Satellite Techniques for Studying
Ocean Circulation

by

Matthew Stephen Jones

Thesis submitted for the degree
of Doctor of Philosophy of the
University of London

Mullard Space Science Laboratory
University College London
Holmbury St. Mary
Dorking
Surrey RH5 6NT

January 1997
To Zöe and my parents
Ships that pass in the night,
and speak each other in passing;
Only a signal shown
and a distant voice in the darkness;
So on the ocean of life
we pass and speak one another,
Only a look and a voice;
then darkness again and a silence.

Henry Longfellow, “The Theologian’s Tale”. 
The work described in this thesis would either not have happened, or would be much poorer in content, were it not for the help and advice of many people. I am particularly grateful to the James Rennell Division and Mullard Space Science Laboratory for the exciting and supportive working environment they provided. I would like to express specific thanks to the following people:

To Trevor Guymer and Peter Challenor for thinking up this project and for their constant support, advice and interest.

To my university supervisor Mark Saunders for his conscientious support, keen scientific advice - and his diligent approach to improving my grammar!

To Helen Snaith and Dave Cotton for providing the basic altimeter processing software and for Helen's constant advice on altimetric and programming matters.

To Graham Quartly and Meric Srokosz for many interesting discussions related to altimetry.

To Tom Forrester for advice on sea surface temperature measurements and for providing the basis of the ATSR processing software.

To David Cromwell and Paolo Cipollini for being such pleasant office-mates and for the many interesting discussions with them about a variety of matters.

To Raymond Pollard for providing the opportunity to participate in a 6-week oceanographic research cruise in the Southwest Indian Ocean.

To Peter Taylor for his advice on the distribution of cloud types.

To Andy Harris for his tireless stream of informative discussion on topics related to SST measurements and to Sabu Kim for his constantly entertaining stories.

To Ian Mason and Seymour Laxon for their helpful advice following presentations of parts of this work at group meetings.

To Myles Allen for his advice on the last parts of the work in this thesis and for his role in ensuring that this work will be continued in the future.
Acknowledgements

To ESA, Chris Mutlow and the ATSR science team for providing the ATSR data. Particular thanks goes to Albin Závody for his helpful and open advice on cloud clearing matters and for permission to use several figures from his papers in this thesis.

To the TOPEX/POSEIDON team and CNES for provision of the altimeter data.

To Brian King for providing the SST data from the WOCE A11 cruise and to Robin Tokmakian for providing SSH data from the Parallel Ocean Climate Model.

To Dudley Chelton for allowing figures from several of his papers to be used in this thesis and to Richard Rapp for providing the JGM3/OSU91A hybrid geoid model.

To Zöe for her constant support and love over the last three years, for helping me to keep my perspective on things and for her endless proof reading of the numerous revisions of this thesis.

Lastly, but most importantly, I thank God, because without Him, what's the point?
Abstract

Satellites provide a unique semi-synoptic view of the world's oceans. In recent years, two forms of remotely sensed data have been particularly useful in providing information about ocean circulation, namely altimetric measurements of sea surface height (SSH) and infrared radiometric measurements of sea surface temperature (SST). However, in order to interpret new types of data correctly and obtain meaningful results, new techniques must be developed. In this thesis, techniques to process TOPEX/POSEIDON radar altimeter SSH data and Along-Track Scanning Radiometer (ATSR) SST data are developed. These techniques are tested in the South Atlantic Ocean. The effectiveness of an existing technique to correct for across-track variations in altimeter sampling and the associated SSH errors due to across-track mean sea surface variation is studied. The effects of orbit error removal and interpolation on altimeter data are investigated using ocean model data from the Parallel Ocean Climate Model (POCM). A technique to obtain absolute velocities from altimetry alone is implemented and its accuracy assessed through use of the POCM data. Remnant cloud contamination in the ATSR 0.5° night SST data is discovered and a new technique to remove the cloud contamination is proposed and tested. The seasonality of this cloud contamination is investigated and is found to coincide with the occurrence of marine stratiform clouds. Finally, the relationship between SST and SSH data is examined. It is found that spatial cross-correlations between SST and SSH are surprisingly high (~0.7) in regions associated with fronts and mesoscale variability such as the Agulhas, the Antarctic Circumpolar Current and the Brazil/Falkland regions. In these areas, coherency analysis reveals that the cross-correlations peak at wavelengths of 400-600 km. The strength of the cross-correlations is found to be seasonal, peaking in the winter and minimising in summer.
Contents

List of Figures 11
List of Tables 17
Acronyms 18

1. Introduction 19

2. The Oceanography of the South Atlantic 24
   2.1 Introduction 24
   2.2 Overview of the South Atlantic circulation 25
   2.3 Agulhas 27
   2.4 Benguela 30
   2.5 South Equatorial Current 31
   2.6 Brazil Current 31
   2.7 South Atlantic Current 33
   2.8 Antarctic Circumpolar Current 33

3. Altimetry 36
   3.1 Introduction 36
   3.2 A brief history of altimetry 38
   3.3 Why are altimeters useful for studying ocean circulation? 41
   3.4 Obtaining an accurate range measurement 43
      3.4.1 How does an altimeter work? 43
      3.4.2 Correcting for the effects of the intervening medium 48
      3.4.3 Measuring the distance to the geostrophic surface 50
      3.4.4 Error budget 53
   3.5 Collocation of data 54
### Contents

3.5.1  Along-track collocation techniques 55  
3.5.2  Correcting for across-track variations 56  

3.6  Orbit error removal 65  
3.6.1  Different orbit error removal techniques 67  
3.6.2  Comparison of different techniques 69  

3.7  Interpolation of altimeter data 74  
3.7.1  Interpolation techniques 74  
3.7.2  Optimising interpolation parameters 77  
3.7.3  The error caused by interpolation 81  

3.8  Determining the large scale flow field 86  
3.8.1  Theory 88  
3.8.2  Observation 90  
3.8.3  Validation 95  

3.9  Conclusions 95  

4. Measuring Sea Surface Temperature from Infrared Radiometry 100  
4.1  Introduction 100  
4.2  A brief history of spaceborne infrared SST measurement 101  
4.3  Sources of error in the ATSR ASST data 103  
4.3.1  Instrument and data description 103  
4.3.2  Calibration 106  
4.3.3  Instrument noise 108  
4.3.4  Atmospheric correction 109  
4.3.5  Aerosol contamination 111  
4.3.6  Cloud contamination 113  
4.3.7  Surface effects 117  
4.3.8  Validation studies 118  

4.4  Regional cloud contamination of the ASST data 120  
4.4.1  Different day and night signals 120  
4.4.2  Quantification of the night cloud contamination 126
<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.3 Comparison with thermosalinograph SST data</td>
</tr>
<tr>
<td>4.4.4 Reducing remnant cloud contamination</td>
</tr>
<tr>
<td>4.4.5 Results of filtering</td>
</tr>
<tr>
<td>4.4.6 Cause of contamination</td>
</tr>
<tr>
<td>4.5 Discussion and conclusions</td>
</tr>
<tr>
<td>5. Correlations between SST and SSH</td>
</tr>
<tr>
<td>5.1 Why could SST and SSH be related?</td>
</tr>
<tr>
<td>5.2 Why would a relationship be useful?</td>
</tr>
<tr>
<td>5.2.1 Relating surface parameters to subsurface</td>
</tr>
<tr>
<td>5.2.2 Model validation</td>
</tr>
<tr>
<td>5.2.3 New techniques</td>
</tr>
<tr>
<td>5.3 Previous studies</td>
</tr>
<tr>
<td>5.4 Observations from TOPEX/POSEIDON and ATSR data</td>
</tr>
<tr>
<td>5.4.1 Data</td>
</tr>
<tr>
<td>5.4.2 SST overlaid on dynamic topography</td>
</tr>
<tr>
<td>5.4.3 Comparisons of ΔSST with ΔSSH</td>
</tr>
<tr>
<td>5.4.4 Geographical dependence of the ΔSST ΔSSH relationship</td>
</tr>
<tr>
<td>5.4.5 The seasonal cycle in the ΔSST ΔSSH correlations</td>
</tr>
<tr>
<td>5.4.6 The spatial scale of the ΔSST ΔSSH correlations</td>
</tr>
<tr>
<td>5.5 Conclusions</td>
</tr>
<tr>
<td>6. Conclusions and Future Work</td>
</tr>
<tr>
<td>6.1 Introduction</td>
</tr>
<tr>
<td>6.2 TOPEX/POSEIDON altimetry (Chapter 3)</td>
</tr>
<tr>
<td>6.2.1 Validation of across-track correction technique (Section 3.5)</td>
</tr>
<tr>
<td>6.2.2 Investigating the effects of interpolation (Section 3.7)</td>
</tr>
<tr>
<td>6.2.3 Obtaining a qualitative estimate of the absolute dynamic height</td>
</tr>
<tr>
<td>6.2.4 Altimetry summary</td>
</tr>
<tr>
<td>6.3 Along-Track Scanning Radiometer SST (Chapter 4)</td>
</tr>
<tr>
<td>6.3.1 Cloud contamination in the ATSR ASST data (Sections 4.4.1 - 4.4.3)</td>
</tr>
</tbody>
</table>
6.3.2 A new filtering scheme is applied and tested (Sections 4.4.4 - 4.4.5) 175

6.3.3 The cause of the cloud contamination is identified (Section 4.4.6) 176

6.3.4 Sea surface temperature summary 176

6.4 Correlations between SST and SSH (Chapter 5) 177

6.4.1 Qualitative evidence for a relationship between SSH and SST is discovered (Section 5.4.2) 177

6.4.2 The strength and geographical variation of the SST-SSH relationship is investigated (Section 5.4.3 - 5.4.4) 178

6.4.3 The seasonal variation in the strength of the correlations is investigated (Section 5.4.5) 178

6.4.4 The spatial dependence of the correlations is investigated (Section 5.4.6) 179

6.4.5 Potential uses of a SST-SSH relationship (Section 5.2) 179

6.5 Final remarks 180

References 181
List of Figures

Figure 1.1  Bar chart showing the number of hydrographic stations collected by the major oceanographic research expeditions.

Figure 1.2  The WOCE survey sections; red lines denote sections that have been completed, blue lines denote sections yet to be completed.

Figure 2.1  Schematic showing the upper level geostrophic circulation of the South Atlantic ocean (taken from Peterson and Stramma, 1991).

Figure 2.2  The bathymetry of the South Atlantic, with the names of the major features shown (taken from Peterson and Stramma, 1991).

Figure 2.3  Schematic depicting the circulation in the Agulhas Region (taken from Peterson and Stramma, 1991).

Figure 2.4  Schematic showing the circulation in the Brazil-Falkland Confluence Region (taken from Peterson and Stramma, 1991).

Figure 2.5  Schematic showing the banded structure of the Antarctic Circumpolar Current (taken from Peterson and Stramma, 1991).

Figure 3.1  Altimetry Schematic.

Figure 3.2  Schematic showing the area illuminated by an altimeter pulse as a function of time (taken from Chelton et al., 1989).

Figure 3.3  Schematic showing the difference between (a) beam limited and (b) pulse limited altimetry (taken from Chelton et al., 1989).

Figure 3.4  An idealised graph of the power received by the altimeter versus time (a "waveform") (taken from Chelton et al., 1989).

Figure 3.5  Collocation schematic. The data points along an arbitrary repeat are collocated to the reference grid points using a perpendicular bisector approach.

Figure 3.6  The mean and r.m.s. across-track distance (relative to cycle 18) for T/P cycles 1-52. The error bars are ±1 standard deviation of the across-track distance for each cycle.

Figure 3.7  The r.m.s. across-track correction together with the r.m.s. across-track distance for T/P cycles 1-52. The "spikes" correspond to Poseidon cycles.

Figure 3.8  The across-track correction for T/P cycle 24. The largest values are in regions of steep mean sea surface slope.
List of Figures

Figure 3.9 T/P SSH variability with across-track correction minus T/P SSH variability without across-track correction. The variability is calculated over T/P cycles 1-52.

Figure 3.10 Histogram of the T/P SSH variability with across-track correction minus T/P SSH variability without across-track correction. The bottom graph is an expanded version of the upper graph. The variability is calculated over T/P cycles 1-52.

Figure 3.11 An example of a height residual profile with and without the across-track correction. Note the region between 54°S and 56°S where an eddy-shaped feature is removed when the across-track correction is applied.

Figure 3.12 The geometry used to calculate the error in different collocation techniques.

Figure 3.13 The r.m.s. error inherent in several commonly used collocation techniques; collocation in latitude, in longitude, with perpendicular bisector method and with nearest point method.

Figure 3.14 Graph demonstrating the effect of collinear orbit error removal on a height residual profile from the Parallel Ocean Climate Model.

Figure 3.15 Graph showing the error in three different methods of orbit error removal as a function of track length. The error bars are ± 1 standard error.

Figure 3.16 The T/P SSH variability for 1993 and 1994 without any orbit error removal.

Figure 3.17 The T/P SSH variability for 1993 and 1994 with orbit error removal.

Figure 3.18 A height anomaly map from the Parallel Ocean Climate Model corresponding to T/P cycle 11.

Figure 3.19 The weighting functions and transfer functions for several smoothers. The dashed lines represent one sided weighting functions, as would occur near land and data boundaries (taken from Schlax and Chelton, 1992).

Figure 3.20 A schematic showing the difference between an ideal transfer function and a typical transfer function for irregularly spaced data (taken from Chelton and Schlax, 1992).

Figure 3.21 The r.m.s. error in Gaussian interpolation as a function of the Gaussian full-width at half-maximum.

Figure 3.22 The correlation between an interpolated field and the original field as a function of the Gaussian full-width at half-maximum.

Figure 3.23 The SSH variability from the Parallel Ocean Climate Model corresponding to 1993 and 1994.
Figure 3.24  The T/P SSH variability from the POCM re-interpolated SSH fields for 1993 and 1994.

Figure 3.25  The r.m.s. error in the Gaussian interpolation.

Figure 3.26  The percentage r.m.s. error in Gaussian interpolation.

Figure 3.27  The percentage r.m.s. error after adding in 5 cm r.m.s. orbit error and interpolating with optimal Gaussian parameters.

Figure 3.28  The percentage r.m.s. error after adding in 5 cm r.m.s. orbit error, applying a collinear orbit error removal method and interpolating with optimal Gaussian parameters.

Figure 3.29  Spectrum showing the JGM-2 geoid error together with the dynamic topography from T/P and an ocean model (taken from Nerem et al., 1994).

Figure 3.30  The T/P height anomaly map for January 1993.

Figure 3.31  The mean dynamic topography using the JGM3/OSU91A hybrid geoid.

Figure 3.32  The dynamic topography using the JGM3/OSU91A hybrid geoid and smoothed with a Gaussian smoother to eliminate wavelengths smaller than 2600 km.

Figure 3.33  The sum of the T/P smoothed mean dynamic topography and the height anomaly map for January 1993.

Figure 3.34  The dynamic topography in the Agulhas Region for January 1993, overlaid with geostrophic flow vectors.

Figure 3.35  The dynamic topography in the Drake Passage Region for January 1993, overlaid with geostrophic flow vectors.

Figure 3.36  The error in the method of obtaining absolute dynamic height.

Figure 3.37  The magnitude of the error in the method of obtaining absolute geostrophic velocities.

Figure 4.1  The response functions of the four ATSR channels (upper) together with the atmospheric transmission for three different atmospheres (lower) (taken from Závody et al., 1995).

Figure 4.2  The ATSR scan geometry, as depicted in Závody et al. (1995).

Figure 4.3  The sampling density of the ATSR ASST data calculated from the ATSR data from January 1992 - December 1994.

Figure 4.4  The atmospheric temperature deficit for several different atmospheres (taken from Deschamps and Phulpin, 1980).
Figure 4.5 ATSR mean ASST for the South Atlantic region from January 1992 - December 1994.

Figure 4.6 The variability (standard deviation) of the ATSR ASST data in the South Atlantic region from January 1992 - December 1994.

Figure 4.7 The SST variability (standard deviation) in the South Atlantic from Levitus (1984) data

Figure 4.8 The ATSR day ASST variability minus the ATSR night ASST variability.

Figure 4.9 The ATSR day mean ASST minus the ATSR night mean ASST.

Figure 4.10 (a) ATSR ASST time series for a location (27.25°S, 30.25°W) sampled only at night during the ERS-1 3 day repeat phase (Julian days 0-90). (b) ATSR ASST time series for a location (27.25°S, 25.75°W) sampled only during the day in the 3 day repeat phase.

Figure 4.11 International Satellite Cloud Climatology Project mean cloud cover percentage for 1988.

Figure 4.12 Difference between the day and night SST variability prior to the failure of the 3.7 μm channel on 26 May 1992. Negative values indicate that the night variability is greater than the day.

Figure 4.13 Histogram of ATSR ASST residuals from an annual and semi-annual model, separated into day and night data. The lower panel is an expanded view of the upper histogram, clearly demonstrating the negative tail in the night SST data.

Figure 4.14 Histograms showing the ATSR minus the TSG comparisons for (a) day and night ASST data, (b) only night ASST data and (c) only day ASST data.

Figure 4.15 The ATSR ASST variability after filtering the data.

Figure 4.16 The day ATSR ASST variability minus the night ATSR ASST variability after filtering the data.

Figure 4.17 The day ATSR mean ASST minus the night ATSR mean ASST after filtering.

Figure 4.18 Percentage of data rejected by the filtering scheme for (a) June, July, August, (b) September, October, November, (c) December, January, February, (d) March, April, May.

Figure 5.1 T/P "absolute" geostrophic flow vectors overlaid on the ATSR ASST for January 1993.
Figure 5.2  Examples of collocated ATSR ASST residuals and T/P SSH residuals for (a) T/P pass 174, cycle 11, (b) T/P pass 174, cycle 30, (c) T/P pass 174, cycle 49.

Figure 5.3  Time series showing the correlation between (i) SST residuals and SSH residuals, (ii) SST gradients and SSH gradients, (iii) SST gradients and SSH residuals and (iv) SST residuals and SSH gradients. The correlations are calculated for each cycle over the Agulhas Region.

Figure 5.4  The cross-correlation function between SST residuals and SSH residuals, calculated separately for ascending and descending tracks within the Agulhas Region.

Figure 5.5  This illustrates the technique of using a sliding zero-lag cross-correlation to quantify the regional dependence of the SST SSH relationship. (a) The SST residuals are overlaid on the SSH residuals for T/P pass 174, cycle 45. (b) Zero lag cross correlations are calculated within a sliding 250 km window.

Figure 5.6  The mean zero-lag cross correlation for 1993 and 1994. The correlations are calculated with a 250 km moving window with both day and night ASST data.

Figure 5.7  The standard deviation of the zero-lag cross correlations for 1993 and 1994. The correlations are calculated using a 250 km moving window with both day and night ASST data.

Figure 5.8  The mean gradient for 1993 and 1994 of the linear fit to ΔSSH as the independent variable with ΔSST as the dependent variable. The gradients are calculated using a 250 km moving window with both day and night ASST data.

Figure 5.9  Time series of the mean correlation over the South Atlantic Region for 1993 and 1994. Also shown is the fit of an annual model with a drift.

Figure 5.10  Mean correlation in the austral summer (January, February, March).

Figure 5.11  Mean correlation in the austral winter (July, August, September).

Figure 5.12  The wavelength dependence of the ΔSST ΔSSH correlations for the entire South Atlantic region (80°W-40°E; 70°S-10°N). Error bars correspond to ± one standard error.

Figure 5.13  The wavelength dependence of the ΔSST ΔSSH correlations for the Drake Passage region (75°W-50°W; 60°S-55°S). Error bars correspond to ± one standard error.
Figure 5.14  The wavelength dependence of the ΔSST ΔSSH correlations for the Antarctic Circumpolar Current region (32°W-10°E; 50°S-40°N). Error bars correspond to ± one standard error.

Figure 5.15  The wavelength dependence of the ΔSST ΔSSH correlations for the Agulhas Return Current region (21°E-40°E; 43°S-35°S). Error bars correspond to ± one standard error.
List of Tables

Table 1.1  Total number of events and deaths for each type of natural disaster 1967-1991 (WMO, 1995).

Table 3.1  Major aliasing periods and wavelengths corresponding to the six most dominant tidal constituents for T/P sampling (taken from Schlax and Chelton, 1994b).

Table 3.2  A summary of the T/P error sources.

Table 3.3  A summary of the optimal interpolation parameters.

Table 4.1  The algorithm noise for NEAT values spanning the ATSR mission lifetime.

Table 4.2  The intrinsic error in the ASST algorithms (taken from Závody et al., 1995).

Table 4.3  A summary of the peer reviewed ATSR validation studies to date. Only results for the RAL algorithms are shown.

Table 4.4  A summary of the percentage of data rejected by the filtering scheme.

Table 4.5  Seasonality of Marine Stratiform clouds (adapted from Klein and Hartmann 1993).
Acronyms

ACC Antarctic Circumpolar Current
ASST Averaged Sea Surface Temperature
ATSR Along Track Scanning Radiometer
AVHRR Advanced Very High Resolution Radiometer
BT Brightness Temperature
CASOTS Combined Action for Study of the Ocean Thermal Skin
DJF December, January, February
ECMWF European Centre for Medium range Weather Forecasting
EKE Eddy Kinetic Energy
EM Electromagnetic
FRAM Fine Resolution Antarctic Model
FWHM Full Width at Half Maximum
GDR Geophysical Data Record
ISCCP International Satellite Cloud Climatology Project
JGM Joint Gravity Model
JJA June, July, August
LSC Low Stratiform Cloud
MAM March, April, May
MSS Mean Sea Surface
MSSL Mullard Space Science Laboratory
PF Polar Front
r.m.s. Root Mean Square
POCM Parallel Ocean Climate Model
RAL Rutherford Appleton Laboratory
SAC South Atlantic Current
SADIST Synthesis of ATSR Data Into Surface Temperature
SAF Sub-Antarctic Front
SD Standard Deviation
SEC Southern Equatorial Current
SOC Southampton Oceanography Centre
SON September, October, November
SSB Sea State Bias
SSH Sea Surface Height
SST Sea Surface Temperature
STF Sub-Tropical Front
STG Sub-Tropical Gyre
SWH Significant Waveheight
T/P TOPEX/POSEIDON
TSG Thermosalinograph
VHRR Very High Resolution Radiometer
WCRP World Climate Research Programme
WMO World Meteorological Office
WOCE World Ocean Circulation Experiment
Chapter 1

Introduction

In the last twenty five years, more than four million people have been killed by natural disasters and more than a billion have been adversely affected in some way (WMO, 1995). Such natural disasters may be caused by events directly related to weather and climate such as droughts and hurricanes, or by events indirectly related to weather and climate such as famine and bush fires, or by geological events such as earthquakes and volcanic eruptions. Analysis of disaster statistics over the period 1967-1991 (WMO, 1995) shows that 63% of deaths are caused by events directly related to weather/climate. This figure rises to 84% if events indirectly related to weather and/or climate are also included (see Table 1.1).

Table 1.1  Total number of events and deaths for each type of natural disaster 1967-1991 (WMO, 1995).

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of Events</th>
<th>Number Killed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather Events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hurricanes, Typhoons</td>
<td>894</td>
<td>896,063</td>
</tr>
<tr>
<td>Flood</td>
<td>1358</td>
<td>304,870</td>
</tr>
<tr>
<td>Storm</td>
<td>819</td>
<td>54,500</td>
</tr>
<tr>
<td>Cold and Heat wave</td>
<td>133</td>
<td>4,926</td>
</tr>
<tr>
<td>Drought</td>
<td>430</td>
<td>1,333,728</td>
</tr>
<tr>
<td>Associated with weather</td>
<td></td>
<td></td>
</tr>
<tr>
<td>events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avalanche</td>
<td>29</td>
<td>1,237</td>
</tr>
<tr>
<td>Landslide</td>
<td>238</td>
<td>41,992</td>
</tr>
<tr>
<td>Fire</td>
<td>729</td>
<td>81,970</td>
</tr>
<tr>
<td>Famine</td>
<td>15</td>
<td>605,832</td>
</tr>
<tr>
<td>Food Shortage</td>
<td>22</td>
<td>252</td>
</tr>
<tr>
<td>Epidemic</td>
<td>291</td>
<td>124,338</td>
</tr>
<tr>
<td>Geological</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earthquake</td>
<td>758</td>
<td>646,307</td>
</tr>
<tr>
<td>Volcano</td>
<td>102</td>
<td>2,764</td>
</tr>
<tr>
<td>Tsunami</td>
<td>20</td>
<td>6390</td>
</tr>
</tbody>
</table>
As well as the tragedy of such deaths, the economic impact of natural disasters is huge. The global economic cost of such disasters rose from US$ 44 billion in 1991 to US$ 62 billion in 1992 (WMO, 1995). This cost has a major impact on the economic prosperity of affected countries and in turn on the welfare of their population. Recent examples of climatic related natural disasters include the 1991-1992 Southern African drought which was the most severe drought in this region for 50 years. Many South African countries experienced a seasonal deficit of up to 80% in their normal rain and temperatures in several regions reached 47°C. The combination of searing temperatures and little rainfall resulted in crop failures and the collapse of local industry due to water shortages and the failure of hydropower plants (WMO, 1995). A second recent example is that of Hurricane Andrew, which inflicted US$ 25 billion worth of damage on Florida in late August 1992. Over a million individuals were left without power, thousands lost their homes and 12 were killed (WMO, 1995).

It is clear from the above examples that changes in climate from the "norm" can have devastating effects on our lives and on our social, economic and ecological well being. In order to minimise the negative effects of such changes, it is necessary to predict the changes in advance. The accurate prediction of climate change is only possible if the mechanisms influencing climate change are understood. Unfortunately this is not yet possible. One of the biggest uncertainties in understanding climate change is the effect that the oceans have on the climate system. The oceans' influence can be appreciated by realising that the heat capacity of the top metre of the ocean is equivalent to that of the entire atmosphere. The ocean is therefore a tremendous heat sink, regulating the temperature of our climate over much longer time periods than the atmosphere. The most important factor in the anthropogenic forcing of climate change is carbon dioxide (Trenberth et al., 1995). The ocean contains approximately 50 times more carbon than the atmosphere (WMO, 1995) and it is known that the ocean acts as both a source of and a sink for carbon dioxide, although the exact location and action of sources and sinks is not well understood. A knowledge of the way in which the ocean transports heat and gases, and hence of ocean circulation, is therefore essential before climate change can be properly understood and predicted. This was the impetus behind the establishment of the World Ocean Circulation Experiment (WOCE) by the World Climate Research Programme (WCRP) as the major thrust of its research into decadal climate change. The fieldwork phase of WOCE started in 1991 and continues to 1997 when the analysis interpretation, modelling and synthesis (AIMS) phase begins. The magnitude of WOCE is put into perspective by comparing it to previous oceanographic experiments.

---

1 Decadal climate change is the third stream of the WCRP's research programme into climate change.
Chapter 1  Introduction

The first extensive survey of the world ocean was undertaken by the Challenger expedition in 1872. This survey took four years to complete and much was discovered about the chemical, biological and physical properties of the ocean. The data from this expedition provided an impetus for oceanographic research, both in terms of observation and theory. Since the Challenger expedition, there have been several major fieldwork campaigns to obtain more information about the ocean. These include expeditions on the Meteor (1925-1927), the Snellius (1929-1930), the Discovery (1924-1949), and the Eltanin (1962) research vessels, as well as the IGY (1957-1958) and the GEOSECS (1972-1976) campaigns. The comparative size of these campaigns is illustrated in Figure 1.1, where the number of stations\(^2\) occupied by each campaign is shown. It is clear that WOCE is by far the largest oceanographic experiment to date. The locations of the completed and proposed WOCE observations are shown in Figure 1.2.

Although an impressive experiment by previous oceanographic standards, how adequate is this sampling for mapping the ocean circulation? This question can be considered by some simple calculations. Assuming that the desired data resolution to resolve most of the physical oceanographic processes is 50 km (the WOCE standard for station spacing), and knowing that the surface area of the world's oceans is \(\sim 3.61 \times 10^8\) km\(^2\) (e.g. Gill, 1982), the length of a 50 km wide strip with the same area as the world's oceans is \(7.22 \times 10^6\) km\(^2\). If it is assumed that the maximum speed of a ship is \(\sim 22\) km/hour (equivalent to 12 knots), the total area of the ocean could be mapped at 50 km resolution in \(\sim 55\) years by one ship\(^3\). An instrument such as the Along-Track Scanning Radiometer (ATSR) mounted on the ERS-1 satellite can measure the global sea surface temperature (SST) of the ocean at 1 km resolution in a time period of \(\sim 10\) days\(^4\). Ocean variability occurs on timescales as short as several days (e.g. Brown, 1989) and hence a temporal data resolution of several days is necessary to adequately map the ocean circulation. Obtaining this temporal resolution is impossible on a global basis with conventional \textit{in situ} measurements. Therefore adequate sampling of the world's oceans requires spaceborne remote sensing. Of course, remote sensing can only directly provide measurements of ocean surface parameters whereas for most studies of ocean circulation information at depth is required. Therefore a combination of \textit{in situ} measurements with remote sensing is likely to be the most promising way forward.

Since remote sensing is such an important part of WOCE, it is necessary to use the data from remotely sensed instruments to their full advantage. This involves both developing

\(^2\) A station is a survey of the water-column at a particular location.

\(^3\) This estimate is for measurement of the surface characteristics of the ocean with a towed or hull mounted instrument. For measurements deeper than \(\sim 400\) m, the ship must stop to lower an instrument which would increase the amount of time required to complete the global survey.

\(^4\) In cloud free conditions, ATSR could provide global coverage in approximately 3 days (Mutlow et al., 1994).
Figure 1.1 Bar chart showing the number of hydrographic stations collected by the major oceanographic research expeditions.

Figure 1.2 The WOCE survey sections; red lines denote sections that have been completed, blue lines denote sections yet to be completed.
optimal techniques to utilise such data and combining different datasets in order to obtain the maximum amount of information about the ocean. Two parameters of large importance to ocean circulation, and measurable by satellite, are sea surface height (e.g. Chelton et al., 1990) (SSH) and SST (e.g. Olson et al., 1988).

In view of this, the aims of this thesis are:

(i) To study and improve techniques for examining ocean circulation using spaceborne altimetric measurements of SSH and satellite infrared radiometric records of SST.

(ii) To use the techniques developed in (i) to study the extent to which SST and SSH are related.

Emphasis is given to datasets from two experimental instruments: (1) sea surface temperature (SST) from the ATSR on board the ERS-1 satellite and (2) sea surface height (SSH) from the TOPEX/POSEIDON (T/P) altimetric satellite. Data from these instruments are analysed in the South Atlantic Ocean region. This is an area of major importance to WOCE (WCRP, 1988b) and containing a variety of dynamical regimes. A description of the upper level circulation of this region is given in Chapter 2. This is followed by a study of the T/P and ATSR datasets and techniques appropriate to their use in Chapters 3 and 4 respectively. These processed datasets are brought together in Chapter 5, where the relationship between SST and SSH is investigated. Finally, conclusions are drawn in Chapter 6.
Chapter 2

The Oceanography of the South Atlantic

2.1 Introduction

Since the focus of this thesis is on satellite techniques and their application to oceanography, it is necessary to choose a particular oceanographic region in order to illustrate these techniques. An area of particular interest to oceanographers is the Southern Ocean. An idea of the importance of the Southern Ocean can be gained by realising that 55-60% of the world's oceans owes its water mass characteristics to the Southern Ocean (WCRP, 1988a,b). The Southern Ocean is, in effect, a "pipeline" linking the Indian, Pacific and Atlantic Oceans. Thus, oceanic heat fluxes and water formation events in particular areas are transformed from being regional to global phenomena. The Southern Ocean is a region where heat, supplied to the ocean at low latitudes, is lost to the atmosphere. Despite the importance of this ocean, it is a remote and hostile place. Historical in situ observations are sparse in comparison to other oceans, and the lack of shipping routes means that even opportunistic measurements of the surface characteristics of the Southern Ocean are lacking. For these reasons, a campaign to study the Southern Ocean is the second core project (WCRP, 1988a,b) (out of only three core projects) in the World Ocean Circulation Experiment (WOCE). To obtain a representative picture of the Southern Ocean, WOCE core project 2 relies heavily on remote sensing as well as in situ observations (WCRP, 1988a).

The Southern Ocean is therefore a logical place to perform a remote sensing study. To reduce the region still further, it was decided to focus on the South Atlantic region of the Southern Ocean. From a UK perspective this seemed the most sensible region to study since several UK cruises were planned in this region to occupy the WOCE A11 and A23 sections. Hence data from these cruises could be used to provide "ground truth" data for the remote sensing. Sea surface temperature (SST) data from the A23 cruise is used to validate SST data from the Along-Track Scanning Radiometer (ATSR) in Chapter 4. To provide the context for the results discussed in Chapters 3, 4 and 5, a description of the state of our present knowledge of the South Atlantic Ocean follows.

---

1 At the outset of this project, computing resources were insufficient to tackle the global case; today this is possible.
2.2 Overview of the South Atlantic circulation

In this chapter the upper level circulation of the South Atlantic is described. This circulation is mainly wind driven geostrophic\(^2\) circulation. The deeper thermohaline circulation\(^3\), although interesting, is not described here. The reason is that this deep circulation cannot be resolved by remote sensing, and hence has no relevance to this thesis. The majority of the material in this chapter is taken from the excellent review paper by Peterson and Stramma (1991), although where appropriate recent results have also been included.

A schematic of the surface geostrophic circulation of the South Atlantic and the major bathymetric features are shown in Figures 2.1 and 2.2 respectively (taken from Peterson and Stramma, 1991). It is evident from Figure 2.1 that the features dominating the circulation are the Subtropical Gyre (STG) and the Antarctic Circumpolar Current (ACC). Starting from the eastern point of the STG, the Benguela Current is fed by water from the South Atlantic Current (SAC) and the Agulhas. The Benguela Current flows to the northwest where it becomes the South Equatorial Current (SEC). When the SEC reaches the coast of South America (at ~10°S) it bifurcates into the southward flowing Brazil Current and the northwestward flowing Northern Brazil Coastal Current. The Brazil Current flows south, down the coast of South America, gaining in volume transport, until it reaches ~36°S where it meets the cold northward flowing Falkland Current and leaves the continental shelf. The transition zone between the subtropical waters and the subantarctic waters is the Subtropical Front (STF) which extends across the South Atlantic. In most parts of the South Atlantic, this front drives the South Atlantic Current until the region near South Africa. The saline Indian Ocean water of the Agulhas then causes the South Atlantic Current to turn northwest into the Benguela Current, thus completing the STG.

The ACC is comprised of three main fronts: (1) the Subantarctic front (SAF), (2) the Polar Front (PF) and (3) the Continental Water Boundary. After passing through the Drake Passage, the SAF turns to the north and flows along the South American continental shelf until it meets the southward flowing Brazil Current. The SAF is termed the "Falkland Current" in this region. The SAF leaves the continental shelf at ~40°S, where it turns southward, and then eastward to flow across the South Atlantic. Unlike the SAF, the PF does not flow around the continental shelf after passing through the Drake Passage, but turns northward at 50°W before continuing eastward. The SAF

\(^2\) The geostrophic approximation is where steady state dynamics are assumed, and all the terms in the equation of motion for water are ignored apart from the pressure gradient and the coriolis force. For a fuller description see Section 3.3.

\(^3\) The thermohaline circulation is conventionally thought of as the deep circulation, caused by the formation of dense water masses at high latitudes which "spread" towards the tropics.
Figure 2.1  Schematic showing the upper level geostrophic circulation of the South Atlantic ocean (taken from Peterson and Stramma, 1991).

Figure 2.2  The bathymetry of the South Atlantic, with the names of the major features shown (taken from Peterson and Stramma, 1991).
and PF can be separated by 500 km in places, although in other locations they can merge together to form one intense front. The Continental Water Boundary is the front between the subantarctic waters and the Weddell Gyre.

One field that has dramatically benefited from satellite altimeter measurements is the study of Rossby waves. These planetary waves exist due to conservation of potential vorticity (e.g. Gill, 1982) and are of fundamental importance in ocean circulation. They are a mechanism by which forcing (e.g. wind stress) in one region of the ocean can influence an area of ocean tens of thousands of kilometers away. Rossby waves were observed in the South Atlantic by the studies of Forbes et al. (1993) and Le Traon and Minster (1993). Forbes et al. (1993) observed baroclinic Rossby waves with periods of 400-500 days and wavelengths of 250 km, propagating west-southwest from near the southern tip of South Africa. Le Traon and Minster (1993) found baroclinic Rossby waves with periods of 180 days and wavelengths of 500 km, propagating southwest from a region to the east of 5°E, between latitudes 25°S and 30°S. These observations, however, may well be due to aliased tidal model errors (Schlax and Chelton, 1994a) since the orbit configuration of Geosat results in an aliasing of several of the tidal constituents into Rossby wave-like signals. Hughes (1996), however, in a study of FRAM and TOPEX/POSEIDON data has shown that Rossby waves exist throughout the Southern Ocean. Moreover, he shows that these Rossby waves are advected by the mean flow and propagate eastward rather than westward in many regions.

A more detailed description of the different components of the surface geostrophic South Atlantic circulation is given in the following sections.

2.3 Agulhas

The Agulhas Current is the western boundary current of the southernmost of two cells that comprise the Indian Ocean Subtropical Gyre (Gordon et al., 1987). It flows southwestward along the continental shelf of South Africa and usually stays within 10-15 km of its mean position between the latitudes of 28°-34°S (Grundlingh, 1983). An exception to this stability is the occurrence of large meanders that propagate down the shelf; the "Natal Pulses" (Lutjeharms and Roberts, 1988; van Leeuwen and de Ruijter, 1996). These pulses can have spatial scales of ~200 km and propagate southwestward along the shelf at a speed of ~20 cm/s at 28°S, decreasing to ~5 cm/s at 34°S (Lutjeharms and Roberts, 1988).

The Agulhas Current separates from the coast at ~34°S and continues southwestward along the Agulhas Bank until 36°S. The volume transport of the Agulhas Current is
largest near the Agulhas Bank, its value reaching 95 Sv (1 Sv = 1x10^6 m^3 s^-1) relative to
the bottom (Gordon et al., 1987) compared to 71 Sv at 32°S (Beal and Bryden, 1996). The
largest current velocities observed within the Agulhas are ~250 cm/s (Pearce,
1977), although this is from direct observation and includes ageostrophic contributions.
The largest geostrophic current speeds observed are ~110 cm/s (Gordon et al., 1987).

At 36°S, the Agulhas Current leaves the continental shelf and retroflects to flow
eastwards back into the Indian Ocean (Figure 2.3). This retroflection usually occurs
between 16°-20°E (Lutjeharms and van Ballegooyen, 1988). The retroflection loop
includes a pool of Indian Ocean surface water with a temperature ~5°C warmer than
South Atlantic water at the same latitude (Gordon, 1985). As the Agulhas Return
Current flows eastward, it interacts with bottom topography in the vicinity of the
Agulhas Plateau. Rossby waves are thereby generated (Lutjeharms and van
Ballegooyen, 1988), resulting in the large meanders visible in both the sea surface
temperature and the dynamic topography shown in Figure 5.1. The Agulhas Current is
highly baroclinic. More than 80% of its transport occurs in the upper 1000 m of the
ocean (Peterson and Stramma, 1991). In the retroflection loop, waters with Indian ocean
properties are only present in the upper 1500-2000 m of the ocean (Gordon et al., 1987).
Transfers of Indian Ocean water to the South Atlantic can occur by two methods. Firstly
by the transfer of water by Agulhas eddies (Gordon and Haxby, 1990; Naeije et al.,
1992). In a study using Geosat data, Gordon and Haxby (1990) observe five eddies per
year, generated in the Agulhas retroflection region and propagating into the South
Atlantic with a translation speed of 5-8 cm/s. The eddies are typically 300 km in
diameter, have rim speeds of 20-80 cm/s, and have centres which stand 40-60 cm above
the surrounding sea surface. Gordon and Haxby (1990) calculate a value of 10-15 Sv for
the time averaged transport of Indian Ocean water (in the upper 1000 m) to the South
Atlantic by this means. Agulhas eddies can either propagate northwestward from the
Retroflection region along the northern limb of the STG, or they can enter the ACC.
Eddies that propagate northwestward have been observed as far west as 30°W (Gordon
and Haxby, 1990) and there is tentative evidence that these eddies can be recirculated in
the STG and can return to the Agulhas vicinity ~4 years later (Smythe-Wright et al.,
1996). The second way in which Indian Ocean water can be transported into the South
Atlantic is by a breakdown of the Agulhas Retroflection. Several modelling studies
indicate that if the volume transport of the Agulhas weakens sufficiently, it may not
retroflect, but rather could feed directly into the Benguela Current (Ou and de Ruijter,
1986; Boudra and Chassignet, 1988). Observational evidence indicates that this has
occurred on at least one occasion for several months in 1986 (Peterson and Stramma,
Chapter 2  The Oceanography of the South Atlantic

Figure 2.3  Schematic depicting the circulation in the Agulhas Region (taken from Peterson and Stramma, 1991).

Figure 2.4  Schematic showing the circulation in the Brazil-Falkland Confluence Region (taken from Peterson and Stramma, 1991).
The Agulhas region dominates the variability of the South Atlantic, both in terms of its eddy kinetic energy (EKE - the variance of the velocity) and its height variability (the standard deviation of the sea surface height (SSH)). EKE estimates range from 1000-4000 cm$^2$s$^{-2}$ (Patterson, 1985; Johnson, 1989; Morrow et al., 1992) and SSH variabilities of -40 cm have been measured in Geosat altimeter data (Wakker et al., 1990; Snaith, 1992; Quartly and Srokosz, 1993). Quartly and Srokosz (1993) show that the seasonal variation in the mesoscale variability associated with the Agulhas retroflection is non-existent in a primitive equation model (the Fine Resolution Antarctic Model - FRAM), but weakly existent in Geosat altimeter data. Quartly and Srokosz (1993) find that the variability associated with the retroflection is further west in the austral summer and further east in winter.

2.4 Benguela

The Benguela Current (Figure 2.1) is the Eastern Boundary current of the South Atlantic STG. It begins as a northward flow off the Cape of Good Hope (Stramma and Peterson, 1989) before bending towards the northwest and separating from the South African coast at about 30°S (Stramma and Peterson, 1989) (Figure 2.1). Once separated from the coast, the current rapidly widens to a diffuse northwestward flow. The Benguela Current is fed primarily by the SAC (Stramma and Peterson, 1990) (Figure 2.1), although Agulhas and Subantarctic surface water sometimes contribute (Peterson and Stramma, 1991). Unlike the Agulhas Current, the Benguela Current is fairly shallow, only extending to about 600 m depth (Reid, 1989).

Off the west coast of South Africa, the winds are from the south and southeast. This wind stress causes a westward surface Ekman flux$^4$, the result being that the surface waters are transported offshore and are replaced with much cooler nutrient rich waters. This upwelling occurs all along the southwestern coast of Africa, from Cape Point (34.3°S) to Cape Frio (18.4°S) and supports one of the world's richest fisheries. The strength of the upwelling is subject to large spatial and temporal variability (Jones, 1971). It occurs mainly along the southern portion of the coast during southern summer (when the sharpest fronts and most intense northward jets occur) and moves further north during winter (Jones, 1971). The jets associated with the upwelling are close to the coast (typically in 200-300 m of water) and can have peak velocities up to ~120 cm/s, widths of 20-30 km and volume transports of ~7 Sv (Bang and Andrews, 1974).

$^4$ The Ekman flux is caused by the transport of water in the surface (upper few tens of metres) ocean by the action of the wind. The Ekman transport is perpendicular to the direction of the wind, and is orientated to the right of the wind in the Northern hemisphere and to the left in the Southern hemisphere (e.g. Gill, 1982).
In the literature, as Peterson and Stramma (1991) point out, there is confusion in the naming of the Benguela Current; some authors refer to it as the large scale northward flow at the eastern side of the STG, while several studies refer to the northward coastal upwelling jets as the Benguela Current. In this study, the term Benguela Current is that adopted by Peterson and Stramma (1991) and refers to the eastern boundary current of the STG (Figure 2.1).

Volume transport estimates of the Benguela Current are few. Stramma and Peterson (1989) use historical data and find that at 32°S, the Benguela current is located near to the coast and has a northward transport of 21 Sv in the upper 600 m. Near 30°S, 18 Sv of the Benguela Current turns northwestward to flow over a deep section of the Walvis ridge (just south of the Valdivia bank). The remaining 3 Sv of flow do not turn northwestward but leave the Cape Basin to the north and enter the Angola Basin (Stramma and Peterson, 1989) (Figure 2.1).

2.5 South Equatorial Current

After passing over the Walvis Ridge, the Benguela Current feeds into what is usually termed the southern branch of the SEC (Figure 2.1). The equatorial current system is complex and is beyond the scope of this brief description chapter. The SEC is the main westward current in the South Atlantic (Fu, 1981). It makes up the northern part of the STG and generally lies to the south of 10°S, reaching this latitude only at the Brazilian coast (Fu, 1981) (Figure 2.1). Stramma et al. (1990) calculate the transport of this current across 30°W to be 16 Sv in the upper 500 m. When the current reaches 10°S, it bifurcates into the North Brazilian Coastal Current and the Brazil Current (Reverdin and McPhaden, 1986) (Figure 2.1). The North Brazilian Coastal Current is the stronger of these, with a volume transport of 12 Sv, compared to the one of 4 Sv for the Brazil current (Stramma et al., 1990). Little is known about the variability of the strength and bifurcation location of the SEC.

2.6 Brazil Current

As stated in the previous section, the southward flowing Brazil Current begins at ~10°S, with a transport of 4 Sv from the SEC (Figure 2.4). Between 10°S and 20°S, there is little evidence of increase in this transport (Stramma et al., 1990). Hence, compared to other western boundary currents, the Brazil Current is very weak. Transport estimates between 20°S and 25°S are always less than 11 Sv, with maximum velocities around 70 cm/s (Evans et al., 1983). This northern section of the Brazil Current is located in very shallow water, typically lying over the 200 m isobath (Evans and Signorini, 1985). As
the Brazil current flows southward along the continental shelf from 24°S, it grows in strength (Gordon and Greengrove, 1986). South of 30°S, this growth seems to be linked to a recirculation cell (Gordon and Greengrove, 1986; Olson et al., 1988) (Figure 2.4). The maximum transport values for the Brazil Current are 19-22 Sv relative to 1400-1500 m, at 38°S (Gordon and Greengrove, 1986). The Brazil Current then encounters the northward flowing Falkland Current and flows offshore (Figure 2.4).

The separation point of the Brazil Current from the continental shelf varies from 33°-38°S, with an average separation location of ~36°S (Olson et al., 1988) (Figure 2.4). Tentative evidence also exists that the separation location varies seasonally, with a more northerly separation point in the austral winter and a more southerly separation point in summer (Olson et al., 1988). It is postulated by Peterson and Stramma (1991) that this may be related to the movement in the large-scale atmospheric pressure fields. In winter, the peak of these pressure fields are further north than in summer. Furthermore, the northern line of zero wind stress curl lies further north, suggesting that the STG may shift to the north in winter. This may be responsible for the Brazil Current separation location being further north in winter.

After separating from the coast, the Brazil Current flows southward with the Falkland Current until ~43°S (Gordon, 1989) where it loops around to the east and then north (Figure 2.4). The southern limit to the warm water bounded by the Brazil Current is 38°S-46°S (Legeckis and Gordon, 1982). This loop of warm water is in some ways analogous to the warm water loop formed by the Agulhas retroflection. Like the Agulhas Retroflection, warm core eddies are shed from the Brazil Current loop (Legeckis and Gordon, 1982). These eddies are less energetic than the Agulhas eddies, typically being 150 km in diameter. They are generated at a rate of ~1 per week (Legeckis and Gordon, 1982), a rate more frequent than that of the Agulhas eddies.

The Brazil/Falkland confluence region is second only to the Agulhas in terms of its strength and variability. Provost and Le Traon (1993) in a study using Geosat data, show that in this region the Falkland Current has SSH variability levels of 8 cm, the Brazil Current has levels of 16 cm and the Confluence region has levels of 30 cm (c.f. 40 cm in Agulhas). The corresponding values for the EKE are 150, 800 and 1700 cm²/s respectively.

After looping around, the Brazil Current flows eastward into the interior of the South Atlantic, before recirculating back towards the South American coast (Figure 2.4).
2.7 South Atlantic Current

To complete the STG, attention is turned to the SAC. This is a current distinct from the ACC. Antarctic waters are separated from subtropical waters by the STF. No subtropical water exists in the Drake Passage and hence the STF must originate at the eastern coast of South America. As described in the previous section, the Brazil Current separates from the coast at ~36°S. The Falkland Current (the SAF) however, generally separates from the coast several degrees to the south of this location (Olson et al., 1988). The transition region between these two fronts (the SAF and the Brazil Current) is about 300 km wide and is filled with eddies (Olson et al., 1988). This occurs until a longitude of about 42°W, where the Brazil Current is thought to turn northward to recirculate (Stramma, 1989), leaving just the STF. Across most of the South Atlantic, the STF is coincident with the SAC. As the SAC nears the Agulhas Retractifaction region, the saline waters of the Indian Ocean form a density gradient that causes the SAC to turn northward and feed into the Benguela Current (Stramma and Peterson, 1990). The STF however, continues to the west as the demarcation between the subantarctic and the subtropical waters of the Indian Ocean. The surface speeds of the SAC are not large in comparison with the Brazil or Agulhas Currents. They are largest in the Argentine Basin where they reach values of ~20 cm/s (Stramma and Peterson, 1990). In the Cape Basin, these velocities decrease to ~10 cm/s (Stramma and Peterson, 1990). The same is true for volume transport; in the Argentine Basin, transport estimates are ~30 Sv (in the top 1000 m, relative to 3000 m), compared with ~15 Sv in the Cape Basin (Stramma and Peterson, 1990). The precise reason for this decrease in volume transport is unclear; the mid-Atlantic ridge may play some part, or it may be due to the flow gradually recirculating into the interior of the STG. As the SAC nears the Agulhas Current, the 15 Sv flows northwestward to feed into the Benguela Current, thus completing the STG.

2.8 Antarctic Circumpolar Current

The ACC is the major link between the world's oceans. Although of great importance in the global ocean circulation, its remote and inhospitable location means that relatively small amounts of data exist to describe this current. The ACC has been studied in greatest detail in the Drake Passage. In particular, results of the International Southern Ocean Studies (ISOS) programme have provided a large amount of information about the complicated structure of the ACC. Figure 2.5 shows the vertically averaged geostrophic speeds in the upper 2500 m relative to 2500 m (taken from Peterson and Stramma, 1991). A banded velocity structure is clearly evident. Three major fronts make up the ACC. From North to South these are (1) the SAF, (2) the PF and (3) the
Figure 2.5 Schematic showing the banded structure of the Antarctic Circumpolar Current (taken from Peterson and Stramma, 1991).
Continental Water Boundary. The vertically averaged speeds of these fronts is typically 12-17 cm/s (Peterson et al., 1982). Upper level geostrophic speeds of \(~30-45\) cm/s are present for the SAF and PF and of \(~15-30\) cm/s for the Continental Water Boundary (Nowlin and Clifford, 1982). Although the flow through the Drake Passage is predominantly eastward, recent results by Challenor et al. (1996) have shown that persistent westward flow is present in the Drake Passage at 57°S between the SAF and the PF\(^5\). The reason for this is unclear, a topographic effect being a possible cause. Although these fronts are only \(~40-60\) km wide each and make up only 20% of the width of the Drake passage, they account for about 75% of the geostrophic transport through the passage (Nowlin and Clifford, 1982). The geographical location of these fronts in the Drake passage is shown in Figure 2.5. Their location throughout the ACC changes dramatically with location. After passing through the Drake Passage, the SAF turns sharply northward around the Patagonian shelf (in this region the SAF is called the Falkland Current), until it reaches the Brazil-Falkland confluence at \(~40°S\). The Falkland Current then retroflects back to the south, before turning east into the Argentine Basin (Peterson and Whitworth, 1989). Unlike the SAF, the PF executes an "S" shaped turn as it passes between the Falkland Islands and South Georgia. In places, the PF and the SAF are separated by 500 km, whilst elsewhere they merge together in one focused jet. Quantifying the location and variability of these jets is an area of active research and is a part of WOCE core project 2.

The volume transport of the ACC was first estimated by Clowes (1933). He derived a volume transport figure of 110 Sv. Since then, estimates for the volume transport have ranged from over 200 Sv eastward to 1 Sv westward. The results are summarised in Peterson and Stramma (1991). Much of the disparity in the results probably stems from different choices of reference levels. The fact that the majority of the volume transport variability is barotropic (Whitworth and Peterson, 1985) also causes problems for conventional geostrophic estimates of the volume transport. Whitworth and Peterson (1985) use this fact to calculate a volume transport time series for January 1977-February 1980 and March 1981-March 1982. They obtain a mean transport of 123 Sv, with a range of 63 Sv.

More information about the circulation of the ACC and the South Atlantic as a whole will be emerging in the next few years as results from the observational phase of WOCE start to appear in the literature.

\(^{5}\) The Challenor et al. (1996) section is further to the east than the one shown in Figure 2.5 and hence 57°S is the location between the SAF and the PF in their study.
Chapter 3

Altimetry

3.1 Introduction

Of all remote sensing techniques, the one most likely to revolutionise our understanding of ocean circulation and its variability is satellite altimetry (Wunsch and Gaposchkin, 1980). Altimetry is conceptually simple, involving only the measurement of the distance between the satellite and the sea surface (Figure 3.1) by timing the return trip of an electromagnetic pulse. Given knowledge of the position of the altimeter, it is possible to measure the sea surface height (SSH). The SSH at any given location depends on the gravity field at that location as well as on the ocean dynamic topography, which is an integrated quantity depending on the full depth density structure of the ocean. Altimetry is therefore a technique which has the potential to provide an indirect measurement of the subsurface density structure of the ocean. (At present this potential is limited by our inadequate knowledge of the earth's gravitational field, and hence only the variability of the dynamic topography can be accurately measured). Unlike parameters measured by other remote sensing techniques, dynamic topography always depends on the subsurface ocean regardless of surface conditions, and therefore a measurement of dynamic topography provides information about the underlying density structure of the ocean.

In terms of climate change, global sea level trends can be indicators of anthropogenic induced changes in climatic parameters. For example, a rise in global ocean temperature would cause an increase in global sea level, due to thermal expansion of water. The latest figures available suggest that global sea level has increased by 10-25 cm over the last century, mainly in response to an ocean surface temperature increase of 0.3-0.6°C over the same time period (IPCC, 1996). Changes in regional sea level can have serious consequences for local inhabitants. Altimetry provides a means of detecting and monitoring such changes (e.g. Nerem, 1995; Minster et al., 1995).

In this chapter, the ability of altimeters to monitor ocean circulation is investigated. Data from the state-of-the-art TOPEX/POSEIDON (T/P) altimeter are used. This instrument has revolutionised the field of altimetry by making measurements of SSH to a root mean square (r.m.s.) accuracy of 5 cm (for a 1 second average; Fu et al., 1994),
Figure 3.1 Altimetry Schematic

Figure 3.2 Schematic showing the area illuminated by an altimeter pulse as a function of time (taken from Chelton et al., 1989).
nearly an order of magnitude better than that of previous altimeters. This accuracy is necessary to gain accurate velocity information on useful length-scales. An r.m.s. accuracy of 3 cm is required to achieve an r.m.s. geostrophic velocity accuracy of 11 cm/s over a distance of 50 km (see Section 3.4.4). In practice, many of the errors leading to the 5 cm r.m.s. accuracy are correlated and hence this stringent requirement is reduced. Nonetheless, accurate velocity calculation requires extremely accurate SSH measurements. T/P provides such accurate measurements. Conventional techniques for processing altimeter data need to be re-evaluated in the light of such an accuracy and new techniques should be developed to take full advantage of the increase in accuracy. This chapter studies several such techniques. The results gained from the research in this chapter guide the way in which the T/P altimeter data are used in the comparison with sea surface temperature data in Chapter 5. The results will also be of value to the oceanographic community wishing to obtain the most accurate processed dataset possible. The structure of the chapter is outlined below.

A brief history of altimetry is given in Section 3.2, followed in Section 3.3 with a description of the reason why altimetric measurements are useful for ocean circulation studies. Section 3.4 describes the method and accuracy of measuring the distance between the altimeter and the ocean surface. The original work in this chapter commences in Section 3.5.2 with the validation of a technique to correct for errors induced by across-track variations in the altimeter sampling. Until T/P, the largest error source in altimetric measurements was the determination of the position of the altimeter (the orbit error). Methods to remove this orbit error are described in Section 3.6, together with an investigation of the optimal technique to use given a specific track length. Altimeter measurements, although global, are anisotropic and irregular in space and time. Most conventional methods of analysing, displaying or assimilating data require regular gridded fields. The method of obtaining gridded fields from irregular data by interpolation is investigated in Section 3.7 through the use of output from the Parallel Ocean Climate Model (POCM). A method for determining the absolute circulation from altimetry alone is presented in Section 3.8. This method is validated using model data from the POCM. Finally, conclusions are drawn in Section 3.9.

3.2 A brief history of altimetry

Spaceborne altimetric measurements began with the Skylab S-193 altimeter, launched in 1973. To measure SSH relative to a reference ellipsoid, it is necessary to measure both the distance from the instrument to the sea surface (the altimeter "range") and the distance from the reference ellipsoid to the instrument (the "orbit height"). Errors in the orbit height (see Section 3.6) usually have very long wavelengths (~40000 km; Tai,
1989), and hence oceanographic information on smaller spatial scales is obtainable despite large orbit errors. The errors in the range measurement are at a variety of length-scales (see Section 3.4 for a full discussion) and hence accuracy of this measurement is the relevant parameter for determining whether or not useful oceanographic information can be obtained. Accuracies in the range measurement of better than ~0.5 m are required (Stewart et al., 1965) for any oceanographic research. The Skylab altimeter range accuracy was ~1 m. Hence, oceanographically useful results could not be obtained and the Skylab altimeter's most useful purpose from an oceanographic point of view was to provide information to aid the design of subsequent spaceborne altimeters.

In 1975, the GEOS-3 altimeter was launched. With a range accuracy of ~0.5 m, it was possible to measure oceanographic signals in areas of high variability such as the Gulf Stream (e.g. Huang and Leitao, 1978; Leitao et al., 1979; Gordon and Baker, 1980; Douglas and Cheney, 1981). Due to the lack of onboard data recorder, however, coverage was restricted to areas near ground receiving stations (mainly around the USA).

Seasat, launched in September 1978, with a range accuracy of ~10 cm (e.g. Bernstein et al., 1982; Byrne and Pullen, 1983), was the first altimeter to provide global information on the ocean variability (Cheney et al., 1983; Fu, 1983). Unfortunately SEASAT failed after only 100 days. Nonetheless it conclusively demonstrated that spaceborne altimeters can obtain useful information about the ocean variability.

In March 1985, the fourth altimetric satellite, Geosat, was launched. Until September 1986, Geosat was kept in a non-repeating orbit designed to measure the mean sea surface with high horizontal resolution. Data from this "Geodetic-Mission" phase were initially classified by the US Navy, but in September 1986 GEOSAT was manoeuvred into an orbit phase that repeated the same ground track pattern every 17 days. The data from this phase were unclassified and distributed to the oceanographic community. Geosat started to degrade in 1989, providing almost three years of unclassified global exact-repeat data. This long time span of accurate (range accuracy ~5-8 cm) altimetric data led to numerous scientific papers. New results were found on oceanic variability and its frequency-wavenumber spectra (e.g. Zlotnicki et al., 1989; Sandwell and Zhang, 1989; Nerem et al., 1990; Stammer and Boning, 1992; Quartly and Srokosz, 1993). Mesoscale eddies were detected and tracked (Gordon and Haxby, 1990; Naeije et al., 1992) and Rossby waves were observed for the first time (Tokmakian and Challenor, 1993; Le Traon and Minster, 1993; Forbes et al., 1993). Although an abundance of information was obtained on the ocean mesoscale variability (wavelengths ~100-1000

---

1 Errors in the tidal models available, coupled with the orbit characteristics of Geosat, led to a debate in the scientific literature regarding the validity of these measurements. The T/P altimeter with different orbit characteristics effectively resolved this debate and confirmed the observations of Rossby waves in the Geosat studies.
km, periods ~10-100 days), orbit error (~2 m for Geosat) limited the amount of information that could be obtained on the large scale variability (~ several 10000 km).

To address this issue, two altimetric satellites, ERS-1 and TOPEX/POSEIDON, were launched, in July 1991 and August 1992 respectively. Mounted on the ERS-1 satellite was a new system for precise orbit determination, PRARE. This was designed to yield orbit accuracies better than 10 cm. Unfortunately PRARE failed to function and was never used. As a result of this failure, ERS-1 orbits were initially only accurate to ~50 cm; not a great improvement over Geosat. However, improvements in gravity models led to improved orbit determination and ERS-1 orbits are now accurate to 6-8 cm (Remko Scharroo, pers. comm., 1996). ERS-1 was placed in a variety of orbit phases throughout its lifespan. It provided five years of data until June 1996 when it was shut down and replaced by ERS-2.

The joint US/French TOPEX/POSEIDON altimetric satellite was the result of over a decade of careful planning and design. Comprising of two altimeters, the US TOPEX dual-frequency altimeter and the French POSEIDON experimental solid-state altimeter, its objective was to measure the large scale global surface circulation of the ocean (TOPEX/POSEIDON Science Investigations Plan, 1991). To accomplish this, small amplitude signals (~5 cm) on basinwide length-scales (~ 10000 km) must be resolvable. Hence, both orbit and range accuracies had to be an order of magnitude better than previous altimetric missions. The pre-launch requirements for the T/P altimeter were an absolute accuracy in SSH measurement of 13 cm and a relative accuracy of 5 cm (TOPEX/POSEIDON Science Investigations Plan, 1991). This was achieved by implementing several design improvements. Firstly, the satellite was placed in a relatively high orbit (~1300 km compared to ~800 km for ERS-1 and Geosat) to reduce the atmospheric drag and enable more accurate orbit modelling. Secondly, DORIS and GPS systems onboard the satellite resulted in more accurate orbit tracking. Thirdly, the range measurement was improved by on-board measurement of the effects of the intervening ionosphere and the atmospheric water vapour, both of which reduce the speed of the altimeter signal. TOPEX/POSEIDON has exceeded all its pre-launch requirements and can measure SSH to an unprecedented r.m.s. accuracy of 5 cm (for a full discussion of the error budget see Section 3.4). This accuracy, coupled with a choice of orbit that minimises the potential for aliasing tidal signals into geophysical periods, has resulted in an abundance of new research on many aspects of the ocean variability and climate change (for a comprehensive selection of such research the reader is referred to the T/P JGR 1996 special issue (vol 100, issue C12)).
In April 1995, ERS-2 was launched, with essentially the same instruments as ERS-1. The PRARE system on ERS-2 operated correctly, however, resulting in accurate orbits (~5 cm) from the outset of the mission.

At present, two altimetric satellites, ERS-2 and TOPEX/POSEIDON, are providing altimeter data on a regular basis. Three altimeters are scheduled for launch in the next few years; the US Navy Geosat Follow-On (GFO) mission, the T/P Follow-On (JASON) and a radar altimeter on the Envisat platform. Hence, altimetric measurements will be available for at least the next decade and research into the best techniques for exploiting these data is vital.

3.3 Why are altimeters useful for studying ocean circulation?

The equation of motion for water is:

\[
\frac{dV}{dt} = -\frac{1}{\rho} \nabla p - 2\Omega \times V + g + F \tag{3.1}
\]

where \( V \) is the velocity vector, \( \rho \) is the density, \( p \) is the pressure, \( g \) is the gravitational force and \( F \) represents the effects of tides and frictional forces. \( \Omega \) is the earth's angular velocity vector. Resolving [3.1] into eastward, northward and upward components (x, y, and z directions) with velocities \( u, v, w \) respectively leads to:

\[
\begin{align*}
(x) \quad \frac{du}{dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + 2\Omega v \sin \phi - 2\Omega w \cos \phi + F_x \\
(y) \quad \frac{dv}{dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial y} - 2\Omega u \sin \phi + F_y \tag{3.2} \\
(z) \quad \frac{dw}{dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} + 2\Omega u \cos \phi - g + F_z
\end{align*}
\]

where \( \phi \) is the latitude. It can be shown by scaling arguments that, for most dynamical regimes in the open ocean\(^2\), the dominant terms in [3.2] are \( g \), the terms involving pressure and the terms involving \( \sin \phi \). (e.g. Pond and Pickard, 1983). The equation for the z-component then reduces to the hydrostatic equation:

\[
\frac{dp}{dz} = -\rho g dz \tag{3.3}
\]

\(^2\) Near the ocean margins and the ocean floor, frictional effects play an important role and therefore geostrophy may not be a good approximation. Furthermore, near the ocean surface (top few tens of metres), wind stress induces Ekman velocities and hence the geostrophic approximation will also be poor here. The geostrophic approximation is only valid for spatial scales similar to or larger than the Rossby radius of deformation (~10-50 km at mid-latitudes).
The equations for the x and y components are:

\[
\begin{align*}
(x) \quad \frac{\partial p}{\partial x} &= \rho f v \\
(y) \quad \frac{\partial p}{\partial y} &= -\rho f u
\end{align*}
\]  

where \( f \) is the Coriolis parameter (\( 2\Omega \sin \phi \)).

A fluid satisfying [3.4] is said to be in geostrophic balance. If a fluid is in geostrophic balance, [3.4] imply that by measuring the horizontal pressure gradient at a particular level, the fluid's velocity at that level may be derived. Hence, if the three dimensional pressure structure of the ocean could be measured, the full geostrophic flow field could be derived. In practice, it is impossible to measure pressure accurately enough to do this. Hence, the density structure of the ocean is derived by measuring temperature and salinity instead and then by converting to density using the equation of state for seawater. It can be shown that a vertical change in the horizontal density gradient leads to a vertical change in the horizontal velocity, therefore the velocity shear can be derived from density alone. To obtain absolute measurements from density however, the geostrophic velocity at some reference level must be known. The assumption is usually made that at some level the geostrophic velocity is zero. The resulting geostrophic velocity profile is then calculated by integrating the velocity shear downwards and upwards from this level of no motion. The choice of the level of no motion is probably the largest single problem in conventional hydrography today. The use of absolute velocities from acoustic doppler current profiler (ADCP) instruments together with inverse modelling is a potential method of overcoming this problem (e.g. Bacon, 1994; Saunders and King, 1995).

Defining \( \zeta \) as the height of the sea surface relative to the geoid (which is the equipotential surface that the ocean surface would describe in the absence of ocean currents - Figure 3.1) and using the hydrostatic equation [3.3], leads to:

\[
\begin{align*}
(x) \quad \frac{d\zeta}{dx} &= -\frac{fv}{g} \\
(y) \quad \frac{d\zeta}{dy} &= \frac{fu}{g}
\end{align*}
\]  

[3.5]

Hence the horizontal sea surface height gradient relative to the geoid gives the surface geostrophic velocity. Unfortunately the geoid height is not well known on small spatial scales (<2000 km), and thus absolute geostrophic velocities cannot usually be determined on small scales. This is the largest single problem in oceanographic
altimetry today. Until a satellite mission is flown to map the surface gravity field of the earth to an accuracy of several cm on scales as short as 50 km, altimetry is not realising its full potential. A method of obtaining the most information about the absolute geostrophic circulation possible from altimetry alone is described in Section 3.8, although this is shown to be error prone in some regions. The majority of oceanographic altimetric studies use the fact that the geoid is to first order time independent. Therefore relative geostrophic currents may be calculated from variations in the sea surface height gradient from the mean gradient. Hence altimetry may be used to study variations in the surface geostrophic ocean circulation.

Attention is now turned to the means by which a radar altimeter orbiting at a height of ~1000 km can measure the distance between itself and the sea surface to an accuracy of several cm.

3.4 Obtaining an accurate range measurement

In order to obtain any useful oceanographic information, the distance between the altimeter and the ocean surface (the range measurement - see Figure 3.1) must be measured to a precision of better than 50 cm. T/P measures this distance with a precision of ~3 cm (Fu et al., 1994). Achieving such a precision requires careful instrument design as well as correction for the effects of the intervening atmosphere. These are described in Sections 3.4.1 and 3.4.2. As described in Section 3.3, a surface of constant pressure is required for studies of ocean circulation. Variations in atmospheric pressure cause the sea surface to respond as an inverse barometer. Hence this effect must be removed if geostrophic currents are to be measured accurately. Furthermore, the effects of tides on the sea surface must be removed, since these are not in geostrophic balance. Although not strictly corrections to the range measurement, it is appropriate to discuss these effects in the context of the overall error budget of the altimetric measurement of SSH. These effects are therefore reviewed in Section 3.4.3.

3.4.1 How does an altimeter work?

Assuming that there is no intervening medium between the altimeter and the sea surface, the process by which an altimeter measures the range distance is conceptually very simple. The two-way travel time of a pulse of electromagnetic radiation is measured and, knowing the speed of light, the distance can be calculated. In practice, the process by which this is achieved is very complicated and a detailed description of the intricacies of altimeter design and operation is outside the scope of this thesis. For an excellent description of the fundamentals of range measurement, the reader is
referred to Chelton et al. (1989). For a description of the TOPEX altimeter, Hayne et al. (1994) and Rodriguez and Martin (1994) should be consulted. The aim of this section is to provide the basis, not the detail, of range measurement.

Before considering the method of range measurement, the desired size of the spatial average (the "footprint" size) must be decided. The footprint size must be large enough to average out the effects of surface gravity waves (several 100 m) and small enough to resolve geostrophically balanced ocean currents. The parameter which governs the length-scale at which currents can be in geostrophic balance is the internal Rossby Radius (e.g. Gill, 1982). Typical values at high latitudes are ~10 km (e.g. Houry et al., 1987) and this is therefore a desirable altimeter footprint size (larger footprints will smooth out geostrophic scales).

Two methods of limiting the footprint size are available, pulse-limited and beamwidth-limited altimetry. In beamwidth-limited altimetry, the electromagnetic pulse is focused to a narrow beam by the altimeter antenna. The smaller the desired beamwidth, the larger the antenna required. For example at an altitude of 1000 km, a 5 km diameter footprint could be achieved with an antenna beamwidth of 0.3°. This requires an antenna diameter of ~5 m for a 13.6 GHz signal. A further consequence of beamwidth-limited altimetry is that mispointing errors can cause large range errors. Assuming a flat surface, a mispointing error of 0.05° would correspond to a height error of 25 cm for a satellite at 1000 km altitude. Such a mispointing error is to be expected and an error of 25 cm is unacceptable. Hence, beamwidth-limited altimetry is both unfeasible because of the sensitivity of the height error to mispointing errors and impractical because of the antenna size required to achieve the desired spatial resolution.

A footprint diameter of several km can be achieved by allowing a fairly large beamwidth (~1-2°) but a very short pulse duration (~ few ns). The pulse expands spherically from the altimeter. In the case of a flat surface, the area illuminated by the pulse appears as an expanding circle followed by an expanding annulus (Figure 3.2) of constant surface area. The footprint size at which the circle becomes an annulus is controlled by the duration of the pulse, and hence this form of altimetry is known as "pulse-limited". Since the pulse is an expanding sphere, rather than a narrow beam (Figure 3.3), this geometry is far less sensitive to mispointing errors than beamwidth-limited geometry. This is because, providing the mispointing angle (γ) does not exceed the beamwidth half-angle (φ) (Figure 3.3), a pulse return is always obtained from the altimeter nadir-point. Furthermore, since a large beamwidth is allowed, a smaller antenna can be used. (A beamwidth of 2° requires an antenna diameter smaller than 1 m for a pulse at 13.6 GHz from a satellite at 1000 km.) For these reasons, pulse-limited altimetry has always been used on the altimeters described in Section 3.2.
Figure 3.3  Schematic showing the difference between (a) beam limited and (b) pulse limited altimetry (taken from Chelton et al., 1989).

Figure 3.4  An idealised graph of the power received by the altimeter versus time (a "waveform") (taken from Chelton et al., 1989).
The footprint diameter achieved for a satellite at an altitude of 1300 km (the T/P altitude) with a pulse duration of 3.125 ns is 2 km for a flat surface (no waves). The effect of waves is to increase the effective footprint diameter because the point at which the illuminated circle becomes an annulus is the point at which the trailing edge of the pulse coincides with the wave troughs. A significant wave height (SWH)\(^3\) of 5 m increases the effective footprint diameter to 6.9 km, while a SWH of 10 m increases the effective footprint diameter to 9.6 km. Typical SWH values range from 2 m in the tropics to 5 m in high latitude winters (Challenor et al., 1990).

An idealised graph of the power of the return pulse (which is proportional to the surface area illuminated by the pulse) versus the time of reception (the "waveform") is shown in Figure 3.4. The two-way time corresponding to mean sea level\(^4\) (for a Gaussian waveheight distribution) is defined by the half power point on the waveform, whilst the slope of the leading edge provides a measure of the SWH (a steeper slope implies a smaller SWH). Hence, the distance to the mean sea surface is calculated by multiplying half of the two-way time by the speed of light in a vacuum (the effects of the intervening medium will be considered in Section 3.4.2). In practice, it is difficult to generate a pulse duration as small as 3.125 ns with enough power to ensure a sufficient signal to noise ratio. To circumvent this problem, a relatively long pulse (~ \(\mu s\)) is used, together with a pulse compression technique, in order to obtain information equivalent to that of a short pulse. Describing this technique is complex and the reader is referred to Chelton et al. (1989) for a detailed (30 page) explanation.

The above discussion has assumed that the wave heights have a Gaussian distribution and that the mean scattering surface is coincident with mean sea level. If this is not true, the half power point of the waveform does not correspond to mean sea level and a correction is necessary.

### Sea State Bias correction

The error in the determination of mean sea level due to sea state consists of three components (1) the electromagnetic (EM) bias, (2) the skewness bias and (3) the tracker bias.

The EM bias is due to the mean scattering surface not corresponding to mean sea level. Surface gravity waves are not sinusoidal, but trochoidal (sharper crests and longer

---

\(^3\) Defined as four times the SD of the surface elevation; approximately equal to the mean crest to trough height of the 1/3 largest waves in the footprint.

\(^4\) In this section, mean sea level refers to the spatial mean over the altimeter footprint. In general throughout the rest of this thesis, mean sea level refers to the time mean sea level over an area the size of or larger than the altimeter footprint.
troughs). The power backscattered from a wave facet to an altimeter is proportional to the long-wavelength local radius of curvature, and hence more power is reflected from the wave troughs than from the crests. This biases the backscattered power towards the wave troughs. A second effect is due to enhanced small scale roughness on the wave crests compared to the wave troughs due to the troughs being more sheltered from the wind. This small scale roughness scatters radiation away from the incidence angle and again results in the backscattered power being biased towards the wave troughs. Typical values of the EM bias effect are 1-4% of SWH (Gaspar et al., 1994).

The skewness bias is due to the mean scattering surface not coinciding with the median scattering surface. The half power point on an altimeter waveform corresponds to the two-way return time to the median scattering surface rather than the mean scattering surface\(^5\). The difference between the mean scattering surface and the median scattering surface is called the skewness because it is related to the skewness of the sea surface height distribution. The skewness bias is generally smaller than the EM bias. It is possible to have an EM bias effect without a skewness bias effect. If, for example, the sea surface height distribution is perfectly Gaussian, a finite EM bias effect is still possible due to increased roughness on the wave peaks relative to the troughs caused by the wind. The tracker bias is due to the error in the tracker's determination of the median height.

Initial attempts to compensate for the effects of Sea State Bias (SSB) used a proportion of the SWH as the correcting factor. The most recent results from an analysis of T/P data (Gaspar et al., 1994) suggest that a four parameter model is more appropriate:

\[
\text{SSB} = \text{SWH}[a_1 + a_2 \text{SWH} + a_3 U + a_4 U^2]
\]  

[3.6]

where \(U\) is the 10 m windspeed (measurable from the altimeter; e.g. Chelton and Wentz, 1986), and \(a_1, \ldots, a_4\) are coefficients determined by a linear least squares regression that minimises the variance of global T/P sea surface height data (Gaspar et al., 1994). The magnitudes of the residual errors after applying this correction are poorly known but are of the order of 1% of the SWH (Rodriguez and Martin, 1994), with wavelengths of 500-1000 km.

The above is a description of how the distance between a spaceborne altimeter and the mean sea surface height is measured, assuming no intervening medium. In reality of course, an atmosphere is present and this retards the speed of a radar pulse. The ionosphere also has an effect on the speed of the electromagnetic radiation. Hence

\(^5\) In the discussion on the way in which an altimeter works, it is assumed that the sea surface height distribution is Gaussian; hence the mean and median are identical.
corrections are needed to compensate for these effects. These corrections are described in the next section.

3.4.2 Correcting for the effects of the intervening medium

(i) Ionospheric correction

Free electrons in the ionosphere interact with the altimeter radar pulse, resulting in a reduction in its propagation speed. This effect is frequency dependent and the way in which it is measured for T/P varies according to whether TOPEX or POSEIDON is operational.

TOPEX is a dual frequency altimeter, making measurements at 13.6 GHz and at 5.3 GHz. Since the ionospheric delay is inversely proportional to the square of the frequency of the pulse, the difference in the range values obtained at the two different frequencies can be used to derive the ionospheric correction. The errors in this approach are due to the noise in the altimeter measurements and to the frequency dependent sea state bias errors. The noise can be reduced by averaging the correction over a spatial scale of ∼100 km (ionospheric variations are over larger spatial scales), but the sea state bias effects cannot be reduced in this way since they are on larger spatial scales. The overall error in the ionospheric correction is 0.2 cm (after smoothing) due to the altimeter noise, 0.2 cm due to EM bias errors, and 0.45 cm due to skewness errors. This leads to a total error of 0.5 cm (Fu et al., 1994). The spatial scale of the errors in this measurement (after smoothing) are of the order of 1000-40000 km (Imel, 1994).

POSEIDON is a single frequency altimeter and hence the ionospheric correction cannot be derived in the same way as for TOPEX. The dual frequency measurements from DORIS are interpolated to provide the ionospheric correction for POSEIDON. By comparisons with the corrections from the TOPEX dual frequency measurements, the accuracy of this correction is found to be ∼1.7 cm (Fu et al., 1994). The wavelengths of the error are the same as for the TOPEX correction.

(ii) Dry tropospheric correction

The dry tropospheric correction is the largest of all the corrections that need to be applied to obtain an accurate range measurement. The mass of air in the atmosphere reduces the speed of light, and hence a correction is necessary to compensate for this. The correction is well modelled by equation [3.7] which requires knowledge of the surface air pressure ($P_{\text{atm}}$) and the latitude ($\theta$) of the measurement.
\[ \text{Dry} = 2.277 \, \text{p}_\text{atm} \, (1 + 0.0026 \cos 2 \theta) \]  

where \( \text{p}_\text{atm} \) is in mbars, and \( \text{Dry} \) is in mm.

Although the magnitude of this correction is large (~2 m), its variability is small (~3 cm) and it varies over large temporal and spatial scales. The surface air pressure cannot be measured by the altimeter and hence some other source of such data is required. For T/P, surface air pressure information is from the ECMWF analysis which is a combined observational and model output. The r.m.s. accuracy of this correction is ~0.7 cm. This is based on an accuracy in the atmospheric pressure product of 3 mbar r.m.s. and is likely to be an underestimate in Southern Hemisphere regions where observational measurements are scant. The wavelengths of the errors in the correction are the same as the wavelengths of atmospheric pressure variations (~500-2000 km).

(iii) **Wet tropospheric correction**

The radar pulse is also delayed by the effects of water vapour in the atmosphere. This can be corrected for given measurements of the total column water vapour content. Although the absolute magnitude of this correction (0-40 cm) is less than the dry correction, its variability is greater and it varies on shorter length and time scales. On T/P, the most accurate water vapour correction is provided by a microwave radiometer. Total column water vapour is derived by measuring brightness temperatures at 18, 21 and 37 GHz. Stum (1994) shows that this is better than the ECMWF analysis correction. The wet tropospheric correction has an r.m.s. accuracy of 1.1 cm (Ruf et al., 1994) with error wavelengths of 100-1000 km (Stum, 1994).

(iv) **The effects of rain and clouds**

Unlike water vapour, the largest effect of liquid water in the atmosphere is not refraction of the radar pulse, but rather inhomogeneous attenuation of the pulse. An attenuation that is constant over the altimeter footprint will affect all parts of the waveform equally. Hence the waveform shape will not be effected. In the case of clouds and rain, however, variations in liquid water content can occur on small spatial scales (compared to the altimeter footprint size) and hence unequal attenuation can occur. This attenuation affects the shape of the return waveform and can affect the altimeter tracker performance. The effect of clouds is largest for parallel cloud streets oriented orthogonal to the satellite ground track (Walsh et al., 1984), and is ~2 cm. The effects of rain can be much larger than the effects of clouds, depending on the spatial scale and structure of the rain cells and on the rain rate. Errors due to rain can be as large as
several tens of cm over along-track distances of 5-10 km (Monaldo et al., 1986). At present, the effects of rain are not understood well enough to be corrected for. Hence the method of avoiding rain contamination in T/P altimeter data is to flag data as rain contaminated, based on rejection criteria applied to the Topex microwave radiometer measurements. Data are rejected if the liquid water path content is greater than 1 mm, or if the 37 GHz brightness temperature is greater than 250 K (AVISO, 1992). The extent to which these rejection criteria are valid and the effects of rain on altimetry are topics of current research (e.g. Guymer et al., 1995; Quartly et al., 1996).

3.4.3 Measuring the distance to the geostrophic sea surface

Sections 3.4.1 and 3.4.2 describe how the distance from a spaceborne altimeter to the sea surface is measured. Most of the circulation in the global oceans is near to geostrophic balance (Section 3.3). Hence for these regions, measuring the horizontal pressure gradient (relative to an equipotential surface) at a particular level in the ocean yields the current velocity at that location. The pressure along an equipotential surface an infinitesimally small distance below the sea surface is the sum of the pressure due to the sea surface height (equation [3.3]) and the atmospheric pressure. In order to remove the effects of atmospheric pressure, so that the pressure along the equipotential surface is directly related to the sea surface height, the inverse barometer correction is used.

It is only possible to calculate geostrophic velocities from pressure gradients if the ocean is in geostrophic balance. In most regions, the largest non-geostrophic effect is that of ocean tides and hence this must be removed before attempting to infer geostrophic velocities from sea surface height. The corrections which reduce the measured sea surface height to a geostrophic sea surface are described below.

(i) Inverse barometer correction

A first order approximation of the way in which the ocean surface responds to atmospheric pressure can be made by assuming that the sea surface responds as a simple inverse barometer, with a 1 mbar increase (decrease) in surface pressure corresponding to a 1 cm decrease (increase) in the surface height (i.e. -1 cm/mbar). A study by vanDam and Wahr (1993) however, sheds doubt on this relationship. Using Geosat data, they obtain a global figure of -0.6 to -0.7 cm/mbar. This seems to suggest that the ocean is not responding as a simple inverse barometer. In their study using T/P data, Fu and Pihos (1994) obtain a figure of -0.96 cm/mbar and show that this simple relationship is generally valid at periods from 20 to 300 days (the periods resolvable in their study). They attribute the discrepant results of vanDam and Wahr (1993) to
inaccuracies (due to orbit error and poor ionospheric and wet tropospheric corrections) in the Geosat data. Hence throughout this study the standard inverse barometer correction of -1 cm/mbar is applied. Assuming that this relationship is valid, the wavelengths of the errors in this correction will be identical to that of the dry tropospheric correction. However, although the magnitude of the correction is less, the variability of the correction is approximately four times greater than that of the dry correction. Hence, assuming a pressure error of 3 mbar r.m.s. leads to a height error of 3 cm r.m.s..

(ii) Ocean tide correction

To remove the non-geostrophic effects of ocean tides, a tide model is necessary. Shortly after the launch of T/P, errors in tide models (~6 cm r.m.s.; Molines et al., 1994) were the largest source of error in the T/P SSH measurements. Reducing this error source gave the impetus to a large amount of research that produced no less than 12 new tidal models (Anderson et al., 1995). These models are all an improvement over the previous models and, in a comparison of these models by the T/P science working team, one was selected to be the chosen tidal model for the T/P GDR production. This model is the University of Texas CSR 3.0 tidal model (Eanes and Bettadpur, 1995), and is the one used for all the T/P data throughout this study. The global r.m.s. error of the CSR 2.0 tidal model derived from a comparison to tide gauges is 2.9 cm r.m.s. (Anderson et al., 1995). The error in the CSR 3.0 tidal model is certainly less than this (T/P Science Working Team, pers. comm.), although no results have yet been published.

Errors in tidal models, together with the non-random altimeter sampling can cause the aliasing of tidal errors. The periods and wavelengths into which the tidal errors are aliased, depends on the sampling (and therefore the orbit) of the altimetric satellite. The orbit of the T/P satellite was chosen very carefully to minimise the aliasing of tidal signals into geophysical wavelengths and periods, as can be seen from Table 3.1.
Table 3.1  Major aliasing periods and wavelengths corresponding to the six most dominant tidal constituents for T/P sampling (taken from Schlax and Chelton, 1994b). The direction of propagation of the aliased signals is denoted by E (East) or W (West).

<table>
<thead>
<tr>
<th>Tide</th>
<th>Tidal Period (hours)</th>
<th>Alias Period (days)</th>
<th>Alias Wavelength (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_2$</td>
<td>12.42</td>
<td>62.11</td>
<td>9.00 E</td>
</tr>
<tr>
<td>$S_2$</td>
<td>12.00</td>
<td>58.74</td>
<td>179.95 W</td>
</tr>
<tr>
<td>$N_2$</td>
<td>12.66</td>
<td>49.53</td>
<td>9.00 W</td>
</tr>
<tr>
<td>$K_1$</td>
<td>23.93</td>
<td>173.19</td>
<td>359.9 W</td>
</tr>
<tr>
<td>$O_1$</td>
<td>25.82</td>
<td>45.71</td>
<td>9.23 E</td>
</tr>
<tr>
<td>$P_1$</td>
<td>24.07</td>
<td>88.89</td>
<td>359.90 W</td>
</tr>
</tbody>
</table>

It can be seen that the only tidal constituent to alias into a geophysical period is the $K_1$ constituent which aliases into a period of 173 days, close to that of the semiannual period. (This is unlike Geosat and ERS-1 where several of the tidal constituents alias into geophysical periods).

Errors in tidal models have led to a debate in the literature over the detection of Rossby waves using satellite altimetry. These westward propagating (wavelength ~500 km, period ~200 days) waves are a means by which the ocean passes information from one location to another and can influence ocean circulation and climate. Tokmakian and Challenor (1993) discover westward propagating signals in Geosat altimetry in the North Atlantic between 30-35°N. These have the characteristics of baroclinic Rossby waves with an annual period. Schlax and Chelton (1994a), however, show that tidal errors can be aliased into westward propagating signals with wavelengths and periods very similar to Rossby waves. Hence, it is not possible to unambiguously show with Geosat data that Rossby waves are present. In a study using T/P data however, Schlax and Chelton (1994b) show that westward propagation in the North Atlantic between 30-35°N is indeed due to Rossby waves.

The extent to which tidal aliasing contaminates altimetry data depends upon the energy in the tidal errors. With the present generation of tidal models obtaining accuracies of ~2 cm, the aliasing effect must now be small.
3.4.4 Error Budget

The error sources in the measurement of the distance between the altimeter and the geostrophic sea surface have been discussed in the preceding sections. To obtain the distance between a reference ellipsoid and the SSH, it is necessary to know the position of the altimeter. The error in measuring this distance is the orbit error. This is discussed in Section 3.6, although the magnitude of the error is summarised here. The final error source is in obtaining a time series of data at the same geographical location, the collocation error. This is discussed in Section 3.5, although the magnitude of the error is summarised in Table 3.2 below in order to present the entire error budget in one location.

Table 3.2 A summary of the T/P error sources. The total range error includes the errors due to instrument noise and sea state bias, ionospheric, dry and wet tropospheric corrections. The collocated SSH error adds the collocation error, whilst the geostrophic SSH error includes all the error sources.

<table>
<thead>
<tr>
<th>Error Source</th>
<th>T, cm (2m SWH)</th>
<th>P, cm (2m SWH)</th>
<th>T, cm (4m SWH)</th>
<th>P, cm (4m SWH)</th>
<th>Wavelength (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument Noise</td>
<td>1.7</td>
<td>2.0</td>
<td>2.0</td>
<td>2.5</td>
<td>6-50</td>
</tr>
<tr>
<td>Sea State Bias</td>
<td>2.0</td>
<td>2.0</td>
<td>4.0</td>
<td>4.0</td>
<td>500-10,000</td>
</tr>
<tr>
<td>Ionosphere</td>
<td>0.5</td>
<td>1.7</td>
<td>0.5</td>
<td>1.7</td>
<td>1000-40,000</td>
</tr>
<tr>
<td>Dry</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>500-2000</td>
</tr>
<tr>
<td>Wet</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>100-1000</td>
</tr>
<tr>
<td>Tide</td>
<td>2.9</td>
<td>2.9</td>
<td>2.9</td>
<td>2.9</td>
<td>~10,000</td>
</tr>
<tr>
<td>Inverse Barometer</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>500-2000</td>
</tr>
<tr>
<td>Orbit Error</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>~40,000</td>
</tr>
<tr>
<td>Collocation</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>100-10,000</td>
</tr>
<tr>
<td>Total Range Error</td>
<td>3.0</td>
<td>3.5</td>
<td>4.7</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>Collocated SSH Error</td>
<td>4.3</td>
<td>4.7</td>
<td>5.6</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>Geostrophic SSH Error</td>
<td>6.0</td>
<td>6.3</td>
<td>7.0</td>
<td>7.3</td>
<td></td>
</tr>
</tbody>
</table>

---

6 Fu et al. (1994)  
7 Rodriguez and Martin (1994)  
8 Imel (1994)  
9 Ruf et al. (1994); Stum (1994)  
10 Anderson et al. (1995)  
11 Fu and Pihos (1994)  
12 Section 3.5
The most serious noise effect for velocity measurements is the instrument noise, since the wavelength of this noise is small and hence affects the measurement of geostrophic velocity to a greater extent than the other errors. This is demonstrated by examining equation [3.5] (page 42), which yields an r.m.s. error for the velocity of:

\[
\Delta v = \frac{\Delta h g \sqrt{2}}{fx}
\]

where \( x \) is the distance between the two sample measurements and \( \Delta h \) is the r.m.s. height error, decorrelated between samples. Assuming a velocity measurement over a distance (\( x \)) of 50 km is required (the WOCE requirement for in situ geostrophic velocity measurements is \(~50\) km), Table 3.2 demonstrates that only instrument noise is uncorrelated at this length-scale. Using the value of 1.7 cm r.m.s. (appropriate to a SWH of 2 m) gives a velocity error of 6.5 cm/s at a latitude of 45°. A SWH of 4 m leads to an instrument noise error of 2 cm and a velocity error of 7.6 cm/s. Although velocities in energetic regimes can reach \(~100\) cm/s, 50 cm/s is more typical (Brown, 1989). In regions other than the most energetic regimes, typical velocity values are 5-10 cm/s (Brown, 1989); the noise level of the altimetric measurements.

### 3.5 Collocation of data

For most kinds of altimeter data analysis, the first step after applying the aforementioned corrections is usually to collocate the altimeter data to a set of fixed geographical locations. The reason for this is that most purposes require removal of the time mean sea surface height at each geographical location since this also removes the geoid which is unknown on small scales. It is impossible to remove the time mean at a particular location if the data are not collocated to this location. Furthermore, for time series analysis it is also important to have data collocated to the same position. Altimeter data are not provided at a consistent set of geographical locations. Even if the altimeter is flown in a repeat orbit, the altimetric along-track measurements are provided every fixed time interval \( \Delta t \) (1.029 s for T/P). Hence measurements separated by a repeat period (1 cycle - 9.9156 days for T/P ) will not necessarily be at the same geographical location, but can be within an along-track distance of \( \pm v \Delta t \) where \( v \) is the ground track speed of the satellite. For T/P, \( v=5.8 \) km/s, hence nearest repeat measurements will be within 3.1 km of each other. Furthermore, even if altimetric missions are in a repeat orbit phase, the orbits are usually only held to within an across-track distance of \( \pm 1 \) km from the nominal ground track position.
These sampling variations would not cause a problem if the SSH did not vary on such small scales. The SSH (H) relative to a reference ellipsoid is the sum of the geoid (G) and the ocean dynamic topography (h) due to ocean currents:

\[ H = G + h \]  \[3.8\]

Consider the hypothetical situation of the difference in H between two locations (a) and (b) separated by a distance ~1 km and zero time difference:

\[ H_1 - H_2 = (G_1 - G_2) + (h_1 - h_2) \]  \[3.9\]

This difference \( \Delta H \) is caused purely by geographical sampling differences. However, since the sampling differences are time dependent, \( \Delta H \) will also manifest itself as a time dependent error. To examine whether or not the magnitude of this error is sufficient to warrant attention, the size of the bracketed terms in equation [3.9] must be examined. Assuming that (a) and (b) are separated by a distance of 1 km, the largest values of the two terms are \( \Delta G \sim 30 \text{ cm} \) and \( \Delta h \sim 1 \text{ cm} \). The value of 1 cm for the dynamic topography errors are derived by assuming a 1 m height change over a 100 km distance (a typical value for the Gulf Stream region). The value for the maximum geoid gradient is taken from Rapp et al. (1994), who also show that the global r.m.s. geoid gradient is 2.4 cm/km. Hence the r.m.s. error due to sampling variations is of the order of several cm. Comparing this to the T/P error budget (Section 3.4.4) shows that this is greater than or equal to most errors in the T/P error budget, and is therefore an error that must be reduced. Such errors can be split into two different components; the along-track errors, and the across-track errors. The along-track errors can be reduced to a minimum by along-track collocation of the data. Across-track errors, however, can only be reduced by using a mean sea surface model. Correcting for the along-track and across-track errors is the subject of the following two sections.

### 3.5.1 Along-track collocation techniques

There are four common methods for along-track collocation of altimeter data; collocation in latitude, longitude, time and the perpendicular bisector approach. Collocation in latitude (e.g. Tokmakian and Challenor, 1993) is accomplished by defining a series of latitudes at which to collocate data and then interpolating the along-track data onto these reference latitudes. This approach is easy to implement, but breaks down at higher latitudes where the altimeter is travelling predominantly east-west. Collocating in longitude is similar to collocating in latitude, except reference longitudes are used. This method has the advantage that it does not break down at high latitudes, although at low latitudes it is less accurate than latitude collocation due to the satellite
travelling predominantly north-south. Collocation in time (e.g. Cromwell et al., 1996) is accomplished by finding the time of the crossing of a line of latitude (e.g. the equator) for each pass. The data are then collocated to a series of reference times (relative to the latitude crossing time). This method of collocation is equivalent to collocation in latitude, except that it will remain valid at high latitudes.

The problem with the above three methods is that the along-track interpolation point is not the point closest in space to the reference location. The best method of along-track collocation is that which collocates to the nearest point in space. This is accomplished by defining a reference grid based on the nominal ground track position. For a particular repeat, the along-track latitude and longitude of the closest point in space to each point on the reference grid is then found by obtaining the intersection of a line joining two consecutive altimeter points, so that the perpendicular to this line passes through the reference point. This is illustrated in Figure 3.5. Collocation is then accomplished by linear interpolation to the intersection point (e.g. Cheney et al., 1983; Snaith, 1993). This is the method used in this thesis for the collocation of T/P data. The accuracy of this technique in comparison to the other techniques is discussed in the next section.

3.5.2 Correcting for across-track variations

All the techniques described above only correct for along-track variations in the altimeter sampling. It is widely recognised that such techniques must be used for time series analysis because, as mentioned previously, points only repeat to within an along-track distance of ~3 km (for 1 second averages). An r.m.s. MSS gradient of ~2 cm/km can cause errors of ~6 cm if a collocation technique is not used. Yet, correcting for across-track variations in the sampling is not a widely used technique. The reason is that across-track variations are usually kept within ±1 km of the nominal track position. Hence the error induced by this is ~2 cm. For altimeters prior to T/P, such a correction was deemed pointless when such large errors existed from other sources. However, a 2 cm error is large compared to most of the errors in the T/P error budget (Section 3.4.4) and hence application of this correction is certainly justified. Although 2 cm might seem fairly small, in specific regions the MSS gradient can be as large as 20-30 cm/km (Rapp et al., 1995) and hence not applying a correction could lead to significant errors. Whether or not an across-track correction is necessary was investigated by Brenner et al. (1990) for Geosat data. They find that the large orbit errors in the Geosat data (~several m) make distinguishing the effect of an across-track variation very difficult. Given the precise nature of the T/P measurements, it is worth reinvestigating the value of this correction.
Figure 3.5 Collocation schematic. The data points along an arbitrary repeat are collocated to the reference grid points using a perpendicular bisector approach.

Figure 3.6 The mean and r.m.s. across-track distance (relative to cycle 18) for T/P cycles 1-52. The error bars are ± 1 standard deviation of the across-track distance for each cycle.
Figure 3.6 shows the mean and r.m.s. across-track distances for the first 52 T/P cycles, relative to cycle 18 (cycle 18 was chosen as a cycle close to the nominal ground track (Fu et al., 1994)). It is clear that the T/P altimeter is indeed kept within its specification of ±1 km from the nominal ground track. The r.m.s. across-track difference from cycle 18 is 400 m. Hence, with an r.m.s. geoid gradient of ~2 cm/km, it may be concluded that the r.m.s. across-track correction will be ~0.8 cm. Figure 3.5 shows the rationale behind the correction. To correct for the variation in across-track sampling, the difference between the mean sea surface (MSS) at location (A) and the MSS at location (B) is added to the SSH value at location B in order to remove the effect of the across-track sampling. The MSS at points A and B is given by:

\[
\text{MSS}_A = G_A + \overline{h}_A + h_A(t_{\text{ms}}) + \epsilon_A
\]  
\[
\text{MSS}_B = G_B + \overline{h}_B + h_B(t_{\text{ms}}) + \epsilon_B
\]

where \( G \) is the geoid, \( \epsilon \) represents the errors in the MSS model, an overbar denotes a time mean and \( t_{\text{ms}} \) denotes time periods longer than the time period over which the MSS is measured.

The SSH values, as measured by the altimeter at points A and B, are:

\[
H_A = G_A + \overline{h}_A + h_A(t_A)
\]  
\[
H_B = G_B + \overline{h}_B + h_B(t_B)
\]

From [3.12] and [3.13] it is clear that \( H_A \) and \( H_B \) are different not only because \( t_A \) and \( t_B \) are different, but also as a result of the sampling variations. To correct for these sampling variations, so that the SSH value at B is representative of that at A, the difference between the MSS values at A and B is added to the measurement at B. [3.10]-[3.11]+[3.13] gives:

\[
H_B + \text{MSS}_A - \text{MSS}_B = G_A + \overline{h}_A + h_A(t_{\text{ms}}) - h_B(t_{\text{ms}}) + h_B(t_B) + \epsilon_A - \epsilon_B
\]

The SSH value at location A at time \( t_B \) is:

\[
H_A(t_B) = G_A + \overline{h}_A + h_A(t_B)
\]

Comparing [3.14] with [3.15] shows that the sampling variations caused by different geoid (\( G \)) and mean dynamic topography (\( \overline{h} \)) values at the different locations have been corrected for. The difference between [3.14] and [3.15] is due to three effects.
(i) The difference between $h_A(t_B)$ and $h_B(t_B)$. If the short period (compared to the length of the MSS averaging period) dynamic topography varies on scales smaller than $\pm 1$ km, then this will be an error that is impossible to correct for. Mesoscale eddies and meandering fronts can give rise to maximum gradients $\sim 1$ cm/km, and hence this is an upper limit on the size of this error.

(ii) The terms involving $t_{\text{msss}}$ exist because a "mean sea surface" model is only a mean over a limited time period. Hence, any long period variations in the dynamic topography will result in errors in this approach because the MSS model is not strictly representative of the MSS at all times. The long period variability of the ocean dynamic topography is a subject of ongoing research. However, an upper limit can be placed on this effect by assuming that since dynamic topography gradients rarely exceed 1 cm/km, the magnitude of this effect will not exceed 1 cm/km.

(iii) The final source of error is due to errors in the measurement of the MSS model. These are caused by errors in the altimetric measurement of the MSS and in the interpolation of the MSS model to the sub-satellite points. The MSS model used in this study is the model of Basic and Rapp (1992) as given on the T/P GDR. This is a 0.125° resolution model, and bicubic spline interpolation is used to derive the MSS values at the sub-satellite points.

Rapp et al. (1994) assess the accuracy of their model by comparing the MSS gradients to MSS gradients measured by T/P cycle 18. They find that the r.m.s. difference of the MSS gradients is 0.9 cm/km. This figure includes the error sources (ii) and (iii) described above, but is also due to short period ocean variability since only one cycle of T/P data is used.

Having described the method and its limitations, this correction is applied to T/P cycles 1-52. The r.m.s. correction for each cycle is overlaid on the r.m.s. across-track distance for each cycle, as shown in Figure 3.7. It is clear that apart from three distinct spikes, the r.m.s. value of the correction is proportional to the r.m.s. value of the across-track distance. This gives confidence that the method is being implemented correctly. The three spikes at cycles 20, 31 and 41 are for the POSEIDON cycles. This suggests that there is a bias in the MSS value given on the POSEIDON GDR. CNES were notified of this problem. The r.m.s. correction over the first 52 T/P cycles is 0.9 cm, with maxima and minima around $\pm 15$ cm. The geographical pattern of this correction is shown in Figure 3.8. It can be seen that the largest values ($\sim 10$ cm) are in regions of high across-track geoid gradient such as the South Sandwich Trench region at $54^\circ$S, $25^\circ$W.
Figure 3.7 The r.m.s. across-track correction together with the r.m.s. across-track distance for T/P cycles 1-52. The "spikes" correspond to Poseidon cycles.

Figure 3.8 The across-track correction for T/P cycle 24. The largest values are in regions of steep mean sea surface slope.
Although the magnitude of the correction is sensible and is proportional to the across-track distance, as expected, no evidence that the correction is working has yet been presented. The usual way to determine whether or not a correction is working is to study the effect that the correction has on the ocean variability. If the variability is reduced, this is evidence that the correction may be working. Figure 3.9 shows the difference between the T/P variability with and without the across-track correction. It is clear that although throughout most of the region the effect is minimal, where the correction is large (Figure 3.8), the variability is reduced. This is also confirmed by studying the histogram of the variability differences shown in Figure 3.10. Close inspection of this figure shows that the frequency distribution on the negative side of the histogram is larger than on the positive side, corresponding to a reduction in the variability. The mean value (± standard error) of the difference between the variabilities is 0.186±0.007 mm. Ignoring values within ±1 cm of zero difference gives a value of 9±0.6 mm, demonstrating that the correction is indeed working. To study the effect of applying this correction, an example of a height residual profile with and without the correction is shown in Figure 3.11. It is clear that applying the across-track correction removes a "bump" between 55°S and 56°S that could otherwise be interpreted as a mesoscale eddy. Since such eddies often occur in regions of steep sea floor topography and hence of steep MSS gradient, one might take for granted that such a feature is real when it is just a manifestation of the across-track variations in the sampling.

One has to be cautious about interpreting a reduction in variability as evidence that a correction is working. If the correction is correlated with the ocean variability in any way, then real variability may be removed by the correction. It is difficult to think of a reason, however, why this correction could be correlated with the ocean variability. It is true that the correction is largest in regions of steep geoid gradients which is where one might expect the ocean variability to be largest. However, for the correction to be correlated with the variability, the variability would have to be correlated with the across-track distance of the satellite relative to the nominal track. This is obviously absurd, and hence the reduction in variability demonstrated in Figures 3.9 and 3.10 is evidence that the correction is working and that it is necessary in a few specific regions.

Since the utility of such an across-track correction has been demonstrated, the error in different along-track collocation techniques caused by not applying such a correction will now be derived. The relative accuracy of the different approaches can be assessed by studying the geometry in Figure 3.12, along with the figures given for the r.m.s. across-track distance and the r.m.s. correction. An r.m.s. across-track distance of x km results in an r.m.s. distance between the nominal point and the point on the same line of latitude of $d_1 = x / \cos \theta$ where $\theta$ is the acute angle between the altimeter ground track and a meridian. For longitudinal collocation, the r.m.s. distance between the nominal point
Figure 3.9  T/P SSH variability with across-track correction minus T/P SSH variability without across-track correction. The variability is calculated over T/P cycles 1-52.
Figure 3.10  Histogram of the T/P SSH variability with across-track correction minus T/P SSH variability without across-track correction. The bottom graph is an expanded version of the upper graph. The variability is calculated over T/P cycles 1-52.
Figure 3.11 An example of a height residual profile with and without the across-track correction. Note the region between 54°S and 56°S where an eddy-shaped feature is removed when the across-track correction is applied.

Figure 3.12 The geometry used to calculate the error in different collocation techniques.
and the point on the ground track is \( d_2 = \frac{x}{\sin \theta} \). The value of \( \theta \) can be determined from [3.16] below (taken from Parke et al., 1987).

\[
\theta = \tan^{-1}\left( \frac{V_S \sin \alpha \pm V_E \cos \phi}{V_S \cos \alpha} \right)
\]

where \( \phi \) is the latitude, \( V_E \) is the equatorial rotation velocity of the earth and \( V_S \) is the ground track speed of the satellite. The two terms in the numerator are added for prograde orbits (satellite orbiting in the same direction as the earth e.g. T/P) and subtracted for retrograde orbits (satellite orbiting in the opposite direction to the earth e.g. ERS-1/2). \( \alpha \) is the angle between the orbit and a meridian and is given by:

\[
\sin \alpha = \left| \frac{\cos i}{\cos \phi} \right|
\]

From the first 52 T/P cycles, the geoid gradient that corresponds to an r.m.s. across-track distance of 400 m and an r.m.s. correction of 0.9 cm is 2.25 cm/km. Applying this value to the r.m.s. distances calculated for the different collocation methods results in Figure 3.13. This demonstrates the errors inherent in using different collocation schemes without compensating for across-track MSS gradients. The crossover latitude at which collocation in longitude becomes more accurate than collocation in latitude is 56.25°.

In conclusion, it has been demonstrated that an across-track correction for mean sea surface slopes is necessary in some regions of the South Atlantic. The r.m.s. correction over the entire region for cycles 1-52 is 0.9 cm. Although small, this correction is still sufficient to justify its application, given the T/P error budget. In regions of large MSS gradient, not applying this correction can lead to features manifesting as mesoscale eddies when they are purely due to across-track variations in the sampling. The errors inherent in several commonly used collocation techniques are also derived.

### 3.6 Orbit error removal

The corrections described in Section 3.4 allow the distance between the altimeter and the sea surface to be measured accurately. This measurement, however, is of little practical use since the altimeter is moving relative to the sea surface. To obtain measurements of value, it is necessary to determine the position of the sea surface height relative to some fixed reference. The reference normally used is a reference ellipsoid, defined by its semi-major axis (a) and flattening coefficient (f). Values for the
Chapter 3  Altimetry

Figure 3.13  The r.m.s. error inherent in several commonly used collocation techniques; collocation in latitude, in longitude, with perpendicular bisector method and with nearest point method.

Figure 3.14  Graph demonstrating the effect of collinear orbit error removal on a height residual profile from the Parallel Ocean Climate Model.
TOPEX/POSEIDON mission are $\sqrt{f/298.257}$ and $a=6378.1363$ km (AVISO, 1992). In order to obtain the sea surface height relative to this reference ellipsoid, it is necessary to know the orbit height (the position of the altimeter relative to the reference ellipsoid). Until TOPEX/POSEIDON, the error in the orbit height was the dominant error term in the altimetric error budget. For example, orbit height accuracies were $\sim10$ m for GEOS-3, $\sim1$ m for GEOSAT and $\sim20$ cm for ERS-1. The T/P orbits are accurate to $\sim3$ cm (Fu et al., 1994); more than two orders of magnitude better than the orbits of the early altimeters. The reason that these large errors did not prevent the use of altimetry for the study of ocean mesoscale variability, is that these errors are predominantly very long wavelength. The dominant term in the orbit error spectrum is usually at 1 cycle per revolution (wavelength $\sim40000$ km) (e.g. Tai, 1989; Blanc et al., 1995). Hence, methods can be used to separate the long wavelength signals from the shorter wavelength signals and hopefully remove the orbit error contamination. These methods are reviewed in the following sections where the efficacy of several of the methods is tested by using ocean model data. It is shown that for some purposes it is necessary to apply an orbit error correction, even with the highly accurate T/P orbits.

### 3.6.1 Different orbit error removal techniques

Two different methods of removing orbit error have been predominant over recent years: crossover analysis and collinear analysis. These are described below.

(a) **Crossover analysis**

This technique uses the information given at the locations where two satellite ground tracks cross each other (the crossover point). The difference between two measurements at the same geographical location but at different times is due to three factors; (1) the difference in the sea surface height, (2) the difference in the residual errors in the geophysical corrections and (3) the difference in the orbit error. The crossover method parameterises the orbit error with a set of long wavelength functions (one for each track) and finds the parameters of these functions that minimise the crossover differences within a specified time range. The first crossover methods use a bias and tilt representation of the long wavelength functions (e.g. Rapp, 1983). These methods produced acceptable results over small regions. However, over large regions, a linear representation of the orbit error is inaccurate (Tai, 1989) because the orbit error is predominantly sinusoidal with a wavelength of $\sim40000$ km. To improve on this, Douglas et al., (1984) use a Fourier series representation of the orbit error. Whilst an improvement in terms of accuracy is evident, the method is computationally expensive to implement. Tai (1988) suggests the use of a sine wave with a wavelength equal to...
one orbital revolution as a compromise between the accuracy of the Fourier series approach and the simplicity of the tilt-bias approach. This approach yields very promising results (Tai, 1988) and is now the usual method of implementing the crossover differences scheme. The advantage of crossover minimisation schemes over other methods of orbit error removal is that the sea surface height relative to the reference ellipsoid is retained. Other methods require sea surface height residuals from the mean sea surface height, and hence cannot provide an absolute SSH measurement. The disadvantage with all crossover minimisation schemes is that the solution obtained is not a unique solution, because the whole solution can be raised or lowered by a constant amount whilst keeping the crossover differences the same. Tai (1988) addresses this problem and shows that the indeterminacy can be removed by requiring that both the crossover differences and also the amplitudes of the sine waves in the solution minimise.

Crossover analysis is invaluable in attempting to measure the mean sea surface height. Geoid uncertainties generally prevent the use of this mean sea surface to obtain dynamic height measurements, and so SSH variability\(^\text{13}\) is usually the parameter of interest. If only the variability is required, a simpler and far less computationally expensive method is used to correct for orbit error, namely that of collinear analysis.

(b) **Collinear Analysis**

Cheney et al. (1983) suggest an alternative method to that of crossover analysis when the altimeter is in an exact repeat orbit. Firstly, a mean SSH value is calculated by temporally averaging the collocated SSH data (see Section 3.5) at each location. Secondly, a series of height residual profiles are obtained by subtracting the mean SSH from each SSH profile. Finally, a long wavelength function is fitted to each height residual profile and the resulting orbit error corrected profile is obtained by subtracting the function from the height residual profile.

The error in such an approach has two components; (1) the ability of the long wavelength function to model the orbit error and (2) the accuracy of the estimation of the long wavelength function by least squares fitting. The ability of the chosen long wavelength function to model the orbit error can be calculated given knowledge of the spectral characteristics of the orbit error. Assuming the orbit error is a 1 cycle per revolution sine wave, Tai (1989) calculates the errors due to bias, tilt+bias and quadratic orbit error representations and shows that the errors are smallest for a quadratic representation and largest for a bias representation. On this basis, one would always use

\(^{13}\) In this thesis a specific reference to the variability of a parameter refers to the SD of that parameter
a quadratic function to model the orbit error. However, the accuracy of the fitting of these functions varies according to the function in use. Le Traon et al (1991) outline the theory behind the accuracy in obtaining a least squares fit of any function. The accuracy with which a function can be estimated depends on the number of degrees of freedom on the track concerned, as well as on the homogeneity of the variability along the track. Using their theory together with the results of Tai (1989), Le Traon et al (1991) show that a tilt-bias correction is optimal for arcs shorter than 5000 km and that for longer arcs a quadratic representation is better. To obtain these results they assume an orbit error of 1 m r.m.s. and a homogenous mesoscale variability of 15 cm. Le Traon et al. (1991) also show that the use of weighted least squares method (where the weighting is the reciprocal of the mesoscale variability) gives more accurate results than that of the conventional least squares method.

Chelton et al. (1990) in a study of the Southern Ocean, use a collinear analysis. However they represent the orbit error with a sinusoidal rather than a polynomial function. This is sensible given that the dominant terms in the orbit error are sinusoidal, but it is not clear whether this is actually more accurate than the polynomial methods due to the inaccuracies in obtaining the functional fit.

Faced with these different methods, it is very difficult to make a decision as to the appropriate choice of long wavelength function. The ocean variability, the magnitude of the orbit error and the track length are all parameters which affect the choice of function. The only paper that the author is aware of that attempts to answer this question is Le Traon et al. (1991). As discussed above, they only compare the quadratic and tilt+bias methods for one specific example. The next section presents research that attempts to answer these questions with parameters appropriate to T/P in the South Atlantic region.

### 3.6.2 Comparison of different techniques

To resolve the question of which orbit error technique is optimal for the region of interest, ocean model data from the Parallel Ocean Climate Model (POCM) are used. The POCM is a primitive equation model based on the Semnter and Chervin (1992) code. It has a nominal resolution of 0.25° in latitude and longitude and a global domain. The version used here (POCM_4B) is forced by daily European Centre for Medium Range Weather Forecasting (ECMWF) wind stress fields, together with monthly mean surface heat fluxes by Barnier et al. (1995). A comparison between the model and oceanic observations can be found in Stammer et al. (1996). This model is arguably the most realistic global model to date, although the model variability is roughly a factor of two lower than the observed T/P variability (Stammer et al., 1996).
SSH fields from the POCM_4B run, averaged into 10 day fields corresponding to the T/P cycles were kindly provided by Robin Tokmakian. The POCM mean SSH is subtracted from each SSH field to obtain height residual maps. These are then interpolated onto the T/P ground track reference grid defined by cycle 18 (see Section 3.5) using bilinear interpolation. To simulate the effects of orbit error for every ground track in each cycle, a sine wave with a wavelength of 40000 km, an amplitude of $5/\sqrt{2}$ cm and random phase is added to each model height residual profile. Three different orbit error techniques are used in an attempt to remove the simulated orbit error; (1) bias (B), (2) tilt+bias (TB) and (3) bias+$\cos(2\pi x/L)+\sin(2\pi x/L)$ (CS), where $L=40000$ km. An example of a profile before and after these methods of orbit error removal have been applied is shown in Figure 3.14 for a 3400 km length pass. It is clear that although the orbit error correction methods are an improvement on no orbit error correction, they do not perfectly reconstruct the original height profile. Surprisingly, the CS method is no better than the other functions, even though it is an identical function to the simulated orbit error. The reason is that the short track length, together with the ocean variability on the track, do not allow an accurate fit to be obtained. The r.m.s. differences between the profiles with orbit error and the original profile are: no orbit error removal - 6.0 cm, B - 2.5 cm, TB - 2.8 cm, CS - 2.9 cm. Although this example illustrates the principle, the results from one profile are not statistically significant and thus it is necessary to examine different track lengths. To increase the statistical significance of the results, 9 T/P cycles are used. For each cycle a 5 cm r.m.s. sinusoidal orbit error with random phase is added into every track and the three different orbit error removal methods are applied. The r.m.s. error is calculated for each track for the nine cycles. These results are presented in Figure 3.15. The error bars on the graphs are the standard error of the r.m.s. errors and hence provide a measure of the statistical significance of the results. Figure 3.15 illustrates several points; (1) that for tracks less than 6500 km, the B method is superior to the other two methods and (2) that for tracks longer than 6500 km, the TB method is more accurate than the other two methods. The CS method is always worse than the TB method, and is worse than the B method for tracks shorter than 8300 km. This shows that it is not only the appropriateness of the orbit error model that must be considered, but also the ability to obtain an accurate model fit. It might seem strange that the accuracy of all the orbit error removal methods degrades at track lengths between 800-1000 km. This is due to irregular distribution of mesoscale variability with track length. That is, the tracks between 800-1000 km are generally those that pass through regions of high variability, and hence the error in fitting the orbit error model is increased.

The effects of orbit error removal on real T/P data are now demonstrated. The variability map of the South Atlantic from T/P is shown in Figure 3.16. This is obtained from two years (1993 and 1994; cycles 11-84) of T/P altimetry, using the corrections
Figure 3.15 Graph showing the error in three different methods of orbit error removal as a function of track length. The error bars are ±1 standard error.

Figure 3.16 The T/P SSH variability for 1993 and 1994 without any orbit error removal.
described in Section 3.4 and the collocation technique described in Section 3.5. The SD values are calculated on the reference grid locations and are interpolated to a 0.5° grid for display purposes, using Gaussian interpolation with a FWHM of 110 km (see Section 3.7 for a justification of these parameters). It is clear that the variability is not homogeneous, but rather is concentrated in specific areas associated with the dynamical regimes discussed in Chapter 2. For example, high variability areas associated with the Brazil/Falklands confluence (peak variability ~40cm), the Agulhas Retroflection (peak variability ~45cm) and the Antarctic Circumpolar Current (peak variability ~20cm) are clearly visible. The region of very low variability (~5cm) in the centre of the South Atlantic places an upper bound on the residual orbit error in the T/P data in this region. Figure 3.17 shows an equivalent variability map, but with the orbit error reduced by the collinear analysis described in Section 3.6.1. Tracks less than 6200 km long have a bias removed and tracks longer than 6200 km have a linear trend removed. Although Figures 3.16 and 3.17 are qualitatively very similar, differences can be seen in some regions. The most noticeable difference is the reduction in variability by ~2 cm throughout the region. In particular, the area of moderate variability in the north-eastern section of the basin is drastically reduced by orbit error removal. The pattern of slightly higher variability in the north-eastern part of the basin is similar to that of SST variability (Figure 4.15) and it is therefore likely that, in this case, the orbit error removal is eliminating real oceanographic variability. A slight increase in variability is present in Figure 3.17 in the region just south of the tip of South Africa. This increase in variability can occur if the variability along a track is inhomogeneous; the variability can be increased in the lower variability regions and decreased in the higher variability regions (Le Traon et al., 1991). A benefit of applying the orbit error correction is that the areas of small spatial scale variability are more clearly resolved and new features can be seen. For example, at 30.25°S, a zonal channel of high variability can be seen in Figure 3.17, whereas this is not as clear in Figure 3.16, demonstrating that even if the orbit error is as low as 3-5 cm r.m.s., the application of an orbit error removal scheme can still be useful for certain purposes.

In conclusion, it has been shown that for the collinear method in the South Atlantic, the bias and tilt-bias methods are more accurate than the sinusoidal method for parameters relevant to T/P. It may be questioned whether it is really necessary to worry about applying an orbit error correction given that the errors are only ~3 cm r.m.s.. Indeed many research groups do not apply any orbit error removal to T/P data. Whether an orbit error removal scheme is necessary for T/P data depends on the application of the data. In the study of large scale variability (such as the large scale seasonal effect; e.g. Gill and Niiler (1973)), the use of orbit error removal can remove or reduce this signal (Cheney and Miller, 1990), and is therefore undesirable. In the study of the ocean mesoscale, the application of an orbit error removal can highlight features that would
Figure 3.17  The T/P SSH variability for 1993 and 1994 with orbit error removal.

Figure 3.18  A height anomaly map from the Parallel Ocean Climate Model corresponding to T/P cycle 11.
not have been clearly observable without any orbit error removal. Furthermore, it is demonstrated in the following section that when interpolating altimeter data to a regular grid, the accuracy of the interpolation is very sensitive to long wavelength errors such as orbit error. For this reason, studies of mesoscale variability requiring gridded fields should still apply an orbit error correction to the T/P altimeter data. The work discussed above will aid the choice of correction method.

3.7 Interpolation of altimeter data

Unlike sensors such as scanning radiometers, the altimeter sampling is essentially one dimensional. In the along-track direction, the sampling is continuous whereas in the across-track direction only regions within ~2 km (depending on footprint diameter) of the nadir point will contribute to the averaged SSH value. The ground track pattern shown in Figure 3.18 demonstrates the anisotropic nature of altimeter measurements. The distance between tracks for the T/P orbit can reach 315 km at the equator. For purposes such as displaying data and performing some statistical calculations, regular gridded fields are necessary. The process by which regular gridded fields are formed from irregular data is called interpolation. Numerous techniques to interpolate data exist, ranging from simple linear interpolation to the complex and computer intensive techniques of optimal interpolation. A comprehensive study of different interpolation techniques with a view to discovering the most effective technique for interpolation of altimeter data does not yet exist. Such a study would take several years to complete and is outside the scope of this thesis. However, for the analysis of T/P data in Section 3.8, it is necessary to interpolate the data to a regular grid, and hence a method must be chosen to accomplish this. The purpose of this section is to describe the interpolation technique used, to justify the choice of parameters and to investigate the errors that are introduced by interpolation.

3.7.1 Interpolation techniques

Numerous interpolation techniques exist and deciding upon the correct technique is difficult. The simplest technique is bilinear interpolation (the two dimensional equivalent of linear interpolation), where the straight line through two points separated in space is used to derive data values elsewhere in space. An extension to this is bicubic interpolation which uses a cubic function through four data points to derive interpolated values between the points. These, and other spline techniques which fit functions exactly through data points, are very susceptible to noise present in the data and care must be taken when using these methods.
Another method of interpolation is to obtain a weighted average of data points within a certain search radius (SR) in order to find the interpolated value. The weights can be unity, in which case this reduces to a simple moving average. Alternatively they can be some function of the distance away from the interpolated point, such as a Gaussian weighted function.

A third method of interpolating data is to use weighted least squares to fit a surface to the data points; the interpolated value is given by the surface at the interpolation location. These are called "locally weighted regression (loess) smoothers" and are described in detail by Cleveland and Devlin (1988). The weights used in the least squares estimate vary according to the distance from the interpolation point. The surface can be any specified function, but is typically a linear or quadratic surface (e.g. Chelton et al., 1990).

The most computationally intensive form of interpolation is that of optimal interpolation, where the signal and error covariance functions are specified \textit{a priori}. If these covariance functions are the true covariance functions, then the estimate is optimal in that it has the lowest r.m.s. error of any linear estimate. In practice, the signal and noise covariance functions are not exactly known and other methods can provide equivalent results without the computational expense attached to optimal interpolation.

Successive correction methods have also been used to interpolate altimeter data (e.g. Tokmakian and Challenor, 1993). These are essentially the same as weighted average methods, except that several passes through the data are made and the parameters of the weighting function are changed with every pass. With idealised data, the SCM give good results (e.g. Voesspoel, 1995). However with real (noisy) data, the advantage of several passes reduces (Voesspoel, 1995).

Schlax and Chelton (1992) compare several methods of interpolating data. These methods include the running mean, Gaussian average, linear loess, quadratic loess and cubic spline methods. They investigate the frequency domain characteristics of these smoothers and conclude that away from edge effects, the quadratic loess smoother has more desirable filter characteristics than other techniques. Near data edges, Gaussian smoothing has the most desirable characteristics. They show that these results hold for both regular and irregular datasets. An example of the filtering characteristics of these smoothers, taken from Schlax and Chelton (1992) and of the transfer function of an ideal filter are shown in Figures 3.19 and 3.20 respectively. It is clear that, away from data edges, the quadratic loess estimate is the nearest to the ideal shape, whereas the Gaussian estimate is better for locations near data edges.
Figure 3.19 The weighting functions and transfer functions for several smoothers. The dashed lines represent one sided weighting functions, as would occur near land and data boundaries (taken from Schlax and Chelton, 1992).

Figure 3.20 A schematic showing the difference between an ideal transfer function and a typical transfer function for irregularly spaced data (taken from Chelton and Schlax, 1994).
In attempting to decide which interpolation method to use, a compromise has to be made between computational speed (since large datasets are being used) and accuracy. Furthermore, since large geographical regions will not contain any data due to the presence of land, a method that does not introduce spurious signals at edges is desirable. With these considerations in mind, single pass Gaussian interpolation was chosen to produce gridded SSH fields. The best set of parameters to use in this interpolation and the errors introduced by Gaussian interpolation are the subjects of the next two sections.

3.7.2 Optimising interpolation parameters

For a set of irregular data points, \( h_i \), the interpolated value, \( h^{\text{int}} \), is found by:

\[
h^{\text{int}} = \frac{\sum_{i=1}^{N} w_i h_i}{\sum_{i=1}^{N} w_i}
\]  

where

\[
w_i = e^{-\frac{d_i^2}{\sigma^2}}
\]

The distance from the interpolation point to the data point is given by \( d_i \) and \( \sigma \) is the e-folding distance. Parameters that are commonly used to describe the Gaussian profile are the e-folding distance and the Full Width Half Maximum (FWHM). These are related by: \( \text{FWHM} = 2\sigma\sqrt{\ln 2} \). Throughout this study, the FWHM parameter rather than the e-folding distance will be referred to. To decrease the computational time in performing this interpolation, only data points within a search radius (SR) of the interpolation point are used in the weighted average.

What value of the FWHM will give the most accurate interpolation? In many altimetric studies using interpolation, the choice of interpolation parameters is rarely justified properly. The standard approach is either to try several different values and then to objectively assess which looks best, or to copy another study and cite this as the justification. The reason that interpolation parameters are rarely rigorously justified is because it is very hard to do so. Interpolation is necessary because the answer is not known and hence there is no way of assessing the performance of the interpolation. In an attempt to inject some rigour into the choice of interpolation parameters for this study, SSH residual data from the POCM are used.

The methodology is to sample the ocean model data as an altimeter samples the ocean, to interpolate the sub-sampled data back to the model grid and then to assess the
accuracy of the interpolation by comparing it to the original model grid. The FWHM can be varied and the interpolation can then be optimised. Figure 3.18 shows a height anomaly map (nominally 0.25° resolution) from the POCM with the T/P ground tracks overlaid. These SSH residual data are interpolated using bilinear interpolation to obtain model SSH residual data along the T/P ground tracks. It may be argued that interpolation at this stage defeats the point of the study, since error is immediately introduced by the linear interpolation. The T/P ground track points are 6.2 km apart, compared to the model resolution of ~25 km. The errors introduced by this step will be negligible compared to errors introduced by interpolating over distances of several 100 km. This is also shown by the results discussed in the following section. Two years (1993 and 1994 corresponding to cycles 11-84) of model data are sampled onto the altimeter tracks and reinterpolated onto the model grid in order to enable comparisons with the original data. The FWHM of the Gaussian interpolator is varied from 25 km to 700 km to obtain the optimal FWHM. The SR is always twice the distance of the FWHM; varying the SR has little effect on the results as long as it is not reduced to a value corresponding to a significant Gaussian weighting. For example, the Gaussian weighting at the SR distance is \( \exp(-161n^2) \approx 0.000015 \), a negligible value as long as there are data points close to the interpolation point.

It is pointless to discuss the optimisation of interpolation parameters without representing in some manner the noise in the raw data. To assess the different effects of noise, five scenarios are implemented; (1) no error, (2) 3 cm r.m.s orbit error, (3) 5 cm r.m.s. orbit error, (4) 5 cm r.m.s. orbit error + 2 cm r.m.s. white noise and (5) 5 cm r.m.s. orbit error, removed with the optimum technique described in Section 3.6. In all cases the orbit error is simulated as a cosine function with a wavelength of 40000 km (see Section 3.6). The rationale for this choice of parameters is that orbit error is still a dominant error source in the T/P data. Since it is only ~3 cm, many studies do not apply any form of orbit error correction; hence it is relevant to assess its effect on altimeter interpolation. The 5 cm r.m.s. orbit error scenario is to assess the sensitivity of the results to different amounts of orbit error, as well as simulating the effect of residual tidal errors to some extent (these are also predominantly long-wavelength). The 5 cm orbit error + 2 cm noise is designed to be representative of the actual T/P error budget with instrument noise ~2 cm and orbit and tide errors dominating the error budget. Finally, the effect of performing an orbit error removal is investigated by using a bias correction for tracks less than 6200 km in length and a tilt+bias correction for tracks greater than 6200 km. These parameters are justified in Section 3.6.

Figure 3.21 shows the r.m.s. difference (or error) between the altimeter interpolated maps and the original model data over a range of values for the FWHM for the five scenarios described above. Without any simulated error, the interpolation error
Figure 3.21  The r.m.s. error in Gaussian interpolation as a function of the Gaussian full-width at half-maximum.

Figure 3.22  The correlation between an interpolated field and the original field as a function of the Gaussian full-width at half-maximum.
minimises at \(-80\) km with an r.m.s. error of \(-2\) cm over the entire SA region. Initially this seems an extremely small value. The reason is that the optimisation point is a balance. On the one hand, it is necessary to obtain a sensible estimate between tracks, whilst on the other, care must be taken not to oversmooth the data. In the case of no noise, a greater FWHM has a large effect on oversmoothing, but little effect on obtaining a sensible estimate between tracks. Hence the optimal FWHM is fairly small. The effect of a 3 cm r.m.s. large scale error is to increase the optimal FWHM to 110 km where the r.m.s. error is 3.1 cm; a 50% increase on the noise free interpolation estimate. Increasing the large scale noise to 5 cm r.m.s. increases the optimal FWHM to \(-170\) km with an associated r.m.s. error of 4.2 cm. Adding 2 cm white noise to the 5 cm large scale error does not significantly change the optimal FWHM or the r.m.s. error. The reason for the large scale error having such an effect on the interpolation error is that to reduce this error it is necessary to average over a large number of tracks, since each track is essentially only 1 degree of freedom. Hence, the optimal FWHM enlarges to increase the effect of all the tracks within the SR, rather than emphasising only the effect of nearby tracks. The beneficial effect of removing orbit error prior to interpolation is demonstrated in Figure 3.21. It is clear that the results with 5 cm orbit error removed are more accurate than those with 3 cm r.m.s. orbit error present.

These results clearly demonstrate the effect of noise on the optimal FWHM, where optimal is defined by the minimum r.m.s. error. Another common method of assessing the similarity between two fields, \(x_i\) and \(y_i\), is the zero lag cross-correlation, \(\sigma_{xy}\) given by:

\[
\sigma_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y n}
\]

where \(\sigma_x\) and \(\sigma_y\) are the standard deviations of \(x\) and \(y\) respectively, \(n\) is the number of samples and an overbar denotes the mean.

Figure 3.22 is similar to Figure 3.21, with the exception that the cross correlation between the original model field and the interpolated field is optimised, rather than the r.m.s. difference between the two fields. It is clear that the FWHM corresponding to maximum correlation are far less sensitive to the orbit error; ranging from \(-80\) km with no orbit error (correlation = 0.94), to 110 km with 5 cm orbit error (correlation=0.75). The effect of removing orbit error is demonstrated clearly in Figure 3.22, where the optimisation curve lies close to the curve with no orbit error, peaking at a correlation of 0.91. In terms of the percentage of variance explained, the noise free interpolated field can explain 88% of the variance of the original field, the 5 cm orbit error field 56% and the orbit error removed field 83%. The reason why the optimal FWHM differ from
those given by minimising the r.m.s. difference is that for the correlation minimisation, the regions of high energy are effectively given more weight due to the product of x and y in [3.20]. In the r.m.s. difference this is not so. The high energy regions are those associated with mesoscale variability. Hence the correlation minimisation gives a smaller optimal FWHM than the r.m.s. minimisation. For the purposes of this thesis, it is the regions of mesoscale variability that are most interesting. Hence correlation, rather than r.m.s. error, will be the statistic used to define the FWHM for use with real T/P data in Section 3.8. A summary of the optimal interpolation parameters is given in the table below.

Table 3.3 A summary of the optimal interpolation parameters.

<table>
<thead>
<tr>
<th></th>
<th>FWHM (min. corr.)</th>
<th>Correlation</th>
<th>FWHM (min. r.m.s.)</th>
<th>R.M.S. Error (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise free</td>
<td>80</td>
<td>0.94</td>
<td>80</td>
<td>2.1</td>
</tr>
<tr>
<td>3 cm Orbit Error</td>
<td>100</td>
<td>0.86</td>
<td>110</td>
<td>3.0</td>
</tr>
<tr>
<td>5 cm Orbit Error</td>
<td>110</td>
<td>0.75</td>
<td>160</td>
<td>4.2</td>
</tr>
<tr>
<td>2 cm noise</td>
<td>110</td>
<td>0.75</td>
<td>160</td>
<td>4.2</td>
</tr>
<tr>
<td>5 cm Orbit Error removed</td>
<td>95</td>
<td>0.91</td>
<td>90</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Although the optimisation of the interpolation parameters has been investigated, it is essential to investigate the geographical variation of the errors due to the interpolation of altimeter data. This is the subject of the next section.

### 3.7.3 The error caused by interpolation

The results of the previous section are derived from model SSH residual data within the South Atlantic region corresponding to one 10 day T/P cycle and are insensitive to the particular cycle used. Hence, if the optimal values of the FWHM are used for each cycle within the two year period from 1993-1994, a measure of geographical variation of the accuracy of the interpolation can be obtained. Figure 3.23 shows the SSH variability from the POCM for 1993-1994. Comparing Figure 3.23 to the results from T/P shown in Figures 3.15 and 3.16 reveals that the fields are qualitatively similar. A comprehensive comparison between T/P data and the POCM is given by Stammer et al. (1996). Figure 3.24 shows the SSH variability derived by sampling the model data along T/P ground tracks and by interpolating the data from each cycle back onto the original model grid with a FWHM parameter of 80 km (the optimal FWHM for a dataset with no noise). The effects of the altimeter sampling can be seen clearly in the
Figure 3.23 The SSH variability from the Parallel Ocean Climate Model corresponding to 1993 and 1994.

Figure 3.24 The T/P SSH variability from the Parallel Ocean Climate Model re-interpolated SSH fields for 1993 and 1994.
Figure 3.25 The r.m.s. error in the Gaussian interpolation.

Figure 3.26 The percentage r.m.s. error in Gaussian interpolation.
resulting variability map. Eddy-like structures have become apparent in the variability in locations where such structures were not apparent in the original variability. For example, the tongue of high variability at 0°E, 30°S is continuous in the original variability, but is broken up into circular structures in the interpolated variability. Another effect of the interpolation is that the variability is lower throughout. Figure 3.25 shows the r.m.s. error in the variability. It is clear that in the regions of high variability, the r.m.s. error is also higher. Furthermore, in regions close to the satellite ground tracks, the r.m.s. error is extremely low, whereas between ground tracks the error is much higher. The percentage r.m.s. error is shown in Figure 3.26. The mainsimilarity to Figure 3.25 is that the percentage r.m.s. error is lowest along points corresponding to the T/P ground track location, and highest in the centre of these tracks. That the r.m.s. and percentage r.m.s. errors are so low at locations near to ground track regions is evidence that the first step of this process, the linear interpolation from the model grid locations to the T/P ground track locations, is not contributing any substantial amount of error to this analysis. The cause of this has been discussed in the previous section.

At first sight, it also appears that the regions of highest percentage r.m.s. error are coincident with the regions of highest variability. After a careful study of Figures 3.23 and 3.26, it becomes apparent that this is wrong. The regions of highest percentage r.m.s. error are not the regions of highest variability, but rather the regions adjacent to them. The reason for this is that in a region of low variability next to a region of high variability, error in the interpolated values in the region of low variability will be introduced by the adjacent area of high variability. Hence the regions with the worst signal to noise ratio are quiescent regions adjacent to energetic regions.

Figure 3.27 shows the percentage r.m.s. error for the case of 5 cm r.m.s. orbit error added to the model data before interpolation. The striking feature of Figure 3.27 is the vast geographical location where any oceanographic signals are dominated by errors due to interpolation and orbit error. Figure 3.28 demonstrates the effect of removing the orbit error before interpolating. It is clearly beneficial to perform this orbit error removal.

To place this work in context, mention of the results of Chelton and Schlax (1994) and Greenslade and Chelton (1996) should be made. They investigate the resolution capability of an irregular dataset and conclude that the T/P dataset is only capable of resolving scales of about 4.2°. They suggest that interpolation parameters should be large enough to remove smaller scales and ensure a homogeneous interpolated dataset. It seems that their results differ from the results of this section, since the optimal FWHMs obtained are far smaller than 4.2°. The difference in results lies in the
Figure 3.27 The percentage r.m.s. error after adding 5 cm r.m.s. orbit error and interpolating with optimal Gaussian parameters.

Figure 3.28 The percentage r.m.s. error after adding 5 cm r.m.s. orbit error, applying a collinear orbit error removal method and interpolating with optimal Gaussian parameters.
difference in philosophy between the two studies. Chelton and Schlax (1994) and
Greenslade and Chelton (1996) use a method based on obtaining the transfer function of
their interpolator, comparing the transfer function to that of an ideal filter and
integrating the difference between these transfer functions (Figure 3.20) for varying
interpolation parameters. From this difference between the transfer functions, they
obtain the relative error introduced by interpolating. As the interpolation parameters get
larger, the aliasing bands in the transfer function of their interpolator get smaller, and
the relative error therefore gets smaller. The problem with this is that the relative error
never minimises, but continues to reduce as the interpolation parameters enlarge. To
obtain a sensible value for the interpolation parameters, Greenslade and Chelton (1996)
 impose an error threshold of 10% and state that the interpolation parameters when the
relative error reaches 10% are the resolution capability of the dataset. This method is
fundamentally different from the analysis in this section in that Greenslade and Chelton
(1996) are essentially comparing their interpolated field with that of a perfectly
smoothed field. In this analysis however, the interpolated field is compared with the
true field. Which approach is correct depends largely on the intended application of the
data. If the application is to study large scale features and/or to use statistical techniques
that require a stationary dataset, then the approach of Chelton and Schlax (1994) and
Greenslade and Chelton (1996) has some merit. However, for studies attempting to gain
the most accurate picture of the real state of the ocean (e.g. eddy tracking studies and
any global ocean observing system applications), the methodology outlined in this
section is the better approach.

The value of the work in this section lies in investigating the sensitivity of the
interpolation error to varying interpolation parameters, in understanding the way in
which interpolation can cause spurious variability structures and in demonstrating the
need for orbit error removal before interpolation. Statements about the absolute
magnitude of the errors in interpolating real altimeter data cannot be made from this
analysis because this depends sensitively on the extent to which the ocean model
represents the real ocean. Stammer et al. (1996) conclude that the energy in the model is
approximately 50% lower than the energy observed in the real ocean. This should not
affect the results of the percentage r.m.s. errors, but will only influence the absolute
values.

3.8 Determining the large scale flow field

It has already been stated that the absolute ocean dynamic topography cannot be
measured on small (<2000 km) spatial scales due to inaccuracies in the geoid on these
scales. This is demonstrated by Figure 3.29 taken from Nerem et al. (1994), which
Chapter 3 Altimetry

Figure 3.29  Spectrum showing the JGM-2 geoid error together with the dynamic topography from T/P and an ocean model (taken from Nerem et al., 1994).

Figure 3.30  The T/P height anomaly map for January 1993.
shows the spectrum of the geoid error together with the ocean dynamic topography, in terms of spherical harmonics\(^4\). The point at which the dynamic topography power spectrum crosses the geoid error power spectrum is the wavelength at which the dynamic topography cannot be resolved from the geoid error. The most recent estimates of this point is degree 14 (wavelength of 2800 km). Hence dynamic topography with wavelengths below 2800 km cannot be resolved. This is the reason why altimetric studies of the ocean usually concentrate on SSH variability. A method to derive the absolute dynamic topography on spatial scales smaller than 2800 km is described in the following sections. It relies on the assumption that there is little energy in the small spatial scale mean dynamic topography and hence will not work in regions where this assumption is not valid. The theory of this method is described in Section 3.8.1. This theory is applied to T/P data in Section 3.8.2 and the technique is validated in Section 3.8.3.

### 3.8.1 Theory

The aim of the method is to establish the best estimate of the dynamic topography from altimetry alone. As discussed in Section 3.3, dynamic topography is the SSH relative to the geoid:

\[
D = H - G \tag{3.21}
\]

where \(H\) is the sea surface height relative to a reference ellipsoid, \(G\) is the geoid height relative to a reference ellipsoid and \(D\) is the dynamic topography. The dynamic topography can be split into time dependent terms and into large and small scale terms (denoted \(L\) and \(S\)) where large scale refers to wavelengths greater than 2800 km and small scale refers to wavelengths smaller than 2800 km.

\[
D(t) = D_L + D_S + D_L'(t) + D_S'(t) \tag{3.22}
\]

Unfortunately, this is not measured by the altimeter. The parameter derived from the altimeter is the SSH relative to the reference ellipsoid:

\[
H(t) = D_L + D_S + D_L'(t) + D_S'(t) + G_L + G_S + \varepsilon(t) \tag{3.23}
\]

where the \(\varepsilon(t)\) term represents the errors inherent in the measurement. The way in which an estimate of [3.22] is obtained from altimetry alone is described in the following analysis.

\(^{14}\) The degree and order of a spherical harmonic is equal to the circumference of the earth divided by the wavelength.
Averaging $H$ in time, such that: $\overline{D_L}(t) = 0$, $\overline{D_S}(t) = 0$, $\overline{\epsilon(t)} = 0$. [3.23] becomes:

$$\overline{H} = \overline{D_L} + \overline{D_S} + G_L + G_S$$  \[3.24\]

The measured geoid ($g$) can be represented by:

$$g = G_L + (G_S + \overline{D_{s2}})$$  \[3.25\]

where $\overline{D_{s2}}$ is the small scale mean dynamic topography. The S2 indicates that it is different to the small scale mean dynamic topography in the previous equations, because it is measured over a different time period. The small scale mean dynamic topography is bracketed with the small scale geoid term because it is impossible to separate these terms.

Subtracting [3.25] from [3.24] gives:

$$\overline{H} - g = \overline{D_L} + (\overline{D_S} - \overline{D_{s2}})$$  \[3.26\]

Averaging [3.26] in space, such that: $\langle \overline{D_S} - \overline{D_{s2}} \rangle = 0$.

This leaves:

$$\langle \overline{H} - g \rangle = \overline{D_L}$$  \[3.27\]

the "mean large scale dynamic topography".

Subtracting [3.24] from [3.23] gives:

$$H(t) - \overline{H} = D_L + D_S(t) + \epsilon(t)$$  \[3.28\]

the "height anomaly". Finally, adding [3.27] and [3.28] gives the estimate (denoted "est") of the absolute dynamic topography.

$$D^{\text{est}}(t) = \overline{D_L} + D_L'(t) + D_S'(t) + \epsilon(t)$$  \[3.29\]

When comparing [3.29] with [3.23], it is clear that the term missing in [3.29] is the small scale mean dynamic topography. Hence, this method will work in regions where the power in the small scale mean dynamic topography is weak. This is discussed in more detail in Section 3.8.3. This method is now applied to T/P data in the South Atlantic.
3.8.2 Observation

Figure 3.30 shows a height anomaly map for January 1993, obtained by subtracting the time mean SSH from each cycle and then averaging the residuals. Interpreting such maps is difficult because the mean component of the dynamic topography is not present. Hence it is impossible to distinguish a meander in a front from a mesoscale eddy. This is the main obstacle in interpreting altimeter data and the discovery of a method for obtaining the mean flow would revolutionise the oceanographic use of altimeter data. Various methods have been proposed to solve this problem. If in situ hydrographic data are available at the same time and location as the altimeter data, it is possible to "calibrate" the altimeter data with the in situ data in order to derive the mean flow (e.g. Challenor et al., 1996; Cromwell et al., 1996). This method is, however, limited to those regions with coincident in situ data and such regions are few.

Another method of determining the mean flow is to use a synthetic geoid. This usually means using model and/or hydrographic data to estimate the dynamic topography and then subtracting this estimate from the altimetric SSH measurement to leave the synthetic geoid estimate. This can then be used to obtain the absolute dynamic topography. These methods always depend on the accuracy of the model and/or in situ data. In some regions such as the Gulf Stream, where an abundance of hydrographic data are present, this method seems to give promising results (e.g. Porter et al., 1992). However, in data sparse regions, this method will not work.

The final method of obtaining the mean flow is to parameterise the flow as a Gaussian jet and then to estimate the parameters of the Gaussian by using the SSH variability (e.g. Kelly and Gille, 1990; Tai, 1990). This method seems to give promising results, although it relies on the mean jet meandering. If the current does not meander, it is impossible to estimate the parameters of the Gaussian and the method does not work.

Hence, in the very regions where in situ data are sparse, there is no satisfactory method of obtaining the absolute dynamic topography. Although useful results, such as eddy statistics and frequency wavenumber spectra, can be calculated from the ocean variability alone, obtaining a picture of the absolute dynamic height field could yield new information about the ocean circulation. Section 3.8.1 outlined the theory behind a method to obtain the best estimate of the absolute dynamic topography available from altimetry alone. This method is now applied to T/P data for 1992 and 1993.

The first step is to establish the mean large scale dynamic topography field (equation [3.27]). This is achieved using the state-of-the-art geoid, kindly made available by Richard Rapp. This is a hybrid geoid, based on the JGM-3 gravity model to degree and
order 70 and the OSU91A geoid from degree and order 71 to 360 (Rapp et al., 1991). The point at which the errors in the geoid overwhelm the dynamic topography is degree and order 14 (Figure 3.29), which corresponds to a wavelength of 2800 km. For each 10 day T/P cycle, the geoid is subtracted from the SSH, and the resulting maps are temporally averaged to obtain the mean dynamic topography (including small scale geoid error). This is shown in Figure 3.31. It can be seen that the general sense of the circulation is consistent with the ideas gained from in situ data, although the small scale features may be attributed to geoid error.

To eliminate this geoid error and obtain the smoothed mean dynamic topography, a Gaussian smoothing scheme is used. The interpolation parameters are a FWHM of 1600 km and a SR of 2000 km. These parameters result in the removal of wavelengths below ~2800 km. The resulting smoothed mean dynamic topography map is shown in Figure 3.32. It can be seen that there is a strong slope from a high at ~30S to the high latitudes. This slope drives the ACC. The northern branch of the Subtropical Gyre and the Agulhas Return Current can also be seen. This general pattern is consistent in a qualitative sense with that observed from in situ data in the region (e.g. Reid, 1989; Peterson and Stramma, 1991).

To obtain the estimate of the absolute dynamic topography, the height anomaly maps (e.g. Figure 3.30) are added into this smoothed mean dynamic topography map. An example of the resulting dynamic topography map for January 1993 is shown in Figure 3.33. This map is also shown in Figures 3.34 and 3.35 where examples from the Agulhas region and the Drake Passage region are shown and are overlaid with the geostrophic velocity field obtained from equation [3.5]. The Agulhas example clearly shows both meanders (e.g. at 35°E, 37°S) and eddy shaped features (e.g. 25°E, 40°S; 19°E, 40°S) with diameters ~300 km. Such structures are known to occur in this region (e.g. Gordon and Haxby, 1990; Naeije et al., 1992; Smythe-Wright et al., 1996). The importance of Figure 3.34 is that it distinguishes between meanders and eddies which have identical signatures in height anomaly maps. Figure 3.35 is a similar plot for the Drake passage region. A meandering flow with a wavelength of ~400 km is visible through the Drake Passage. It is possible that this meandering is the signature of westward propagating Rossby waves being advected eastward by the strong flow through the Drake Passage (Hughes, 1996).

Although the maps resulting from this method are interesting and provide a qualitative picture of the ocean dynamics, it is very difficult to assign error bars to the dynamic topography maps or to the geostrophic velocity fields. The accuracy of this method is investigated in the next section.
Figure 3.31  The mean dynamic topography using the JGM3/OSU91A hybrid geoid.

Figure 3.32  The dynamic topography using the JGM3/OSU91A hybrid geoid and smoothed with a Gaussian smoother to eliminate wavelengths smaller than 2600 km.
Figure 3.33  The sum of the T/P smoothed mean dynamic topography and the height anomaly map for January 1993.

Figure 3.34  The dynamic topography in the Agulhas Region for January 1993, overlaid with geostrophic flow vectors.
Figure 3.35 The dynamic topography in the Drake Passage Region for January 1993, overlaid with geostrophic flow vectors.

Figure 3.36 The error in the method of obtaining absolute dynamic height.
3.8.3 Validation

Attempting to validate these dynamic topography maps is extremely difficult. The data are so useful because there is no other method of obtaining such information about the ocean on a regular basis over such a large area. Furthermore, one would expect the accuracy of the results described in Section 3.8.2 to be dependent on geographical area, since the method will work well in areas where the small scale mean flow is weak. The method will not give good results in areas where a small scale mean flow exists however.

The errors in results from this method derive from three sources; (1) altimetric error (instrument, geophysical correction errors, collocation errors, orbit error correction errors), (2) interpolation and (3) inaccuracies in the method. The errors from (1) and (2) have been discussed in Sections 3.4, 3.5, 3.6 and 3.7. However the accuracy of the above method, regardless of other errors, needs to be assessed.

To obtain a "feel" for the accuracy of this method, data from the POCM is used. A mean dynamic topography map, is obtained by averaging the dynamic topography data from the model over 1993 and 1994. This is the "true" mean dynamic topography. To obtain a smoothed mean dynamic topography map the mean dynamic topography is smoothed in the same way as the altimeter data (Gaussian FWHM of 1600 km, SR of 2000 km). The difference between these two maps is the error in the method. This is shown in Figure 3.36 and the magnitude of the error in the geostrophic velocity is shown in Figure 3.37. It can be seen that in the regions associated with sharp stationary fronts, the method does not perform well. However in regions such as the larger scale northward branch of the Subtropical Gyre, the performance is better. Hence, this method is only really useful for obtaining a qualitative picture of the ocean circulation and must be interpreted with care. Nonetheless, it provides a more informative picture of the ocean circulation than can be obtained from height anomaly maps alone.

3.9 Conclusions

The aim of this thesis is to improve techniques for satellite remote sensing of the ocean. This chapter focuses on techniques relating to satellite altimetry, which is one of the most important sources of remotely sensed oceanographic data. To place the work in perspective, a review of the history of altimetry is given and the reason why altimetry is important for ocean circulation studies is discussed. This is followed by a full discussion of the error budget of TOPEX/POSEIDON which includes a review of instrument errors and geophysical correction errors. It is shown that the resulting r.m.s.
Figure 3.37  The magnitude of the error in the method of obtaining absolute geostrophic velocities.
error of the measurement of the distance between a reference ellipsoid and the
gEOSpheric sea surface is 6.0 cm. This precision is unprecedented. With such a
precision, corrections that were previously considered unnecessary due to their small
magnitude should be reconsidered. One such correction is the compensation for the
variation in the across-track sampling of the altimeter and the errors induced by the
across-track mean sea surface gradients. The r.m.s. correction over T/P cycles 1-52
within the South Atlantic region is shown to be 0.9 cm, with extreme values of ~10-20
cm. Although these corrections seem small, the effect is largest in regions of large
across-track mean sea surface gradient. These regions are likely to be regions associated
with large sea floor topography gradient, and hence with ocean currents and ocean
current variability. Therefore, although over most of the SA region the correction is not
worth applying, in the most interesting regions it is important to apply such a correction
in order to avoid interpreting across-track variations as mesoscale signals. It is
demonstrated that the correction reduces SSH variability in regions of strong MSS
gradient, which is evidence that the correction is working. An example is given where a
mesoscale eddy like structure is shown to be purely an artefact of across-track
 variations in the altimeter sampling. Applying the across-track correction removes this
artefact and, now that the error budget of T/P is so small, it is an important and relevant
technique.

In all of the measurements associated with satellite altimetry, the one that has shown the
largest improvement over the past few years is that of orbit error determination. T/P
orbits are almost two orders of magnitude better than Geosat orbits. This vast
improvement means that the standard orbit error correction techniques should be re-
evaluated. The most common techniques to remove orbit error in a non-global region
are collinear techniques. These involve modelling the orbit error as a long-wavelength
function and fitting this function to height residual profiles to remove the long-
wavelength orbit error. Although theory exists to predict the errors in such an approach,
there has not been a study clearly stating which long-wavelength function should be
used for the orbit error removal and how this depends on the altimeter track length. This
is investigated with the use of ocean model data from the POCM for several common
orbit-error removal methods. It is demonstrated that, within the South Atlantic region,
the most accurate method to use is a bias correction for tracks shorter than 6200 km and
a tilt+bias correction for longer tracks. Even though the artificial orbit error is
introduced as a sinusoid with a wavelength of 40000 km, the orbit error correction
method that attempts to fit such a function is generally less accurate than the bias and
tilt+bias methods.
For purposes such as performing statistical tests or visualising data, a regular gridded dataset is essential. Altimeter data is irregular and it is necessary to apply interpolation in order to obtain a regular grid. Although interpolation is a very important topic, little research has been conducted on the most effective interpolation techniques and on the effects of interpolation on altimeter data. Gaussian interpolation is used in this study as a good compromise between the computational intensiveness of optimal interpolation techniques and the simplicity of moving average interpolation. The effects of varying interpolation parameters are investigated by using model data from the POCM. It is found that the optimal interpolation parameters are very sensitive to the amount of long-wavelength (e.g. orbit error) noise in the altimeter data. Removing the orbit error prior to interpolation results in lower r.m.s. errors and higher correlations and is therefore beneficial. The optimal FWHM with 5 cm orbit error, removed using bias and tilt+bias techniques, is ~100 km. The geographical variation of the interpolation errors is also studied. It is found that these errors can result in eddy like structures in SSH variability maps that are not present in the control data. The highest r.m.s. errors in the interpolation are in regions of high variability away from altimeter tracks. The lowest r.m.s. errors are in regions of low variability coincident with altimeter tracks. In terms of percentage r.m.s. error (i.e. the noise to signal ratio) the largest areas are found in quiescent regions directly adjacent to very energetic regions. All of these results will aid in determining interpolation parameters and in interpreting interpolated altimeter data, as well as in highlighting the errors present in interpolation.

The main obstacle for the use of altimeter data in oceanographic studies is the inability to determine absolute geostrophic velocities in the absence of in situ data. A method is presented that will yield the best possible method of obtaining absolute velocities using altimetry alone. The method is to add the large scale mean dynamic topography to height anomaly maps. The resulting dynamic topography is accurate in areas where the small scale mean flow is weak. Realistic flow structures are observed in the Drake Passage and Agulhas regions. The validity of this method is tested by applying it to SSH data from the POCM. It is demonstrated that the errors are large and in the most interesting regions. Hence this method is only useful for qualitative studies.

Across-track correction validation demonstrates the utility of this correction and hence this correction is used for the all the T/P data processing described in this thesis. The method for obtaining absolute velocities, although qualitative, inspired the research in Chapter 5 by providing the first evidence that SST and SSH may be related. As a result of the study into the effects of interpolation on altimeter data, the work described in Chapter 5 does not use interpolated data and therefore avoids the errors inherent in interpolation.
All of the above research into altimetric techniques should aid the oceanographic remote sensing community. It should ensure that the highest quality altimeter data possible is used, promote awareness of the effects of interpolation error, enable choice of an appropriate orbit error removal method and lead to the attainment of a qualitative picture of the absolute surface geostrophic circulation.
Chapter 4

Measuring Sea Surface Temperature from Infrared Radiometry

4.1. Introduction

Sea surface temperature (SST) is one of the most important parameters in climate and ocean circulation research. In combination with near surface parameters such as humidity, wind speed and air temperature, SST controls the release of heat and water from the ocean to the atmosphere. SST is also an indicator of events such as the Indian Monsoon and the El Niño Southern Oscillation (ENSO) (Bigg, 1995) which can have devastating effects on regional weather and climate (WMO, 1995). Hurricane activity causes thousands of deaths every year and SST is thought to be a dominant factor in the abundance of hurricanes (Saunders and Harris, 1996). SST is also an extremely important parameter in detecting global warming; the large heat capacity of water means that fluctuations in the SST are much less than those in air, and hence estimation of long term trends in temperature from SST is more efficient. Measurements of SST are therefore essential for climate research and global absolute accuracies of $\sim 0.3$ K averaged over an area of $2^\circ$ and a time period of 15 days are required (WCRP, 1985). Conventional methods of measuring SST using ship or buoy mounted instruments can meet the accuracy requirement, but not the sampling requirement. These methods provide data that are usually concentrated near shipping lanes, but are sparse in the southern hemisphere (e.g. Bottomley et al., 1990).

SST has many uses in the study of ocean circulation. It can be used to track features and therefore provide advective velocities (e.g. Kelly, 1989), to provide statistics on the spatial and temporal variability of the ocean circulation (e.g. Olson et al., 1988) and to identify ocean fronts (e.g. Legeckis, 1978; Vazquez et al., 1990). In contrast to the requirements for climate research, the absolute accuracy of SST measurements is not usually as important as the relative accuracy. To observe the ocean mesoscale, high relative accuracies over a length-scale of 10-50 km and a time scale of 10-20 days are needed. The strength of the surface signature of ocean fronts and eddies depends on the geographical location, season and local meteorological conditions (Legeckis, 1978) and therefore a high relative accuracy is required to detect such phenomena. A lower limit
on the accuracy required can be derived by considering typical conditions. In areas of mesoscale variability, typical values for the SST gradients are $\sim 0.01$ K/km$^1$. Hence to distinguish these signatures over a distance of 50 km requires a relative SST accuracy of at least $0.01 \times 50 / \sqrt{2} = 0.3$ K. Moreover, global measurements at such a resolution and accuracy are required to aid campaigns such as WOCE and TOGA and to place in context the rather blurred "snapshot" gained by the in situ data.

The only source of global SST measurements approaching the accuracies required for ocean circulation and climate research are spaceborne infrared measurements of SST$^2$. A brief review of the history of infrared SST measurements is given in Section 4.2. Until 1991, the only source of infrared SST data was from the Advanced Very High Resolution Radiometer (AVHRR) instruments mounted on the series of operational NOAA satellites. Data from these instruments has an accuracy of $\sim 0.6$ K, after calibrating the SST retrieval algorithms against buoy data (McClain et al., 1985). In July 1991, a new source of SST data became available through the Along-Track Scanning Radiometer (ATSR) instrument mounted on the ERS-1 satellite. This instrument has several design advantages over AVHRR and has the potential to yield an extremely accurate high resolution global SST dataset. However, the ATSR instrument is experimental and therefore the accuracy of the data must be assessed before using them for ocean circulation or climate research. A review of the error sources in the measurement of SST from infrared radiometers, with reference to the ATSR spatially Averaged SST (ASST) data is given in Section 4.3. Results from the ATSR validation studies to date are also summarised in this section. Section 4.4 presents evidence for regional cloud contamination of the ASST data. In some regions, this cloud contamination is the dominant error source in these data and must therefore be removed. A scheme to eliminate this cloud contamination is proposed and tested (Jones et al., 1996a), and the seasonality of the contamination is investigated (Jones et al., 1996b). Conclusions and the implications of the cloud contamination are discussed in Section 4.5.

4.2 A brief history of spaceborne infrared SST measurement

Infrared measurements of SST began in the late 1960s with the High Resolution Infrared Radiometer (HRIR) on the Nimbus satellites (Rao, 1968; Curtis and Rao, 1969; Smith et al., 1970). These radiometers had one infrared channel at 3.5-4.1 $\mu$m (limiting

---

$^1$ This figure is obtained by assuming a 1 K temperature change over 100 km; typical values for a mesoscale eddy.

$^2$ Microwave measurements of SST can be made and have the advantage that cloud free conditions are not required. However microwave SSTs are only available at low spatial resolution ($\sim 2^\prime$) and are only accurate to $\sim 2$ K (Grankov and Shutko, 1991), mainly due to emissivity variations caused by changes in wind speed. Such accuracy and resolution does not provide a large advantage to climatology in most regions.
useful measurements to nighttime), a noise equivalent temperature error (NEAT) of ~2 K and a maximum spatial resolution of ~8 km. The high noise and poor spatial resolution limited the use of the infrared images that these instruments provided. In 1972 the Very High Resolution Radiometer (VHRR) was launched on the NOAA-2 platform. These instruments provided increased spatial resolution (IFOV ~ 1km) and reduced instrument noise (NEAT ~ 0.5K) (Legeckis, 1978). Furthermore, the VHRR instruments had two channels, a visible channel at 0.6-0.7 µm and an infrared channel at 10.5-12.5 µm, thus making daytime measurements feasible. As a result of these improvements, ocean fronts could be detected in the infrared BT images (Legeckis, 1978); the first practical use of satellite derived infrared images in oceanography. The VHRR instruments were flown on successive NOAA satellites (NOAA-2 through NOAA-5) until 1978. At about this time, theoretical methods were emerging to correct for the atmospheric attenuation of the infrared radiation emitted from the sea surface (McMillin, 1975). The most promising of these methods was correction by measuring the top of atmosphere radiance at two different infrared wavelengths and using the differential absorption at these two wavelengths to correct for the atmospheric effect - the so called "split-window" method (see Section 4.3.4). Although this method was first proposed in the early 1970s (Anding and Kauth, 1970), it was not until late 1978 that an infrared radiometer with a true "split-window" was launched.

The first AVHRR was launched on the TIROS-N platform in 1978. It had four channels at (1) 0.6-0.9 µm, (2) 0.7-1.1 µm (3) 3.6-3.9 µm and (4) 10.3-11.3 µm. This instrument had NEAT values of ~0.1 K at 11 µm and ~0.2 K at 3.7 µm and the two infrared channels allowed the dual-window technique to be used for nighttime measurements. Unfortunately, the 3.7 µm channel noise increased rapidly after several months, limiting the use of this channel and therefore rendering the split-window technique useless. TIROS-N was followed by NOAA-6 (launched in 1979) which also carried a four channel AVHRR instrument, however the same problems were experienced. In 1981, NOAA-7 carried the first five channel AVHRR into space. This had essentially the same channels as the previous two AVHRR instruments, but had an extra one at 11.5-12.5 µm, allowing the split-window technique to be used (with the 11 µm and 12 µm channels) in both the daytime and nighttime. Furthermore, the accurate atmospheric correction from the split-window technique was no longer reliant on the problematic 3.7 µm channel. Following the launch of this satellite, the multichannel sea surface temperature (MCSST) method became operational and has provided SSTs with accuracies of about 0.6 K (McClain et al., 1985).

The method for obtaining MCSSTs from AVHRR data is to regress infrared BTs against collocated buoy SST measurements to obtain the coefficients of the SST retrieval algorithm. The error variance from such regressions is typically ~0.6 K
(McClain et al., 1985; for a fuller discussion see Section 4.3.4). It is necessary to derive SST in such a way because the calibration of AVHRR is only accurate to ~0.55 K (mainly due to temperature gradients within the calibration blackbodies (Weinreb et al., 1990)).

It was to improve upon the accuracy of the AVHRR instruments and to obtain a stable dataset for the monitoring of SST variability and trends that the ATSR instrument was designed and built. This instrument and its error sources are described in the following section.

4.3 Sources of error in the ATSR ASST data

4.3.1 Instrument and data description

ATSR was designed and built by a consortium led by the UK Rutherford Appleton Laboratory (RAL) together with the Mullard Space Science Laboratory (MSSL), Oxford University, the UK Meteorological Office, the French CRPE and the Australian CSIRO. ATSR is mounted on ERS-1 which was launched on 17th July 1991. (A very similar instrument to ATSR, ATSR-2 is now operating on ERS-2 which was launched on 21st April 1995).

The ATSR instrument has four spectral channels centred at wavelengths of 1.6 μm, 3.7 μm, 10.8 μm and 12 μm, as shown in Figure 4.1. These are located in high atmospheric transmission regions of the atmospheric spectrum. The 1.6 μm channel is uncalibrated and operates only during the day. Its main purpose is in cloud identification. The 3.7 μm channel operates only at night but failed prematurely on May 26th 1992. However, when available, it is always used in SST derivation and in cloud clearing tests.

ATSR has several design improvements over previous infrared radiometers that should lead to a more accurate SST measurement. These are:

(i) Two stable onboard calibration blackbodies designed and built by the Mullard Space Science Laboratory (Section 4.3.2).

(ii) The use of a novel Stirling Cycle cooler to maintain the detectors at a temperature of ~85 K to reduce instrument noise (Section 4.3.3).

(iii) Measurement of the radiance from the same geographical location on the earth's surface through two different path lengths separated by in time by 150 s. This
Figure 4.1 The response functions of the four ATSR channels (upper) together with the atmospheric transmission for three different atmospheres (lower) (taken from Zavody et al., 1995).

Figure 4.2 The ATSR scan geometry, as depicted in Zavody et al. (1995).
results in a more accurate atmospheric correction (Section 4.3.4 and 4.3.5) and allows new cloud screening techniques to be used (Section 4.3.7).

The ATSR scan geometry is shown in Figure 4.2, where it can be seen that two views of the earth's surface are accomplished by sampling the surface in two curved swaths (AB and CD). This is achieved by means of a scanning mirror which rotates at a rate of one revolution every 150 ms. The incidence angle of the nadir scan varies from 0° to 24°, while the incidence angle of the forward scan varies from 53° to 55° (Zavody et al., 1995). The swath width of ATSR is nominally 500 km, with an IFOV resulting in a pixel resolution of 1 km×1 km at nadir and 1.5 km × 2 km at 55°. Measurements of the onboard calibration targets are made in each scan to obtain accurate radiance values from the detector counts.

Processed ATSR data are generated at the Rutherford Appleton Laboratory (RAL) using the SADIST (synthesis of ATSR data into sea-surface temperature) processing scheme. SST data from ATSR are available at two different spatial resolutions: 1 km and 0.5° in latitude and longitude (Zavody et al., 1994b). The 0.5° ASST data are calculated by averaging the cloud free 1 km resolution brightness temperatures from each channel into 10 arcminute cells and obtaining a SST value from these for each 10 arcminute cell (Zavody et al., 1995). These SSTs are then averaged to give the ASST value for the 0.5° cell. Three different averages are provided for each 0.5° cell; the average of the nadir only SSTs, the average of the dual-view only SSTs and a mixed average where dual-view SSTs are used where possible and nadir-view SSTs are otherwise used. Using only the dual-view average is desirable since such ASSTs are more robust to aerosol contamination and have a better atmospheric correction (Sections 4.3.4 and 4.3.5). However, there are approximately 15% fewer dual-view ASSTs than nadir-view only (or mixed) ASSTs because of the requirement for cloud free conditions in both views. Hence the mixed product is used here, which incorporates the advantage of the dual-view whenever possible. The ASST data are more suitable for research over large spatial areas than the 1 km data since the sheer volume of 1 km data prevents use of these data for such studies. Furthermore, the ASST data is at a sufficient spatial resolution to resolve most of the ocean variability. It is these data (version 500) that are used here.

The temporal resolution of the ATSR data depends largely on the ERS-1 orbit phase and cloud cover. Figure 4.3 shows the sampling density in the South Atlantic region for three years of ASST data, from January 1992 to December 1994. In the tropics where the cloud coverage is generally lower (except in the Intertropical Convergence Zone (ITCZ)) than at high latitudes, typical values for the sampling density are two measurements every 10 days. At higher latitudes, the increase in cloud coverage is such
that the sampling density is typically one measurement every 20 days. The effect of the 3 day repeat orbit phase is evident in the diamond shaped regions of alternating high and low sampling density in the tropics. In this orbit configuration, the ATSR scan width is insufficient to provide complete coverage in the tropics and some regions are never sampled. Other regions however are sampled much more frequently, resulting in the pattern shown in Figure 4.3. This sampling has implications for the study of day/night SST differences. These implications are discussed in Section 4.4.

Detailed descriptions of the error sources pertinent to the ATSR data are given in the following sections.

4.3.2 Calibration

The detectors on board an infrared radiometer measure the photon flux integrated over some time period (75 μs for ATSR). The detector "counts" corresponding to this photon flux are converted to radiance by a calibration based on two onboard calibration blackbodies. Each blackbody has seven platinum resistance thermometers mounted just under its surface so that the temperature of the blackbody can be determined to a high accuracy. In each ATSR scan, the detector counts from the cold (263 K) and the warm (303 K) blackbodies are measured. These are converted to radiances by look-up tables that also correct for the small non-linearity of the detectors. The detector counts from the scene measurements for that scan can then be converted to radiance values by linear interpolation between the two calibration values. The radiance values can then be converted to brightness temperatures by the Planck function:

$$B_\lambda(T)d\lambda = \frac{2hc^2}{\lambda^5(e^{hc\lambda/kT} - 1)}$$  \[4.1\]

where h is Planck's constant (6.626×10^{-34} Js), k is Boltzmann's constant (1.381×10^{-23} JK^{-1}), c is the speed of light, $B_\lambda(T)d\lambda$ is the radiance emitted between wavelengths $\lambda$ and $\lambda+d\lambda$ and T is the temperature of the blackbody.

Mason et al. (1996) show that the maximum error in the ATSR SST measurement due to inaccurate calibration is 0.1 K, a vast improvement over the 0.55 K calibration accuracy of AVHRR (Weinreb et al., 1990).
Figure 4.3 The sampling density of the ATSR ASST data calculated from the ATSR data from January 1992 - December 1994.

Figure 4.4 The atmospheric temperature deficit for several different atmospheres (taken from Deschamps and Phulpin, 1980).
4.3.3 Instrument noise

The instrument noise in the ATSR detectors is dependent on the temperature of the focal plane assembly. For this reason, it is essential that the assembly is cooled. Cooling mechanisms on previous missions have used passive radiative coolers. The Stirling cycle cooler on ATSR maintained the focal plane assembly at temperatures of 85 K (Albin Zavody, pers. comm., 1996) at the beginning of the mission to 97 K at the end of 1994 (the dates corresponding to the data used in this study). The resulting noise equivalent temperature errors (NEATs) are 0.03-0.07 K at 11 µm, 0.05-0.13 K at 12 µm and 0.02-0.04 K at 3.7 µm (whilst the 3.7 µm channel was working). Although these figures appear small, for high resolution SST data it is essential that the NEAT figures are as low as possible. The reason is that the atmospheric correction technique uses a linear combination of BTs from the different channels\(^3\) to correct for the effect of the intervening atmosphere and retrieve a SST value (see Section 4.3.4). The errors in the retrieved SST values due to the NEATs depend upon the coefficients in [4.3] on page 109. The NEAT induced error in the SST value is given by:

\[ \sigma^2_{\text{NEAT}} = \sum_{i=1}^{n} (a_i \sigma_i)^2 \]  

[4.2]

Using the coefficients from Zavody et al. (1995) for the high resolution product, together with the NEAT values above, results in the SST errors due to instrument noise shown in Table 4.1.

Table 4.1 The algorithm noise for NEAT values spanning the ATSR mission lifetime.

<table>
<thead>
<tr>
<th></th>
<th>Two Channel (\sigma_{\text{NEAT}}, \text{(K)})</th>
<th>Three Channel (\sigma_{\text{NEAT}}, \text{(K)})</th>
<th>Four Channel (\sigma_{\text{NEAT}}, \text{(K)})</th>
<th>Six Channel (\sigma_{\text{NEAT}}, \text{(K)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tropics</td>
<td>0.19-0.47</td>
<td>0.03-0.08</td>
<td>0.37-0.91</td>
<td>0.06-0.13</td>
</tr>
<tr>
<td>Mid Latitudes</td>
<td>0.17-0.44</td>
<td>0.05-0.12</td>
<td>0.34-0.85</td>
<td>0.05-0.12</td>
</tr>
<tr>
<td>High Latitudes</td>
<td>0.12-0.29</td>
<td>0.08-0.21</td>
<td>0.10-0.25</td>
<td>0.05-0.12</td>
</tr>
</tbody>
</table>

Table 4.1 shows that the NEAT induced SST errors vary according to algorithm and geographical region. The errors are smallest for the three channel and six channel algorithms and largest for the two and four channel algorithms in the tropics. These errors are only correct for the ATSR high resolution product. For the spatially averaged

---

\(^3\) When referring to atmospheric correction algorithms, the word "channel" refers to a source of BT, rather than a spectral region. In this context, ATSR has eight channels corresponding to the two different views of each wavelength region.
product, the BTs are averaged into 10 arcminute cells before applying the atmospheric correction. Hence the NEATs are reduced to such small values that even the noise amplification in the retrieval algorithm does not cause significant SST errors. For example, the largest noise induced SST error in Table 4.1 is for the four channel algorithm in the tropics. Assuming that each 10 arcminute cell is 25% full, results in approximately 70 1 km pixels (out of a maximum of 324 and a minimum of 1) contributing to the SST retrieval for that cell. Using retrieval coefficients pertinent to the ASST data, the noise induced SST error in the four channel algorithm in the tropics reduces to 0.14 K. Assuming that three 10 arcminute cells contribute to the ASST average (out of a maximum of 9 and a minimum of 1) results in a further reduction in the error to 0.07 K. Although this error is a function of the amount of 1 km pixels contributing to the ASST value, for the majority of the ASST dataset the error from instrument noise can be ignored.

4.3.4 Atmospheric correction

In the absence of an atmosphere, the radiance emitted from the earth’s surface could be measured by a spaceborne sensor and converted directly to temperature by inverting the Planck function [4.1]. In reality, an atmosphere exists which absorbs and re-emits the surface radiation. This re-emitted radiation consists of an upwelling component (which is emitted to space) and a downwelling component (a proportion of which is reflected from the Earth’s surface and also emitted to space). The net result of this atmospheric effect is that the radiation emitted to space is less than that emitted from the earth’s surface.

The amount of radiation absorbed and re-emitted depends upon the atmospheric temperature profile and composition. At infrared wavelengths, the dominant and most variable absorber at most latitudes is water vapour. Other gases that have an effect include carbon dioxide (CO$_2$), ozone (O$_3$) and nitrogen (N$_2$) (Závody et al., 1995). Typical values for the temperature deficit due to the atmosphere, as computed by Deschamps and Phulpin (1980) from realistic temperature profiles and a radiative transfer model, are shown in Figure 4.4. It can be seen that there are several wavelength “windows” where the atmospheric effect is small. For example, between 3.5-4.1 $\mu$m and 10.5-12.5 $\mu$m, the atmospheric temperature deficit is generally less than ~5 K. This varies seasonally and with latitude, depending mainly on the water vapour content of the atmosphere. In the tropics, the atmospheric deficit can be as large as ~10 K at 12.5 $\mu$m, whereas in high latitude winters the deficit can be as low as ~1 K.

Early attempts to measure SST from infrared BTs relied on climatological estimates of the vertical distribution of water vapour and temperature. Such estimates of SST have
accuracies of ~2 K which do not provide a great advantage over SST climatologies in many regions. A method to improve on the climatology based corrections was first proposed by Saunders (1967a). Using measurements from an aircraft mounted radiometer, he found that the atmospheric effect could be deduced by observing the same area of ocean at two different angles (and therefore through two different atmospheric path lengths). Anding and Kauth (1970,1972) proposed a method also based on differential absorption, but using different wavelength "windows" rather than different viewing angles. These papers, together with work of Prabhakara et al. (1974) led to the now commonly used technique of using a linear combination of BT channels (T_i) to derive SST (T_0) (Deschamps and Phulpin, 1980).

\[ T_0 = a_0 + \sum_{i=1}^{n} a_i T_i, \quad \text{where } \sum_{i=1}^{n} a_i = 1 \]  \[4.3\]

This equation is the basis of SST retrieval in nearly all operational SST algorithms. The coefficients in [4.3] are derived for ATSR by using a radiative transfer model and a set of representative temperature and water vapour profiles (obtained from radiosondes). The coefficients follow by using a multiple linear regression technique, minimising:

\[ \sigma_T^2 = \frac{1}{m-n-1} \sum_{j=1}^{m} [T_0^{\text{mod}}(j) - T_0^{\text{true}}(j)]^2 \]  \[4.4\]

Here m is the number of samples, n is the number of channels, T_0^{\text{mod}} is the model generated SST given by [4.7] and T_0^{\text{true}} are the SSTs associated with each water vapour/temperature profile (typically T_0^{\text{true}} are the profile air surface temperature, the air temperature ± 3 K and the air temperature ± 5 K, hence m = 5 times the number of profiles in this case). The resulting \( \sigma_T^2 \) is the internal accuracy of the algorithm.

The \( \sigma_T \) for the ASST algorithms, taken from Závody et al. (1995) and averaged over the range of across-track distances, are given in the table below.

**Table 4.2** The intrinsic error in the ASST algorithms (taken from Závody et al., 1995).

<table>
<thead>
<tr>
<th></th>
<th>Two Channel ( \sigma_T ) (K)</th>
<th>Three Channel ( \sigma_T ) (K)</th>
<th>Four Channel ( \sigma_T ) (K)</th>
<th>Six Channel ( \sigma_T ) (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tropics</td>
<td>0.50</td>
<td>0.12</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Mid-Latitudes</td>
<td>0.37</td>
<td>0.10</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>High-Latitudes</td>
<td>0.08</td>
<td>0.07</td>
<td>0.04</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 4.2 shows that the most accurate algorithm is the six channel ASST algorithm. This gives $\sigma_T$ values of 0.01-0.03 K depending on the latitude band. In comparison, the standard split-window algorithm has $\sigma_T$ values of 0.08 K at high latitudes and 0.5 K in the tropics (Závody et al., 1995). The triple-window and the split-window dual-view algorithms give similar performances with $\sigma_T$ values of ~0.1 K.

The total error in the SST retrieval from instrument noise and atmospheric correction effects is the sum of [4.2] and [4.4]. In order to reduce the total error, rather than just [4.4], the noise on each channel is included in the regression for the high resolution ATSR product. The resulting coefficients are those that minimise the sum of [4.2] and [4.4]. If a SST estimate over a large area (e.g. 18 km by 18 km) is acceptable, then the method is to average the BTs first (which reduces the random noise to a small value - see Section 4.3.3) and then to determine the coefficients by minimising [4.4] alone. Such is the case for the ATSR ASST product.

The errors in the atmospheric correction given in Table 4.2 are the "internal errors" of the algorithm and are due partly to the assumptions that led to the derivation of [4.3] not being completely valid, and partly due to the validity of the radiosonde profiles used. In particular, the assumption of a high transmission "thin" atmosphere is invalid in the tropics (e.g. Harris and Mason, 1992) and causes a large error (~0.5 K) in the two channel algorithm at these latitudes. If more channels are available, then the addition of extra channels, whether in the form of extra wavelength windows or different viewing angles, minimises the effect of non-linearities caused by the breakdown of the "thin" atmosphere assumption, and a more accurate atmospheric correction results.

The discussion above assumes a clear sky (i.e. no aerosols or clouds). If clouds are present it is impossible to correct for the effect they have on the BTs and a SST retrieval cannot be made. Hence the problem is identifying when clouds are present, rather than correcting for them. Cloud clearing is discussed in Section 4.3.6.

4.3.5 Aerosol contamination

Aerosols (tropospheric and stratospheric) also have an effect on the observed BTs, and hence on the SST, through scattering of the infrared radiation. Walton (1985) in a modelling study, investigates the effect that stratospheric aerosols have on SST retrieval when three different wavelength channels are available (3.7 µm, 11 µm and 12 µm). He finds that the BT deficit due to stratospheric aerosols is largest in the 11 µm channel and smallest in the 12 µm channel. He compares three algorithms, the standard split-window (11 µm and 12 µm), the triple window (3.7 µm, 11 µm and 12 µm) and the dual window (3.7 µm, 11 µm) and finds that aerosol induced errors are greatest (~3 K)
Chapter 4 Measuring Sea Surface Temperature from Infrared Radiometry

in the split-window and smallest (~1 K) in the dual window. From these results, he proposes a correction scheme that utilises the difference in the aerosol robustness of different algorithms. He proposes:

\[ T_0 = T_{alg1} + C(T_{alg1} - T_{alg2}) \]  

where \( T_{alg1} \) and \( T_{alg2} \) are the SST estimates from different algorithms, \( T_0 \) is the true SST and \( C \) is a constant. Equation [4.5] has a very similar form to the split-window algorithm. The difference here is that instead of using differential absorption in the two BT channels to correct for water vapour effects, the differential effect on the SST algorithms is used to correct for aerosol contamination. Walton (1985) finds that if the split-window is algorithm-1 and the 3.7 \( \mu \text{m} \)-11 \( \mu \text{m} \) is algorithm-2, the value of \( C \) (from a modelling study) that minimises the error due to stratospheric aerosols is 0.4-0.5. This relationship is confirmed by studying real data when a value of \( C \) of 0.42±0.15 is obtained. Unfortunately, the high noise on the 3.7 \( \mu \text{m} \) channel for the majority of the AVHRR instruments prevents the use of this technique for most of the AVHRR dataset.

Závody et al. (1994a) perform a similar study to that of Walton (1985), but they use a more realistic radiative transfer model and channels more appropriate to ATSR. They find that the six channel algorithm is least affected by the presence of stratospheric aerosols; aerosol contamination reducing the SST by ~0.3 K. The split-window algorithm is affected most with the SST decreasing by ~1 K. An algorithm identical to that proposed by Walton (1985) to correct for the aerosol error is given in Závody et al. (1994a). In this case however, the dual-view capability of ATSR allows differences between dual-view and single-view algorithms to be used to reduce aerosol contamination and the 3.7 \( \mu \text{m} \) channel is not necessary. This method needs to be validated however.

The above discussion has concentrated on stratospheric aerosols such as those caused by volcanic eruptions. Závody et al. (1995), in a modelling study, show that tropospheric aerosols affect SST retrievals in a different way. To simulate the effect of an increased tropospheric aerosol content, they decrease the ground level visibility in their model from 100 km to 23 km. This shows that, unlike the stratospheric aerosols, the effect is largest in the 3.7 \( \mu \text{m} \) channel and least at 12 \( \mu \text{m} \). The resulting error in the SST retrievals is least in the six channel algorithm (SSTs decreased by ~0.08 K) and greatest in the triple-window algorithm (SSTs decreased by up to 0.7 K in the tropics). SSTs retrieved from the split window algorithms are decreased by ~0.2 K.

Finally, a recent study by Brown et al. (1996) indicates that it is possible to obtain a completely aerosol robust algorithm from the dual-view capability of ATSR. Their
method is to vary the aerosol loadings as well as the temperature and water vapour content, to obtain the coefficients in [4.3] by the standard regression technique. This gives aerosol robust SSTs without the need to calculate SST from two different algorithms, as is necessary in the method proposed by Walton (1985) and Závody et al. (1994a). Furthermore, this should be more accurate than such differencing techniques, because differencing the SST from two different algorithms must always increase the noise from that in either algorithm alone.

4.3.6 Cloud contamination

Of all the error sources in the infrared measurement of SST, the one with the largest potential for error is inadequate identification of cloudy pixels. Numerous cloud screening methods have been developed for both daytime and nighttime infrared imagery but the elimination of cloud remains a notoriously difficult problem. A review of several cloud clearing methods is given in this section, starting with the cloud clearing scheme used by RAL to process the ATSR data.

(a) Cloud clearing in the ATSR ASST data

The cloud screening tests implemented by RAL are based on the work of Saunders (1986) and Saunders and Kriebel (1988a,b). There are eight tests (Závody, pers. comm., 1994), as summarised below:

(i) Gross cloud test (day and night). This is a basic test which compares the 12 μm BT of each pixel with a threshold value determined by the time of year, latitude and viewing angle. Data with temperatures below the threshold are rejected as cloudy.

(ii) Spatial coherence test (day and night). Cloud contamination can produce high spatial variability in BTs, whereas usually in small areas of the ocean such temperature variability is small. This property can be used to eliminate cloud by examining the SD of the 10.8 μm BTs within a 3 km by 3 km array of pixels. If the SD exceeds a certain threshold, all nine pixels are flagged as cloudy. To prevent rejection of data over areas of high natural SST variability (for example frontal areas), a second scan through the data is performed. If more than half the pixels surrounding the flagged pixels are cloud free, the mean 11 μm-12 μm BT difference is calculated for the 3 km by 3 km array. The pixels are only rejected as cloudy if this mean exceeds a threshold.
(iii) 1.6 $\mu$m channel test (day only). This channel is sensitive to reflected solar radiation and therefore radiation reflected from liquid water is detected and a high 1.6 $\mu$m signal is used as an indicator of cloud presence. A histogram of 1.6 $\mu$m detector counts over a 33 km by 33 km area is obtained. If the histogram has a clear cloud free peak and is uncontaminated by sunglint, then all points lying further than a certain distance from the cloud free peak are rejected. If the histogram has no cloud free peak or is contaminated by sunglint, a spatial coherence test on 2 km by 4 km pixel arrays is used to distinguish cloud contaminated data. (The 4 km is in the along-track direction since sunglint can cause the 1.6 $\mu$m counts to vary in the across-track direction.)

(iv) Thin cirrus test (day and night). The optical properties of cloud change with wavelength and hence differences in brightness temperatures between channels can be used to infer the presence of cloud. This is a powerful test for detecting semi-transparent cirrus cloud as the emissivity of cirrus is different at 10.8 $\mu$m and 12 $\mu$m. In order to eliminate thin cirrus, the 10.8 $\mu$m-12 $\mu$m BT difference is compared to a threshold value that varies with the viewing angle and the 10.8 $\mu$m BT.

(v) Forward/Nadir view difference test (day and night). This test is applied to each full swath image (512 km by 512 km). A graph of the nadir 10.8 $\mu$m - forward 10.8 $\mu$m brightness temperature difference against nadir 10.8 $\mu$m - nadir 12 $\mu$m brightness temperature difference is obtained. In cloud free conditions, this graph gives a straight line corresponding to a correlation between the view difference and the 10.8 $\mu$m - 12 $\mu$m difference caused by atmospheric absorption. In cloudy conditions, clusters of cloudy points appear lying off the cloud free line due to cloud being observed in one view but not the other. A correct line is obtained from an atmospheric model, and if points deviate by more than a certain amount from this line they are rejected as cloudy. (Prior to the failure of the 3.7 $\mu$m channel, the nadir 3.7 $\mu$m - forward 3.7 $\mu$m BT difference was plotted against the nadir 11 $\mu$m - nadir 3.7 $\mu$m BT difference for nighttime cloud screening.)

(vi) Infrared histogram test (day and night). A histogram of the 11 $\mu$m -12 $\mu$m BT difference is obtained for each 512 km by 512 km image area. Small differences with respect to the peak position are caused by differing atmospheric absorption, whilst larger differences are caused by the optical properties of clouds changing with wavelength. Histogram points with too large a difference are thus rejected as cloudy.
(vii) Fog/low stratus test (night only, before the 3.7 μm channel failure). This is the same as test (vi) except that the 10.8 μm-3.7 μm value is used. This test is effective at detecting fog and low cloud since the emissivity of these is 10% less at 3.7 μm than at 10.8 μm and the resulting BT difference is usually larger than the difference due to atmospheric absorption.

(viii) Medium/high level cloud test (night only, before the 3.7 μm channel failure). This is the same as test (vi) except that the 12 μm-3.7 μm value is used.

(b) Other cloud clearing techniques

The techniques to eliminate cloud contaminated data described in the previous section are arguably the most comprehensive set of cloud clearing tests applied in a global operational SST retrieval scheme. Improving upon these tests at the BT level is difficult, although new tests are currently being developed at RAL (Závody, pers. comm., 1996). One technique which shows promise, but is probably too computationally expensive to implement in a global operational sense, is the technique of Gallaudet and Simpson (1991). This is developed specifically for AVHRR data and involves finding the principal components of the differences between the BTs of the three different infrared channels for each image. The resulting principal component transformed image is then segmented using a split-and-merge clustering algorithm. From this, clusters corresponding to cloudy data can be identified, and by retransforming the segmented, transformed, differedenced image, cloudy pixels can be identified.

Although the techniques described in Section 4.3.6a are comprehensive, no checks beyond these are applied to the ATSR data. In using AVHRR data, investigators rarely rely on such cloud clearing techniques alone. Typically some post-processing technique is applied to the SST data as a final check for cloud contaminated data. Several methods for such post-processing are described below.

The warm pixel composite approach (e.g. Olson et al., 1988) involves taking data over a short time period (for example, 10 days) and only using the warmest SST measurement within that time period. Since cloud contamination usually lowers the SST measurement, the warmest SST value is least likely to be cloud contaminated. This approach has a number of problems, one of which is that this method will give a positive bias to the data, since the coolest SST data are rejected rather than contributing to a mean. Another problem for systematic removal of cloud contaminated data is that the number of SST measurements within each time period will vary, from one extreme where there is only one SST value so a composite can not be formed and the value must
be rejected, to the other extreme where there are many SST values, yet only the information from the warmest value is used. The number of values one expects within a certain time period will depend on the sampling pattern and cloud cover. Hence when and where one expects to find more clouds and therefore more chance of cloud contamination, there is less likely to be a sufficient number of SST values within a given time period to enable use of this warm pixel composite approach.

Comparison of satellite SST values with a climatological SST value for that time and location is another method of eliminating cloud contaminated data. This is a good method for rejecting the worst of the cloud contaminated data. The problem in comparing with ship-based climatologies is that in regions where satellite derived SSTs are most informative (e.g. the southern hemisphere), the climatology is likely to be fairly sparse and heavily interpolated. Furthermore, ATSR measures the skin SST whilst ship measurements are of the bulk SST. This would introduce a bias between the two SST values. Comparison with other satellite based SST climatologies, such as that obtained by AVHRR, does not suffer from the sampling problem, but has problems caused by calibration drifts of AVHRR data, biases introduced due to the different SST retrieval algorithms used for AVHRR and cloud contamination of the AVHRR dataset. Hence for eliminating anything but the worst of the cloud contaminated data, comparison with climatological SST is a poor approach.

A more satisfactory method would be to derive a climatology from ATSR day ASST data and then to use this to compare to night and day ASSTs. The rejection criterion could be based on the SD of the monthly ASST value. This would eliminate any biases between climatology and dataset and would be based on the assumption that day data are relatively uncontaminated compared to night data. The only problem with this approach is that a lengthy time series is necessary to obtain enough data in each month to make the climatological mean and SD valid. Two years of data are insufficient for this purpose since the maximum number of ATSR day ASST observations in the region of interest is approximately 120, giving an average coverage of 10 values per month. However this will mean less values in winter months and more in summer due to seasonal variation of cloud cover. This small quantity of ASST measurements could be helped by increasing the grid size of the climatology.

A diagnostic that could be used to detect cloud contamination is the number of 10 arcminute cells contributing to each 0.5° average. One might expect that where there are few 10 arcminute cells contributing to the average, this could indicate that the region is cloudy and that the data are therefore untrustworthy. No correlation between the number of 10 arcminute cells used and the presence of cloud contaminated data could
be found in this study, and hence this method could not be used as a method of rejecting cloud contaminated data.

Another possible approach would be to compare day data within a time series to the most coincident night data and then to reject the night data if the difference is greater than a certain threshold. This approach has the disadvantage that where conditions are most cloudy there is less likely to be day and night data close in time. It is also susceptible to any small amount of day cloud contamination as well as differences in the daytime data due to diurnal warming of the sea surface.

In summary, there are disadvantages to most traditional methods of post-processing SST data to remove remnant cloud contamination. In Section 4.4 a new technique for post-processing the data is described and tested.

4.3.7 Surface effects

The ATSR SST measurements are of the temperature of the top few 100 microns of the ocean surface. Experimental studies show that a temperature difference can exist between this top layer of the ocean and the layer at some depth. This is the "skin" effect (Robinson et al., 1984). Although not an error source as such, if measurements of the bulk temperature are required (for example, for use in flux formula based on bulk SST measurements), then the skin effect must be corrected for. For ocean circulation studies the absolute value of the skin effect is not as important as its spatial and temporal variability. If the spatial and temporal scales on which the skin effect occurs are decoupled from the variability of the bulk temperature, this will cause problems for ocean circulation studies. Unfortunately, the skin effect remains essentially an unsolved problem. A study by Kent et al. (1996) has shown that efforts to model the skin effect have not improved since the work of Saunders (1967b). An international effort to improve this situation is underway in the Combined Action for Study of the Ocean Thermal Skin (CASOTS) programme. The most recent experimental results (Donlon, 1994; Wick, 1995) suggest that the magnitude of the skin effect is 0.3-0.4 K (skin cooler than bulk). Work by Donlon (1994) also shows that the skin temperature is more strongly correlated with the bulk temperature on larger length scales (>150km).

An effect distinct from the skin effect, but related and often confused with it, is that of the daytime warming of the top few metres of the ocean. This is the "diurnal thermocline" (Stommel et al., 1969; Price et al., 1986) effect and occurs in regions of low wind speed and high solar flux. The magnitude of this effect can be of the order of several degrees in rare situations, although it is typically of the order of a few tenths of a degree (Stramma et al., 1986). Unlike the skin effect, the diurnal thermocline is usually
only present during the daytime. Strong cooling at night and the associated convective mixing rapidly eliminates the effect. Hence for studies of ocean circulation, nighttime SST data is likely to more accurately represent the temperature structure at depths greater than 1-2 m.

4.3.8 Validation studies

The above sections have discussed the error sources in the spaceborne infrared measurement of SST. The exact magnitude and variation of several of the error sources is not accurately known (e.g. aerosols and cloud contamination) and it is therefore extremely important to obtain ground truth data to validate the satellite measurements. This section focuses on validation studies relevant to the SADIST algorithms, since these are the ATSR products most widely used by the oceanographic community. Indeed, it is impossible to apply different algorithms to the ASST dataset because the BTs are not yet available. Hence the results described in this section represent the actual accuracy of the SADIST 500 products, rather than the potential accuracy of ATSR with improved algorithms.

The ATSR SSTs are skin temperatures and this should be remembered when attempting to validate ATSR by comparison to *in situ* bulk SSTs. Since the behaviour of the skin effect is not well known, it is impossible to accurately validate ATSR with bulk measurements. The best method for validation is by *in situ* measurements of skin temperature using ship or aircraft mounted radiometers. Unfortunately high accuracy (~0.1 K) infrared radiometers are scarce and the strict temporal and spatial match-up conditions (<10 km, <2 hours; Minnett, 1991), coupled with the requirement for cloud free skies, makes the attainment of a statistically meaningful set of validation data extremely difficult. The ATSR validation attempts to date are summarised in Table 4.3.

It can be seen from Table 4.3 that, to date, there have only been three validation studies using radiometers. The remaining three use bulk SSTs, and in these studies the skin effect is an inherent geophysical limit to the accuracies that can be obtained. The spread of different results demonstrates the need to perform global validation over a suitable period of time. For example, the skin results of Barton et al. (1995) indicate that the bias in any RAL ATSR algorithm is no greater than 0.21 K and the scatter no greater than 0.50 K. Smith et al. (1994), in contrast, present results indicating that a large bias ~2 K exists in the ATSR single view SSTs. They deduce that stratospheric aerosol contamination is the cause. The results of Barton et al. (1995) indicate that the RAL six channel algorithm may be biased warm. This is also suggested by the results of Mutlow et al. (1994). Thomas and Turner (1995) do not find a warm bias in their results for the six channel algorithm. This may be due to the effects of cloud contamination in their
Table 4.3 A summary of the peer reviewed ATSR validation studies to date. Only results for the RAL algorithms are shown.

<table>
<thead>
<tr>
<th>Study</th>
<th>Region</th>
<th>Time</th>
<th>in situ Data</th>
<th>ATSR Data (Ver.500)</th>
<th>Algorithm (RAL)</th>
<th>Day result ATSR-in situ (K)</th>
<th>Night Result ATSR-in situ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutlow et al., 1994</td>
<td>Global</td>
<td>15/4/92-15/5/92</td>
<td>Bulk</td>
<td>Dual</td>
<td>-0.36 ± 0.42</td>
<td>-0.03 ± 0.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5°</td>
<td>Nadir</td>
<td>-0.58 ± 0.47</td>
<td>-0.46 ± 0.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5°</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith et al., 1994</td>
<td>10°S-0° 25°W-5°W</td>
<td>1/11/91-8/11/91</td>
<td>Skin</td>
<td>Dual</td>
<td>-0.63 ± 0.15</td>
<td>-2.12 ± 0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td>Nadir</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forrester and Challenor, 1995</td>
<td>62°N-64°N 7°W-4°W</td>
<td>20/9/91-24/9/91</td>
<td>Bulk</td>
<td>Dual</td>
<td>-0.41 ± 0.38</td>
<td>-0.18 ± 0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td>Nadir</td>
<td></td>
<td>-0.48 ± 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5°</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barton et al., 1995</td>
<td>31°S-17°S 145°E-168°E</td>
<td>2/9/91-5/12/91</td>
<td>Bulk</td>
<td>Dual</td>
<td>-0.43 ± 0.33</td>
<td>-0.37 ± 0.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td>Nadir</td>
<td></td>
<td>-0.33 ± 0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td></td>
<td></td>
<td>0.20 ± 0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td></td>
<td></td>
<td>-0.06 ± 0.36</td>
<td></td>
</tr>
<tr>
<td>Harris et al., 1995</td>
<td>Global</td>
<td>1/2/92-1/4/92</td>
<td>Bulk</td>
<td>Dual</td>
<td>-0.19 ± 0.62</td>
<td>-0.39 ± 0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thomas and Turner, 1995</td>
<td>45°S-39°N 57°W-14°W</td>
<td>16/10/91-19/5/92</td>
<td>Bulk</td>
<td>Dual</td>
<td>-0.60 ± 0.62</td>
<td>-0.60 ± 0.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td>Nadir</td>
<td></td>
<td>-1.42 ± 0.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td></td>
<td></td>
<td>-1.05 ± 0.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td></td>
<td></td>
<td>-0.65 ± 0.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1km</td>
<td></td>
<td></td>
<td>-1.10 ± 0.71</td>
<td></td>
</tr>
</tbody>
</table>
results, as suggested by the presence of a larger scatter in the night results compared to the day comparisons. The large improvement in the results of Thomas and Turner (1995) when using the dual-view algorithm suggest that, like Smith et al. (1995), aerosol contamination is also degrading their results.

Differences can be seen in Table 4.3 between skin comparisons and bulk comparisons. A reduction in the bias of ~0.2 K is present in the skin results for the two studies using both skin and bulk measurements (Barton et al., 1995; Thomas and Turner, 1995). The scatter in the skin comparisons is similar or worse than the scatter in the bulk comparisons for these two studies. This can be attributed to in situ radiometer error.

In summary, Table 4.3 shows that in some regions with cloud free conditions and no aerosol contamination, the SADIST ATSR SST products can provide point measurements of SST with a bias of <0.3 K and a scatter (SD) of < 0.3 K.

If BTs are available, one is free to use any algorithm to produce a SST measurement. In this case, a global comparison to drifting buoys using a variety of algorithms by Harris and Saunders (1996) show that ATSR can achieve a global accuracy of better than 0.3 K (both in terms of bias and scatter).

### 4.4. Regional cloud contamination of the ASST data

The quality of the ASST product is examined by studying ASSTs produced for the South Atlantic region for the three year period from the 1st January 1992.

#### 4.4.1 Different day and night signals

Figure 4.5 shows the mean SST for January 1992 - December 1994. Most of the dynamical regimes discussed in Chapter 2 are visible in this figure. For example, the Brazil/Falklands confluence can be seen, where the cold northward flowing Falklands current meets the warm southward flowing Brazil current. The upwelling off the west coast of South Africa is clearly visible as a region of cooler SSTs. The Agulhas retroflection is also visible, as is the strong meridional temperature gradient associated with the Antarctic Circumpolar Current (ACC).

The variability (defined as the standard deviation) plot for the whole South Atlantic region is shown in Figure 4.6, with shading indicating the level (°C) of the ASST variability. Several arc-shaped structures of high variability are immediately visible, prominent examples being in the Southern Ocean around 55°S,25°E and 55°S,50°W.
Figure 4.5  ATSR mean ASST for the South Atlantic region from January 1992 - December 1994.

Figure 4.6  The variability (standard deviation) of the ATSR ASST data in the South Atlantic region from January 1992 - December 1994.
Less pronounced examples occur at 20°S, 30°W and at 45°S, 30°E. As the width of each of these structures corresponds to the ATSR scan width (512 km), the high variability in these regions is almost certainly due to rogue data. This is confirmed by close examination of the data which shows that the high variability is caused by a rogue point several degrees higher in temperature than the mean temperature for that area. It is later demonstrated that filtering out such rogue data is straightforward. The reason for such data is unclear, but is probably caused by a "bug" in the SADIST software allowing incorrectly geolocated data to slip through.

Another striking feature of Figure 4.6 is the area of high variability present in the South Atlantic between 30°S and 50°S extending northwards to the equator along the west coast of Africa. This is investigated further by comparing Figure 4.6 with the SST variability computed using the Levitus climatology (Levitus, 1984) shown in Figure 4.7. Inspection reveals similarity between the two figures in many areas (e.g. the high variability band of about 3 K between 25°S and 35°S). However, off the west coast of Africa the ATSR ASST variability is clearly much higher than the Levitus SST variability (e.g. at 10°S, 5°W the ATSR ASST variability is 4.5 K compared to a Levitus variability of just 2.4 K). This difference is examined by plotting the day - night variability difference (Figure 4.8). Large areas where the night variability is higher than the day variability by greater than 1°C are clear. These correspond to the areas where the Levitus variability is much smaller than the ATSR SST variability shown in Figure 4.6. Hence, it seems that the discrepancy between the Levitus variability and the ATSR variability is caused by an increased nighttime variability. In the absence of rogue data this difference is opposite to expectations; one would normally expect the day variability to be slightly greater than the night due to effects of the diurnal thermocline (Stommel et al., 1969; Price et al., 1986; Price et al., 1987; Yokoyama et al., 1995) and sunglint (Cracknell, 1993; Williams, 1993) which are present only during the day. The magnitude of the difference would certainly mask oceanographic signals in the affected regions.

Figure 4.9 shows the difference between the day and night mean ASST. In the same affected regions as Figure 4.8, the night mean ASST is clearly lower (by more than 1°C) than the day mean. Diurnal heating can cause day SSTs to be larger than night SSTs by up to approximately a degree (e.g. Price et al., 1986), however this effect is largest in areas of strong solar heating and low wind speeds. Hence one would expect a latitudinally dependent effect, being largest in the tropics, and least in the higher latitudes. This is not what is observed in Figure 4.9 and therefore some other explanation is required. The diamond shaped regions of alternating positive and negative day-night mean which are most prominent between 20°S and 30°S are an artefact of the period January 1992 to March 1992 when ERS-1 was in its 3 day repeat
Figure 4.7  The SST variability (standard deviation) in the South Atlantic from Levitus (1984) data.

Figure 4.8  The ATSR day ASST variability minus the ATSR night ASST variability.
Figure 4.9  The ATSR day mean ASST minus the ATSR night mean ASST.
Figure 4.10 (a) ATSR ASST time series for a location (27.25°S, 30.25°W) sampled only at night during the ERS-1 3 day repeat phase (Julian days 0-90). (b) ATSR ASST time series for a location (27.25°S, 25.75°W) sampled only during the day in the 3 day repeat phase.
orbit. To achieve such an orbit, the distance between ground tracks is large and the 512 km swath is insufficient to provide complete coverage within this period. Thus some regions are covered only during the night, and some only during the day. Figure 4.10 illustrates this by showing ASST time series extracted at points having, at certain times, only night coverage (Figure 4.10a) and only day coverage (Figure 4.10b). The bias between the night and day mean ASST is caused by the peak in the SST annual cycle coinciding with the 3 day repeat phase (j-day 0-90).

The most plausible explanation for the regions of higher night ASST variability and lower night mean ASST is cloud contamination of the night data. This explanation is reinforced by examination of the International Satellite Cloud Climatology Project (ISCCP) percentage cloud cover for 1988 shown in Figure 4.11. Regions corresponding to large differences between night and day ASST variability (Figure 4.8) exhibit a marked similarity to the regions of highest percentage cloud cover. The only exception is the area south of 50°S where the cloud percentage is high but the difference between the day and night variability is small. This is almost certainly due to the high cloud percentage south of 50°S being caused by a type of cloud which is handled well by the cloud clearing algorithms described in Section 4.6.1, while the high cloud percentage cover elsewhere contains cloud that is more difficult to detect (such as marine stratocumulus which is low lying and uniform in appearance). This is described in more detail in Section 4.9, where it is demonstrated that the seasonality of the cloud contamination coincides remarkably with the seasonality of marine stratiform clouds. The reason why night data has more cloud contamination than day data comes from the cloud clearing tests described in Section 4.3.6a. In particular, the 1.6 \textmu m channel which is used only in daytime cloud clearing, is a major reason why daytime cloud clearing is better. A further reason is the failure on 26th May 1992 of the 3.7 \textmu m channel used in several nighttime cloud clearing tests. Confirmation of this second reason comes from plotting, in Figure 4.12, a difference plot of the day - night variability for data before the failure of the 3.7 \textmu m channel. If the 3.7 \textmu m channel is a factor, one would expect an improvement in the night cloud clearing before May 1992. Figure 4.12 shows that while this is indeed the case, there is still evidence of higher night variability in the same affected areas evident in Figure 4.8, implying that using the 3.7 \textmu m channel for cloud clearing does not fully compensate for being unable to use the 1.6 \textmu m channel at night.

4.4.2 Quantification of the night cloud contamination

It is generally agreed that, in most parts of the world’s oceans, a time series of SST variability consists mainly of an annual signal and, especially in the southern hemisphere and tropics, a semiannual signal (e.g. Provost et al., 1992). Cloud
Figure 4.11 International Satellite Cloud Climatology Project mean cloud cover percentage for 1988.

Figure 4.12 Difference between the day and night SST variability prior to the failure of the 3.7 \( \mu m \) channel on 26 May 1992. Negative values indicate that the night variability is greater than the day.
contamination manifests itself in such a time series as high frequency noise. In order to eliminate this noise, a model consisting of annual and semi-annual periods is fitted to the ASST time series at each 0.5° cell and residuals from this model are calculated. The model used is:

\[
\text{Model} = a + b \cos\left(\frac{2\pi t}{T_{\text{ann}}}\right) + c \sin\left(\frac{2\pi t}{T_{\text{ann}}}\right) + d \cos\left(\frac{2\pi t}{T_{\text{s-ann}}}\right) + e \sin\left(\frac{2\pi t}{T_{\text{s-ann}}}\right) \tag{4.6}
\]

where \( T_{\text{ann}} \) is the annual period, \( T_{\text{s-ann}} \) is the semi-annual period, \( t \) is the time and \( a, b, c, d \) and \( e \) are coefficients determined by linear least squares regression.

Figure 4.13 displays the resulting separate day and night SST residual histograms obtained using a histogram interval of 0.25 K. For residual levels less than -1.5°C, the night histogram shows a long tail which is absent in the day histogram. This is consistent with cloud contamination of the night data. If one assumes therefore that where the night histogram is higher than the day histogram this is due to night cloud contamination, a lower limit for this cloud contamination can be obtained by integrating the difference between these histograms \textit{where the night histogram is greater than the day histogram}. This is found to be 5.7% of the night data.

### 4.4.3 Comparison with thermosalinograph SST data

Between 22nd December 1992 and 1st February 1993, the RRS \textit{Discovery} occupied a section in the South Atlantic from Punta Arenas in Chile, along 45°S until 15°W, and then northeast to Cape Town in South Africa. This cruise (denoted A11) (Saunders et al., 1993) was conducted by the Institute of Oceanographic Sciences Deacon Laboratory and is a UK contribution to the World Ocean Circulation Experiment (WOCE). Continuous temperature measurements at 5 m depth were made by a thermosalinograph (TSG) accurate to hundredths of a degree (Brian King, pers. comm., 1994).

ATSR ASST data are compared with TSG data by first binning the latter into 0.5° degree cells and then proceeding with a comparison only if more than 20 TSG measurements exist in a cell. TSG data are provided approximately every minute, so only using 0.5° cells containing more than 20 measurements is a simple way of excluding cells with insufficient TSG data to provide a reliable temperature. TSG data are also rejected between January 23rd to 26th as during this period the TSG was producing unreliable results (Saunders et al., 1993). Finally it is noted that a time criterion of ±2 days is used as the match-up time limit if ATSR and TSG data are to be compared. The results are summarised below in terms of mean difference and standard
Figure 4.13 Histogram of ATSR ASST residuals from an annual and semi-annual model, separated into day and night data. The lower panel is an expanded view of the upper histogram, clearly demonstrating the negative tail in the night SST data.
Figure 4.14 Histograms of the ATSR minus TSG temperature difference for (a) day and night ASST data, (b) only night ASST data and (c) only day ASST data.
deviation, with histograms of the ATSR-TSG temperature differences given in Figure 4.14.

Number of comparisons

<table>
<thead>
<tr>
<th>Description</th>
<th>Equation</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day and Night ATSR data:</td>
<td>ATSR - TSG = -0.93±1.89K</td>
<td>164</td>
</tr>
<tr>
<td>Night ATSR data only:</td>
<td>ATSR - TSG = -1.36±2.69K</td>
<td>74</td>
</tr>
<tr>
<td>Day ATSR data only:</td>
<td>ATSR - TSG = -0.59±0.61K</td>
<td>90</td>
</tr>
</tbody>
</table>

Figure 4.14b shows that the cause of the large bias and standard deviation (scatter) in the night analysis is the presence of six rogue points which lie well away from the main distribution. In contrast, the distribution of SST differences for the day analysis is compact and has no such outliers, leading to a much better result. These rogue data are almost certainly due to cloud contamination and correspond to 8.1% of the night ATSR data; a figure consistent with the lower limit for the cloud contamination derived in Section 4.5.2. (One should note that the figure of 5.7% derived in Section 4.5.2 applies to the whole South Atlantic, whereas the TSG comparison is restricted to a localised region where cloud contamination is more predominant; hence one would expect to obtain a higher percentage of rogue data.) The sign of the bias between the ATSR and the TSG SSTs is consistent with the skin effect, although the scatter in the day result of ATSR-TSG = -0.59±0.61 K is slightly worse than previously published results for the ASST data (see Section 4.2.2). This difference could be due to the comparisons being in a more cloud contaminated region which is likely to increase the error in the SST retrieval. The above results are largely insensitive to the time criterion of ±2 days used to select data. Increasing the time criterion makes the scatter in the comparisons worse, but decreasing the time criterion produces no improvement.

4.4.4 Reducing remnant cloud contamination

A new algorithm to filter SST data is described below, together with a brief discussion of its merits compared to other methods of filtering the data. There are two steps to the filtering scheme. The first step is elimination of SST data with values below -2.5°C or above 40°C, these being outside sensible data limits. For comparison, Bottomley et al. (1990) use -2.0°C and 37°C, while Reynolds and Smith (1994) use -2.0°C and 35°C.
Wide limits are deliberately chosen in order to remove only the worst data. The second step is the important part of the filtering algorithm and involves the fitting of a mean, annual and semiannual model [4.6] to the day data. Data are rejected as cloud contaminated if they differ by more than three times the SD of the residuals from this model.

In practice, it is necessary to filter the day data further to remove gross outliers before fitting the model. This is done by rejecting data more than 12°C from the median SST for that location. The reason for such rogue data is unclear; they occur randomly during both day and night and affect a complete ATSR scan width, rather than just a specific 0.5° cell. The criterion of 12°C was chosen after careful study of the rejected data and is used only to obtain a good model fit. After obtaining the model fit, the SST data are filtered using the two steps above.

The occurrence of persistent cloud cover and/or ice cover means that in some high latitude regions there are no data during winter months. To avoid the resulting poor model fits, it is specified that no model is fitted to the data if there are less than 19 ASST values in the three year time series. As a further check the values of the model mean are examined and are rejected if they are less than -2.5°C or greater than 40°C. While it is plausible to have a model mean less than -2.5°C, in practice this only occurs when the model fit is poor because of uneven distribution of the data. This results in the model fit being rejected in a few regions at high latitudes.

The rejection criteria for the second step of the filtering algorithm are obtained by calculating the SD of the daytime residuals within 2° (in latitude and longitude) bins. It is necessary to use 2° bins rather than 0.5° since the map of the SD of day residuals on a 0.5° grid appears noisy. Calculating the SD in 2° bins reduces this noise substantially. For 2° bins with no rejection criterion due to lack of data points or poor model fit, the rejection criteria are determined by using the SD of the day SSTs. Data where the SST difference from the median SST for each location exceeds three times the day SST SD are rejected. The rationale here is that locations with no model fit occur mainly in regions with very few data points in winter. Therefore a criterion based on the SD of the day SSTs is an adequate proxy for the SD of the day residuals.

The rejection criterion of three times the SD was selected after experimentation using two and four times the SD. The quantity and geographical location of the rejected data was studied and it was found that tightening the criterion to two times the SD resulted in rejection of some day data as well as large amounts of night data in regions with small day/night variability differences. Relaxing the criterion to four times the SD resulted in much less rejected data. Even in regions with large day/night variability
differences, few data were now rejected. Thus the criterion of three times the SD is used as a good compromise.

This approach is preferable to one based on rejection of data farther than a given amount from a climatological value, since climatologies derived from *in situ* data are usually poorly sampled and heavily interpolated in the very regions where remotely sensed SST data are most useful. Comparison with another infrared SST climatology, such as that obtained from AVHRR data, is also not ideal. As ATSR measures the skin temperature, while AVHRR SST retrievals incorporated in global climatologies are empirically adjusted to give the bulk temperature, offsets are likely.

This filtering algorithm also has the desirable feature that the rejection criterion changes with region. Thus in regions with high natural SST variability on small timescales, the rejection criterion will be less severe because the SD of the daytime residuals is larger. This will reduce the possibility of filtering out real oceanographic features such as mesoscale eddies. It should also be noted that the scheme is robust to a small amount of cloud contamination in the daytime data.

In the following section the results of applying this filtering scheme to the ASST data in the South Atlantic region for 1992, 1993 and 1994 are described.

### 4.4.5 Results of filtering

The results of applying the filtering algorithm to the ASST data in the SA are summarised in Table 4.4. In total 4.71% of the data are rejected, this corresponding to 8.5% of the nighttime data and 0.9% of the daytime data. The seasonality and geographical distribution of the rejected data is discussed in Section 4.4.6.

| Table 4.4 | A summary of the percentage of data rejected by the filtering scheme. |
|---|---|---|---|
| **Criterion** | **Day (\% of total number)** | **Night (\% of total number)** | **Day and Night (\% of total number)** |
| SST < -2.5°C | 0.10 | 0.73 | 0.83 |
| SST > 40°C | 0.00 | 0.00 | 0.00 |
| No valid rejection criterion | 0.02 | 0.01 | 0.03 |
| $|\text{Residual}| > 3\text{SD}$ of day residuals | 0.29 | 3.56 | 3.85 |
| Total rejected | 0.41 | 4.30 | 4.71 |
To investigate further the effect of this filtering, differences between the day and night signals are re-examined after filtering and comparisons with the A11 cruise TSG data are repeated.

The ASST variability after filtering is shown in Figure 4.15. Comparison with Figure 4.6 reveals that there is a large reduction of variability in the area previously considered to be cloud contaminated (across the entire region from 30°S to 50°S and also in the northeastern section of the South Atlantic off the west coast of Africa). Figure 4.16 shows that the day and night variabilities are now very similar. The characteristic swath shape of high night variability has disappeared and the resulting difference map is almost homogeneous across the entire region. Figure 4.17 shows the difference between the day and night mean ASSTs after filtering; again the swath shape has disappeared. The largest differences between the day and night mean ASST are now due to the sampling effect when ERS-1 was in its 3 day repeat orbit, as discussed in Section 4.4.1.

Differences between day and night signals can only reveal the internal homogeneity of the dataset. To investigate any improved accuracy of the data after filtering, the TSG comparisons of Section 4.4.3 are repeated with the filtered ATSR dataset. Results of this analysis are given below in terms of mean difference and standard deviation.

**Number of comparisons**

<table>
<thead>
<tr>
<th>Data Type</th>
<th>ATSR - TSG</th>
<th>Number of Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day and Night ATSR data</td>
<td>ATSR - TSG = -0.61±0.67K</td>
<td>158</td>
</tr>
<tr>
<td>Night ATSR data only</td>
<td>ATSR - TSG = -0.63±0.76K</td>
<td>68</td>
</tr>
<tr>
<td>Day ATSR data only</td>
<td>ATSR - TSG = -0.59±0.61K</td>
<td>90</td>
</tr>
</tbody>
</table>

The filtering removes the six rogue points present in the ATSR nighttime ASST which caused the poor night comparisons noted in Section 4.4.3. As a result, the night comparison result is now similar to the day comparison, the difference in scatter being just 0.15 K rather than 2.08 K. The overall result of ATSR-TSG = -0.61±0.67 K shows a factor of three reduction in scatter and a 0.3 K improvement in bias. Histograms of the temperature differences between ATSR and TSG are identical to those shown in Figure 4.14 but without the six gross outliers present in Figures 4.14a and 4.14b.

In summary, the filtered dataset appears both more realistic in terms of the differences between day and night signals and more accurate in terms of comparisons with in situ data. To narrow down the type of cloud causing the contamination, the seasonality of the data rejected from the filtering scheme is investigated. This is described in the following section.
Figure 4.15  The ATSR ASST variability after filtering the data.

Figure 4.16  The day ATSR ASST variability minus the night ATSR ASST variability after filtering the data.
Figure 4.17 The day ATSR mean ASST minus the night ATSR mean ASST after filtering.
4.4.6 Cause of contamination

The rejected data are split into four seasons (December, January and February (DJF); March, April and May (MAM); June, July and August (JJA); and September, October and November (SON)). Figure 4.18 shows the percentage of the total number of data points rejected after filtering at each 0.5° location. Two conclusions may be drawn from Figure 4.18. First, the regions associated with a large percentage of rejected points are those characterised by the large day/night SST variability differences shown in Figure 4.8, proving that the filtering scheme is indeed rejecting data where there are problems. Second, the percentage of data points rejected is clearly seasonal. For example, the maximum amount of data rejected in the latitude band 40°S-60°S occurs in the austral summer (DJF). Alternatively, the minimum amount of data rejected in the area off the west coast of South Africa occurs in MAM.

The reason for both the location and the seasonality of the rejected data lies in the type of cloud causing the contamination. Klein and Hartmann (1993) investigate the seasonal cycle of low stratiform clouds (LSC) (fog, stratus and stratocumulus), using data from an ocean cloud atlas based on ship observations (Warren et al., 1988). Klein and Hartmann (1993) define nine ocean regions that have climatologically high amounts of LSC. Furthermore, Klein and Hartmann (1993) give the seasonal variation in LSC amount at each location. Their results are summarised below:

Table 4.5 Seasonality of Marine Stratiform clouds (adapted from Klein and Hartmann 1993).

<table>
<thead>
<tr>
<th>Region</th>
<th>Location</th>
<th>Season and amount (%) of maximum cloudiness</th>
<th>Season and amount (%) of minimum cloudiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peruvian</td>
<td>10°-20°S, 80°-90°W</td>
<td>SON 72</td>
<td>DJF 42</td>
</tr>
<tr>
<td>Namibian</td>
<td>10°-20°S, 0°-10°E</td>
<td>SON 75</td>
<td>MAM 48</td>
</tr>
<tr>
<td>Californian</td>
<td>20°-30°N, 120°-130°W</td>
<td>JJA 67</td>
<td>DJF 45</td>
</tr>
<tr>
<td>Australian</td>
<td>25°-35°S, 95°-105°E</td>
<td>DJF 45</td>
<td>JJA 41</td>
</tr>
<tr>
<td>Canarian</td>
<td>15°-25°N, 25°-35°W</td>
<td>JJA 35</td>
<td>SON 17</td>
</tr>
<tr>
<td>North Pacific</td>
<td>40°-50°N, 170°-180°E</td>
<td>JJA 82</td>
<td>DJF 54</td>
</tr>
<tr>
<td>North Atlantic</td>
<td>50°-60°N, 35°-45°W</td>
<td>JJA 68</td>
<td>DJF 51</td>
</tr>
<tr>
<td>ACC</td>
<td>50°-65°S</td>
<td>DJF 62</td>
<td>No data available</td>
</tr>
</tbody>
</table>
Within the South Atlantic region there are two areas of LSC formation, (1) Namibian, and (2) ACC. The locations of these regions are identical to those with a high percentage of rejected data shown in Figure 3.18. Furthermore the seasonality of the cloud percentage is the same as the seasonality of the data rejected. Jones et al. (1996b) in a global study confirm that the location and seasonality of cloud occurrence in the other regions also coincides with that of data rejection.

Visual comparisons between the percentage of ASST data rejected and the percentage amounts of LSC shown in Table 4.5 and Figure 1 of Norris and Leovy (1994) show that, although the phase of both is almost identical, the percentage of ASST data rejected is generally less than the amount of LSC. There are several reasons why this should be so. The main reason is that the percentage amounts of LSC are for both day and night, whereas the percentage of data rejected is dominated by the rejection of night data. If no day data are rejected, the percentage of data rejected will be 50%, or less, of the percentage of LSC. There are three other reasons why it would expected that the percentage of data rejected would not coincide with the percentage of LSC as shown by Klein and Hartmann (1993) and Norris and Leovy (1994). First, the presence of other cloud types at the same time as the LSC would enable the ASST data to be identified as cloud contaminated. However, it appears that LSC occurs in places where other cloud types are less common. For example, many clouds have strong maxima in the intertropical convergence zone, whereas LSC do not. Cumulonimbus cloud amounts tend to decrease from west to east within an ocean basin (Warren et al., 1988), whereas amounts of LSC increase from west to east. After careful study of the Warren et al. (1988) cloud atlas, no areas where large amounts of LSC correspond significantly with large amounts of a differing cloud type could be found. Therefore the effect of other cloud types present at the same time as LSC is not substantial, although it would act to decrease the percentage of any data contaminated. The second reason why the amount of LSC does not correspond exactly with the percentage of data rejected is that some of the LSC will be detected by the cloud clearing tests implemented by RAL. For example, the spatial coherence test (Section 4.3.6a) could detect the edges of the LSC, even if the LSC are fairly uniform in general. The third reason for the differences between the LSC amount and the percentage of data rejected is that the cloud climatology may be in error in some regions. Interannual variability in LSC amount could also make the climatological mean unrepresentative of the LSC amount for the specific three years here. In summary, it is found that the locations and seasonality of large amounts of LSC correspond well with the locations and seasonality of large percentages of data rejected.

That LSC seems to be the cause of the cloud contamination is not entirely surprising. Fog and low stratus are extremely hard to detect at infrared wavelengths over the sea surface, since they are reasonably uniform and their temperatures are fairly close to
Figure 4.18  Percentage of data rejected by the filtering scheme for (a) June, July, August, (b) September, October, November, (c) December, January, February, (d) March, April, May.
those of the underlying surface (Saunders, 1986). Differences between 3.7 μm and 10.8 μm brightness temperatures can reveal these cloud types because the emissivity of such cloud at 3.7 μm is approximately 10% less than at 10.8 μm and the resulting brightness temperature difference is more than the difference in atmospheric absorption between the two channels (Saunders, 1986). However, as mentioned earlier, the 3.7 μm channel failed on May 26th 1992 and therefore 3.7 μm data after this date cannot be used for nighttime cloud clearing.

4.5 Discussion and conclusions

This chapter highlights the presence of cloud contamination in the ATSR night ASST data within the South Atlantic area. The cloud contamination concentrates mainly in a latitude band between 30°S and 50°S and stretches northwards up the west coast of Africa to the equator. It appears as an increased night ASST variability and as a decreased night ASST mean. A lower limit for the amount of cloud contamination of the night ASST data is derived and found to be 5.7% for this dataset. Further confirmation of the cloud contamination comes from comparing ATSR data with in situ TSG data which yields the relationship ATSR-TSG = -0.93±1.89 K, the large scatter and bias again being due to cloud contamination of the night data.

A filtering algorithm is proposed based on the assumptions that the day ASST data are free of significant cloud contamination and that cloud contamination manifests itself as high frequency noise. Applying this filtering algorithm to the ATSR data results in a much closer agreement between day and night. Furthermore, comparing the filtered ATSR data with the TSG data, yields ATSR-TSG = -0.61±0.67 K, thus improving the r.m.s. scatter by a factor of three and the mean bias by 0.3 K.

It is shown that the percentage of data rejected is seasonal and that both the location and seasonality of the percentage of data rejected correspond well with the amount of LSC shown by Klein and Hartmann (1993). This is strong evidence that this cloud type is causing the contamination in the ASST dataset.

The ASST data from the ATSR instrument have the potential to provide global SST measurements from space of unprecedented accuracy and stability. These data could increase our knowledge of ocean/atmosphere heat fluxes and thus our ability to model the ocean/atmosphere system. By combining these data with SSTs from the ATSR-2 instrument mounted on the recently launched ERS-2 satellite and with SST data from the advanced ATSR instrument scheduled for launch on Envisat in 1999, it should be possible to obtain a decadal length SST dataset of high accuracy and low calibration.
drift. Such a dataset would permit investigation of any anthropogenic induced climatic warming trend. Before using the ASST dataset for such studies, it is important to ensure that they are of the highest quality. For example, consider that a portion of the ASST dataset is affected by cloud contamination and that the amount of cloud type causing this contamination increases as the climate changes. This not unrealistic scenario would result in more cloud contamination as the amount of cloud type increased, and therefore a negative SST trend would result. Such a trend would eliminate any chance of detecting a true SST trend. It is therefore extremely important to ensure that infrared derived temperatures are SSTs and not cloud temperatures.

For the study of ocean circulation, accurate cloud clearing is also essential. Of all the error sources in the infrared measurement of SST, cloud contamination is the one most likely to cause localised errors on small spatial and temporal scales (for the high resolution ATSR data, instrument noise is also an issue). These errors can be confused with mesoscale ocean circulation features such as cold core eddies. In attempting to eliminate cloud, it is very important that oceanographic features are not also eliminated. The scheme proposed in this chapter is based on a priori statistics from the essentially cloud free daytime ASST data. Hence the rejection criteria vary with geographical region, which reduces the extent to which real features are rejected. This scheme is already in use by research groups at MSSL, RAL, SOC, University of East Anglia and the Proudman Oceanographic Laboratory. It has therefore demonstrated its benefit as a remote sensing technique.

The following chapter describes quantitative comparisons between ATSR SST data and sea surface height (SSH) data from the TOPEX/POSEIDON altimeter. For this stringent use it is essential to have as accurate a dataset as possible. Errors in the SST data will obscure any relationship between SST and SSH and therefore must be reduced. The work described in this chapter aids the study in the following chapter by reducing a key source of error in the ATSR ASST data.

As WOCE moves from its observational phase to its Analysis, Interpretation, Modelling and Synthesis phase, it is important that the datasets gathered from the observational phase are of the highest possible quality. The filtering scheme described here will be of benefit in this regard.
Chapter 5

Correlations between SST and SSH

In this chapter, techniques and results from the previous two chapters are used to process TOPEX/POSEIDON (T/P) sea surface height (SSH) data and Along Track Scanning Radiometer (ATSR) sea surface temperature (SST) data. These two data sources are brought together to investigate whether any relationship can be found between the two different parameters. The first section of this chapter outlines some basic theory which describes the reasons why SST and SSH could be related. This is followed in Section 5.2 by a description of the reasons why a relationship between SST and SSH would be useful. A review of previous work relating to this subject is given in Section 5.3 and the new research describing the relationship between SST and SSH is described in Section 5.4. Conclusions are made in Section 5.5.

5.1 Why could SST and SSH be related?

To study the reasons why SST and SSH could be related, it is first necessary to describe the processes affecting SSH. As described in Chapter 3, it is not feasible to measure SSH relative to the geoid. Hence deviations (represented by a prime) of SSH from the time mean SSH are studied throughout this chapter. Gill and Niiler (1973) divide variations in SSH into four components:

\[ \Delta \text{SSH} = \eta_s' + \eta_i' + \eta_b' + \eta_a' \]  

[5.1]

where \( \Delta \text{SSH} \) is the deviation of the SSH from the time mean SSH at a particular geographical location.

The first of the four terms, \( \eta_a' \) represents the effect that atmospheric pressure has on the sea surface. SSH variations caused by changes in atmospheric pressure are dynamically uninteresting and do not affect SST variations. To remove such variations, a first order approximation can be made by assuming that the sea surface responds as a simple inverse barometer, with a 1 mbar increase (decrease) in surface pressure corresponding to a 1 cm decrease (increase) in \( \eta_a' \) (i.e. -1 cm/mbar). This is discussed in more detail in Chapter 3, Section 3.4.3(i).
The second of the four terms, $\eta_s$, represents the effects of tides on the sea surface height. Tidal height variations result from the combined effect of the sun and moon's gravitational force and the earth's rotation. A deeper discussion of tidal effects and the problems they can cause in altimetry is given in Chapter 3. Tidal variations in SSH are treated as noise and removed by the CSR 3.0 tidal model (Eanes and Bettadpur, 1995).

The term, $\eta_b$, represents barotropic changes caused by the effect of the wind on the ocean surface. These changes do not affect the density structure of the water column in any way and therefore do not affect SST. Unlike the two previous terms, however, $\eta_b$ cannot be modelled simply and therefore remains in the altimetric measurement of $\Delta$SSH. If changes in $\Delta$SSH in a region are purely barotropic, one would expect no relationship between SST and $\Delta$SSH. The extent to which barotropic or baroclinic changes affect $\Delta$SSH is largely unknown because measurement of barotropic motions is extremely difficult.

The fourth term, $\eta_s$, represents variations which are caused by changes in the density structure of the water column (steric variations):

$$\eta_s = -\frac{1}{\rho} \int_{-H}^{0} \rho' \, dz$$

[5.2]

Here $z=-H$ is the depth of the water, $z=0$ is the mean sea surface height, $\bar{\rho}$ is the time mean density of the water column and $\rho'$ is the variation of density from the time mean density at a particular depth. The way in which $\rho'$ varies depends upon the mechanism causing the change. Two mechanisms are important. These are (1) changes in the upper several hundred metres caused by variations in ocean-atmosphere heat and water fluxes, and (2) changes due to advection. Changes in the density of the upper few hundred metres of the water column can be caused by changes in the ocean-atmosphere heat and water fluxes. Gill and Niiler (1973) show that on large spatial scales (~1000 km), this is the dominant mechanism for changes in SSH; seasonal changes in these fluxes can cause SSH variations of several cm. On smaller spatial scales (~100 km), the effects of advection are important. A mesoscale eddy advected by the mean flow can cause changes in the density of the water column that result in changes in SSH of several tens of cm (e.g. Gordon and Haxby, 1990; Smythe-Wright et al., 1996); an order of magnitude larger than SSH variations caused by ocean-atmosphere heat fluxes. The meandering of an ocean front can also cause such changes. Indeed Vazquez et al. (1990), in a study of the SSH variability in the Gulf Stream area, conclude that the majority of the seasonal variability can be explained by a meandering front. The reason that changes in SSH caused by advection are larger than those due to ocean-atmosphere heat fluxes is that although these two mechanisms cause similar changes in density for a
particular water parcel, advection changes can occur over a large depth range (~1000 m), depending on the structure of the feature being advected. SSH changes induced by surface fluxes, however, are limited to the depth of the mixed layer of the ocean (~100 m) and are therefore usually an order of magnitude smaller than SSH changes caused by advection.

Irrespective of the mechanism causing the density change, variations in SSH will only be related to changes in SST if the variations in density are related to changes in SST. Changes in density at a particular pressure ($\delta p = 0$) are related to variations in temperature and salinity by:

$$\delta \rho = \frac{\partial \rho}{\partial T} \delta T + \frac{\partial \rho}{\partial S} \delta S$$  \[5.3\]

Hence a change in density will only correlate with temperature variations if the change in density caused by salinity is either small, or correlates with the change in density caused by temperature. Assuming that changes $T'$, $S'$ of temperature and salinity from mean values are sufficiently small that they can be used in equation [5.3], and substituting into [5.2] gives:

$$\eta_s = -\frac{1}{\rho} \int_{-h}^{0} \left( \frac{\partial \rho}{\partial T} T' + \frac{\partial \rho}{\partial S} S' \right) dz$$ \[5.4\]

Thus for a relationship to exist between SST and SSH, three criteria must be satisfied. These are (1) that the variation in ASSH is caused to some extent by a change in $\eta_s$, (2) that the variation in the surface temperature corresponds to a change in the temperature at depth and (3) that there is either no salinity change, or the variation in temperature is related to it.

As an example of this, assume that for a given location any changes in temperature are linearly related to changes in salinity and that for the range of $T'$ and $S'$ encountered, the equation of state can be linearised about $\bar{T}$ and $\bar{S}$. This gives:

$$S' = c_i T'$$ \[5.5\]

and

$$\rho' = \bar{\rho} \alpha (\bar{T}, \bar{S}, \bar{z}) T' + \bar{\rho} \beta (\bar{T}, \bar{S}, \bar{z}) c_i T' = \bar{\rho} T' (\alpha + c_i \beta)$$ \[5.6\]

$$\alpha = \frac{1}{\rho} \frac{\partial \rho}{\partial T}, \text{ and } \beta = \frac{1}{\rho} \frac{\partial \rho}{\partial S}$$ \[5.7\]
Chapter 5 Correlations between SST and SSH

where $\alpha$ and $\beta$ are the thermal expansion coefficient and the expansion coefficient for salinity respectively.

SST and SSH may be explicitly related by assuming a relationship between $T'$ and SST'. For example, a linear relationship between $T'$ and SST' with a maximum at $z=0$ and disappearing at $z=-h$ gives:

$$T'(z) = \Delta\text{SST}(1 + \frac{z}{h}), \quad z \geq -h$$

$$T'(z) = 0, \quad z < -h$$

Substituting [5.8] and [5.6] into [5.2] gives

$$\eta_h = \rho(\alpha + c,\beta)\Delta\text{SST} \int_{-h}^{0} (1 + \frac{z}{h})dz$$

[5.9]

and integrating [5.9] leads to:

$$\eta_h = \rho(\alpha + c,\beta)\Delta\text{SST} \left(\frac{h}{2}\right)$$

[5.10]

So if the large assumptions made to obtain [5.10] are valid (in general they will not be valid - this is purely used as a simplistic example to explain the fundamentals of why changes in SST and SSH could be related), changes in SST are linearly related to changes in SSH. Furthermore, if the linear relationship is established, it may be possible to obtain dynamical parameters from measurements of SST and SSH. For example, in [5.10], if the linear relationship is measured, a value of $h$ can be obtained. The utility of this depends on what $h$ represents. If it is found in some particular region that $h$ represents the mixed layer depth, or the depth of the bottom of the thermocline, then it would be very useful to be able to measure these parameters from remote sensing. The extent to which dynamical parameters such as these can be inferred from a combination of $\Delta\text{SST}$ and $\Delta\text{SSH}$ requires study of the evolution of the density structure of the water column and is outside the scope of this research. However, if it is found that there does not exist any relationship between $\Delta\text{SST}$ and $\Delta\text{SSH}$, then it would be pointless to proceed with a study of this kind. A first attempt at relating SST and SSH anomalies is presented in Section 5.4.
5.2 Why would a relationship be useful?

As the observational phase of WOCE draws to a close, a key emphasis is the analysis and synthesis of the large dataset that has been collected. In terms of hydrographic data alone, WOCE has been by far the largest oceanographic experiment to date. In addition to this, over the last few years satellite sensors have been providing global semi-synoptic measurements of parameters such as wind speed, significant wave height, SST and SSH. The amount of such data collected is huge and analysing these data presents many challenges. One important task is to establish the extent to which these different parameters are related, and the extent to which extra information about the ocean can be obtained from a combination of parameters. An initial attempt at relating two of these parameters, SST and SSH, is presented here. However, first the reasons why establishing a relationship between SST and SSH would be of interest are described.

5.2.1 Relating surface parameters to subsurface

In most forms of remote sensing of the ocean, a major difficulty is obtaining information about the subsurface ocean from purely surface parameters. SST, as measured by infrared radiometers, is the temperature of the top few microns of the sea surface, and any relationship between SST and SSH is evidence that the temperature of the top few microns of the ocean surface contains information about the entire water column. In particular, the SST structure must be correlated with the temperature structure to some depth before it will correspond to a similar SSH structure. Establishing the regions and seasons where there is any relationship between SST and SSH will show when and where to expect the SST structure as observed from infrared radiometers to represent more than just the surface layer. Furthermore, by examining the T-S structure of the water column (from ocean model data, or wherever possible from in situ data) in regions where a relationship holds, it may be possible to infer dynamical information. Examples of such information are the baroclinicity of the variability, or the depth to which surface temperature structure is manifested in subsurface temperature structure.

5.2.2 Model validation

As global ocean models become more sophisticated and are able to represent the state of the ocean more accurately (Semtner, 1995), it becomes more important to test these models to assess their performance. Remote sensing provides the only method for

---

1 AVHRR SST retrieval algorithms are usually 'tuned' by incorporating buoy SST data and therefore have the mean skin-bulk difference subtracted, but nonetheless variations in the skin temperature also cause variations in the AVHRR SSTs.
Chapter 5  Correlations between SST and SSH

global validation of models. Traditionally altimeter SSH data are used and statistics such as SSH variability and eddy kinetic energy (EKE) are compared to the corresponding model statistics (Stammer et al., 1996). A more stringent test of a model, however, would be to compare the relationship between SST and SSH as observed from satellites with the corresponding SST-SSH relationship from the model. To achieve a realistic coherency between SST and SSH, the model must correctly represent both the near-surface thermodynamics and the deeper structure of the water column. This is a challenging task but is especially important for coupled ocean-atmosphere models where SST is an important factor.

5.2.3 New techniques

Establishing a relationship between SST and SSH would open up the way for new techniques based on any such relationship. One example is interpolation of altimeter data. It is shown in Section 3.6 that various problems can arise when attempting to interpolate altimeter data onto a regular grid. For an altimeter with the sampling characteristics of T/P, it is impossible to consistently resolve scales with wavelengths shorter than about 400 km (Chelton and Schlax, 1994). However, if there is any relationship between SST and SSH it may be possible to use SST data to interpolate the SSH data, therefore providing pseudo-SSH fields of much higher resolution than that given by altimetry alone. This would be useful for tracking small wavelength features such as mesoscale eddies which blink in and out of view as they disappear between altimeter tracks. It would also be helpful for assimilating into ocean models where a high resolution gridded SSH field would be much easier to use than along-track altimeter data.

A second example where a relationship between SST and SSH would be of use is in the calculation of eddy kinetic energy (EKE) (the variance of the geostrophic velocity anomalies). To calculate EKE from altimetry, both horizontal components of the velocity field are required. Hence, either EKE calculations are limited to crossover points, or isotropy is assumed and EKE is calculated at the full along-track resolution of the altimeter. However, the assumption of isotropy is known to be poor in certain regions (Morrow et al., 1994), especially the Southern Ocean where the bottom topography exerts a large influence on the eddy field. Hence EKE maps calculated at the full resolution of the altimeter will be in error. SST fields are fully two dimensional however and if there is any relationship between SST and SSH it may be possible to use the SST data to remove the assumption of isotropy and calculate more accurate high resolution EKE fields.
5.3 Previous studies

Previous attempts to relate SST and SSH are scarce. Vastano and Reid (1985) apply an interactive feature tracking technique (Vastano and Borders, 1984) to two AVHRR BT images, separated by a day, in the Oyashio frontal zone in the north-western Pacific, in order to obtain surface current velocities. From these velocity vectors, they determine the stream function by representing it as a series of trigonometric basis functions and obtaining the coefficients of these basis functions by a least squares fit of the derivatives of the stream function to the velocity field. The resulting stream function is then related to SSH by assuming geostrophy and integrating the geostrophic relationship. By overlaying a map of the SSH obtained in this way on the first AVHRR BT image, Vastano and Reid (1985) show that an anticyclonic warm core eddy features in both the image and the derived SSH. Unfortunately no in situ data exist to validate these SSH calculations. The method is certainly novel, but has limitations. A problem present in all feature tracking methods is the requirement for two consecutive (within ~24 hours) BT images which are mostly cloud free. In most areas of the world's oceans, cloud cover (Warren et al., 1988) is such that this requirement limits the occasions when feature tracking can be used (although using BT images from multiple radiometers alleviates this to some extent). A more fundamental limitation to the method is that the streamfunction is obtained from surface velocities which may not be in a near-geostrophic balance due to the effects of winds and inertial motions. Hence making the geostrophic assumption to convert from the streamfunction to SSH may not be appropriate. Furthermore, since the SSH is obtained by integrating the streamfunction, the absolute SSH is not known due to the constant of integration. Without proper validation against either altimeter data or in situ data, this technique can not be assumed to work in anything other than a qualitative sense.

Several studies comment on similar patterns in both infrared imagery and satellite altimeter data. Scott and McDowall (1990) conduct a thermistor chain survey along a section of a GEOSAT track across the Iceland-Faeroes Frontal Zone. They use several AVHRR BT images to provide the context for the thermistor and the GEOSAT data. Several cold cross-frontal jets are observed in both the AVHRR SST and the GEOSAT measurements of ΔSSH, with cold SST corresponding to a lowering in ΔSSH. Vasquez et al. (1990) analyse two years of GEOSAT data in the Gulf Stream region, together with the position of the north wall of the Gulf Stream as measured by AVHRR. They find several maps of ΔSSH where the Gulf Stream front is visible (due to the front meandering from its mean position and therefore appearing in maps of ΔSSH). From this, they measure the distance between the front from the two different data sources. They find that between 75°W and 50°W the r.m.s. difference is 92 km. van Woert and Price (1993), in a study of Rossby waves near the Hawaiian islands, observe cusp
shaped features in AVHRR SST fields which match up well with gridded ΔSSH data from GEOSAT. By studying the propagation of these features in the ΔSSH data, they deduce that these cusp shaped patterns are the surface manifestations of baroclinic Rossby waves. If this is the case, it would be the first time that such a phenomenon has been observed in both altimeter and SST data. However, care must be taken in interpreting westward propagating signals in GEOSAT data; Schlax and Chelton (1994) have shown that aliasing of tidal errors can manifest as westward propagating features. van Woert and Price (1993) also compare ΔSSH data along four GEOSAT tracks with AVHRR SST data interpolated onto the GEOSAT tracks. They find that in most of the region 16-26°N, 162-152°W in April 1988, water cooler than 24°C corresponds to negative ΔSSH and warmer water corresponds to positive ΔSSH. The exception to this is an area southwest of the Hawaiian islands, where there is no relationship between SST and ΔSSH. They attribute this to the weaker winds in this area allowing surface heating to obscure any underlying horizontal structure, whereas elsewhere in the region the winds are stronger and this decoupling mechanism does not occur.

Knudsen et al. (1996) take a different approach to relating SST and SSH by using T/P altimeter data and ATSR SST data. Rather than studying spatial variations in SST and SSH, they average both SSH and SST data into 2° by 5° bins and compare time series of SST and SSH at each location. By averaging the data into such large bins, they effectively reduce any mesoscale signal and emphasise the large scale seasonal signal in both SST and SSH due to the seasonal change in heat and water fluxes (Gill and Niiler, 1973). In most regions, they find similarities between the amplitudes and phases of the annual cycles of SST and SSH, with largest amplitudes in the northern hemisphere being off the eastern coasts and in the southern hemisphere being in a latitude band 30°S-50°S.

Aside from direct observations of similarities between SST and SSH, there have been several studies that address issues linked to quantifying any relationship between SST and SSH.

The question of whether or not the SST structure is related to the temperature structure at depth is crucial in establishing any relationship between SST and SSH. Legeckis (1978), in a global survey of ocean fronts from infrared imagery, concludes that the ability to detect fronts in such imagery depends upon the season and the latitude of the observations. At high latitudes (|latitude| > ~35°), the ocean surface layer is well mixed and horizontal temperature gradients can be seen at the sea surface. This is also the case for latitudes between 25° and 35° in winter months. In summer months at these latitudes, however, an isothermal layer develops that can obscure horizontal temperature gradients. In tropical oceans (|latitude| < 25°), these isothermal conditions persist during
all seasons making detection of fronts difficult or impossible. Fiedler (1988), in a study of the California Current system, analyses a set of hydrographic data to determine whether or not subsurface thermal structure manifests at the surface. He finds that mixed layer depth and seasonal thermocline depth are not significantly correlated with SST, but that the seasonal thermocline strength is significantly correlated ($r^2 \approx 0.3$) with it. He attributes this to the fact that the temperature of the ocean beneath the seasonal thermocline stays fairly constant whereas the mixed layer temperature varies with season. The SST is a good proxy for the mixed layer temperature and therefore gives information on the strength of the seasonal thermocline. He finds that correlations between SST and seasonal thermocline strength are greatest in winter and summer and smallest in spring and autumn. As well as studying correlations between SST and subsurface parameters, Fiedler (1988) also studies correlations between the surface temperature structure and the temperature structure at depth. Surface-subsurface correlation profiles show two types of pattern; either a monotonic decrease in correlation with depth or a minimum correlation at an intermediate depth between 50 and 125 m with correlation increasing below these depths. Explaining the increase in correlation with depth is difficult. One explanation is by intrusion of a different water type at the intermediate depths of low correlation, but a long-term mean profile with a correlation minimum is hard to explain in this way. Fiedler (1988) concludes that correlation between surface and subsurface structure is most likely to extend below the mixed layer in summer when it is at its most shallow. At first this seems to contradict the conclusions of Legeckis (1978). However, the surface-subsurface correlation could be stronger and penetrate deeper in winter and still be consistent with the results of Fiedler (1988). His conclusion that the correlation is deeper than the mixed layer more often in summer could be due purely to the variation in mixed layer depth, rather than to any strengthening or weakening of the correlation structure.

Carnes et al. (1990) describe a promising technique to infer subsurface thermal structure from dynamic height in the Gulf Stream region. Their approach requires a large amount of historical data to establish a relationship between dynamic height and subsurface structure. Once this relationship is established, they use XBT data and feature modelling to calculate the geoid along several GEOSAT tracks. From this relationship they derive subsurface thermal structure to a depth of 1000 m. They obtain r.m.s. accuracies in isotherm depths of about 85 m, differences between inferred and measured temperatures of $\approx 2$°C and differences between XBT dynamic heights and GEOSAT (feature modelled geoid) dynamic heights of $\approx 15$ cm. Although this technique is new, the errors in the derived parameters are significant and, as previously mentioned, a large amount of historical data is required to establish the relationship between dynamic height and thermal structure. An estimate of the geoid is also required which is not available in most areas of the world oceans. In addition, something not mentioned by
Carnes et al. (1990), but which might explain some of the large errors, is that this technique will not work in areas where the variability in dynamic height is mainly barotropic, for the reasons described in Section 5.1. Carnes et al. (1994) extend the Carnes et al. (1990) study to include measurements of SST. In the northwest Atlantic and northwest Pacific they compare five different models using different combinations of dynamic height, geographical location, SST and time of year. They find that the models using dynamic height and SST provide better estimates of T and S profiles than do traditional climatologies, although the same reservations about the Carnes et al. (1990) study still apply.

The question of whether temperature can be used to infer salinity was first studied by Stommel (1947) in an attempt to infer dynamic topography from temperature profiles alone. He shows that by using temperature profiles and a T-S relationship, dynamic height can be obtained to an accuracy of ~5 cm. Emery (1975) continues this work by studying data from three weather stations in the Pacific Ocean. He finds that for two of the three stations, dynamic height can be estimated from temperature profiles and a mean T-S relationship to within 2 cm. At the third location this approach gave errors of 10 cm due to the presence of a temperature inversion which resulted in the same temperature at two different depths having different salinities. In a more comprehensive study, Emery and O'Brien (1978) compare the use of a mean T-S relationship to that of a mean salinity profile. They find that in some regions (such as those with a temperature inversion), a mean salinity profile leads to more accurate dynamic height computations, whereas in others the use of a mean T-S relationship is better. Overall they find that by using a combination of these techniques, they can derive dynamic height to an accuracy of ~4 cm.

In summary, direct observations of any relationship between SST and SSH are limited to a handful of qualitative studies. Several studies have contributed towards answering questions pertinent to whether or not SSH and SST are related, but the answers are limited to specific regions. Before using model or in situ data to attempt to establish some relationship between SST and SSH, the approach taken here is to see if the available remotely sensed data suggests any relationship. If it does not, it is pointless to progress further. If it suggests a relationship in some regions but not in others, then this guides one's choice for further research using in situ and/or model data. For the first time, the relationship between SST and SSH is quantified in the South Atlantic area for 1993 and 1994. This work is described in the following sections.
Chapter 5  Correlations between SST and SSH

5.4. Observations from TOPEX/POSEIDON and ATSR data

5.4.1 Data

The altimetry data used for this study are T/P SSH data spanning 1993 and 1994 (T/P cycles 11-84 inclusive). These data are processed with the techniques described in Chapter 3. In particular, a perpendicular bisector approach together with an across-track correction is used to collocate the data to a reference grid defined by the sub-satellite points of cycle 18. The CSR3.0 tidal model is used to remove the effect of tides, an inverse barometer correction is applied and the Gaspar et al. (1994) four parameter sea state bias correction is used. Data are rejected if the rejection criteria specified by Vincent et al. (1994) are exceeded.

The SST data used are the spatially averaged 0.5° data (ASST) from ATSR over the same time period as the T/P data. To investigate a relationship between SST and SSH, it is necessary to ensure that the SST data are as accurate as possible. Hence the filtering algorithm described in Chapter 4 is applied to all of the ASST data. Throughout this study, the mixed ASST data are used, which incorporate the advantage of the dual view whenever possible, together with the increased coverage provided by not requiring dual-view only data.

5.4.2 SST overlaid on dynamic topography

Initially, an attempt is made to compare patterns of absolute SST and SSH. To obtain an estimate of absolute SSH, the technique described in Chapter 3, Section 3.8 is used. This involves adding the smoothed (over a length-scale of ~2600 km) mean dynamic topography to height anomaly maps (interpolated to a 0.5° grid with a Gaussian FWHM of 110 km - see Chapter 3, Section 3.7) in order to obtain an estimate for the mean dynamic topography. This is accurate in regions where the small spatial scale mean flow is weak.

The dynamic topography from T/P for January 1993 is converted into geostrophic velocity (see Chapter 3, Section 3.3) and overlaid on the ATSR SST averaged within the same month. An example from the Agulhas region is shown in Figure 5.1. It is clear that in some areas a striking similarity exists between the SST contours and the T/P flow field. For example, the large meanders (wavelength ~ 600 km) in the Agulhas Return Current at 37°S, 30-40°E are present in both the flow field and the SST field; regions of high dynamic topography corresponding to regions of high temperature, and

---

2 Cycle 18 is close to the nominal T/P ground track (see Section 3.5).
Figure 5.1  T/P "absolute" geostrophic flow vectors overlaid on the ATSR ASST for January 1993.
vice-versa. At 41°S, 18°E however, there is an anticyclonic eddy present in the altimetric flow field which is not clearly visible in the SST. Similar patterns in these two fields can be observed in many energetic regions in this way, such as the Brazil/Falklands confluence and the northern area of the Drake Passage.

Given that these two data sources are completely independent, that the altimetry is interpolated, and that the method of estimating absolute dynamic topography is error prone, it is surprising to see such marked similarity in certain areas between the two fields. Although this is proof that there certainly is mutual information in both data sources, establishing a quantitative estimate of the relationship between SST and SSH from these comparisons is difficult because of the large errors involved. Thus a different approach is taken which reduces the errors, both in interpolating the altimetry and calculating absolute dynamic topography. This is described in the following section.

### 5.4.3 Comparisons of ΔSST with ΔSSH

To narrow the uncertainties, a more quantitative attempt is made to relate SST and SSH by using SSH anomalies (from the 1993-1994 mean SSH) along the T/P ground tracks. ATSR SST anomalies (from the 1993-1994 mean SST) are then collocated onto the T/P ground tracks by using a match-up criteria between the SST and SSH data of ±2 days and ±60 km. The criterion of 60 km was selected by considering that the spatial resolution of the ASST dataset is 55 km at the equator. This immediately sets a minimum spatial match-up criterion of 32 km (the distance of the corner of an ASST cell from its centre), since if the criterion is less than this there will be locations where it is not possible to collocate the SST even with full coverage. It is also desirable that each collocation point has the potential to obtain information from more than one ASST cell. This is to reduce the effects of noise and increase the amount of match-ups available. To achieve this, a criterion of greater than 55 km is necessary and 60 km was chosen. The criterion of 2 days was arrived at by experimenting with various criteria and choosing a compromise between a small time criterion, allowing more precise match-ups, and a large time criteria, allowing a larger quantity of match-ups and therefore improving the statistics of the SST SSH correlations.

An estimate of the SST at the collocation point \( T_0 \) is then obtained from the SSTs within the match-up criteria by using the following equation:

\[
T_0 = \frac{\sum_{i=1}^{N} w_i T_i}{\sum_{i=1}^{N} w_i} \tag{5.11}
\]
where $T_1$ are the SSTs within the match-up criteria, $N$ is the number of SSTs within the match-up criteria and the weights $w_i$ are given by:

$$w_i = e^{-\left(\frac{(x_i-x_0)^2}{\sigma_d^2} + \frac{(t_i-t_0)^2}{\sigma_t^2}\right)}$$  \[5.12\]

where $(x_i-x_0)$ are the distance of the SSTs from the collocation point, $(t_i-t_0)$ are the time between the SST measurements and the altimeter SSH measurement at the collocation point and $\sigma_d$ and $\sigma_t$ are 30 km and 1.2 days respectively.

The purpose of such a weighting function is to attach more weight to data nearer in time and space to the collocation point and less weight to data near the edge of the match-up criteria. The parameters of 30 km and 1.2 days were chosen after experimentation with several different parameters. The results presented here are largely insensitive to the choice of parameters. To ensure this, the results described below were also calculated with $w_i=1$ (i.e. just a simple average). It was found that the correlations between SST and SSH are only marginally smaller using the simple average compared to using the Gaussian weighting.

The collocation of the SST anomalies onto the T/P ground tracks does not present the same problems as interpolation of altimeter data onto a regular grid because the sampling of the SST data is regular and well-posed for this purpose. The alternative method of interpolating the altimeter data onto a 0.5° grid and comparing with the ASST data would incur the interpolation errors described in Chapter 3 and has therefore been avoided. These collocated data are analysed over the entire South Atlantic region for 1993 and 1994 to identify the regions and seasons when correlations between SST and SSH exist.

Three profiles of collocated $\Delta$SST and $\Delta$SSH along T/P ground tracks are shown in Figure 5.2. In certain places, there are close similarities between $\Delta$SST and $\Delta$SSH and it is clear that the shapes of $\Delta$SST and $\Delta$SSH profiles are related. For example in Figure 5.2a, a similarity in the region near 37°S is evident. In Figure 5.2b the structure of the SST and SSH profiles between 38°S and 45°S is similar. In Figure 5.2c the SST and SSH structure at around 40°S and 25°S seems to be related. It is also clear, however, that the relationship is far from simple. Indeed it is impossible to deduce from these few profiles any statistically significant relationship between $\Delta$SST and $\Delta$SSH.

---

3 For some purposes it may be perfectly acceptable to perform SST-SSH comparisons by interpolating the SSH data onto a regular grid. The purpose of the study in this chapter is to quantify as accurately as possible the SST SSH relationship; hence the SSH data is not interpolated.
Figure 5.2 Examples of collocated ATSR ASST residuals and T/P SSH residuals for (a) T/P pass 174, cycle 11, (b) T/P pass 174, cycle 30, (c) T/P pass 174, cycle 49.
To quantify the extent to which ΔSST and ΔSSH are related, data in the Agulhas region (0-40°E, 10-50°S) are considered. Restricting the geographical area reduces the amount of data for comparison and therefore reduces the amount of processing time required to perform the statistical tests. Zero lag cross-correlation coefficients (r) are calculated for each T/P cycle using [5.13] below:

$$\sum_{i=1}^{N} (\eta_i - \bar{\eta})(\sigma_i - \bar{\sigma}) \over N\sigma\eta\sigma_i$$  

[5.13]

where N is the number of collocated data points in each cycle, s denotes SST, η denotes SSH, σ_x is the standard deviation of x and a bar denotes the mean over each T/P cycle.

The results of these calculations for 1993 and 1994 (T/P cycles 11-84) are shown in Figure 5.3. The cross-correlations between ΔSST and ΔSSH, ΔSST gradients and ΔSSH gradients, between ΔSSH and ΔSST gradients and between ΔSSH gradients and ΔSST are shown. It is clear from the temporal variation of the time series shown in Figure 5.3 that the correlations between ΔSST and ΔSSH, and between the gradients of ΔSST and ΔSSH, are all different from zero, whereas correlations between the gradients of one parameter and the actual values of the other parameter are much weaker. The correlations between gradients of ΔSST and gradients of ΔSSH are nearly always slightly less than the correlations between ΔSST and ΔSSH. This can be attributed to two factors; (1) removal of any large scale correlation between ΔSST and ΔSSH, since taking the gradient is effectively high pass filtering the data and (2) increasing the noise in the data due to the calculation of gradients, and therefore reducing the correlations. It is likely that both of these contribute to the slight lowering of the gradient correlations. The magnitude of the correlations, although significantly different from zero is fairly small. Typical values of about 0.35 imply that SST variability can only explain about 13% of the SSH variability. Even the maximum value of 0.53 means that only 28% of the variability of one field can be explained by the other. One of the reasons why the correlations are so low is that these are correlations over the entire Agulhas region. If the relationship between SST and SSH is strong in some areas, but weak in others, the net result over the entire region will be an intermediate level of correlation. The geographical variation of the correlations is explored in the next section.

An interesting aspect of the correlations between ΔSST and ΔSSH shown in Figure 5.3 is the seasonal signal. The largest correlations occur in the austral winter and the smallest in the austral summer. This could be quantification of the result that Legeckis (1978) observed qualitatively, that in summer a warm pool of water can mask the
Figure 5.3 Time series showing the correlation between (i) SST residuals and SSH residuals, (ii) SST gradients and SSH gradients, (iii) SST gradients and SSH residuals and (iv) SST residuals and SSH gradients. The correlations are calculated for each cycle over the Agulhas Region.

Figure 5.4 The cross-correlation function between SST residuals and SSH residuals, calculated separately for ascending and descending tracks within the Agulhas Region.
underlying horizontal temperature structure and therefore obscure ocean fronts from
detection by infrared radiometers. Alternatively, the lower correlations in summer could
be due to the diurnal thermocline effect (e.g. Price et al., 1986; Hawkins et al., 1993)
where in regions with high net heat flux and low winds, the top few metres of the ocean
can be warmed by several degrees. This can mask the underlying density structure and
could cause poor correlations between ∆SST and ∆SSH. The seasonal signal in the
correlations is investigated in detail for the entire South Atlantic region in Section 5.4.5.

The mean cross-correlation function (CCF) for the Agulhas region for T/P cycles 11-84
is shown in Figure 5.4. It has been split into CCFs for both ascending and descending
tracks so that any differences due to different sampling orientations can be resolved.
The CCFs peak at about zero lag with a value of ~0.35 and first zero crossings are at
~200 km. This means that the average decorrelation length-scale is ~200 km (with an
associated wavelength of 800 km). It is clear from this cross-correlation function that a
quantitative relationship between ∆SST and ∆SSH exists. If the relationship between
these two variables was purely random, the cross-correlation function would not be
structured with a dominant peak at almost zero-lag.

From this preliminary investigation in the Agulhas region, it is clear that (1) a
relationship does exist between ∆SST and ∆SSH, (2) the relationship is strongest
between ∆SST and ∆SSH rather than between the gradients of these parameters or any
combination thereof and (3) a seasonality exists in the strength of the correlations. This
study is extended in the following sections by using data for the whole of the South
Atlantic region and by studying in more detail the geographical and seasonal variation
of the correlations between ∆SST and ∆SSH

5.4.4 Geographical dependence of the ∆SST ∆SSH relationship

The results of the previous section are interesting, but in order to fully investigate any
relationship between ∆SST and ∆SSH, it is vital to quantify the geographical
dependence. To do this, a technique is developed to calculate the zero-lag cross
correlations as a function of geographical location. The technique is to use an along-
track moving window and to calculate the zero-lag cross-correlations within this
moving window. The correlation value is assigned to the data point at the centre of the
window, the window is then moved on and the process repeated. Figure 5.5
demonstrates this technique. The ∆SST and ∆SSH profiles for a section of T/P pass
174, cycle 45 are displayed in Figure 5.5a. The aim is to quantify the places along these
profiles where ∆SST and ∆SSH are correlated and the places where they are not. Figure

4 The relationship between the gradients will not be considered because the ∆SST - ∆SSH relationship has been
shown to be stronger.
Figure 5.5 This illustrates the technique of using a sliding zero-lag cross-correlation to quantify the regional dependence of the SST SSH relationship. (a) The SST residuals are overlaid on the SSH residuals for T/P pass 174, cycle 45. (b) Zero lag cross correlations are calculated within a sliding 250 km window.
Chapter 5  Correlations between SST and SSH

5.5b shows the windowed zero-lag cross-correlation profile resulting from a window size of 41 points (250 km). It is clear that the correlations are high in the places where a good correspondence can be observed visually. This is therefore a quantitative way of measuring the strength of a relationship between $\Delta$SST and $\Delta$SSH as a function of location.

This technique is applied to all the collocated data within the South Atlantic region for 1993 and 1994. For every T/P cycle, a cross-correlation map is obtained. The utility of a correlation map for one cycle is limited because the number of degrees of freedom in a 250 km moving window is only ~5, and therefore the level at which the correlations are significantly different from zero is very high. To obtain a statistically significant picture of the correlations, the 74 correlation maps corresponding to all the T/P cycles in 1993 and 1994 are averaged. The resulting mean correlation for each location is shown in Figure 5.6. It is clear that the correlations between $\Delta$SST and $\Delta$SSH are very regional, with high correlations (~0.7) in regions associated with mesoscale variability and with strong fronts. For example, the region of high correlation stretching through the northern part of the Drake Passage is in the vicinity of the Subantarctic and Polar Fronts, as is the broad swath of high correlation extending across the latitude band 40-50°S. The region of high correlation just southeast of the tip of South Africa is associated with the Agulhas Return Current and the region to the west of South Africa is associated with the Benguela Current. A rather strange area of high correlation exists extending from 30°S at the east coast of South America to 35°S at 0°E. This region is not associated with any frontal region and is neither a region of intense mesoscale variability (see Figure 3.17). The reason for this is unclear and is an area for future research.

Away from the areas of high mesoscale variability, the correlations are low. Figure 5.7 shows the standard deviation (SD) of the correlations at each location. It can be seen that the SDs are low in areas where the correlations are high, and vice-versa. This is evidence that the correlations are consistently high in the regions of high mean correlation. The reason that the SD of the correlations is higher in the areas of low correlation is that in most of these areas, the geophysical signals are below the noise levels of the $\Delta$SST or $\Delta$SSH measurements. Hence the correlations are purely random, and should have zero mean and a high SD. The implication of Figures 5.6 and 5.7 demonstrates that there is indeed a relationship between $\Delta$SST and $\Delta$SSH in specific geographical regions associated with mesoscale variability. The square of the correlation values in Figure 5.6 is the percentage of variance in one parameter that can be explained by the other parameter in a linear relationship. Values in many regions are as high as 50%.
Figure 5.6  The mean zero-lag cross correlation for 1993 and 1994. The correlations are calculated using a 250 km moving window with both day and night ASST data.

Figure 5.7  The standard deviation of the zero-lag cross correlations for 1993 and 1994. The correlations are calculated using a 250 km moving window with both day and night ASST data.
To investigate the linear relationship between ΔSST and ΔSSH, the mean of dSST/dSSH is plotted in Figure 5.8. This shows that in the regions associated with high correlations, the gradient is ~0.1-0.2 mK⁻¹. These values are sensible when one considers that a typical mesoscale eddy might have a temperature contrast of ~2°C and a height contrast of ~40 cm. With knowledge of the correlation structure of the relationship between SST and temperature at depth and of the effect of salinity, it would be possible to obtain information about the depth to which these features persist. This, however, is outside the scope of this thesis.

5.4.5 The seasonal cycle in the ΔSST - ΔSSH correlations

The preliminary study of the ΔSST ΔSSH correlations in Section 5.4.3 in the Agulhas region clearly showed the existence of an annual cycle; the correlations being lower in the austral summer and higher in the austral winter. Figure 5.9 shows a time series of the spatial mean cross-correlations over the entire South Atlantic region. It is clear that an annual signal exists in the correlations.

In order to quantify the seasonality, a model of an annual cycle plus trend is fitted to this time series. This is also shown in Figure 5.9. The peak of this cycle is ~0.28 in mid-August, the middle of the austral winter, compared to a trough of ~0.18 in mid winter. Hence, the correlations in the winter are almost 60% stronger than in the summer. The reason for this cycle is unclear. It is possible that development of a diurnal thermocline might be stronger in summer and hence the correlations might be lower because of this effect decoupling the surface from the subsurface. It is also possible that a seasonal cycle exists in the skin effect and this could also decouple the surface from the subsurface. A third possibility is the effect that Legeckis (1978) observed, where summer heating creates a warm pool of water that masks the underlying frontal structures. Legeckis observed that between latitudes of 25° and 35° frontal structures are visible in winter but not in summer, whereas for higher latitudes frontal structures are visible throughout the year.

To investigate whether or not the diurnal thermocline effect is causing a seasonal cycle in the correlations, the above analysis was undertaken for day and night SST data separately. The results (not shown), rather surprisingly, indicate that there is virtually no difference between the correlations according to whether day or night SST data are used. The diurnal thermocline effect, in some circumstances, is known to obscure underlying thermal structure (e.g. Stramma et al., 1986). The fact that no difference in the day and night correlations occurs suggests that although the diurnal thermocline effect may occur occasionally, it is not the dominant effect in terms of obscuring underlying structure. This result also validates, to some extent, the cloud filtering
Figure 5.8  The mean gradient for 1993 and 1994 of the linear fit with ΔSSH as the independent variable and ΔSST as the dependent variable. The gradients are calculated using a 250 km moving window with both day and night ASST data.

Figure 5.9  Time series of the mean correlation over the South Atlantic Region for 1993 and 1994. Also shown is the fit of an annual model with a drift.
scheme described in Chapter 4. If the effects of cloud contamination were not properly removed, a lower nighttime correlation would be expected. This is not found here.

In order to study the geographical variation of the differing summer winter correlations, the correlation maps are calculated for January, February and March (JFM) and for July, August and September (JAS). These are shown in Figures 5.10 and 5.11 respectively. It is clear that the winter correlations are stronger than the summer correlations, especially in the region of high correlation extending across the South Atlantic at ~30-35°S and in the region associated with the Benguela Current at ~25°S, 10°E. These regions are both at latitudes close to those affected by the homogeneous pool of warm water caused by summer heating (Legeckis, 1978), and it is therefore likely that this is a cause for the differences in correlation between summer and winter.

The effect of any seasonal variation in skin effect is not investigated here, as the skin effect is a complex area of current research.

5.4.6 The spatial scale of the ΔSST ΔSSH correlations

To quantify the structure of the ΔSST ΔSSH relationship and to assess its possible application, it is necessary to investigate the spatial dependence of the correlations. A method of analysing this is to band-pass filter the collocated ΔSST and ΔSSH data before calculating zero-lag cross-correlations within a moving window. Various methods exist for band-pass filtering data, but these are mainly for gap-free data. The collocated data used in this study are extremely "gappy" due to the requirement for cloud free conditions within fairly strict match-up criteria. Hence conventional methods cannot be accurately used. The method used here to calculate the zero-lag cross-correlation at a wavelength λ is firstly to use a sliding window of length 2λ and to taper the data within the first and last λ/2 of the window using a ramp taper\(^5\) (0 at the window ends, 1 at a distance λ/2 into the window). Secondly, the data within each window is detrended. Finally, a sinusoid with a wavelength λ is fitted to both the ΔSST and the ΔSSH data by conventional least squares. The zero-lag cross-correlations are computed from the central λ of the fitted sinusoids. The result is a correlation map for each T/P cycle corresponding to a specific wavelength. These maps are averaged over 1993-1994 to obtain the mean correlation map for a specific wavelength. The spatial dependence can then be examined by performing the above analysis at different wavelengths.

Figure 5.12 shows the wavelength dependence of the correlations averaged over the entire SA region, separated into ascending and descending tracks. It is clear that the

---

\(^5\) Fitting a function by least squares is least accurate near data edges. Hence, to prevent edge effects, a tapered window is used.
Figure 5.10 Mean correlation in the austral summer (January, February, March).

Figure 5.11 Mean correlation in the austral winter (July, August, September).
Chapter 5  Correlations between SST and SSH

Figure 5.12  The wavelength dependence of the ΔSST ΔSSH correlations for the entire South Atlantic region (80°W-40°E; 70°S-10°N). Error bars correspond to ± one standard error.

Figure 5.13  The wavelength dependence of the ΔSST ΔSSH correlations for the Drake Passage region (75°W-50°W; 60°S-55°S). Error bars correspond to ± one standard error.
\(\Delta \text{SST} \Delta \text{SSH}\) correlations increase with wavelength, with the correlations from the descending data being typically 0.015 less than the correlations from the ascending data. This is slightly larger than can easily be attributed to error, as can be seen by the error bars which correspond to \(\pm 1\) standard error (SE) on the mean (calculated by \(\text{SE} = \frac{\text{SD}}{\sqrt{n \times 6.2/100}}\)) where 6.2 km is the data spacing and 100 km is taken to be the integral length-scale of the variability (Le Traon et al., 1991). This is evidence that the basin wide correlations are weakly anisotropic and increase towards larger length-scales.

As is shown and discussed in Section 5.4.4, the relationship between SST and SSH varies with geographical location. Hence, the wavelength dependence of the correlations may also vary with location. To investigate this, three specific regions of the band-passed correlation maps are defined. These regions are (1) the Drake Passage region (75°W-50°W; 60°S-55°S), (2) a region straddling the ACC (Subantarctic and Polar fronts) (32°W-10°E; 50°S-40°S) and (3) the Agulhas Return Current region (21°E-40°E; 43°S-35°S). Figure 5.13 shows the wavelength dependence of the correlations within the Drake Passage region. Although the error bars are larger due to the smaller region size, the structure of Figure 5.13 is clearly different to that of Figure 5.12. A peak in the correlations is evident at wavelengths of around 500 km. This peak is evident in both the ascending and descending data, although the descending correlations are clearly stronger than the ascending correlations. A similar peak at 500 km is also visible in the results for the ACC region (Figure 5.14). In this case however, very little difference exists between the ascending and descending data. Finally, the results for the Agulhas Return Current are shown in Figure 5.15. A weak peak in the correlations is present at 800 km in the descending data, although not in the ascending.

Care must be taken in interpreting these results. A common feature of all the correlation graphs is that the correlations become weaker at smaller wavelengths. This may not be due to geophysical reasons, but may purely be due to the match-up criteria used to collocate the data. For example, assuming an advection speed of 5 cm/s (e.g. Gordon and Haxby, 1990), a feature can move 9 km in 2 days (the match-up criteria). Hence if it had a wavelength of less than \(4 \times 9 = 36\) km it would be completely decorrelated. Obviously the longer the wavelength of the feature, the less the reduction in correlation. The collocation therefore places a limit on the extent to which the actual scale dependence of the relationship between SST and SSH can be established.

Even if the reduction in correlation at smaller wavelengths cannot be trusted, a reduction in correlation at larger wavelengths is seen in Figures 5.13, 5.14 and, to a lesser extent, in Figure 5.15. This cannot be due to poor collocation, and must be geophysical. A peak in the correlations at a wavelength of 600 km corresponds to a
Chapter 5 Correlations between SST and SSH

Figure 5.14 The wavelength dependence of the ΔSST ΔSSH correlations for the Antarctic Circumpolar Current region (32°W-10°E; 50°S-40°N). Error bars correspond to ± one standard error.

Figure 5.15 The wavelength dependence of the ΔSST ΔSSH correlations for the Agulhas Return Current region (21°E-40°E; 43°S-35°S). Error bars correspond to ± one standard error.
decorrelation length-scale of ~150 km or an eddy diameter of ~300 km\(^6\). This decorrelation length-scale is similar to that obtained from the cross-correlation functions in Figure 5.4 and is evidence that the high correlations apparent in Figures 5.6, 5.10 and 5.11 are caused by large eddies, meanders or Rossby waves rather than by the very large scale seasonal response of the ocean to varying heat and water fluxes (see Section 5.1). Distinguishing between these two physical mechanisms is critical if any relationship between SST and SSH is to be used to interpolate altimeter data or provide improved eddy statistics. If the relationship between SST and SSH was purely due to the large scale seasonal response, there would not be a relationship on small enough spatial scales to be useful for the above mentioned purposes. The results in this section are strong evidence that the cause of the high SST SSH correlations is due to advection at mesoscale wavelengths.

5.5 Conclusions

This chapter builds upon the results obtained in the previous two chapters by exploring the relationship between SSH and SST. Initially, the theory behind the reason for any relationship between SST and SSH is laid out and it is concluded that three conditions must be satisfied for a relationship to exist. Namely that (1) the SST structure is related to the temperature structure at depth, (2) the changes in density due to temperature are related to the changes in temperature due to salinity and (3) the motion is not purely barotropic.

The reason why a relationship would be useful is discussed in Section 5.2 and several applications are discussed. These are (1) model validation, (2) relating surface parameters to subsurface and (3) developing new techniques based on such a relationship. Two examples of new techniques are presented; (1) calculating eddy kinetic energy without having to make the commonly used isotropic assumption and (2) improving the resolution of altimetric SSH fields by using SST information within an optimal interpolation framework.

Following the description of the theory and the reasons why a relationship would be useful, a literature review of the previous work on this subject is given. It is shown that the state of the current research into this area is limited to a handful of qualitative studies and several studies into problems related to the relationship between SST and SSH.

\(^6\) A feature with a wavelength of \(L\) has an inverse wavenumber of \(L/2\pi\), a decorrelation length-scale of \(-L/4\) and an eddy diameter of \(-L/2\)
Given the lack of research in this field, an initial attempt to perform a quantitative study into the SST SSH relationship is given in Section 5.4. It is first shown that if the T/P geostrophic flow vectors are overlaid on the SST for January 1993, a remarkable agreement is present in some locations. This qualitative agreement is investigated using collocated SST and SSH anomalies and computing zero-lag cross-correlation coefficients within an along-track 250 km sliding window. The resulting 1993-1994 mean correlation map (Figure 5.6) reveals high correlations (~0.7) in regions generally associated with ocean fronts. Such regions include the Polar Front, the Sub-Antarctic front, the Agulhas Return Current, the Benguela Current and the Brazil-Falkland Confluence.

The strength of the spatial correlations between ΔSST and ΔSSH is shown to vary with season, with correlations in mid-August (austral winter) being almost 60% higher than correlations in mid-February (austral summer). The cause of this seasonality is investigated, firstly by studying the correlations using day and night SST data separately. It is found that using day rather than night SST data, and vice-versa, makes virtually no difference in the strength of the correlations and rules out the diurnal thermocline effect as a source of the seasonal variability. The geographical pattern of the correlation map for summer and winter reveals that the main difference between the correlations is at latitudes of ~30°S. This suggests that the isothermal pool of water caused by summer heating may be obscuring the underlying temperature structure and may result in lower correlations Legeckis (1978).

Finally, an analysis of the spatial dependence of the cross-correlations is performed. It is shown that in areas with high correlations (the Drake Passage and ACC), the correlations peak at wavelengths of ~500 km. This wavelength is typical of mesoscale variability and Rossby waves and is evidence that in these regions, the mechanism for the high correlations is advection at mesoscale wavelengths. This is distinct from the SST SSH correlation at large spatial scales (~10,000 km) caused by the seasonal change in surface fluxes (Knudsen et al., 1996). A relationship between SST and SSH at the smaller spatial scales will allow techniques such as interpolation of altimeter data and improvement of eddy statistics to be developed.
Chapter 6

Conclusions and Future Work

6.1 Introduction

The aims of this thesis were twofold:

(i) To study and improve techniques for examining ocean circulation spaceborne altimetric measurements of sea surface height (SSH) and satellite infrared radiometric records of sea surface temperature (SST).

(ii) To use the techniques developed in (i) to study the extent to which SST and SSH are related.

The datasets used in this study are those from the TOPEX/POSEIDON (T/P) altimetric satellite and the Along-Track Scanning Radiometer (ATSR). Both of these instruments are experimental and thus it is important that the accuracy of their data are investigated. Since new datasets may have different "characteristics" to previous datasets, it is also important to investigate the extent to which techniques developed based on previous data can be applied to these new data. This is the reason for the ordering of aims (i) and (ii). The following sections describe the main conclusions of this thesis and, where appropriate, discuss future avenues of research. The original work in this thesis is found in Chapters 3, 4 and 5; Chapters 1 and 2 being the introduction and the oceanography review respectively.

6.2 TOPEX/POSEIDON altimetry (Chapter 3)

Chapter 3 commences with a brief history of spaceborne altimetric measurements, followed by a description of the basic oceanographic theory which is necessary to understand why altimetry is useful for ocean circulation studies. A discussion of the T/P error budget follows where the resulting r.m.s. error of the measurement of the distance between a reference ellipsoid and the geostrophic sea surface is shown to be 6 cm. With such an accuracy, it is necessary to evaluate whether or not techniques used to process previous altimeter data are necessary and whether new techniques are appropriate.
6.2.1 Validation of across-track correction technique (Section 3.5)

A new technique that needs investigation is compensation for the variation in the across-track sampling of the altimeter (Brenner et al., 1990; Rapp et al., 1994). This variation can induce errors if across-track mean sea surface gradients are present. One method of correcting for this variation in across-track sampling is through use of a mean sea surface model. However the validity of the correction depends upon the accuracy of the mean sea surface model and must therefore be investigated. It is shown that the r.m.s. correction for T/P cycles 1-52 is 0.9 cm, although extremes can reach 10-20 cm. This correction is found to have little effect on the T/P SSH variability over most of the South Atlantic region. In areas of high across-track mean sea surface slope (such as the South Sandwich Trench), however, the T/P SSH variability is reduced by several cm after application of the correction. Regions of high mean sea surface slope are often associated with regions of steeply sloping bottom topography. These regions can be correlated with oceanographic variability and it is therefore useful to apply this across-track correction so that variability caused by across-track sampling variations is not mistaken for real oceanographic variability. An example is shown of a mesoscale eddy like structure which is completely removed by application of the across-track correction.

This work was undertaken with the 1991 Ohio State University mean sea surface model (MSS) (Basic and Rapp, 1992). A new MSS model incorporating T/P, ERS-1 and Geosat data was recently released. A future study should be carried out to determine how this new MSS model improves the accuracy of the across-track correction. Furthermore, the computing power available today makes a global study feasible. This would, of course, be more useful to the altimetric community.

6.2.2 Investigating the effects of interpolation (Section 3.7)

After processing altimeter data to obtain along-track height anomaly profiles, it is often necessary (either purely for display purposes, or in order to perform statistical tests that require regular data) to interpolate these along-track data to a regular grid. Although an important topic, little research has been conducted into the effects of interpolation on altimeter data. In this study, the effects of Gaussian interpolation on altimeter data are investigated using POCM data as a proxy for altimeter data. The effects of using different interpolation parameters are first investigated and it is found that the optimal \(^1\) interpolation parameters (the parameter varied is the full-width at half maximum (FWHM) of the Gaussian weighting function) are very sensitive to the amount of large

---

\(^1\) Defined by minimising the r.m.s. interpolation error or maximising the correlation between the interpolated and the control field.
wavelength error (orbit error, tidal error) present. Removing this long-wavelength error prior to interpolation, results in a smaller optimal FWHM than if no orbit error removal is applied. The optimal FWHM with 5 cm orbit error, removed using bias and tilt+bias techniques, is ~100 km. The effect of interpolation using the optimal FWHMs is then investigated and it is shown that the SSH variability map from the interpolated fields contains eddy-like structures not present in the control field. The largest interpolation r.m.s. errors are in regions of high variability away from altimeter tracks. The smallest r.m.s. errors are in regions of low variability near to or coincident with the altimeter tracks. In terms of percentage r.m.s. error (i.e. the noise to signal ratio), the largest errors are found in quiescent regions directly adjacent to very energetic regions.

The interpolation of altimeter data is a subject needing more research. The research described above focuses on one particular type of spatial interpolation. A comprehensive study, however, would test several types of interpolation and include both spatial and temporal interpolation techniques. A study of this kind would take several years to complete, but a definitive answer to the question of which interpolation technique and which parameters to use would be of tremendous benefit to the community. It is by no means obvious that so called "Optimal Interpolation" techniques are actually the optimal techniques to use, since the a priori signal and error covariance functions required for these techniques are not precisely known.

6.2.3 Obtaining a qualitative estimate of the absolute dynamic height (Section 3.8)

The largest step forward in altimetric research will take place when the geoid is measured to an accuracy that allows absolute geostrophic currents to be obtained on small spatial scales (~50 km). Until this happens, methods to try and obtain an estimate of the absolute dynamic topography are useful. A method to obtain the absolute dynamic topography from altimetry alone is laid out in Chapter 3 (Section 3.8). This method is validated by applying it to SSH data from the POCM. It is found that the method is poor in areas such as western boundary currents, where a small scale flow is constrained by the continental shelf and therefore does not move around much. Away from such small scale mean features, this method provides a useful qualitative picture of the absolute geostrophic flow field.

6.2.4 Altimetry summary

The results from Chapter 3 are used to process the altimeter data in an informed manner for the study in Chapter 5. In particular, the across-track correction is applied
throughout, interpolation is avoided and SST and SSH residuals are compared rather than the absolute values, since the errors involved in obtaining absolute SSH are large. Furthermore, the results should aid the oceanographic altimetry community in terms of intelligent processing of altimeter data.

Future research would be worthwhile in the area of the interpolation of altimeter data and in extending the across-track correction study to a global region with the new mean sea surface model.

6.3 Along-Track Scanning Radiometer SST (Chapter 4)

Chapter 4 focuses on the spaceborne infrared measurement of SST from the ATSR. This instrument is placed in context by a brief review of the history of infrared SST measurements. This is followed by a description of the error sources in the spaceborne infrared measurements of SST, with particular emphasis on the ATSR 0.5° spatially averaged SST (ASST) data. Of all these error sources, the two with the largest magnitude in ATSR ASST data are likely to be aerosol contamination and cloud contamination. Of these, the one most likely to represent a problem for ocean circulation studies is cloud contamination. This is because cloud contamination can manifest on shorter temporal and spatial scales than aerosol contamination (which is often on global scales), and hence may be mistaken for mesoscale variability.

6.3.1 Cloud contamination in the ATSR ASST data (Section 4.4.1 - 4.4.3)

The results of Chapter 3 indicate that there is a problem with the ATSR ASST nighttime cloud clearing in particular geographical locations within the South Atlantic region. The cloud contamination concentrates mainly in a latitude band between 30°S and 50°S and stretches northwards up the west coast of Africa to the equator. It appears as an increased night ASST variability and as a decreased night ASST mean. A lower limit for the amount of cloud contamination of the night ASST data is derived and found to be 5.7% for this dataset. Further confirmation of the cloud contamination comes from comparing ATSR data with in situ TSG data from the World Ocean Circulation Experiment A11 cruise which yields the relationship ATSR-TSG = -0.93±1.89 K, the large scatter and bias being due to cloud contamination of the night data.

6.3.2 A new filtering scheme is applied and tested (Section 4.4.4 - 4.4.5)

A filtering algorithm to remove the contamination is proposed, based on the assumptions that the day ASST data are free of significant cloud contamination and that
cloud contamination manifests itself as high frequency noise. In brief, the filtering algorithm computes the residuals from an annual and semi-annual model fitted to the daytime data and rejects data further than 3SD of the daytime residuals away from the model. Applying this filtering algorithm to the ATSR data eliminates 4.7% of the data and results in a much closer agreement between the day and night SSTs, both in terms of the variability and of the mean values. Furthermore, comparing the filtered ATSR data with the TSG data yields ATSR-TSG = -0.61±0.67 K, a large improvement on the previous comparison.

6.3.3 The cause of the cloud contamination is identified (Section 4.4.6)

The data rejected from this filtering algorithm is split into season and it is found that a strong seasonality is present. Both the location and seasonality of the percentage of rejected data correspond well with the seasonality of low stratiform clouds (fog, stratus, stratocumulus) shown by Klein and Hartmann (1993). This is strong evidence that this cloud type is causing the contamination in the ASST dataset, which is not surprising given that these cloud types are fairly uniform and low-lying. This makes them hard to detect with conventional cloud clearing algorithms.

6.3.4 Sea surface temperature summary

For the study of ocean circulation, accurate cloud clearing is extremely important. Of all the error sources in the infrared measurement of SST, cloud contamination is the one most likely to cause errors on small spatial and temporal scales (for the high resolution ATSR data, instrument noise is also an issue). These errors can be confused with mesoscale ocean circulation features such as cold core eddies. In attempting to eliminate cloud, it is important that oceanographic features are retained. The scheme proposed in Chapter 4 is based on a priori statistics from the essentially cloud free daytime ASST data. Hence the rejection criteria vary with geographical region which reduces the extent to which real features are rejected. This scheme is already in use by research groups at MSSL, RAL, SOC, University of East Anglia and the Proudman Oceanographic Laboratory.

Although a useful technique for post-processing of the data, a method based on SST data rather than brightness temperatures (BTs) is always a last resort because the ability to accurately distinguish clouds from true ocean data is vastly reduced when the information from different BT channels is not available. The ATSR Science team at RAL are investigating new techniques based on BT data and it is hoped that these techniques will solve the problem described here. An interesting study would be to
identify the specific regions of cloud contamination described in this thesis and in Jones et al. (1996b), to pick out daytime images, to identify marine stratiform clouds from the visible channel and to investigate what combination of views and BTs yields a valid detector of the marine stratiform clouds. Even with improved BT tests however, a post-processing technique will always be required for the most stringent applications of the ATSR data. The technique described in Chapter 4 is a suitable post-processing technique to use.

As WOCE moves from its observational phase to its Analysis, Interpretation, Modelling and Synthesis (AIMS) phase, it is important that the datasets gathered from the observational phase are of the highest possible quality. The filtering scheme described in Chapter 4 will be of benefit in this regard.

6.4 Correlations between SST and SSH (Chapter 5)

The work described in Chapters 3 and 4 is brought together in Chapter 5 to address a new and interesting scientific question: to what extent are SST and SSH related? A quantitative answer to this question has not yet been obtained, although, as the review section of Chapter 5 shows, several studies have shown that a qualitative relationship exists between SST and SSH. A brief description of the theory outlining the reasons why SST and SSH could be related is given and it is shown that three conditions must be satisfied before any relationship can be present. Firstly that variations in SSH are baroclinic rather than barotropic, secondly that a variation in SST is related to a change in surface density and thirdly that the change in surface density is correlated with a density change at depth. If any of these conditions are not met, there will be no relationship between SST and SSH.

6.4.1 Qualitative evidence for a relationship between SSH and SST is discovered (Section 5.4.2)

With the present abundance of different satellite remotely sensed datasets, it is important that these datasets are utilised together in order to ensure extraction of the maximum amount of information about the ocean circulation. An initial method of investigating any relationship between SST and SSH is to overlay these variables in order to spot regions where coherent patterns or structures exist in both. This is demonstrated in Chapter 5 in the Agulhas Retroflection region. The absolute geostrophic flow field obtained from the method described in Chapter 3 is overlaid on the filtered ATSR ASST data for January 1993 (Figure 5.1). This map clearly shows regions (such as the Agulhas Return Current) where a relationship between SST and
SSH exists; the geostrophic flow contours follow the isotherms. In other regions, a relationship is not so distinct. The problem with this kind of approach is that it is difficult to obtain quantitative information about the regions and times of a relationship between SST and SSH. It is also difficult to quantify the strength of such a relationship.

### 6.4.2 The strength and geographical variation of the SST-SSH relationship is investigated (Section 5.4.3 - 5.4.4)

A quantification of the relationship between SST and SSH is given by collocating ATSR SST anomalies (from the 1993-1994 time mean SST) onto the T/P reference grid (corresponding to the sub-satellite points of cycle 18) with match-up criteria of ±2 days and ±60 km. These criteria are selected as a compromise between obtaining a sufficient number of match-ups to ensure that the results are statistically significant and that the criteria are tight enough to avoid errors due to phenomena moving in time and/or space. A quantification of the relationship is then obtained by computing zero-lag cross-correlations within an along-track moving window. This yields a field of the cross-correlation coefficients for each cycle. When these are averaged into the mean cross-correlation map over two years of coincident T/P and ATSR data (1993-1994), the resulting picture (Figure 5.6) shows that regions of high correlation (> 0.6) between SST and SSH are indeed present. This quantifies the extent to which SST and SSH are linearly related. Correlation values reach 0.7 in several regions, corresponding to variations in SST explaining 50% of the variance in SSH (or vice-versa). The regions of high correlation correspond mainly to the regions of high variability, such as the Agulhas Return Current, the Brazil-Falklands Confluence and the Antarctic Circumpolar Current (ACC). Interestingly, the ACC is much more visible in the SST-SSH correlation map than in the SSH variability map, demonstrating that using both datasets may provide information that cannot be obtained from either alone.

### 6.4.3 The seasonal variation in the strength of the correlations is investigated (Section 5.4.5)

A seasonal variation in the strength of the correlations is found, correlations being 60% higher in the austral winter (July, August, September) than in the austral summer (January, February, March). To determine whether or not the development of a diurnal thermocline affects the strength of the correlations, computations are performed with day and night SST data separately. The results show that there is very little difference between correlations using day or night SST data. The geographical pattern of the seasonal correlations indicates that the difference is largest at a latitude of ~30°S. A
possible reason for this is an isothermal warm pool of water developing in the summer heating. This warm pool could obscure underlying thermal structure (Legeckis, 1978).

6.4.4 The spatial dependence of the correlations is investigated (Section 5.4.6)

To determine the spatial scales on which the SST and SSH anomalies are correlated, a coherency analysis is performed. The results vary according to region, but generally, in the regions of high correlations, the correlations tend to peak around wavelengths of 500km. This is evidence that the SST and SSH correlations are caused by advective changes due to eddies, meanders or Rossby waves rather than by the large scale seasonal signal due to seasonally varying surface fluxes. A SST SSH relationship at these smaller wavelengths will allow synergistic techniques for studying the ocean mesoscale to be developed.

It should be stressed that the SST-SSH relationship in this study is not complete, but is a precursor to a study that is anticipated to take a further three years. In this subsequent study, the full spatial and temporal SST-SSH dependence will be investigated, as will the effects of parameters such as surface pressure, heat flux and wind speed on the SST-SSH relationship.

6.4.5 Potential uses of a SST-SSH relationship (Section 5.2)

There are many potential uses of a relationship between SST and SSH. In the subject of comparing surface to subsurface characteristics, the existence of a relationship between SST and SSH is evidence that surface characteristics are correlated with subsurface characteristics. This is because SST is a purely surface parameter, while SSH depends on the vertical density structure of the ocean. Whether or not any useful parameters can be inferred from this (such as mixed layer depth or thermocline depth) is the subject of future research and would involve studying both in situ and model data.

Another potential use of the structure of the correlation maps between SST and SSH is in model validation. As global ocean models become "eddy-permitting", the only suitable global validation fields available are those obtained by satellite remote sensing. Typically, eddy kinetic energy or SSH variability maps are used as validation fields. A more stringent test, however, would be to determine whether or not a model has the correct relationship between SST and SSH. To do this, the model must have realistic representations of both the near surface thermodynamics (that govern SST) and the vertically integrated density structure. Comparison of SST-SSH cross correlation maps
with the equivalent fields from a model such as OCCAM should be a valuable model validation tool.

The final, and perhaps most exciting, application of the relationship between SST and SSH is in developing new techniques based on this relationship. One example of this would be to include SST information in an optimal interpolation scheme to generate high resolution SSH fields. The problems in interpolating altimeter data are described in Chapter 3. However if extra information from SST data is used, the errors in interpolating such data may be reduced. The resulting "pseudo-SSH" fields would be of considerable interest to the model assimilation community who require high resolution global fields to assimilate into their high resolution global models. The pseudo-SSH fields could also be used to improve our present knowledge of eddy statistics. With altimetry data, eddy kinetic energy can only be calculated at crossover points, unless an isotropic assumption is made. If it is assumed that the relationship between SST and SSH is isotropic, rather than the conventional assumption that the variability alone is isotropic, then higher resolution eddy kinetic energy fields may be generated.

6.5 Final remarks

The work described in Chapters 3 and 4 investigates the utility and accuracy associated with various techniques for the processing of altimetry and infrared SST data. Chapter 5 applies the knowledge gained from the research in Chapters 3 and 4 to an interesting and little researched question: are SST and SSH related? The work described in Chapter 5 provides the preliminary answer: yes, in certain regions. The refinement and application of this answer is a new and exciting field of research that may provide many benefits to the oceanographic research community. These benefits range from an increased understanding of the near surface ocean, to a useful model validation tool, to increased SSH resolution data. Turning the preliminary results in Chapter 5 into practical tools to enhance our knowledge of ocean circulation leaves plenty of scope for future research.
References


References


References


References


References


WCRP, World Ocean Circulation Experiment Implementation Plan: Scientific Background, 2, 1988b.


Williams, S. W., An investigation of sun glint in the 1.6μm channel of the Along Track Scanning Radiometer, MSc thesis, University College London, 1993.


