. Douglas, and R. W. Agreen, Geosat altimeter crossover difference user handbook, *NOAA Manual NOS NGS 6*, 32 pp, Nat'l. Oceanogr. and Atmos. Admin., Silver Spring, Md., 32pp, 1991.

Cheney, R.E., B.C. Douglas, R.W. Agreen, L. Miller, D.L. Porter, N.S. Doyle, Geosat altimeter geophysical data record user handbook, *NOAA Technical Memorandum NOS NGS-46*, pp28, 1987.

Christiansen, E. J., B. J. Haines, S. J. Keihm, C. S. Morris, R. A. Norman, G. H. Purcell, B. G. Williams, B. D. Wilson, G. H. Born, M. E. Parke, S. K. Gill, C. K. Shum, B. D. Tapley, R. Kolenkiewicz and R. S. Nerem, Calibration of TOPEX/Poseidon at Platform Harvest, *J. Geophys. Res.* **99**, 24465-24485, 1994.

Cooley, S.W. and J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Math. Comput.*, **19**, 297-301, 1965.

Cracknell and Hayes, *Introduction to Remote Sensing*, Taylor and Francis Ltd, London, U.K., 1991.

Denker, H., and R. H. Rapp, Geodetic and oceanographic results from the analysis of 1 year of Geosat data, *J. Geophys. Res.*, **95**, 13,151-13,168, 1990.

Douglas, B.C., and R.W. Agreen, The sea state correction for GEOS-3 and SEASAT satellite altimeter data, *J. Geophys. Res.*, **87**, 3221-3226, 1983

Dumont, J. P., and J. Stum, *Altimeter products user manual*, C11-EX-MUT-A2-07-CN, Issue 1, Version 2, CLS/ARGOS, Toulouse, France, 1992

Ehlers, S., *Various techniques and proceedures for refining ERS-1 orbits*, PhD Thesis, Aston University, 1993.

Engelis, T., On the simultaneous improvement of a satellite orbit and determination of sea surface topography using satellite altimetry, *Manuscr. Geod.*, **13**, 180-190, 1988.

Francis, R., The height calibration of the ERS-1 radar altimeter, ESA SP-339, 381-394, 1993.

Fu, L.-L. and D. B. Chelton, Observing large-scale temporal variability of ocean currents by satellite altimetry with application to the Antarctic Circumpolar Current, *J. Geophys. Res.*, **90**, 4721-4739, 1985.

Gaspar, P., F. Ogor, P.Y. Le Trayon and O.Z. Zanife, Estimating the sea state bias of the TOPEX and Poseidon altimeters from crossover differences, *J. Geophys. Res.*, in press, 1994.

Harangozo, S. A., P. L. Woodworth, R. H. Rapp, and Y. M. Wang, A comparison of monthly sea level variability determined from Geosat altimetry and a global tide gauge dataset, *Int. J. Remote Sensing*, **14**(4), 789-796, 1993.

Hecht, E., *Optics*, 2nd ed., Addison-Wesley Publishing Company, 1987.

Holdridge, M.E., Results from Geosat ERM data gap/attitude analysis, JHU APL memo., Laurel, MD., USA, 1988, as reported by *Cheney et al.*, 1991.

Houry S, and P. Mazzega, Large inversions of altimeter and tied tide gauge data for the Mediterranean Sea mean surface, *J. Geophys. Res*, **96**, B2, 2417-2429, 1991.

Jolly, G.W., and P. Moore, Validation of empirical orbit error corrections using crossover difference differences, *J. Geophys. Res.*, **99**, 1994a.

Jolly, G.W., and P. Moore, Accuracy assesment of TOPEX/Poseidon dual crossover correction of ERS-1 altimetry, *Adv. Space Res.*, in press, 1994b.

Jolly, G.W., and P. Moore, Analysis of global ERS-1 altimetry using optimal Fourier transform interpolation, submitted to *Mar. Geod.*, 1995.

Kaula W.M. *Theory of Satellite Geodesy*, Balisdell Publishing Co, 1966.

Kilgus, C.C., Geosat coverage of the ocean, JHU APL memo., Laurel, MD., USA, 1987, as reported in *Cheney et al.*, 1991

Kingsley, S., and S. Quegan, *Understanding Radar Systems*, McGraw-Hill Book Company, Maidenhead, U.K., 1992.

Kim M.C., C. K. Shum and B.D. Tapley, An altimetric mean sea surface and its spectral analysis, *Proceedings of the second ERS-1 Symposium*, ESA SP-361, 1994.

Lam, C. W., P. Moore and P. L. Woodworth, Calibration of ERS-1 altimetry over the North Sea, Proceedings of the first ERS-1 symposium, Nov. 1992, ESA SP-359, 1993.

Lambek, K., Aristoteles: an ESA misson to study the Earth's gravity field, *ESA Journal*, **14**, (1), 1-21, 1990.

Lelgeman, D., On harmonic properties of interpolation with harmonic kernal functions, *Manuscr. Geod.*, **6**, 157-191, 1981.

Lerch, F. J., C. E. Doll, S. M. Klosko, J. J. McCarthy, W. F. Eddy, R. G. Williamson, T. A. Williams, Precision orbit determination for TOPEX/Poseidon using TDRSS doppler tracking data, Paper No. PSD-024, presented at COSPAR XXX, Hamburg, Germany, 1994.

Marsh, J. G., and T. V. Martin, The Seasat altimeter mean sea surface model, *J. Geophys. Res.*, **87**, 3269-3280, 1982.

Marsh, J.G., et al. The GEM-T2 gravitational model, NASA Tech. Memo., 100746, 94 pp., 1989.

Massemann, F. H., C. Reigber, R. Konig, J. C. Raimondo and C. Rajasenan, ERS-1 orbit information provided by D-PAF, ESA-SP 361, 765-770, 1994.

Mazzega, P. and S. Houry, An experiment of invert SEASAT altimetry for the Mediterranean and Black Sea mean sea surfaces, *Geophys J.* **96**, 2, 259-272, 1989.

McAdoo, D.C., and K. M. Marks, Gravity fields of the Southern Occean from Geosat data, *J. Geophys. Res.*, **97**, 32476-3260, 1992.

Menard, Y., E. Jeansou and P. Vincent, Calibration of TOPEX/Poseiodn altimeters at Lampedusa: additional results at Harvest, *J. Geophys. Res.*, **99**, 24487-24504, 1994.

Milbert, D., B. C. Douglas, R. E. Cheney, L. Miller and R. W. Agreen, Calculation of sea level time series from non-collinear Geosat altimeter data, *Mar. Geod.*, **12**, 287-302, 1988.

Moore, P., and S. Ehlers, Precise orbit determination and sea surface topography from ERS-1, *Adv. Space Res.*, in press, 1995.

Moore, P., and D. A. Rothwell, A study of gravitational and non-gravitational modelling errors in cross-over differences, *Manuscr. Geod.*, **15**, 187-206, 1990.

Moritz, H., *Advanced Physical Geodesy*, 500 pp, Herbert Wichmann verlag, Karlsruhe, Germany, 1980.

Müller, T., and Groten, E., Evaluation of geodetic and geophysical information in a common model for the determination of mean sea level, Paper presented at IUGG, XX, Union Symposiumm U13, Vienna, August 11-24, 1991.

Nerem, R. S., F. J. Lerch, J. A. Marshall, E. C. Pavlis, B. H. Putney, B. D. Tapley, R. J. Eanes, J. C. Ries, B. E. Schutz, C. K. Shum, M. M. Watkins, S. M. Klosko, J. C. Chan, S. B. Lithcke, G. B. Patel, N. K. Pavlis, R. G. Williamson, R. H. Rapp, R. Baincale and F. Nouel, Gravity model development for TOPEX/POSEIDON: Joint Gravity Models 1 and 2, *J. Geophys. Res.*, **99**, 24421-24448, 1994.

Pugh, D. T., *Tides, surges and mean sea-level: a handbook for engineers and scientists*, John Wiley and Sons, Ltd, Chichester, UK, 1987.

Rapp, R. H., and J. Y. Cruz, Spherical harmonic expansion of the Earth's gravitational potential to degree and order 360 using 30' mean anomalies, *OSU report 376*, Columbus, Ohio, U.S.A., 1986.

Rapp, R. H., Y. M. Wang and N. K. Pavlis, The Ohio State 1991 geopotential and sea surface topography harmonic coefficient models, *OSU Rep. 410*, Columbus, Ohio, 1991.

Rees, W.G., *Physical Principals of Remote Sensing*, University Press, Cambridge, U.K., 1990.

Robinson, I.S., *Satellite Oceanography: an introduction for oceanographers and remote sensing scientists*, Ellis Horwood Ltd, Chichester, U.K., 1985.

Rodriguez, E., Y. Kim, and J.M. Martin, The effect of small-wave modulation on the elctromagnetic bias, *J. Geophys. Res.*, **97**, 2379-2389, 1992.

Rothwell, D. A., *Precise orbit determination and analysis from satellite altimetry and laser ranging*, PhD thesis, Aston University, 1989.

Roy, A.E., *Orbital Motion*, Adam Hilger, Bristol, U.K., 1978

Rummel R. and Rapp, R.H., Undulation and anomaly estimation using GEOS-3 altimeter data without precise satellite orbits, *Bull. Geod.*, **51**, 73-88, (1977).

Sandwell D. T., and D. C. McAdoo, High-accuracy, high reolution gravity profices from 2 years of the Geosat exact repeat mission, *J. Geophys. Res.*, **95**, 3049-3060, 1990.

Scharoo, R., K. F. Wakker and G. J. Mets, The orbit detyermination accuracy of the ERS-1 mission, *ESA-SP 361*, 735-740, 1994.

Schrama, E. J. O., Some remarks on several definitions of geographically correlated errors: Consequances for altimetry, *Manuscr. Geod.*, **17**, 282-294, 1992.

Schwarz, K. P., M. G. Sideris and R. Forsberg, The use of FFT techniques in physical geodesy, *Geophys. J. Int.*, **100**, 485-514, 1990.

Schwiderski, E.W., On charting global tides, *Rev. of Geophys. Space Phys.*, **18** (1), 243-268, 1980.

Shum, C.K., D.N. Yuan, J.C. Ries, J.C. Smith, B.E. Shutz and B.D. Tapley, Precision Orbit determination for the Geosat Exact Repeat Mission, *J. Geophys. Res.*, **95**, C3, 2887-2898, 1990

Shum, C.K., B.H. Zhang, B.E. Schutz, B.D. Tapley, Altimeter crossover methods for precision orbit determination and the mapping of geophysical parameters, *J. Astron. Sci.*, **38**, 355-368

Srokosz, M.A., On the joint distribution of surface elevation and slopes for a nonlinear random sea, with application to radar altimetry, *J. Geophys. Res.*, **91**, 995-1006, 1986.

Stewart M. P., and R. G. Hipkin, A high resolution, high precision geoid for the British Isles, 1989.

Tai, C. K., How to observe the gyre to global-scale variability in satellite altimetry: Signal attenuation by orbit error removal, *J. Atmos. Oceanic Technol.*, **8**, 271-288, 1991.

Tai, C. K., Accuracy assessment of widely used orbit error approximations in satellite altimetry, *J. Atmos. Oceanic Technol.*, **6**, 147-150, 1989.

Tai, C. K., Geosat crossover analysis in the tropical pacific, 1, Constrained sinusoidal crossover adjustment, *J. Geophys. Res.*, **93**, 10,621-10,629, 1988.

Tapley, B. D., J. C. Ries, G. W. Davis, R. J. Eanes, B. E. Schutz, C. K. Shum, M. M. Watkins, J. A. Marshall, R. S. Nerem, B. H. Putney, S. M. Klosko, S. B. Luthcke, D. Pavlis, R. G. Williamson and N. P. Zelensky, Precision orbit determination for TOPEX/Poseidon, *J. Geophys. Res*, **99**, 24393-24404, 1994.

Tapley, B. D., C. K. Shum, D. P. Chambers, G. Peterson and J. C. Ries, Large-scale dynamic sea surface topography from TOPEX/Poseidon, presented at the TOPEX/Poseidon verification workshop, JPL, California, U.S.A, 1993

Telford, W.M., L.P. Geldart, R.E. Sheriff, D.A. Keys, Applied Geophysics, University Press, Cambridge, U.K., 1976.

TOPEX, *Satellite altimetric measurements of the ocean*, Report of the TOPEX Science Working Group, NASA, Jet Propulsion Laboratory, California Institute of Technology, California, U.S.A., 1981.

Van Gysen, H., R. Coleman, R. Morrow, B. Hirsch, and C. Rizos, Analysis of collinear passes of satellite altimeter data, *J. Geophys. Res.*, **97**, 2265-2277, 1992.

Visser, P. N. A. M., K. F. Wakker and B. A. C. Ambrosius, Dynamic sea surface topography from Geosat altimetry, *Mar. Geod.*, **16**(3), 1993

Voss, P., and M. Handoll, *UK-ERS-1 reference maual*, DC-MAEOS-DD-0001, ver 1.0, published by Royal Aircraft Establishment, Farnborough, U.K., 1991.

Wagner, C.A., and R.E. Cheney, Global sea level changes from satellite altimetry, J. Geophys. Res., **97**, 1992

Walsh, E.J., F.C. Jackson, D.E. Hines, C. Piazza, L.G. Hevizi, D.J. McLaughlin, R.E. McIntosh, R.N. Swift, J.F. Scott, J.K. Yungel and E.B. Frederick, Frequency dependence of electromagnetic bias in radar altimeter sea surface measurements, *J.Geophys. Res.*, **96**, 20571-20583, 1991.

Wakker, C.A., B.A.C. Ambrosius and L. Aardoom, Precise orbit determination for ERS-1, Delft Univ. of Technology, Dept. of Aerospace Engineering, 1983.

Waterson, W, *The essential Calvin and Hobbes*, Publ. by Universal Press Syndicate, Kansas City, Missouri, USA, 1990.

West, G.B., Data processing system specifications for the Geosat satellite radar altimeter, NSWC TR 86-149, Naval Surface Weapons Center, Dahlgren, VA 22448, U.S.A., 1986

Woodworth, P. L., and C. W. Hughes, Tidal models, paper presented at the *Royal Astronomical Society*, 'Geodesy from Space' meeting, March, 1995.

Wunsch, C., Global-scale sea surface variability from combined altimetric and tide guage measurements, *J. Geophys. Res.*, **96**, 15,053-15,082, 1991.

Wunsch, C., and V. Zlotnicki, The accuracy of altimetric surfaces, *Geophys. J. Roy. Astron. Soc.*, **78**, 795-808, 1984