Data Assimilation Experiments using an Indian Ocean General Circulation Model

A Thesis
Submitted For the Degree of
Master of Science (Engineering)
in the Faculty of Engineering

by
Aneesh C. S.

Center for Atmospheric and Oceanic Sciences
Indian Institute of Science
Bangalore – 560 012

August 2006
The oceans which cover over 70% of the earth’s surface play a primary role in earth’s climate system. Hence, a good understanding of the oceans is necessary to understand the earth’s climate well. Ocean models are tools to understand the oceans better. However, even with more than 50 years of development in the field of ocean modeling, ocean models still have many deficiencies in reproducing the ocean fields accurately due to erroneous forcing datasets, subgrid scale parameterizations and inaccurate boundary and initial condition estimations. To overcome these deficiencies of models, a relatively new tool in ocean modeling termed as data assimilation is adopted. Data assimilation is the melding of observed data of the ocean state with the model estimated ocean state.

In this thesis we have studied the impact of assimilation of temperature and salinity profiles from Argo floats and Sea Surface height from satellite altimeters in a Sigma-coordinate Indian Ocean model. An ocean data assimilation system based on the Regional Ocean Modeling System (ROMS) for the Indian Ocean is established. This model is implemented, validated and applied in a climatological simulation experiment to study the circulation in the Indian Ocean. The validated model is then used for the implementation of the data assimilation system for the Indian Ocean region. This dissertation presents the qualitative and quantitative comparisons of the model simulations with and without temperature and salinity profiles and sea surface height anomaly data assimilation for the Indian Ocean. This is the first ever reported data assimilation study of the Argo temperature and salinity profile data with ROMS in the Indian Ocean.

The model solutions obtained when forced by two different climatological datasets have been compared. The two different climatological forcing datasets used were COADS (Comprehensive Ocean-Atmosphere Data Set (Slutz, Lubker, Hiscox, Woodruf, Jenne, Joseph, Steurer, & Elms, 1985)) and SOC(Southampton Oceanography Center (Josey, Kent, & Taylor, 1999))-ERA (ECMWF Re-Analysis (Källberg, Berrisford, Hoskins, Simmons, Uppala, Lamy-Thépaut, & Hine, 2005)) surface momentum and heat flux datasets. The model is capable of reproducing the Indian Ocean circulation well when compared to climatological observations. However, comparisons of the simulations done with the two different forcing datasets showed that the model forced by the COADS fluxes over-estimates the SST(Sea Surface Temperature) by about 1°C and SSS(Sea Surface Salinity) by about 0.5 psu inspite of the relaxation of the surface tracers to climatological values. The model forced by the SOC-ERA flux performs better in comparison in simulating the SST, SSS and surface currents.

An objective analysis scheme was used to interpolate observed data of temperature and salinity
profiles obtained from the Argo floats and sea surface height data from TOPEX/Poseidon satellites for the year 2004 onto the model grid. The interpolated data and the corresponding error estimate due to this interpolation is used for assimilation in the model simulations of the Indian Ocean. The mathematical framework for the objective analysis also known as the optimal interpolation scheme is presented. The ROMS model was forced by surface fluxes for the year 2004 from NCEP surface flux dataset for simulations with and without data assimilation. The assimilation scheme for this procedure is the sequential suboptimal assimilation. The error in previous literature (Dombrowsky & Mey, 1992) in the error estimate term for this assimilation scheme has been corrected for, in the model.

The assimilation of temperature and salinity fields into the hindcast experiment demonstrates that the sequential suboptimal interpolation data assimilation scheme implemented in this study can efficiently correct the model evolution in time and improve the model simulated fields. The model simulated SST resembles the synoptic observations of SST by the TMI satellite to a good extent. The Root Mean Square Error (RMSE) of the model simulations of SST without assimilation grows significantly over the time of the model simulation and has the steepest increase during the pre-monsoon months. However, the RMSE of the model simulated SST with assimilation does not grow significantly and shows almost a constant value behaviour. The spatial correlation of the model SST field with the satellite observed SST field is plotted to realise the regions of the Indian Ocean where the data assimilation has improved the simulated SST field. The regions with a higher density of available insitu measurements of the tracer profiles are the regions where the model improves the SST more significantly as compared to regions where the observations are sparse. This suggests that with a much more dense and evenly spreadout observation network we can imporve the model simulations over the entire model domain.

The model temperature and salinity profiles for the simulation with assimilation resemble insitu temperature and salinity fields obtained from the Argo floats to a closer extent, where as the tracer profiles in the model simulations without assimilation are more diffuse near the surface. An Argo profile from the Arabian Sea and from Bay of Bengal are taken to compare with the model simulated values of the temperature and salinity profiles at those locations. The model with assimilation simulated the mixed layer depth and also thermocline structure much closer to observed mixed layer depth and thermocline structure than the model without assimilation. In the model without assimilation simulations, the vertical profiles of temperature and salinity are much more diffuse and spread out than observed profiles. A significant improvement in the model simulations due to data assimilation in the model as shown in this study gives us an impetus to increase the available observations in the Indian Ocean further with these improved observations to establish a expedient nowcast-forecast modeling system for the Indian Ocean.
To
my parents who are responsible for all I am
and to
all my teachers who taught me how to extract sweetness from life.

-------------
"Although this may seem a paradox, all exact science
is dominated by the idea of approximation."

BERTRAND RUSSELL
Acknowledgements

No one deserves more thanks for the success of this work than my adviser Dr. P. N. Vinayachandran. I wholeheartedly thank him for his guidance. I thank Dr. P. N. Vinayachandran for his continued support throughout my years as a graduate student. I always looked upon him for advice. He very patiently critiqued my research approach and results. Without his trust and guidance this thesis would not have been possible. I feel that I am more disciplined, simple and punctual after working under his guidance.

The opportunity to watch Dr. P. N. Vinayachandran in action (particularly during the discussions) has fashioned my way of thought in oceanography. He has been a valuable adviser, and I hope my one and half years of working with him have left me with atleast few of his qualities.

I am thankful to the Chairman, Center for Atmospheric and Oceanic Science for providing the facilities available in the department.

I am privileged to learn the basics of meteorology, oceanography and mathematics from the great teachers: Prof. J. Srinivasan, Prof. B. N. Goswamy, Dr. R. S. Nanjundaiah, Dr. P. N. Vinayachandran, Dr. D. Sengupta, Dr. S. K. Satheesh, Dr. V. Venugopal and Dr. Atanu Mohanty.

I wish to thank Jaison Kurien and Praveen for helping me out from being a novice at ocean modeling to gaining some amount of experience in the field by patiently teaching me the various tools involved and also clearing up some of my inane doubts when I began working with ocean modeling. I also extend my gratitude to Dr. V. Venugopal for all the discussions and clearing up a lot of doubts in my mind. I extend my gratitude to all my friends and colleagues at the center for making my two years cherishable.

I wish to thank the CAOS staff Ms. Rama, Ms. Nagarathna and Mr. Mohan for being of very great help in administrative works. I am thankful to my labmates: Dr. Sooraj and J. V. S. Raju for their help. I will always cherish the memories of discussions with Raju which stochastically stepped from meteorology to life and its purposes with neither of us having a clue! I also wish to thank him for all the guidance in my moments of confusion.

I will never forget the time I spent with Asit, Ambi, Doc, Gunthi, Joey, PJ, Ravi, Sai, Sundara. The unending discussions on any sundry topic in the world at the mess tables are etched in my memory forever. Thanks to all IISc Football and Hockey club members for making my evenings so much more fun after a hard day’s work at the lab! I owe a lot to Asit, Ambi and Sai for channelising my futile fervor into fruitful endeavours. A hearty thanks to Sundara and Joey for making life so much more fun at IISc. I owe a special thanks to Sai for taking time out from his busy schedule to check for errors in my thesis.
I am forever indebted to my brother Hemang for his prayers, guidance and inspiration and to my sister Susmitha for all the care.

I am indebted to my father for his continuous encouragement. I owe everything to my mother for taking care of every need of mine. I dedicate this thesis to my parents and to my teachers.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Ocean Models as Tools For Ocean Science</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Ocean General Circulation Models</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Essence of Data Assimilation</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Development of Observations for World’s Oceans</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Motivation</td>
<td>6</td>
</tr>
<tr>
<td>1.4 A reader’s guide to the thesis</td>
<td>8</td>
</tr>
<tr>
<td>2 Seasonal dynamics of the Indian Ocean: A study using ROMS</td>
<td>10</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Model, Data and Experimental Setup</td>
<td>12</td>
</tr>
<tr>
<td>2.2.1 Model Experiments</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Surface Momentum Forcing</td>
<td>15</td>
</tr>
<tr>
<td>2.3.1 Climatological Winds</td>
<td>15</td>
</tr>
<tr>
<td>2.4 The Indian Ocean Circulation</td>
<td>17</td>
</tr>
<tr>
<td>2.4.1 Indian Monsoon Currents</td>
<td>18</td>
</tr>
<tr>
<td>2.4.2 Arabian Sea and Bay of Bengal Circulation</td>
<td>23</td>
</tr>
<tr>
<td>2.4.3 Equatorial Surface Jets</td>
<td>24</td>
</tr>
<tr>
<td>2.4.4 Sea Surface Temperature</td>
<td>24</td>
</tr>
<tr>
<td>2.4.5 Sea Surface Salinity</td>
<td>26</td>
</tr>
<tr>
<td>2.5 Conclusions</td>
<td>27</td>
</tr>
<tr>
<td>3 The Oceanographic Data Assimilation Problem: Overview, Motivation and Recipe</td>
<td>35</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>35</td>
</tr>
<tr>
<td>3.2 Historical Perspective</td>
<td>36</td>
</tr>
<tr>
<td>3.2.1 Objectives of Oceanographic Data Assimilation</td>
<td>41</td>
</tr>
<tr>
<td>3.3 Objective Analysis</td>
<td>42</td>
</tr>
</tbody>
</table>
3.3.1 The Statistical Model ........................................... 43
3.3.2 Correlation Function ........................................... 45
3.4 Forecast and Assimilation Cycle ................................. 46
3.5 Conclusions ....................................................... 49

4 Experiments with Assimilation:ROMS and ARGO data system 50
4.1 Introduction ......................................................... 50
4.2 Regional Ocean Modelling System (ROMS) Ocean Model Configuration ............... 51
  4.2.1 Supplied Forcing ............................................. 53
  4.2.2 Assimilation Data Sets and Methods ....................... 53
  4.2.3 Data used for Sea level Variation Assimilation ............. 53
  4.2.4 Data used for the Temperature and Salinity Assimilation .... 54
4.3 Assimilation Experiment Procedure ............................ 57
  4.3.1 Initial and Forcing Conditions ......................... 57
  4.3.2 Forecast and Assimilation Cycle ....................... 58
4.4 Results of the Assimilation Experiment ....................... 59
  4.4.1 Impact of assimilation on the sea surface temperature .... 60
  4.4.2 Quantitative Analysis of the Ocean SST Field ............. 60
  4.4.3 Impact of assimilation on the Thermocline Depth .......... 64
  4.4.4 Impact of assimilation on the sea surface height anomaly ... 66
4.5 Conclusions ....................................................... 71

5 Summary and Conclusions .......................................... 73
  5.1 Objective of the Dissertation ................................. 73
  5.2 Contributions of the Dissertation .......................... 73
  5.3 Future Directions .............................................. 76
  5.4 Concluding Comments ......................................... 76
Appendix A A brief overview of ROMS ............................... 77
  A.1 Equations of the Model ...................................... 77
  A.2 Boundary Conditions ......................................... 80
  A.3 Model Parameterisation ....................................... 81

Bibliography .......................................................... 82
List of Figures

1.1 Different classifications of ocean model developed over the last two decades with the principle attributes of ROMS ticked. Source: CSEP notes on Ocean Models ©1995  

2.1 Model Domain with the bathymetry values in meters in the Indian Ocean  

2.2 Model vertical grid at 90° longitude in the Indian Ocean  

2.3 Bimonthly Wind Stress fields (Nm⁻²) (shaded) along with the wind vectors depicting the seasonal reversal (derived from SOC Climatology)  

2.4 A Schematic representation of the surface currents in the Indian Ocean during the Summer Monsoon from Shanker et al. (2002)  

2.5 A Schematic representation of the surface currents in the Indian Ocean during the Winter Monsoon from Shanker et al. (2002)  

2.6 Bimonthly Sea Surface current vectors ν' from the model run forced by COADS dataset over the Indian Ocean  

2.7 Bimonthly Sea Surface current vectors ν' from the model run forced by SOC-ERA dataset over the Indian Ocean  

2.8 January-March Surface Temperature fields from the model solution forced by COADS, SOC-ERA dataset compared with WOA01 Climatology field over the Indian Ocean  

2.9 May-July Surface Temperature fields from the model solution forced by COADS, SOC-ERA dataset compared with WOA01 Climatology field over the Indian Ocean  

2.10 Sep-Nov Surface Temperature fields from the model solution forced by COADS, SOC-ERA dataset compared with WOA01 Climatology field over the Indian Ocean
2.11 January-March Surface Salinity fields from the model solution forced by COADS, SOC-ERA dataset and a WOA01 Climatology field over the Indian Ocean ................................. 31

2.12 May-July Surface Salinity fields from the model solution forced by COADS, SOC-ERA dataset and a WOA01 Climatology field over the Indian Ocean ........................................... 32

2.13 September-November Surface Salinity fields from the model solution forced by COADS, SOC-ERA dataset and a WOA01 Climatology field over the Indian Ocean ............................... 33

3.1 A schematic of data assimilation cycles in model forecast run ........ 46

4.1 Schematic diagram of a single cycle in the mission of a profiling float. Source: www.bom.gov.au ................................................................. 55

4.2 Schematic diagram of the cross-section of an Argo Float. Source: www.bodc.ac.uk ................................................................. 56

4.3 The trajectory of Argo Floats in the Indian Ocean region in the year 2004 ................................................................. 56

4.4 Comparison of SST reconstructed by the model with and without Data assimilation with TMI SST data ................................. 61

4.5 Comparison of SST reconstructed by the model with and without Data assimilation with TMI SST data (Values above 33°C are not plotted) ................................. 62

4.6 Comparison of SST reconstructed by the model with and without Data assimilation with TMI SST data (Values above 33°C are not plotted) ........................................... 63

4.7 RMSE between SST reconstructed by the model (with and without Assimilation) and TMI SST data ................................. 64

4.8 Correlation coefficient between model SST and TMI SST data with red regions significant above 99% and blue significant above 95% ................................. 65

4.9 Comparison of Thermocline depth reconstructed by the model with and without Data assimilation for the months of January-February-March 2004 ........................................... 67

4.10 Comparison of Thermocline depth reconstructed by the model with and without Data assimilation for the months of July-August-September 2004 ........................................... 68

4.11 Vertical profiles of Temperature at 65.5E-21N in the Arabian Sea (left panel) and 89.5E-11.5N (right panel) in the Bay of Bengal till 500m ................................. 69
4.12 Vertical profiles of Salinity at 65.5E-21N in the Arabian Sea (left panel) and 89.5E-11.5N (right panel) in the Bay of Bengal till 500m . . . . 70

4.13 Comparison of SSHA reconstructed by the model with and without Data assimilation with Topex/Poseidon SSHA data . . . . . . . . . . 72
Abstract

This chapter serves as an introduction to the thesis. The purpose is to motivate the discussion on the oceanographic data assimilation problem by introducing in broad brush-strokes a picture of the development of ocean modeling and data assimilation fields of research and their main motivations. It illustrates the objectives of combining fully complex ocean general circulation models (OGCM) and oceanographic data. An overview of the problem studied in this thesis of subsurface data assimilation into a basin scale sigma coordinate primitive equation ocean circulation model configured for the North Indian Ocean is presented. This chapter also has a road-map of the thesis, which should serve as a reader’s guide.

1.1 Ocean Models as Tools For Ocean Science

The earth is the only place in the universe known by mankind that supports life. Covered by a thin blanket of air, a thinner film of water and a thin veneer of soil, the three combine to support a web of life of wondrous diversity and continual change. The unique characteristics of the earth in order for it to support life is its climate. The very fine balance of the earth’s climate provides just the necessary condition required to support life. The impact of a changing climate on such a fine balance could be profound. Hence, an indepth study of the climate system and its component subsystems is very essential for humans to survive on this planet without spoiling its fine balance and supported diverse life forms.

The oceans which cover over 70% of the planet’s surface play a primary role in the earth’s climate system (Mellor, 1996). A column of ocean water only 3m thick contains as much heat capacity as the entire atmospheric column above it (Gill, 1982). This and other key properties of the earth’s oceans make it a huge reservoir for heat, water and carbondioxide. The oceans are huge buffers and also have dynamics at very slow time scales which make them the key regulators of the earth’s climate system.

A scientific understanding of the ocean’s time mean state and its variability about this mean and its response to various perturbations represents the key goal of physical
oceanography and climate science. Our scientific and economic limitations to conduct experiments on such a large scale as the entire world ocean necessitates the search for alternate tools to study the world oceans and climate at large. This day, with the technological revolution and multifold increase in computational capability, computer models are becoming the primary tools used to study and predict ocean circulation and climate change. (Kowalik & Murty, 1993).

1.1.1 Ocean General Circulation Models

Models of the ocean range in complexity from theoretical models with mathematical equations that depict ocean circulation to realistic global ocean circulation models which are numerical solvers of the mathematical formulation of ocean circulation.

The first computer models to study climate and weather were developed in the 1950s. The development of complex ocean general circulation models also began parallel to the development of climate models (Bryan, 1969; Griffies, Boning, Bryan, Chassignet, Gerdes, Hasumi, Hirst, Treguier, & Webb, 2000; Griffies, 2004; Haidvogel & Beckmann, 1999).

In the last two decades ocean general circulation models have developed into different classes depending on the desired purpose for which they are applied. Some of the main criteria by which oceanographers classify their models are illustrated in Figure 1.1. The ROMS model used for the present study is a basin-scale hydro-thermodynamic free-surface sigma coordinate ocean model which solves for the barotropic and baroclinic equations of motion.

Though ocean general circulation models have evolved over the last 40-50 years, they are still confronted with having to solve an initial value problem without the right initial data. Even the highest resolution ocean circulation models cannot resolve all of the dynamically important physical processes in the ocean, from the small scale turbulence to basin scale currents. There will always be processes that are not represented directly, but rather are parameterized. These parameterizations are mostly complicated and always quiet uncertain both in form and in value. The uncertainty in these parameterizations is accompanied by an extreme sensitivity of the model results to slight variations in them (Chassignet & Verron, 1998). To overcome the deficiencies of the present ocean general circulation models in their physics as well as computation, the methodology of data assimilation was introduced into ocean modeling.
Figure 1.1: Different classifications of ocean model developed over the last two decades with the principle attributes of ROMS ticked. Source: CSEP notes on Ocean Models ©1995
1.2 Essence of Data Assimilation

The terminology “data assimilation” developed in meteorology about 30 years ago as the methodology in which observations are used to improve the forecasting skill of operational meteorological models. Data assimilation (DA) uses observational data to generate the best estimate of the true state of a system by combining these data with a numerical model of a dynamical system. The analysis is used as the initial condition for a numerical forecast by the model. The chaotic nature of the atmosphere and oceans means that small errors in the initial conditions may amplify rapidly (Lorenz, 1993), and therefore the analysis needs to be as accurate as possible. For this reason, data assimilation is becoming one of the most important parts of numerical ocean modeling. When used in the oceanographic context, the term data assimilation has acquired a much broader meaning. Oceanographic data assimilation denotes a collection of different methodologies such as the inverse methods, variational methods (Sasaki, 1958) and sequential estimation. All of these methods attempt to constrain a dynamical model with available data. The observations of real time ocean state variables have errors and also have a sparse spatial distribution. Hence the DA algorithm must optimally interpolate between the observation points, while ensuring that the fields are oceanographically realistic.

L. F. Richardson (Richardson, 1922), as a part of his historical effort to produce the first numerical forecast in the 1920s used an interpolation technique for relating the value of a field variable of interest between the computational grid and observation location by hand using only mechanical computing devices available at that time. This interpolated field variable on the grid was then used as the initial condition for solving a system of nonlinear model equations to create a forecast. With the introduction of the stored program digital computers in the late 1940s, it was soon realized that much of the efforts in numerical weather forecast can be automated (Charney, Fjörtoft, & von Neumann, 1950).

There were various classical DA algorithms developed during the 1950s and 1960s from linear and polynomial interpolation to successive correction methods. In the former Soviet Union, Gandin championed the use of a technique called optimal interpolation (OI). This technique belongs to a class of methods earlier developed independently by two of the most influential mathematicians of all times - Norbert Weiner (Weiner, 1949) in USA and Kolmogorov (Kolmogorov, 1941) in the former Soviet Union.

The OI scheme, first derived by Gandin (Gandin, 1965) and later developed
and implemented by other oceanographers such as Bretherton (Bretherton, Davis, & Fandry, 1976), Lorenc, Daley (Daley, 1991) among others is used in the present study. The basic reference for the OI equations used here is described by Carter and Robinson (Carter & Robinson, 1987). Comprehensive description of this methodology can also be found in Gandin (Gandin, 1965). The current and expected explosion in remotely sensed and insitu measured oceanographic data is ushering in a new age of physical oceanography. This increased observations of ocean parameters helps us to advance our knowledge of ocean dynamics and also translate this increased understanding into prediction and forecast models.

A good implementation of data assimilation can be achieved with the availability of large number of good quality observations of the oceanic fields as both synoptic and in-situ data. With the technology in satellite meteorology and also sensor floats for oceanography advancing by leaps over the past two decades, good synoptic and insitu observations of oceanic surface fields have been achieved. The current and expected explosion in remotely sensed and insitu measured oceanographic data is ushering a new age of ocean modeling and data assimilation.

### 1.2.1 Development of Observations for World’s Oceans

There had been very few hydrographic observations of large sections in the world’s oceans prior to 1970s. Available observations until the 1980s were dominated by regional and small scale observations of the ocean fields. The World Ocean Circulation Experiment (WOCE), conceived in the late 1970s was the first major effort to observe the world’s oceans and gather insitu observations on a large scale. Neutrally buoyant floats with Current-Temperature-Depth probes (CTDs) were used in the world oceans by a hugely collaborated world oceanographers community. With the completion of WOCE in the 1990s and introduction of satellite measurements of the ocean surface state variables such as sea surface temperature and sea surface height in the 1980s, wide spread data sets of oceanic variables were made available for analysis of the ocean state and its scales of dynamics.

Our only extant true global scale observations come from space. These, however, are currently restricted to surface properties; this restriction in turn means that only a small number of the possible measurements are really of interest, including altimetry, roughness (for the wind field) and temperature. It is difficult to imagine any future ocean observations not requiring satellites of these types. One also needs global insitu measurements. Here the trade off tends to be of accuracy and precision versus the
need for large number of measurements.

The Argo programme (deployment of several thousand robotic profiling floats) is an example of a large scale in situ data measurement experiment that is required at present: the production of thousands of adequate (not perfect or even the best one can do) measurements worldwide for a long period of time.

The major datasets that inspired the field of data assimilation were the synoptic observation fields of Sea Surface Temperature (SST) and Sea Surface Height Anomaly (SSHA) gathered from satellites and the in situ observations of temperature and salinity profiles from the Argo programme.

The availability of observations of oceanic fields increasing by the day has necessitated the development of good tools to make use of these observations and also of the present knowledge of the ocean circulation physics to guide us to a better understanding of the oceans. With the advent of data assimilation as a tool to incorporate observed knowledge of the ocean state into circulation models, the field of ocean state estimation is developing by leaps. There have been various studies on the assimilation of observed data of surface and subsurface data into oceans all around the world to understand their dynamics better. (Halpern, Ji, Leetma, & Reynolds, 1998; Killworth, Dieterich, Le Provost, Oschlies, & Willebrand, 2001; Greinera, Arnaultb, & Morlière, 1998a, 1998b; Fox, Haines, deCuevas, & Webb, 2000)

#### 1.3 Motivation

The North Indian Ocean is quite different from all other oceans of the world primarily due to its seasonally reversing monsoon wind system and its impact on millions of lives in one of the most densely populated and poorest regions in the world which surround the Indian Ocean. The oceanic currents and atmospheric winds of the North Indian Ocean reverse their patterns seasonally, causing dramatic seasonal changes, ranging from flooding rains to dreadful droughts. The Indian Ocean is bound in the north at about 25° N by land. Hence the currents in the Indian Ocean cannot transport and discharge heat from the equatorial region to higher latitudes unlike the Gulf stream in the Atlantic Ocean or the Kuroshio current in the Pacific Ocean. Also, the North Indian Ocean is connected to the western Pacific ocean characterized by its “warm pool” (Vinayachandran & Shetye, 1991). The unique feature of the North Indian Ocean is its two basin-split into the Arabian sea and Bay of Bengal which is a
recipe for its unique oceanographic features. These and many other unique features of
the Indian Ocean have made it a region of considerable interest to oceanographers and
atmospheric scientists alike.

Modeling studies in the Indian Ocean have not been as extensive as in other parts of
the world’s oceans. However, our present understanding of the dynamics of the North
Indian Ocean is determined primarily from the numerical models (Shankar, Vinay-
chandran, & Unnikrishnan, 2002). With the increased interest in the Indian Ocean
dynamics and its inherent complex circulation as compared to the other oceans, the
studies of the Indian Ocean are increasing in the last and present decade. There have
been several efforts to paint a complete picture of the Indian Ocean dynamics, both
spatial and temporal (Potemra, Luther, & O’Brien, 1991; Vinayachandran, Kagim-
oto, Masumoto, Chauhan, Nayak, & Yamagata, 2005; Vinayachandran, Murty, &
Babu, 2002; Vinayachandran, Iizuka, & Yamagata, 2002; Unnikrishnan, Kumar, &
Navelkar, 1997; Behera, Salvekar, & Yamagata, 2000). Some of the major efforts in
this direction are the studies by Schott and McCreary. (2001), Shankar et al. (2002)
and McCreary, Kundu, and Molinari (1993). Though there is a comprehensive review
of the general circulation in the Indian Ocean in these studies, there are still some
less understood features in the Indian Ocean circulation which necessitate some more
study. This thesis is an effort in the direction to setup a good ocean circulation model
for nowcast/forecast of the Indian Ocean and to understand it better. Although these
models offer good qualitative reproduction of the oceanic fields, they have deficiencies
in reproducing tracer as well as flow fields quantitatively. To overcome these deficien-
cies in model studies, it is necessary to assimilate real-time observations into the model
to bring their simulations closer to the observations.

The Indian Ocean used to be a region which was sparsely covered by in situ data,
before the advent of the Argo floats program. With the introduction of the Argo floats
into the Indian Ocean (Ravichandran, Vinayachandran, Joseph, & Radhakrishnan,
2004) and also newer and better satellite altimetry and temperature remote sensors,
a very good estimate of the ocean circulation can ultimately be obtained through as-
simulation systems which represent an optimal way of merging the knowledge about
the ocean obtained independently from models, in situ and remote sensing observa-
tions which have become the norm of the day. With the number of observations in
the Indian Ocean increasing by the day and also their quality improving, we have a
good dataset and an ocean model to implement a nowcast/forecast data-assimilative
modeling system for the Indian Ocean region.

Only three other previous assimilation studies in the Indian Ocean are known
for now. (Greiner & Perigaud, 1994) and (Greiner & Perigaud, 1996) assimilated Geosat altimeter sea-level variations into a nonlinear reduced gravity model of the Indian Ocean using the adjoint method. Another work is by (Lopez & Kantha, 2000), who applied the optimal interpolation method to a version of the sigma coordinate, free surface, primitive equation Princeton Ocean Model (POM) for the North Indian Ocean. (Vibeke & Evensen, 2002) studied the assimilation of Sea level anamoly and sea surface temperature data from satellites into an OGCM for the Indian Ocean.

The main objective of the study in this thesis is to implement and evaluate the impact of data assimilation of surface and subsurface tracer observations into a basin scale sigma coordinate primitive equation multilevel model of the North Indian Ocean and in turn to study the feasibility of building a accurate regional ocean modeling system of the North Indian Ocean. There have been very few applications of a sigma coordinate primitive equation model to the North Indian Ocean basin and no study known to me of the impact of data assimilation of subsurface tracers in an Indian Ocean model simulation. Hence this study of the impact of data assimilation of subsurface data into a ocean model of the Indian Ocean gathers importance in the evolution of a nowcast/forecast system for the Indian Ocean. Subsurface temperature and salinity profiles derived from the Argo robotic floats have been optimally interpolated onto the computational grid and sequentially assimilated in the model run forced by surface fluxes derived for the year 2004.

The experiment model runs with and without data assimilation are compared qualitatively and quantitatively to study the effect of data assimilation on the model simulations. It is observed that the data assimilation significantly improves the model simulations of the surface and subsurface tracer fields and also the overall circulation pattern.

1.4 A reader’s guide to the thesis

Roadmap

Apart from this chapter this thesis contains four other chapters. We now briefly outline a summary of each chapter.

In Chapter 2, we present a brief review of the previous studies of the Ocean General Circulation in the North Indian Ocean and the main features observed. We discuss the model setup for this study and present results obtained in an evaluation study of
the model’s climatological simulation of the North Indian Ocean circulation. Also, we compare the model results obtained when forced by two different climatological datasets and determine which forcing dataset helps the model to simulate the observed climatological circulation better and this dataset is to be used in the data assimilation studies.

In Chapter 3, we discuss various aspects of oceanographic data assimilation, its motivation and purposes. We discuss the implementation methodologies for data assimilation schemes and also give a brief overview of previous studies in data assimilation. We present the derivations of the data assimilation schemes used for the present study and also correction to the existing sequential optimal interpolation error term.

In Chapter 4, we discuss the motivation to setup a data assimilative model for the North Indian Ocean, the available observations for data assimilation and the model forcing and boundary conditions for the study. We present the results obtained for the study on the impact of subsurface data assimilation into the model. The derived oceanic fields from the model runs with and without data assimilation are compared qualitatively and quantitatively to observed oceanic fields to present the improvement in the model simulations due to data assimilation.

Finally, in Chapter 5, we summarize the contributions of this thesis, and discuss possible future directions.
2 Seasonal dynamics of the Indian Ocean: A study using ROMS

Abstract

This chapter gives a general overview of the studies done on Indian Ocean circulation and builds the background for the model setup done for the present thesis study. It presents the main features of the Indian Ocean circulation that need to be studied in a model experiment and presents some of the intercomparison results of our study of the Indian Ocean using different forcing data.

2.1 Introduction

The Indian Ocean is only the third largest of the oceans of the world. Yet it is considered to be highly complex and the least understood of all the oceans. Of the three major oceans - Pacific, Atlantic and Indian - the Indian Ocean is unique in that its northern boundary is located primarily in the tropics. It is also the only ocean with a low-latitude opening in its eastern boundary. The unique geography of the Indian Ocean has important implications for the oceanic circulation and climate of the region. The Indian Ocean cannot transport heat from the low latitudes to the higher latitudes as happens in the Atlantic and Pacific, mainly via the strong western boundary currents in the Northern Hemisphere. It gains additional heat through the Indonesian throughflow from the tropical Pacific Ocean. Heat is carried southward along the western coast of Australia toward the southern subtropics. The Indian Ocean consequently has a complex circulation pattern as compared to the other two major oceans of the world.

A consequence of the unique geography is that the northern Indian Ocean is forced by intense, seasonally reversing monsoon winds. These strong winds force the ocean locally and excite waves that travel large distances to affect the ocean and correspondingly the climate remotely in time and space (Luther & O’Brien, 1985; Shankar et al.,
The intense monsoon winds cool the sea surface temperature significantly, by direct heat loss from the ocean surface and also due to the upwelling of cooler subsurface waters along the western boundaries caused due to the intense parallel winds to the western boundary. This makes the annual variation of the Sea Surface Temperature (SST) in the Arabian Sea one of the largest in any ocean (Lee, Jones, Brink, & Fischer, 2000). Thus, the Indian Ocean is an ideal “laboratory” for studying a variety of phenomena related to ocean dynamics such as coastal, equatorial and sub-tropical ocean circulations and also the time dependent response of the ocean to changing wind forcing.

An intensive coordinated study of the Indian Ocean has taken place every decade over the last 40 years. The first effort to investigate the Indian Ocean circulation was made during the International Indian Ocean Expedition (IIOE) during 1964 - 66. It consisted of a basin-wide survey that subsequently resulted in a comprehensive hydrographic atlas (Wyrtki, 1971) and of a number of regional studies including the first study of the monsoon circulation of the Somali Current (Swallow & Bruce, 1966). The next intensive investigation was the Indian Ocean Expedition (INDEX) during the First GARP Global Experiment (FGGE), which investigated the summer monsoon response of the Somali Current (Swallow, Molinari, Bruce, Brown, & Evans, 1983). The next decade saw some less intensive but more regional investigations of the Indian Ocean circulation.

In parallel to the observation expeditions over the Indian Ocean, number of theoretical and modelling studies have also been carried out, in an effort to explain the observed features of the Indian Ocean circulation (Cox, 1970; Shankar et al., 2002; Schott & McCreary., 2001; McCreary et al., 1993; Haugen, Johannessen, & Evensen, 2002; Murtugudde & Busalacchi, 1999). Beginning with Lighthill’s (Lighthill, 1969) work on the influence of the equatorial wave, various studies have investigated the role of different wind forcing patterns, topography and currents and inertial waves from outside the Indian Ocean basin on the circulation of the Indian Ocean.

The World Ocean Circulation Experiment (WOCE) which was carried out to establish the role of the oceans in the earth’s climate and to obtain a baseline dataset against which all future change can be assessed saw a fresh increase in the research activities involving the Indian Ocean (Reppin, Schott, Fischer, & Quadfasel, 1999). Sophisticated numerical ocean models were developed to provide the framework for the interpretation of the observations and for the prediction of the future ocean state.

The objective of the study presented in this chapter is to review the current state of knowledge on the seasonal cycle of sea surface temperature, sea surface salinity...
and the surface currents in the Indian Ocean and to compare the main features of the circulation with the model solution obtained from ROMS simulations for this region.

We proceed now to introduce the model and the forcing data used and present the comparisons of the mean and the seasonal circulation of the model runs with climatological observations; the relation between the wind forcing and the surface currents and a dynamical framework of the Indian Ocean current system to summarise the results.

2.2 Model, Data and Experimental Setup

We use a free surface primitive equation ocean model called ROMS (Regional Ocean Modeling System), being used by a broad community for applications ranging from basin scale to the coastal and estuarine scales (Song & Haidvogel, 1994a; Haidvogel, Blanton, Kindle, & Lynch, 2000a; Marchesiello, McWilliams, & Shchepetkin, 2003; Chassignet, Arango, Dietrich, Ezer, Ghil, Haidvogel, Ma, Mehra, Paiva, & Sirkes, 2000; Haidvogel, Arango, Hedstrom, Beckmann, Malanotte-Rizzoli, & Shchepetkin, 2000b; Ezer & Mellor, 1994a). (Shchepetkin & McWilliams, 2005, 2003, 1998; Kowalik & Murty, 1993) describe in detail the algorithms and the computational kernel of ROMS model. (Song & Haidvogel, 1994a; Beckmann & Haidvogel, 1993) describe the mathematical and physical equations of the ROMS model also described in Appendix A. ROMS includes a careful formulation of the time-stepping algorithm to allow for both exact conservation and constancy preservation for tracers, while achieving enhanced stability and accuracy in its solutions. Conservative parabolic-spline discretization in the vertical reduces the pressure-gradient truncation error that has previously plagued terrain following coordinate models such as POM (Blumberg & Mellor, 1987) and SCRUM (Haney, 1991). The equations of motion solved in the ROMS model are illustrated in Appendix A along with the boundary condition equations.

The model uses a generalised sigma coordinate system in the vertical (Phillips, 1957) and a curvilinear horizontal grid (Arakawa & Lamb, 1977) (1/2 degree longitude resolution) that extends from 30°S to 30°N and 30°E to 120°E. The model bathymetry is obtained by a smooth interpolation of ETOPO5 analysis (Figure 2.1). The vertical grid has 20 levels with enhanced resolution in the surface and bottom boundary layer (Figure 2.2).

A modified radiation condition Marchesiello, McWillaims, and Shchepetkin (2001), which allows for a stable, long-term integration of the model is used at the southern
Figure 2.1: Model Domain with the bathymetry values in meters in the Indian Ocean

Figure 2.2: Model vertical grid at 90° longitude in the Indian Ocean
and eastern open boundaries, together with a nudging term for relaxation to observed climatologies of temperature and salinity. At the southern and eastern boundaries the volume flux is conserved and free surface gradient condition is applied. A more complete description of the model numerics, open boundary conditions and mixed layer parameterizations can be found in (Shchepetkin & McWilliams, 2005) and (Large, McWilliams, & Doney, 1994). For the surface boundary layer mixing parameterization the model implements the Large/McWilliams/Doney Oceanic Planetary Boundary Layer scheme. In the horizontal, the model uses a biharmonic horizontal mixing scheme (Smagorinsky, 1963). It also solves for the nonlinear equation of state to derive the density.

For the model initial condition and open boundary condition, we use World Ocean Atlas 2001 (WOA01) dataset for Sea Surface Temperature and Sea Surface Salinity monthly climatologies. At the surface, the model is forced with monthly climatologies for heat and freshwater flux derived from two different datasets to test the forcing solutions, Comprehensive Ocean Atmosphere Data Sets (COADS (Slutz et al., 1985)) and SOC (Southampton Oceanography Center (Josey et al., 1999)) heat and salinity fluxes and ERA (European Center for Medium-range Weather Forecast (ECMWF) ReAnalysis) winds. The model surface temperature and salinity fields were relaxed (Killworth, Smeed, & Nurser, 2000) to WOA01 climatological values to correct for the incorrect flux forcing files or the errors developed from the mixing schemes used.

The model was initially driven by the monthly mean climatology of wind stresses, freshwater flux and heat flux of the COADS dataset, (Slutz et al., 1985), determined from ship observations using bulk formulae. Subsequently, the model was forced using a combination of winds, stresses and air-sea fluxes from ERA (Källberg et al., 2005), SOC (Josey et al., 1999) surface flux climatology. The results of the various simulations and their intercomparison have been shown in this chapter.

### 2.2.1 Model Experiments

At the surface, different monthly surface flux climatologies were used to force the model and the model solutions were compared with observed data to study the model performance as compared to observed Indian Ocean circulation features and also to arrive at a good forcing dataset for the model simulations. The two experiment runs done were as follows:

**COADS case:** Monthly climatologies of wind stress and heat flux across the ocean
surface at a 0.5° x 0.5° resolution have been downloaded from the Atlas of Surface Marine Data 1994 (IRI/LDEO Climate Data Library) (http://ingrid.ldgo.columbia.edu; http://iridl.ldeo.columbia.edu/SOURCES/.DASILVA/) (daSilva, Young, & Levitus, 1994), and interpolated to the model grid. The model experiment done by the use of the COADS forcing files is named as COADS Run(CO). The model is integrated for 5 years model time and the fifth year monthly average fields are plotted to compare with the observed climatology fields.

**SOC and ERA case:** Monthly climatologies of heat flux across the ocean surface at a 1° x 1° resolution, developed at the National Oceanography Center (formerly Southampton Oceanography Center) (Josey et al., 1999), were downloaded and have been interpolated onto the model grid. ECMWF (European Center for Medium-Range Weather Forecast) 40 Re-Analysis 10 m wind components were downloaded and the monthly climatologies of the surface wind stress were computed (Källberg et al., 2005). The model experiment done by the use of the SOC and ERA forcing files is named as SOC-ERA Run(SE). The model is integrated for 5 years model time and the fifth year monthly fields are plotted to compare with observed climatology fields.

### 2.3 Surface Momentum Forcing

In this section, we first give an overview of the climatological wind stresses that force the Indian Ocean, their seasonal cycle and interannual variability (Josey et al., 1999). Then we review the large-scale circulation patterns of the tropical Indian Ocean. We also review the regional processes in the subsequent sections.

#### 2.3.1 Climatological Winds

Indian Ocean circulation is mainly a response to the seasonally changing winds that force it (Haines, Fine, Luther, & Ji, 1999; Lee et al., 2000). The wind climatologies over the Indian Ocean are available from different sources and two different forcing fields have been used to study the performance of the model used in this study.

The trade winds in the southern hemisphere of the Indian Ocean persist throughout the year. They have their seasonal maximum and most northerly extent during the southern hemisphere winter. In the northern Indian Ocean, during winter, the winds blow away from the land (Asian continent), causing north easterly wind stresses over the Arabian Sea and the Bay of Bengal, where as during summer monsoon, winds are south westerly over both the basins (Figure 2.3). The summer monsoon in the
Figure 2.3: Bimonthly Wind Stress fields ($Nm^{-2}$) (shaded) along with the wind vectors depicting the seasonal reversal (derived from SOC Climatology)
northern Indian Ocean witnesses a continuation of the trade winds from the southern Indian Ocean into the Arabian Sea and a narrow atmospheric jet called the Findlater jet (Findlater, 1971).

The wind stress field changes in correspondence to the seasonal reversal of wind-stress with large reversals between the seasons in the northern hemisphere (Figure 2.3). During the summer monsoon, an anticyclonic windstress curl resides over most part of the Arabian Sea and equatorial Indian Ocean. Its maximum lies to the right of the Findlater jet axis. It reverses to become cyclonic during the winter monsoon. Over the Bay of Bengal, the curl is cyclonic in summer and anticyclonic in winter.

During April to May, weak along shore winds occur off Somalia. The subsequent onset of the monsoon over the Arabian Sea causes an abrupt change from these weak pre-monsoon winds into full developed Southwest Monsoon winds in early to mid June. The Southwest monsoon becomes most strongly developed in late July.

The equatorial Indian Ocean is subjected to a unique wind forcing pattern unlike that over the other two equatorial oceans. A semi-annual eastward wind over the equator during April-May and October-November generates an annual mean equatorial zonal wind stress that is eastward.

In the section (2.4), a description of the Indian Ocean Circulation system is illustrated along with a description of the model solutions for the two experiments conducted.

## 2.4 The Indian Ocean Circulation

In order to assess the quality of the model integration subjected to the two different forcing experiments and to determine which solution gives a better understanding of the physical processes in the Indian Ocean on the seasonal and mean circulation scale, we need to compare it with observations. Our goal is to check whether the model captures the leading order dynamics of the observed system when driven by adequately realistic atmospheric forcing although the surface tracers are relaxed to climatological values.

The surface circulations of the oceans are primarily in response to the wind patterns which force them throughout the year (Hastenrath & Greischar, 1991; Ganachaud, Wunsch, Marotzke, & Toole, 2000). The wind driven circulation in the oceans are restricted to the upper few hundred meters and is mainly horizontal. In order to describe
the circulation, the Indian Ocean can be mainly divided regionally into the Arabian Sea basin, the Bay of Bengal basin and the southern Indian Ocean basin. The seasonal surface currents from the CO run solution ($v_{CO}$) and the SE run solution ($v_{SE}$) are plotted (Figure 2.6, Figure 2.7), respectively and the seasonal current cycles are compared with the observed currents of the Indian Ocean (Figure 2.4 and Figure 2.5).

Both solutions are able to reproduce the major seasonal currents of the Indian Ocean except that their magnitude and certain fine features such as eddies are not resolved accurately in both model experiment solutions. However, the $v_{SE}$ seems to be closer in magnitude to the observed values of the major currents in this region. A discussion on the various main features observed in the Indian Ocean circulation and a comparison of the model experiment solutions to these observed features is presented in the following sections.

2.4.1 Indian Monsoon Currents

The monsoonal wind system over the North Indian Ocean causes a seasonal reversal of the surface currents. The Somali current reverses annually (Wyrtki, 1973) and the strong equatorial surface jets reverse semi-annually.

![Figure 2.4: A Schematic representation of the surface currents in the Indian Ocean during the Summer Monsoon from Shanker et al. (2002)](image)

Available observation data evidence that the western boundary circulation south of the monsoon regime ($10^\circ$ S) do not witness much of the seasonal variability (Swallow, Fieux, & Schott, 1988). The main western boundary current in the Indian Ocean is
the Somali Current which has a strong seasonal variation. It flows north eastward during the southwest monsoon and south eastward during the North East Monsoon. Both the CO and the SE model solutions capture the strength and spatial extent of the Somali current well compared to the observed values (fig 2.4, fig 2.5) as observed in the central right and bottom left panel of the figures (fig 2.6, fig 2.7) for the months of July, September and also the reversal in its direction is captured well by both simulations for the month of November in the bottom right panels of (fig 2.6, fig 2.7).

During the summer monsoon, the Southern Equatorial Current (SEC)(Fig 2.4) and East African coastal current(EACC)(Fig 2.4) supply the northward flowing Somali current. The Somali current, after crossing the equator, turns offshore at about 4° N forming a cold upwelling region to its left; the other part recirculates back across the equator to form the “Southern Gyre”(SG) (Figure 2.4). In the north, a second gyre is formed, the “Great Whirl”(GW). A third gyre termed as the Socotra Eddy(SE) is noticed to the north of the GW during many summer monsoons (Schott & McCreary., 2001).

![Figure 2.5: A Schematic representation of the surface currents in the Indian Ocean during the Winter Monsoon from Shanker et al. (2002)](image)

An important link between the western and eastern parts of the North Indian Ocean is the zonal Monsoon Current (Shankar et al., 2002). It flows westward as the Northeast Monsoon Current (NMC) during the Northeast Monsoon and eastward as the Southwest Monsoon Current (SMC) during the Southwest Monsoon. The Southwest monsoon current (SMC) south of Sri Lanka flows eastward, curves around Sri
Lanka and into the Bay during the season. Most of the SMC connection to the Somali current outflow appears to be at lower latitudes, but part of it is supplied by the southward flowing West Indian Coastal Current (WICC) (Vinayachandran, Masumoto, Mikawa, & Yamagata, 1999). The Laccadive Low (LL), is a persistent feature of the summer monsoon in the northern Indian Ocean. LL is the location of a persistent low sea surface height anomaly, associated with anticlockwise (cyclonic) circulation off the coast of southwest India, near the Lakshwadeep islands. The East Indian Coastal Current (EICC) bifurcates in the Bay of Bengal. During the summer monsoon, the Ekman flow is southward across the equator throughout the Indian Ocean.

During the winter monsoon, the EACC meets the southward flowing, near surface Somali current and hence, both supply to an eastward-flowing South Equatorial Counter Current (SECC) (Figure 2.5). At the eastern end of SECC, a south eastward current flows as a boundary current called the South Java current. During the northeast monsoon, multiple (2-3) eddies of 200 km in diameter form off the southwest coast of India and propagate westward at 17 cm/s (Bruce & Kindle, 1994). A low latitude supply from the westward flowing Northeast Monsoon Current (NMC) south of Sri Lanka also supplies WICC after circulating around the Laccadive High.

The NMC carries low saline waters from the Bay of Bengal into the Arabian Sea. Its strength and spatial scale determines the amount of freshwater influx into the Arabian Sea, and concomitantly changes the sea surface salinity balance in the northern Indian Ocean. Solution $v_{CO}$ (Figure 2.6) shows a weak NMC compared to both $v_{SE}$ (Figure 2.7) and observed values of the NMC (Figure 2.4, Figure 2.5).

The SMC carries high saline water from the Arabian Sea into the Bay of Bengal (Vinayachandran et al., 1999). The northward bending of the SMC is forced by the westward propagating Rossby waves generated by the reflection of the spring Wyrtki Jets from the eastern boundary and by local positive curl of the wind stress (McCreary et al., 1993), (Siedler, Church, & Gould, 2001). The CO solution SMC is weaker than the SE solution SMC and also weaker than the observed values (fig 2.4, fig 2.5) as observed from the figures (fig 2.6, fig 2.7, fig 2.4, fig 2.5).

There is a broad zonal flow by the Southern Equatorial Current (SEC), driven by the southeast trade winds, which flows ahead as the western boundary current east of Madagascar within latitude range of $12 - 25^\circ$ S. This is the only major current in the Indian Ocean which does not reverse direction seasonally. At about $17^\circ$ S, the SEC splits into a northward and a southward flowing current. The northern branch feeds into the East African coastal current (EACC). The southern hemisphere circulation largely follows Sverdrup dynamics, where the mass transport is related to the curl of the wind...
Figure 2.6: Bimonthly Sea Surface current vectors $v'$ from the model run forced by COADS dataset over the Indian Ocean.
Figure 2.7: Bimonthly Sea Surface current vectors $v'$ from the model run forced by SOC-ERA dataset over the Indian Ocean.
stress within the upper ocean (Shankar et al., 2002; Schott & McCreary., 2001).

2.4.2 Arabian Sea and Bay of Bengal Circulation

The seasonal variations of the Arabian Sea is very significant and results mainly due to the seasonal variations in the monsoon winds which drive the surface circulation. Mesoscale eddies and gyres with associated temperature gradients and current shears ensues the onset of the South West monsoon (SW monsoon) in the Arabian Sea (Bruce, 1973; Luther & O’Brien, 1985; Schott & McCreary, 2001). These eddies are important components of the short-term variability in this region. The central Arabian sea exhibits a marked bowl shaped mixed-layer deepening. The wind stress forcing from the Findlater jet and Ekman pumping leads to the warming in the deeper layers pushing the mixed layer deeper during the summer monsoon months. Since the maximum of the wind stress curl of the Findlater jet is to the right of the jet close to the center of the Arabian sea, the bowl shaped mixed layer is an expected feature (Schott & McCreary, 2001; McCreary et al., 1993; Qian, Hu, & Zhu, 2003; Gordon, 1986).

The West Indian Coastal Current (WICC) in the Arabian Sea changes its direction of flow like the other prominent currents in the Indian Ocean: its poleward during the North East Monsoon and equatorward during the SW monsoon (Lee et al., 2000).

In the Bay of Bengal, the EICC reverses direction twice a year. It flows north eastward from February to September, with a strong peak in March-April and south westward from October to January with the strongest flow in November (Potemra et al., 1991; Vinayachandran et al., 2002; Vinayachandran, 1995).

From February to May, the EICC forms the western boundary current of a basin wide anticyclonic gyre. This gyre disappears during summer monsoon, when the boundary current splits at the 10° N confluence. During winter, there seems to be a cyclonic gyre in the Bay which is coupled to the southward flow of EICC then.

The Red Sea and the Persian Gulf to the northwest of the Arabian Sea also contribute a significant amount of water mass into the Indian Ocean and are identifiable over long distances due to their prominent high salinity at core densities.

The surface circulation in the Bay of Bengal and Arabian Sea show similar structure in the model runs CO and SE and also in the observed features. Both the model runs show anticyclonic circulation in spring months and cyclonic circulation in the monsoon months in the Bay of Bengal. But the anticyclonic circulation in the Arabian Sea during the winter months is captured much better in the SE solution than the CO solution(Figures 2.6, 2.7, 2.4, 2.5).
Similarly the East Indian Coastal Current (EICC) solution in SE run agrees better with observation than the EICC solution in the CO run. Also, the West Indian Coastal Current (WICC) responsible for carrying high saline waters from the Arabian Sea into the Bay of Bengal is solved better in magnitude and spatial extent in the SE run than the CO run.

### 2.4.3 Equatorial Surface Jets

Surface currents in the equatorial Indian Ocean reverse direction four times in a year, flowing eastward during spring and fall, and westward during winter and summer. The eastward currents were first described by Wyrtki (Wyrtki, 1973), and they are now commonly referred to as Wyrtki Jets, (WJ) (Han, McCreary, Anderson, & Mariano, 1999). The eastward flowing WJs bring the saltier water from the Western Indian Ocean into the eastern basin. The SE solution and CO solution agree well with respect to the strength and spatial extent of the WJ during the months of May-July when the WJs are known to have the strength of about 1 m/s.

A unique phenomenon in the Indian Ocean is the presence of the semi-annual eastward winds along the equator which produces a strong eastward surface jet (Wyrtki, 1973) during the transition seasons between the monsoons, i.e., April to June and October to December. This strong jet carries warm surface water eastward and increases the sea level and mixed layer in the east and concomittantly decreases them in the west.

Another unique phenomenon of the Indian Ocean is the flow of the warm equatorial Pacific Ocean water into the Indian Ocean via the Indonesian throughflow. The pathway of the water mass of the Indonesian throughflow in the Indian Ocean is still not known and is a matter of debate among modelers even to this day.

### 2.4.4 Sea Surface Temperature

Sea Surface Temperature (SST) is one of the most important variables of the ocean state because of its significant role in the exchange of heat, moisture and gases across the air-sea interface. SST is indicative of the ocean surface processes such as upwelling, eddies, fronts and current boundaries (Levitus, 1984). In this section we compare the SST simulated by the model runs CO and SE with a valuable SST climatology dataset of World Ocean Atlas 01 (Stephens, Antonov, Boyer, Conkright, Locarnini, O’Brien, & Garcia, 1998). Bimonthly maps of SST for model simulations and observed data are shown in figures (fig 2.8, fig 2.9, fig 2.10).
The surface temperatures in the Tropical Indian Ocean ranges from less than $20^\circ C$ in the southern parts during July-October to more than $30^\circ C$ in the northern parts during April-May (Figure 2.8, 2.9, 2.10). The surface temperature in the Persian Gulf and Red Sea increase above $32^\circ C$ during the summer monsoon.

In the region north of $10^\circ S$, the SST is significantly influenced by the prevailing monsoon winds and associated currents. The east-west gradient during January-February in the region $10^\circ N - 10^\circ S$ is high, where the SST increases from $28^\circ C$ in the west to $29^\circ C$ in the eastern part. The thermal equator extends from $10^\circ S$ in the western Indian Ocean to the equator near $90^\circ E$ in February (fig 2.8). During April-May, it moves towards $10^\circ N$. It then descends south of the equator in the month of July (fig 2.9). The prominent feature of the Pacific and Atlantic Ocean of an equatorial SST minimum is absent in the Indian Ocean due to the absence of any upwelling along the equator (McCreary et al., 1993).

During January-February, the SST decreases northward in the Arabian Sea and Bay of Bengal (from about $28^\circ C$ at $10^\circ N$ to about $24^\circ C$ at the northern part of Arabian Sea and Bay of Bengal)(fig 2.8). The ocean surface warms up gradually in the northern Indian Ocean as the solar insolation beings to increase in the summer of the northern hemisphere(March-April). A zonal band of high SST($>30^\circ C$) is present in the south eastern Arabian Sea including the Lakshwadeep Sea during April. By the month of May, the SST exceeds $30^\circ C$ in most parts from the equator to $15^\circ N$. The onset of summer monsoon in June results in the falling of SST which continues until August due to the strong monsoon winds and the advection of upwelled water off the Somali and Arabian coast.

Though the fall in SST is only marginal ($1^\circ C$), it is significant in the Arabian Sea and equatorial part with distinct low SST off the Somali coast ($<25^\circ C$). Once the southwest monsoon starts retreating in September, both the Arabian Sea and the Bay of Bengal basin begin to warm up until October(fig 2.10). The SST increases to about $28^\circ C$ in the Arabian Sea and to about $29^\circ C$ in the Bay of Bengal. Once the northwest/winter monsoon onsets in November, the SST in both basins begins to decrease until January.

The SST is found to vary from more than $28.5^\circ C$ in the southern and eastern Arabian Sea to less than $18^\circ C$ off the coast of Somali. A unique feature of the Arabian Sea is that the SST in July (post monsoon onset) is $2^\circ C$ lower than that in April and in July, the SST is mainly due to the strong air-sea interactions and the uniquely strong winds during the southwest monsoon (Halpern et al., 1998).
The model integrations of the forcing with COADS dataset (CO) and SOC-ERA (SE) datasets are compared with observed fields. The SST fields generated in the CO run and SE run are compared with the climatological fields of the WOA2001 (World Ocean Atlas 2001) dataset. The bimonthly plots are shown in figures (2.8, 2.9, 2.10).

The bimonthly SST from solution COADS run $[T_{CO}]$ and from SOC-ERA run $[T_{SE}]$ agree well with each other and with observed values as shown. The seasonal high and low SST and also upwelling regions where SST cools lower than the adjacent regions is solved well in both the model integrations (Figures 2.8, 2.9, 2.10).

Some of the finer features such as the upwelling off the Somali Coast during July-September summer monsoon and also the upwelling off the West Indian Coast are solved well in the SE integration. The strong equatorial heating is captured well in the SE solution. This difference between SE and CO solutions in these regions is about 0.5°C which may be due to the smoothing of the COADS data because of which it cannot resolve the narrow features.

### 2.4.5 Sea Surface Salinity

Salinity and temperature are the two variables which determine the density of the water and thereby influence the ocean dynamics primarily. In this section we compare the SSS (Sea Surface Salinity) simulated by the model runs CO and SE with a valuable SSS climatology dataset of World Ocean Atlas 01 (Stephens et al., 1998). Bimonthly maps of SSS for model simulations and observed data are shown in figures (fig 2.11, fig 2.12, fig 2.13). In the open ocean, changes in the evaporation and precipitation mainly affect the Sea Surface Salinity (SSS).

The average surface salinity is mainly distributed zonally and is minimum along the equator and is maximum in the subtropics at 25° N and 25° S. In case of the tropical Indian Ocean, the surface salinity is maximum in the Red Sea and the Persian Gulf due to intense vaporization and minimum in the Bay of Bengal due to high entrainment of freshwater from river discharges. Bimonthly climatological plots of the Sea Surface Salinity distribution over the Indian Ocean are shown in the (Figure 2.11, Figure 2.12, Figure 2.13)

During January to June, the SSS is more than 36 psu in the Arabian Sea, north of about 10° N and west of 70° E and decreases towards south east. The SSS in the Bay of Bengal increases from about 30 psu in the head of the Bay where there is a large discharge of fresh water from the Ganges and Brahmaputra and Irrawady, to about 34 psu towards south upto 5° N. The SSS distribution in January clearly shows the
spreading of low salinity waters (< 34 psu) into the south eastern Arabian Sea under the influence of northeast monsoon. During July to December, the SSS of more than 36 psu in the Arabian Sea (between 50°E and 60°E) extends towards south up to the equator. The SSS in the Bay increases strongly from 28 psu near the head of the Bay to about 34 psu at about 10° N.

In the equatorial region, the SSS decreases from more than about 35 psu in the western part to less than 34 psu in the east reflecting the influence of the low salinity waters of the Indonesian throughflow. South of 10° S, the SSS distribution is mainly zonal and increases towards south reaching a maximum (> 35 psu) around 30° S reflecting the least difference in the precipitation and evaporation quantities in this region.

The sea surface salinity (SSS) fields solved by the model integration using COADS dataset \([S_{CO}]\) and SOC-ERA \([S_{ES}]\) datasets as forcing fields have been plotted in figures (2.11, 2.12, 2.13) and compared with the sea surface salinity fields of the WOA01 climatology dataset.

The bimonthly SSS from COADS forcing run \([S_{CO}]\) and SOC-ERA forcing \([S_{SE}]\) agree well with each other with respect to the seasonal cycle of the SSS fields observed in the WOA01 dataset.

The precipitation-evaporation difference generally forces low salinity in the Bay of Bengal throughout the year, and along the west coast of India during the South West Monsoon. These features are captured well in both CO and SE solutions. But the extent and magnitude of the fields are captured much better in the SE solution than the CO solution. The magnitude of \(S_{CO}\) is generally 0.5 to 1 psu lesser than \(S_{SE}\) in the Arabian Sea where \(S_{SE}\) compares well with the observed field. Also the finer features of low saline waters in the northern Bay of Bengal are well featured in the SE solution rather than the CO solution (Figure 2.11, Figure 2.12, Figure 2.13).

### 2.5 Conclusions

This chapter describes the results from the ROMS terrain-following North Indian Ocean basin model. The primary conclusion is that ROMS is able to reproduce the known features of the surface and subsurface circulation as compared to climatology. Though the surface tracers were relaxed to climatological values to make up for the errors in the mixing and frictional parameters and also uncertainty in forcing surface flux data, the model solutions compared well with observed features of the Indian
Figure 2.8: January-March Surface Temperature fields from the model solution forced by COADS, SOC-ERA dataset compared with WOA01 Climatology field over the Indian Ocean.
Figure 2.9: May-July Surface Temperature fields from the model solution forced by COADS, SOC-ERA dataset compared with WOA01 Climatology field over the Indian Ocean.
Figure 2.10: Sep-Nov Surface Temperature fields from the model solution forced by COADS, SOC-ERA dataset compared with WOA01 Climatology field over the Indian Ocean.
Figure 2.11: January-March Surface Salinity fields from the model solution forced by COADS, SOC-ERA dataset and a WOA01 Climatology field over the Indian Ocean
Figure 2.12: May-July Surface Salinity fields from the model solution forced by COADS, SOC-ERA dataset and a WOA01 Climatology field over the Indian Ocean.
Figure 2.13: September-November Surface Salinity fields from the model solution forced by COADS, SOC-ERA dataset and a WOA01 Climatology field over the Indian Ocean
Ocean circulation and tracer fields.

Also, the model was forced by two different surface flux datasets to compare the respective model solutions obtained and arrive at a good forcing dataset to prepare the model initial conditions for the data assimilation experiment run. The two different datasets were COADS and SOC-ERA surface momentum and heat flux datasets, which are described in detail in Subsection (2.2.1). The model solution comparisons of these two separate runs showed that the model forced by the COADS fluxes consistently over-estimates the SST, SSS fields which in turn affect the circulation features negatively. The SOC-ERA flux forced model performs better in comparison in reconstructing the SST, SSS and surface currents.
Abstract

This chapter builds the background for this thesis and discusses the development of assimilation methodologies in physical oceanography. It presents the mathematical recipe of the assimilation method adapted in the present study and also illustrates the implementation of assimilation of observed data over the oceans in a general circulation ocean model (ROMS).

3.1 Introduction

“Assimilate” is a word that conjures up diverse meanings ranging from physiological metabolism to absorption of knowledge into the mind. Its usage in all contexts is in the sense of absorption of knowledge/nutrients external to the system into the present state of the system.

In recent years, interest in data assimilation into ocean models has increased dramatically after the increase in available observational data of oceanic fields and the concomitant development of more accurate ocean general circulation models. In this new age sophisticated theories about ocean circulation developed up till now from very little field information, will be tested against the incoming data, new theories will be developed to explain the varied phenomena transpiring from the more plentiful data and increased understanding of ocean circulation will be translated into prediction.

The concept of “data assimilation” was developed in meteorology about 50 years ago as the methodology in which observations are used to improve the forecasting skill of operational meteorological models. In operational meteorology, all the observational data available at a given time are “assimilated” into numerical prediction models by melding them with the model-predicted values of the same variables in order to prepare initial conditions for the forecast model run.
In the ocean state estimation context, the term “data assimilation” has acquired a much broader meaning. Data assimilation in ocean state estimation pertains to various different methodologies, originating not only from meteorology but also from solid earth geophysics inverse theories and in engineering control theories. All these methods attempt to constrain a dynamical model with the available data.

Oceanographic data assimilation has three main objectives. One goal is to quantitatively use the data in order to improve the ocean model parameterizations of sub-grid scale processes, boundary conditions, etc. A second goal is to obtain a four-dimensional realization of the oceanic flow that is simultaneously consistent with the observational evidence and with the dynamical equations of motion. This realization can be used for detailed process studies of ocean dynamics of a region. A third major motivation of ocean data assimilation, the closest to the meteorological one, is to provide initial conditions for predictions of the oceanic circulation.

In this chapter, we intend to illustrate the data assimilation scheme used in the study of this thesis along with certain illustrations of the landmarks in the development of ocean data assimilation. The data assimilation scheme used in this study is termed as Optimal Interpolation. Our main focus here is the mathematical framework based on which the optimal interpolation methodology was developed and the objective of using data assimilation in an ocean general circulation model.

Two major differences still prevent the simple “borrowing” of techniques from meteorology. The first is the motivation for oceanic data assimilation which is not as narrowly focused towards short term prediction as are most meteorological efforts. The second reason resides in the major differences between the meteorological and oceanographic data sets. The available oceanographic data are sparse in the interior of the ocean where as the surface measurements of oceans are much more spatially dense. This is not the same in the case of meteorological observations. This implies that the assimilation methodologies, far from being blindly applied to oceanic dynamical problems, must be revisited and sometimes profoundly modified to make them feasible and successful for physical oceanography.

### 3.2 Historical Perspective

The first data assimilation methods were called the “objective analysis” (Cressman, 1959). The idea of applying objective analysis was first proposed by Kibell. In
a lecture in 1949, Kibel pointed out that calculations of barometric and thermal tendencies according to the forecasting scheme developed by him at that time could be carried out quietly conveniently, provided the pressure and temperature fields were first represented by polynomials. In subsequent works, Kibel made use of formulae which described a plane field by means of a quadratic and third order polynomials, on the basis of the method of least squares and using data from points on a square grid.

At the end of 1949, Panofsky’s article (Panofsky, 1949) appeared, in which the term “objective analysis” was apparently used for the first time. He used a technique of fit by least squares in two dimensions. This technique consists basically in expanding the fields (variables), which are to be analyzed, in a series of polynomials about the observation point, minimizing the square of their differences with the observed values. The expansion coefficients are then determined by inverting a matrix.

The studies described above were carried out before numerical prediction by means of electronic computers began to be used in meteorology. The introduction of the latter provided a powerful stimulus to the development of objective analysis methods. Consequently, the main attention began to be focused just on obtaining values of the analyzed elements at regularly spaced grid points (values to be used afterwards as initial data for the numerical prediction) rather than on a construction of the whole field.

This was followed by the development of two fundamentally different methods of objective analysis for the purpose of numerical weather prediction. One was developed by American investigators Gilchrist and Cressman (Cressman & Gilchrist, 1954) and the other by Bergthorsson and Döös (Bergthorsson & Döös, 1955). Gilchrist and Cressman developed a polynomial interpolation scheme termed as “successive correction scheme”. It achieves its results by forcing convergence of data to observed, interpolated values using multiple iterations.

Swedish investigators Bergthorsson and Doos developed a completely different method of objective analysis. The information used in this method includes not only observational data but also the results of a numerical prediction for the given time and the average climatological values of the analyzed elements.

The breakdown in the field of data assimilation was achieved by L. S. Gandin (Gandin, 1965) who introduced the ”statistical interpolation” (or ”optimal interpolation”) method. This method is a 3D data assimilation scheme and is a kind of ”regression analysis”, which utilizes the information about the spatial distributions of covariance functions of the errors of the ”first guess” field (previous forecast) and ”true field”. These functions are never known. However, different approximations were
assumed.

The optimal interpolation algorithm is the reduced version of the Kalman filtering algorithm, when the covariance matrices are not calculated from the dynamical equations, but are pre-determined in advance.

When this was recognised, the attempts to introduce the Kalman Filtering algorithms as a 4D data assimilation tool for Numerical Weather Prediction models were tried. However, this was (and remains) a very difficult task, since the full version of Kalman Filtering algorithm requires solution of the enormous number of additional equations. In connection with that, the special kind of Kalman Filtering algorithms (sub optimal Kalman Filtering) for Numerical Weather Prediction models were developed (Dee, 1991).

Over the past 25 years or so, since the initial efforts to develop three dimensional ocean circulation models (Bryan, 1969), ocean modeling has made a very significant progress. In parallel, oceanic observational techniques have been thoroughly revolutionized. However, the lack of a single focusing motivation of oceanic data assimilation such as provided by the need for Numerical Weather Prediction (NWP) in meteorology, caused ocean models and observational techniques to develop quiet independently from each other. When oceanic models and observations started converging, it happened in different paths, depending on the specific objectives of each effort (Emery & Thomson, 1998).

The early days of oceanography saw dynamic calculations as the main quantitative tool to combine data (temperature and salinity) with the then present ocean models. From this modest beginning, relying on highly simplified models and on no formal assimilation procedure, the next step was to introduce a formal least square inverse methodology imported from solid earth geophysics and add the tracer conservation constraints in order to solve the problem of level of no motion (Wunsch, 1977), (Wunsch & Grant, 1982). This was done in the framework of coarse resolution box models whose dynamics was still very simple although the inverse methodology used was very general. Much of the work done at present on the combination of OGCMs and data stems from the experience obtained in the pioneering work on oceanographic box inverse models.

As the complexities of the models grew, equally more sophisticated assimilation methods needed to be developed. Efforts towards this began with the diagnostic models in which temperature and salinity data were simply inserted into the dynamical equations of fairly complex ocean models in order to evaluate the velocity field (Hol-
land & Hirschman, 1972). The results were very poor due to model-data-topography inconsistencies, and at the next stage, a very simple assimilation methodology was introduced into OGCMs and became known in the oceanographic context as the “robust diagnostic” approach (Sarmiento & Bryan, 1982). The same approach had actually been introduced earlier in meteorology as the “nudging” technique (Anthes, 1974) and the term “nudging” has by now become commonly used in oceanography as well. In this approach, there is no effort to introduce least-square optimality, and the data are just used to nudge the model solution towards the observations at each time step through a relaxation term added to the model equations. The result is far superior to simple diagnostic models, but leaves much to be desired due to the inability to use the information about data uncertainty or to estimate the errors in the solution obtained (Holland & Malanotte-Rizzoli, 1989), (Malanotte-Rizzoli & Tziperman, 1996), (Malanotte-Rizzoli & Young, 1995), (Lermusiaux & Robinson, 1999), (Lu & Browning, 1998).

As the objectives of modeling and observational oceanography began to converge, more formal least square methods taken from meteorology were also used in ocean models, in particular the Optimal Interpolation (OI) method (Mellor & Ezer, 1991), (Derber & Rosati, 1989), (Ezer & Mellor, 1994b). OI may be viewed as a nudging technique in which the amount of nudging of the model solution towards observations depends on the data errors, while also allowing to make error estimates for the solution. This approach, developed in meteorology for NWP, is not capable of improving model parameters or parameterizations, nor is it capable of fitting the entire four dimensional distribution of observations simultaneously to the model solution. However, due to the relatively low computational cost of OI, it is appropriate for higher resolution, short term prediction and state estimation purposes.

Carrying the least squares approach for a time dependent model to its rigorous limit, leads to the "Kalman filter/smooother" assimilation methodology, which is capable of assimilating data into a time dependent model while assuring least-square optimality, full use of a priori error estimates, and calculation of the covariance error matrix for the model outputs. Apart from the fact that the Kalman filter is a formally optimal technique in the least-square sense only for linear models, its high computational cost limits its use at present to simple models, or very coarse OGCMs. Recent developments are directed at developing efficient, even though sub optimal, variants of the Kalman filter that allow the use of a full nonlinear OGCM with this method (Lewis, Lakshmivarahan, & Dhall, 2006).

The ultimate goal of combining a formal least-square optimization approach with
a full complexity OGCM requires the simultaneous solution of hundreds of thousands of coupled nonlinear equations (the model equations at all grid points and all time steps), and therefore requires an efficient approach which can be found in the "optimal control" engineering literature.

The development of assimilation methods in physical oceanography seemed to always trail behind meteorology by a few years. This lag is in spite of the fact that the ocean and atmosphere, even though characterized by some important differences, are at the same time similar enough that they can be treated with the same theoretical approaches and methodologies. It is important, therefore, for the ocean modeler to try and understand the reason for this difference in rate of development of data assimilation methodologies in order to be able to isolate potential obstacles for their future use in oceanography.

Clearly, a primary reason for the delayed development of oceanic data assimilation was the lack of urgent and obvious motivation such as the need of forecasting the weather and of producing better and longer forecasts as in meteorology. This situation has changed rapidly over the last few years with the role of oceans in the climate being understood better which entailed a better ocean state estimation. This in turn necessitated a systematic model improvement. This has become the main motivation to develop a robust data assimilation platform for ocean models. The need for ocean prediction is also arising now on various temporal and spatial scales, from climate change predictions, through regional forecasts of the large scale ocean climate variability, e.g. of the North Atlantic thermohaline circulation (Carton & Hackert, 1990), (Gordon, 1986) or El Nino in the Pacific Ocean, to a few weeks regional mesoscale ocean forecasts in frontal regions such as the Gulf Stream system that are required for example by various Naval applications.

The most profound limitation on the development of oceanic data assimilation may have been, however, the lack of adequate data sets. The number of available oceanographic observations is far smaller than the number of meteorological observations, especially when the different temporal and spatial scales are considered. It is estimated, in fact, that the number of presently available oceanographic observations is smaller than its meteorological counterpart by several orders of magnitude (Ghil, Ide, Bennett, Courtier, Kimoto, & Sato, 1997), (Ghil & Malanotte-Rizzoli, 1991).

New oceanographic data sets, nearly comparable to the meteorological one, i.e. synoptic and with global coverage, are however becoming available. This oceanographic observational revolution of the 90’s has been made possible by the advent of satellite oceanography. A second worldwide major source of oceanographic observa-
tions is the World Ocean Circulation Experiment and the ARGO project that, through basin wide hydrographic sections, meridional and zonal, should provide us with a picture of the large scale global circulation in the World Oceans in the 90’s and the first decade of the 21st century.

Thus, the increased availability of observation data and increased need for data assimilation has necessitated and expedited the process of developing advanced data assimilation techniques to reap the full benefits of the data sets and models combinedly (Bennett, 1992), (Anderson, Sheinbaum, & Haines, 1996), (Malanotte-Rizzoli & Tziperman, 1996), (Anderson & Willebrand, 1988), (Tziperman, Thacker, Long, & Hwang, 1992), (McIntosh, 1977).

3.2.1 Objectives of Oceanographic Data Assimilation

The three main objectives of combining data and complex OGCMs are: model improvement, study of the dynamical processes through state estimation, and, finally, ocean/ climate forecast. Each of these objectives have been met with relevant assimilation methodologies respectively. Even the highest resolution ocean circulation models cannot resolve all of the dynamically important physical processes in the ocean, from small scale turbulence to basin scale currents. There will always be processes that are not represented directly, but rather are parameterized. These parameterizations are sometimes simple, often complicated, and always uncertain both in form and in the value of their tunable parameters. The models are very sensitive to even small changes in these parameterizations. A few examples of such parameterizations are that of the small scale vertical mixing in the ocean interior (Bryan, 1987), the mesoscale eddy parameterizations in coarse ocean models used in climate studies, of mixed layer dynamics (Mellor & Yamada, 1982), and of deep water formation (Schott & Send, 1994). Another set of uncertain yet crucial parameters corresponds to the poorly known surface forcing by wind stress, heat fluxes and evaporation and precipitation, all of which are subject to typical uncertainties of 30-50% (Trenberth, 1989), (Trenberth & Solomon, 1993).

One of the most important goal of ocean data assimilation is to use available observations of the oceans systematically and quantitatively in order to test and improve the various uncertain parameterizations used in OGCMs. A good estimate of these parameters would allow the models to predict the ocean state more accurately in the absence of available data which is the ultimate goal of ocean models: To be able to forecast the ocean state in the absence of any available observational data during forecast runs.
Improvement of internal model parameters and boundary conditions via data assimilation can be complemented by the state estimation of the ocean via data assimilation (Killworth et al., 2001). In ocean state estimation via data assimilation, the model deficiencies are compensated for by using data to force the model nearer to observations during the model run (Woodgate & Killworth, 1997).

### 3.3 Objective Analysis

Objective analysis includes the development and realization of methods which make it possible to use the measurement of meteorological stations to reconstruct objectively, the fields of the meteorological elements, or at any rate to specify their values at the nodes of some kind of regular network. These values may then be used as initial data for a numerical prediction of the meteorological fields.

Another task of objective analysis is the matching of meteorological fields. This term refers to a processing of initial data or of already interpolated quantities in such a way that, to some given degree of accuracy, the relations showing the coupling between fields of different meteorological elements, and also between fields of the same element at different levels or different times, are satisfied. Examples of such relations are the geostrophic equations and the statistical equations.

Finally, the third task of objective analysis consists in the detection of erroneous data and the subsequent elimination, or if possible the correction, of these data. This refers, of course, only to very rough errors, and not to the minor errors which inevitably creep into all measurement data.

Here we shall discuss our statistical model and method used for the optimal estimation of oceanic fields from observational data. The method is based on the estimation theory known as objective analysis (Gandin, 1965) first described in an oceanographic context by Bretherton, Davis and Fandry (Bretherton et al., 1976) and is now an important tool in oceanography for both analysis and observational array design (Carter & Robinson, 1987).

The structure used for the correlation functions is of major importance to the method. When the data are scarce, analytical models are fit or assumed. Issues of importance include choice of statistical model, the choice of correlation function representation, the domain of influence of the data points, the number of data points in the estimation of an analysis point, computational efficiency and stability.
3.3.1 The Statistical Model

Theory

To derive an optimal estimate of a field, \( \theta_x \) from a linear combination of observations, \( \phi_r \). The field variable \( \theta_x \) is the value \( \theta(x, y) \) at a general point \( x = (x, y) \) and the observations, \( \phi_r \) are at a limited number of data points, \( x_r, (r = 1, 2, \ldots, N) \).

Procedure

Choosing \( \hat{\theta}_x \) as a linear combination of \( \phi_r \), we can express \( \hat{\theta}_x \) as:

\[
\hat{\theta}_x = A\phi_r
\]

We need to derive a matrix \( A \), such that the expected mean square error (difference between the estimated field and actual field \( x \)) is minimised.

\[
E(\epsilon\epsilon^T) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T]
\]

should be minimised.

By using (3.1) and (3.2), we have

\[
E(\epsilon\epsilon^T) = E[A\phi_r - \theta][A\phi_r - \theta]^T
\]

\[
= E[A\phi_r\phi_r^T - \theta\phi_r^T A^T - A\phi_r\theta^T + \theta\theta^T]
\]

If we express:

- \( C_\phi \) as the autocorrelation of the observations = \( E[\phi_r\phi_r^T] \)
- \( C_\theta \) as the autocorrelation of the field = \( E[\theta\theta^T] \)
- \( C_{\theta\phi} \) as the cross correlation of the field and observations = \( E[\theta\phi_r^T] \)
- \( C_\epsilon \) as the autocorrelation of the error fields = \( E(\epsilon\epsilon^T) \)

We have, from (3.3)

\[
C_\epsilon = AC_\phi A^T - C_{\theta\phi} A^T - AC_{\theta\phi}^T + C_\theta
\]

We have a matrix identity:

\[
(A - BC^{-1})C(A - BC^{-1})^T - BC^{-1}B^T = ACA^T - BA^T - (BA^T)^T
\]
From (3.4) and (3.5), we have:

\[ C_e = AC_\phi A^T - C_\theta C_\phi^{-1} C_\theta + C_\theta \]  

(3.6)

\[ = (A - C_\theta C_\phi^{-1}) C_\phi (A - C_\theta C_\phi^{-1})^T - C_\theta C_\phi^{-1} C_\phi^T + C_\theta \]  

(3.7)

Since \( C_\theta \) and \( C_\phi^{-1} \) are positive definite, \( (A - C_\theta C_\phi^{-1}) C_\phi (A - C_\theta C_\phi^{-1})^T \) and \( C_\theta C_\phi^{-1} C_\phi^T \) have positive diagonal elements as \( C_e \) is positive definite. This means that the diagonal elements of \( C_e \) are minimised when it is true that,

\[ A - C_\theta C_\phi^{-1} = 0 \]

\[ A = C_\theta C_\phi^{-1} \]  

(3.8)

Hence, we have the estimator matrix

\[ \hat{\theta} = C_\theta C_\phi^{-1} \phi_r \]  

(3.9)

The expected error of the estimator is

\[ C_e = C_\theta - C_\theta C_\phi^{-1} C_\theta^T \]  

(3.10)

Equations (3.10) and (3.9) constitute the Gauss-Markov estimator for the linear minimum least square estimate of a random variable.

The determination of \( C_\phi \) and \( C_\theta \) require having the true field \( \theta \), but all one can have is the observation measurement \( \phi_r \). Hence, we make an assumption of how the observations are related to the actual state of the system. We will assume that the observations are a linear function of the actual state plus random noise.

\[ \phi_{rs} = H\hat{\theta}_s + \nu_s \]  

(3.11)

where the subscript \( s \) signifies the space-time location \( s \), where the observation is made, which is not necessarily where the estimate \( \hat{\theta} \) is made.

\[ :. C_{\theta\phi} = \mathbb{E}[\theta(H\theta_s + \nu_s)^T] \]

\[ = C_{\theta s} H^T + C_{\theta\nu} \]  

(3.12)

\[ C_\phi = \mathbb{E}[(H\theta_s + \nu)(H\theta_s + \nu)^T] \]

\[ = HC_{\theta s} H^T + C_{s\nu}^T H^T + HC_{s\nu} + C_\nu \]

\[ = HC_{s s} H^T + C_{s\nu}^T H^T + HC_{s\nu} + C_\nu \]  

(3.13)
If we suppose the actual state $\theta$ and the noise $\nu$ are uncorrelated, then

$$C_{s\nu} = 0$$  \hspace{1cm} (3.14)

$$C_{\theta\nu} = 0$$  \hspace{1cm} (3.15)

$$\Rightarrow \hat{\theta} = C_{\theta s} H^T [H C_s H^T + C_{\nu}]^{-1} \phi$$  \hspace{1cm} (3.16)

$$C_{\epsilon} = C_{\theta} - C_{\theta s} H^T [H C_s H^T + C_{\nu}]^{-1} [C_{\theta s} H^T]^T$$  \hspace{1cm} (3.17)

If the measurements are the same physical quantity as the estimated data, the $H$ is an identity matrix,

$$\hat{\theta} = C_{\theta s} [C_s + C_{\nu}]^{-1} \phi$$  \hspace{1cm} (3.18)

$$C_{\epsilon} = C_{\theta} - C_{\theta s} [C_s + C_{\nu}]^{-1} C_{\theta s}^T$$  \hspace{1cm} (3.19)

where $[C_s + C_{\nu}]$ is the matrix of covariance between all pairs of observations and $[C_s + C_{\nu}]^{-1}$ is the inverse of this matrix.

$C_{\theta s}$ is the covariance between the quantity $\theta$ to be estimated and the $s^{th}$ measurement. For given position of the observation points, the matrices $[C_s + C_{\nu}], C_{\theta s}$ are constants. Thus for different realizations of the field $\theta(x)$, the estimate $\hat{\theta}_x$ depends linearly on the observations $\phi_r$. Hence, $\hat{\theta}_x$ is a linear estimator.

### 3.3.2 Correlation Function

The correlation function

The properties to be satisfied by a correlation function suitable for objective analysis are:

1. The correlation function must be symmetric. $C(r) = C(-r)$ for $r = (\Delta x, \Delta y, \Delta t)$.
2. The correlation function must be positive definite.

The correlation model used by Arango’s code was

$$C = (1 - r^2)e^{-r^2/2}$$  \hspace{1cm} (3.20)

$$r^2 = \left(\frac{\Delta x}{L}\right)^2 + \left(\frac{\Delta y}{L}\right)^2 + \left(\frac{\Delta t}{L}\right)^2$$  \hspace{1cm} (3.21)

This correlation function is used to calculate the estimated value $\hat{\theta}$ and the error covariance matrix $C_{\epsilon}$. 

45
3.4 Forecast and Assimilation Cycle

Forecast and Assimilation cycle

A simple suboptimal intermittent assimilation scheme has been implemented in this study. This scheme consists of a chain of assimilation cycles. Each cycle starts at a certain restart time and runs up to the next forecast time.

At the beginning of the cycles, the model is initialised using the existing forecast field as well as incoming observed information about the true state of the system at that time. Both these pieces of information, forecast field and observed field, are in terms of a 4-dimensional estimate of the state of the system and its associated error variance.

An assimilation cycle runs from the restart time to the forecast time. At the forecast time, the model is stopped, the forecast fields are updated to reflect the incoming observations, and the model is restarted and run until the next forecast time (Figure 3.1).

Sequential, continuous assimilation

![Figure 3.1: A schematic of data assimilation cycles in model forecast run](image)

The forecast field, observation field estimates and their associated error variance are defined on the same grid and same set of levels.

Let:

1. $\psi_f$ be the forecast field and $e_f$ be the estimate of the associated error variance;
2. $\psi_o$ be the observed field and $e_o$ be an estimate of the associated error variance;
3. $\psi$ be the true field of the system

Both $\psi_f(x, y, z)$ and $\psi_o(x, y, z)$ are assumed to have zero expectation. The problem now is to find a reasonably good estimate of the assimilation stream function $\psi_a$ and its error variance $e_a$ at restart time from those fields.
We calculate the assimilation field as the pointwise linear optimal combination of the forecast field and the observation field.

Using least square approach:

The error in the assimilated field $\psi_a$ with $\psi$ as the true field is given by:

$$\langle \epsilon_a^2 \rangle = \langle (\psi_a - \psi)^2 \rangle$$  \hspace{1cm} (3.22)

Taking a linear optimal combination of $\psi_o$ and $\psi_f$ we have:

$$\psi_a = \mu \psi_o + (1 - \mu) \psi_f$$  \hspace{1cm} (3.23)

From (3.22) and (3.23) we have,

$$\langle \epsilon_a^2 \rangle = \langle ((\mu \psi_o + (1 - \mu) \psi_f) - \psi)^2 \rangle$$  \hspace{1cm} (3.24)

$$= \langle ((\mu \psi_o + (1 - \mu) \psi_f) - (\psi + (1 - \mu) \psi))^2 \rangle$$  \hspace{1cm} (3.25)

$$= \langle (\mu (\psi_o - \psi) + (1 - \mu) (\psi_f - \psi))^2 \rangle$$  \hspace{1cm} (3.26)

$$= \langle (\mu \epsilon_o + (1 - \mu) \epsilon_f)^2 \rangle$$  \hspace{1cm} (3.27)

$$= \langle (\mu^2 \epsilon_o^2 + (1 - \mu)^2 \epsilon_f^2 + 2\mu(1 - \mu)\epsilon_o \epsilon_f) \rangle$$  \hspace{1cm} (3.28)

$$= \mu^2 \langle \epsilon_o^2 \rangle + (1 - \mu)^2 \langle \epsilon_f^2 \rangle + 2\mu(1 - \mu)\langle \epsilon_o \epsilon_f \rangle$$  \hspace{1cm} (3.29)

$$\langle \epsilon_a^2 \rangle = e_a = \mu^2 \epsilon_o + (1 - \mu)^2 \epsilon_f + 2\mu(1 - \mu)\gamma \epsilon_o^{1/2} \epsilon_f^{1/2}$$  \hspace{1cm} (3.30)

where:

$e_o$ - observed error variance

$e_f$ - forecast error variance

$\gamma$ - correlation of forecast and estimate field

To minimise this error, we differentiate $\langle \epsilon_a^2 \rangle$ by $\mu$ and equate it to zero.

$$\frac{d\langle \epsilon_a^2 \rangle}{d\mu} = 2\mu e_o - 2(1 - \mu)e_f + 2(1 - 2\mu)\gamma e_o^{1/2} e_f^{1/2}$$  \hspace{1cm} (3.31)

$$= 2\mu (e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2}) - 2e_f + 2\gamma e_o^{1/2} e_f^{1/2}$$  \hspace{1cm} (3.32)

$$\frac{d\langle \epsilon_a^2 \rangle}{d\mu} = 0$$  \hspace{1cm} (3.33)

(3.34)

which gives:

$$\mu = \frac{e_f - \gamma e_o^{1/2} e_f^{1/2}}{e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2}}$$  \hspace{1cm} (3.35)
Substituting (3.35) in (3.30), we have,

\[ e_o = \mu^2 e_o + (1 - \mu)^2 e_f + 2\mu(1 - \mu)\gamma e_o^{1/2} e_f^{1/2} \]

\[ = (\mu e_o^{1/2} + (1 - \mu)e_f^{1/2})^2 - 2\mu(1 - \mu)(1 - \gamma)e_o^{1/2} e_f^{1/2} \]

\[ \mu e_o^{1/2} + (1 - \mu)e_f^{1/2} = \frac{e_o^{1/2} e_f - \gamma e_o e_f^{1/2} + e_f^{1/2} e_o - \gamma e_o^{1/2} e_f}{e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2}} \quad (3.36) \]

\[ = \frac{e_o^{1/2} e_f^{1/2} (1 - \gamma)(e_o^{1/2} + e_f^{1/2})}{e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2}} \quad (3.37) \]

\[ 2(1 - \gamma)e_o^{1/2} e_f^{1/2} (\mu(1 - \mu)) = 2(1 - \gamma)e_o^{1/2} e_f^{1/2} \left( \frac{e_o e_f + \gamma^2 e_o e_f}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})^2} \right) \]

\[ -\gamma e_o^{1/2} e_f^{1/2} (e_o + e_f) \]

\[ = 2(1 - \gamma)e_o e_f \left( \frac{(e_o^{1/2} e_f^{1/2} (1 + \gamma^2) - \gamma(e_o + e_f))}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})^2} \right) \]

\[ (3.38) \]

\[ e_o = \frac{(1 - \gamma)e_o e_f}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})} \left( \frac{1 - \gamma(e_o^{1/2} + e_f^{1/2})^2}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})} \right. 

\[ \left. - 2(e_o^{1/2} e_f^{1/2} (1 + \gamma^2) - \gamma(e_o + e_f)) \right) \quad (3.39) \]

\[ = \frac{(1 - \gamma)e_o e_f}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})} \left( \frac{(1 + \gamma)(e_o + e_f) + (1 - \gamma)(2e_o^{1/2} e_f^{1/2})}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})} \right. 

\[ \left. - 2e_o^{1/2} e_f^{1/2} (1 + \gamma^2) \right) \quad (3.40) \]

\[ = \frac{(1 - \gamma)e_o e_f}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})} \left( \frac{(1 + \gamma)(e_o + e_f) + 2\gamma e_o^{1/2} e_f^{1/2} (1 + \gamma)}{(e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2})} \right) 

\[ = \frac{e_o e_f (1 - \gamma^2)}{e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2}} \quad (3.41) \]

\[ e_o = \frac{e_o e_f (1 - \gamma^2)}{e_o + e_f - 2\gamma e_o^{1/2} e_f^{1/2}} \quad (3.42) \]
The expected error variance associated with $\psi_a$ in the linear optimal combination has been previously cited in literature (Dombrowsky & Mey, 1992) to be:

$$e_a = \frac{e_o e_f (1 - \gamma)}{e_o + e_f - 2 \gamma e_o \gamma - e_f}$$  \hspace{1cm} (3.43)

The error in the equation (3.43) of the $(1 - \gamma)$ term in the numerator has been corrected to $(1 - \gamma^2)$ in the ROMS sequential data assimilation module too, to reflect the correct error estimate term as derived in (3.42). Though the term $\gamma$ has an inherent uncertainty to its value, the error estimate get affected by this, as $\gamma$ tends to take values close to zero, in which case, $\gamma^2$ can significantly differ from $\gamma$.

### 3.5 Conclusions

A study of the objective analysis scheme to interpolate observed data onto a model grid was made. The mathematical framework for the objective analysis, also known as the optimal interpolation scheme has been illustrated in 3.3. The assimilation scheme for the model assimilation experiments is the sequential suboptimal assimilation as illustrated in 3.4.

The objective analysis scheme is used in the study done in Chapter 4 to interpolate observed subsurface temperature and salinity data in the Indian Ocean region for the year 2004 onto a model grid used for simulations in ROMS. This interpolated data and the corresponding error estimate due to this interpolation is assimilated in the model simulations of the Indian Ocean region using the ROMS model forced by surface fluxes for the year 2004. The error in the previous error estimate term for this assimilation scheme has been corrected in the model. The experimental simulations and model results with assimilation are compared to observed fields in Chapter 4.
4 Experiments with Assimilation: ROMS and ARGO data system

Abstract

This chapter illustrates the main results of this thesis. The results of experiments done with assimilation of Topex/Poseidon Sea Surface Height data and ARGO CTD temperature and salinity profiles data into an Indian Ocean model of ROMS for the year 2004 is presented with some qualitative and quantitative analysis results.

4.1 Introduction

One of the main aims of this thesis is to elicit a case of implementation of data assimilation in an ocean nowcast/forecast model system of the North Indian Ocean. In Chapter 2, a brief overview of the Indian Ocean circulation was given along with an elaboration on an experimental setup of ROMS for the Indian Ocean. The climatological runs on ROMS with separate forcings were compared with observed fields to validate the modeled results. It was seen that the model setup for the Indian Ocean forced by SOC surface heat fluxes and ERA derived surface momentum flux gave a reasonably good climatological solution of the Indian Ocean circulation, temperature and salinity fields. However, the climatological simulation did not solve for certain fine features such as the surface eddies and mixed layer dynamics as accurately as observed in nature. Also the surface temperature field evolved with a growing error in comparison to the observed field, if not relaxed to the climatological field (Killworth et al., 2000).

The availability of good quality synoptic datasets provided by satellites and also increasing availability of insitu measurements from profiling floats and towed bodies deployed in the oceans provides an opportunity to meld observations with models to obtain a good description of the ocean circulation. This has prompted the oceanographic modeling community to adopt data assimilation in ocean models.
A system approach that synthesizes theory, data and numerical computations is essential for rapid and efficient progress in modern interdisciplinary ocean science and technology (Robinson, Lermusiaux, & Sloan-III, 1998). The concept of observing the ocean accompanied by modeling of the ocean processes in order to predict the future states of the ocean and also for parameter estimation of physical processes not understood completely yet has recently crystallised in ocean science and technology. There are three main components of an ocean observation and prediction system: an observational network; a suite of interdisciplinary dynamic models; and data management, analysis and assimilation schemes.

Chapter 3 gave an overview of the history and development of the field of data assimilation in oceanography. It also illustrated the mathematical framework of the data assimilation scheme that is used in this thesis. It was shown theoretically that data assimilation into a numerical ocean model will better the solution (3.4).

In this chapter, the experiments set up to study and implement an Indian Ocean model with real time data assimilation for nowcasting/forecasting are presented. The elements of the modeling and data assimilation system as implemented in the present Indian Ocean circulation study for 2004 are also described in this chapter. Simple yet practical data assimilation schemes are used to reinitialize the ocean model prior to each forecast cycle using data available in real time from a distributed observing network of profiling floats fitted with CTD (Conductivity, Temperature and Depth) sensors deployed as a part of the ARGO project and also altimeter data from the satellite altimeter data. A description of the ARGO float project is given in Subsection 4.2.4 and Subsection 4.2.3 describes the data used from the Merged TOPEX/Poseidon-Jason-ERS1 satellite synoptic measurements of sea-level anomaly for the assimilation experiment. Quantitative validation metrics are formulated and used to evaluate the model solutions.

### 4.2 Regional Ocean Modelling System (ROMS) Ocean Model Configuration

ROMS is a free-surface primitive equation ocean model being used by a broad community for applications from the basin to the coastal and estuarine scales (Haidvogel et al. (2000a), Marchesiello et al. (2003), Peliz, Dubert, Haidvogel, and Cann (2003). Shchepetkin and McWilliams (2005), Shchepetkin and McWilliams (2003), Shchepetkin and McWilliams (1998), Kowalik and Murty (1993)) describe in detail
the algorithms that comprise the ROMS computational kernel. Exact conservation and constancy preservation for tracers has been achieved in ROMS by careful formulation of the time-stepping algorithm using a leap-frog trapezoidal scheme. It also achieves enhanced stability and accuracy in coastal and shallow region applications where the free surface displacement is a significant fraction of the total water depth. Conservative parabolic-spline discretization in the vertical reduces the pressure gradient truncation error that has previously plagued terrain following sigma-coordinate ocean models. The features of ROMS that were implemented for the present study of the ROMS model configured for the Indian Ocean are briefly summarized in Table 4.1.

<table>
<thead>
<tr>
<th>Regional Ocean Modeling System (ROMS) model features</th>
<th>Vertical turbulence closures: KPP (Large et al., 1994) and Generalized Length Scale scheme of Mellor and Yamada (1982)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-surface, hydrostatic primitive equations model in terrain-following coordinates</td>
<td></td>
</tr>
<tr>
<td>Horizontal finite difference and vertical finite volume discretization</td>
<td>- Radiation boundary condition (Marchesiello et al., 2001)</td>
</tr>
<tr>
<td>3rd order upstream-biased advection (Shchepetkin &amp; McWilliams, 1998)</td>
<td>Intermittent sub-optimal melding assimilation</td>
</tr>
<tr>
<td>Pressure gradient and equation of state give reduced σ-coordinate error (Shchepetkin &amp; McWilliams, 2003)</td>
<td>Split-explicit time-stepping of barotropic and baroclinic modes with volume and tracers conserved (Shchepetkin &amp; McWilliams, 2005)</td>
</tr>
<tr>
<td>3rd-order predictor/corrector time stepping</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Regional Ocean Modeling System (ROMS) model features that were implemented for the present study

The model domain covers the Indian Ocean north of $30^\circ S$ (Figure 2.1). The curvilinear boundary fitted grid has an average horizontal resolution of $1/2^\circ \times 1/2^\circ$. The bathymetry is from the National Geophysical Data Center 2-minute relief dataset. All coastal depths less than 10m have been truncated to zero depth, to reduce the CFL constraint on model time step, since the focus of the study was on the large scale ocean circulation of the Indian Ocean. Outflow radiation conditions (Marchesiello et al., 2001) were applied at the open boundaries to active tracers (temperature and salinity) and the nontidal component of the velocities. The southern and eastern boundaries were open boundaries.

Air-Sea fluxes of momentum and heat were computed using the standard bulk formulae (Fairall, Bradley, Godfrey, Wick, & Edson, 1996) applied to the modelled sea surface temperature and atmospheric marine boundary values (10m wind, sea level air temperature, pressure and relative humidity) from datasets of ERA and SOC as
explained in the experiment setup later in this chapter.

4.2.1 Supplied Forcing

The model setup for this study was forced by SOC monthly climatological surface heat fluxes and winds during the first five year run after which the model was restarted and run with monthly climatological heat flux forcing of SOC data but the momentum flux was computed using ERA daily winds from 1996 to 2000 for daily wind forcing. The model was further run forced by NCEP Reanalysis daily winds and monthly wind momentum flux for the years 2001 - 2003. The final experiment done with assimilation was run with a daily wind forcing of NCEP (NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at www.cdc.noaa.gov/) wind stress fields for the year 2004 and daily heat flux from NCEP dataset for the year 2004.

4.2.2 Assimilation Data Sets and Methods

In the assimilation experiments, information from two different sources were used. The altimetry data was sourced from the Merged TOPEX/Poseidon-Jason-ERS1 satellite Mean Sea Level Anamoly data for the year 2004 and the temperature and salinity profile data provided by the Argo floats. A brief description of the datasets follows.

4.2.3 Data used for Sea level Variation Assimilation

It is desirable as done in numerical weather forecasting models, to assimilate observed data into a model to nudge it toward reality. With the advent of satellite meteorology, availability of observation data of the oceans has increased multifold facilitating data assimilation.

In the present study, TOPEX/Poseidon Sea Surface Height Anamoly data (sourced from www.aviso.oceanobs.com/html/donnees/produits/hauteurs/global/msla_uk.html) generated from TOPEX/Poseidon Merged Geophysical Data Record, Generation B (MGDRB) organised as 10 day repeat cycles, have been assimilated into the model on a continuous basis. The TOPEX/Poseidon data for the year 2004 have been used in the present study. The satellite altimetry data of 1° x 1° resolution was objectively analysed onto the model grid of 1/2° x 1/2° resolution using the objective analysis scheme as explained in Section 3.3.
4.2.4 Data used for the Temperature and Salinity Assimilation

Argo is a pilot programme of the Global Ocean Observing System. Salinity and Temperature profiles are collected from a sparse (average $3^\circ \times 3^\circ$ spacing) array of robotic floats that populate the ice-free oceans. This array of robotic floats form the Argo project overseen by an International Argo Steering Team and a Data Management Team that are comprised of representatives of float-providing countries. The first argo floats were deployed in 2000 and the array is expected to be completed by 2006/07. Argo data are made available to users quickly and free of restriction. These data are collected and made freely available by the International Argo Project and the national programmes that contribute to it. (www.argo.ucsd.edu, argo.jcommops.org).

For the first time, the physical state of the upper ocean is being systematically measured. This is a valuable resource for near realtime data assimilation in computational ocean models to setup a good nowcast/forecast system. Argo builds on other upper-ocean observing networks, extending their coverage in space and time, their depth range and accuracy, enhancing them through the addition of salinity and velocity measurements.

ARGO Design and Data:

The design of the Argo network, the number of floats and their density of deployment, is based on experience from the other observing systems, on recent knowledge of ocean circulation variability from the TOPEX/Poseidon altimeter and on the requirements for climate and high resolution ocean models. These floats help to free the large scale oceanographic data collection process from the dependency on ships.

The final array of 3000 floats is expected to provide about 100000 temperature and salinity (T/S) profiles and velocity measurements per year distributed over the global ocean. Floats cycle to 2000m depth every 10 days with 4-5 year life-times for the individual instruments. It continuously performs measurement cycles. Each cycle lasts 10 days and can be divided into 4 phases (fig 4.1):

- A descent from surface to a defined pressure (for eg. 1500 dB)
- A subsurface drift for about 10 days
- An ascending profile with measurements (eg. pressure, temperature, salinity)
- A surface drift with data transmission to a communication satellite
Figure 4.1: Schematic diagram of a single cycle in the mission of a profiling float. Source: www.bom.gov.au

**Argo profiling float**

The Argo floats have an aluminium hull, approximately 6” in diameter and 60” long, with an antenna for data transmission to the satellite system (Figure 4.2). The antenna, conductivity sensor and a temperature probe are mounted on the top end of the float and a pressure sensor is fitted near the bottom. A damper plate is attached near the top to stabilize the float in the surface wave field. The floats can measure temperature to an accuracy of 0.002°C and salinity to an accuracy of 0.05 psu under ideal conditions of a stable water mass. The float is submerged and raised to the surface by the control of the flow of a small amount of oil into and out of an external rubber bladder and into the hull of the float, thus controlling its buoyancy. A small high-power electric pump pumps the oil into and out of the rubber bladder. Typically the weight of the float is about 25 kg and must be accurate to within a few grams. The pressure hulls are designed to withstand pressure of up to 2500 dB and have a design life of more than 4 years with every profiling cycle spanning 10 days.

All Argo data are publicly available in near real time via the Global Data Assembly Centers (GDACs) in Brest, France and Monterey, California after an automated quality control (QC) and in scientifically quality controlled form, delayed mode data, via the GDACs within six months of collection.
Figure 4.2: Schematic diagram of the cross-section of an Argo Float. Source: www.bodc.ac.uk

Figure 4.3: The trajectory of Argo Floats in the Indian Ocean region in the year 2004.
For the present study, 2778 temperature and salinity profiles from 95 ARGO floats present in the Indian Ocean region during the year 2004 were obtained. These profile data were corrected for the operational errors as explained in (Ravichandran et al. (2004) and Vinayachandran (2004)). These corrected profiles were objectively mapped on to the model grid of the Indian Ocean Region (30°S to 30°N and 30°E to 120°E). Figure (4.3) shows the present location of the floats in the Indian Ocean region and their track.

4.3 Assimilation Experiment Procedure

4.3.1 Initial and Forcing Conditions

The ROMS model was initialised with zero momentum and climatological temperature and salinity values at each grid point. The model was run with monthly climatological surface heat, freshwater and momentum forcing for a model time of 5 years using the same forcing files as used in the SE run described in Chapter 2. Subsequently, the model was initialised with this solution and run for a period of 8 model years forced at the surface by daily momentum flux derived from daily winds for the respective years and monthly surface heat and freshwater flux for the period from 1996-2003. The data used for the corresponding surface forcing for these runs have been summarised in Table (4.2). All the model runs were forced by monthly freshwater flux obtained from the SOC dataset. The ROMS model solutions of the Indian Ocean circulation system obtained from the run forced by daily wind momentum data of 2003 were used to initialise the experiment runs in which the assimilation of data was implemented.

<table>
<thead>
<tr>
<th>Model Run</th>
<th>Forcing Data Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Year Climatology Run</td>
<td>SOC climatological monthly heat flux and surface winds</td>
</tr>
<tr>
<td>5 year run</td>
<td>ERA daily winds for the period 1996-2000 and SOC climatological monthly heat flux for the period 1996-2000</td>
</tr>
<tr>
<td>3 Year run</td>
<td>NCEP daily winds for the period 2001-2003 and NCEP monthly heat flux for the period 2001-2003</td>
</tr>
<tr>
<td>1 Year Experiment Run with and without assimilation</td>
<td>NCEP daily winds and NCEP daily heat flux for the year 2004</td>
</tr>
</tbody>
</table>

Table 4.2: Forcing datasets used to setup the assimilation experiment
4.3.2 Forecast and Assimilation Cycle

An overview of the assimilation cycle implemented is presented in this section. The mathematical basis of suboptimal assimilation is given in the Chapter 3 along with the theory of objective analysis used to map the scattered ARGO profiles onto the model grid.

Each 10 day forecast proceeded in the following steps:

The model assimilation cycle initialises with an a-priori estimate of the ocean state, \( \phi^f \), where \( \phi \) can be any model variable (temperature, salinity etc.) forecasted in a previous model assimilation cycle. The difference between \( \phi^f \) and observations \( \phi^{data} \) gathered during the cycle was mapped to the model grid using optimal interpolation at 10 day intervals. The mapped adjustment is given by

\[
\phi' = CA^{-1}(\phi^{data} - \phi^f)
\]  

(4.1)

where the matrix \( C \) is the covariance of each model grid point with each observation location and time, and \( A \) is the covariance of the observations with each other (McIntosh, 1977), (Daley, 1991). The covariance function was assumed to be a gaussian with scales of 5000 km radius of influence. The ratio of observational error to signal variance that augments the diagonal of \( A \) was set constant at \( 10^{-2} \). The gridded observations are then \( \phi^o = \phi^f + \phi' \), and have normalised expected error variance, \( e_o^2 \), given by the diagonal of \( (I - CA^{-1}C^T) \) where \( I \) is the identity matrix (McIntosh, 1977).

In regions of the model domain several covariance length distant from any data, the adjustment \( \phi' \to 0 \), the model state \( (\phi^f) \) is retained and \( e_o^2 \to 1 \). Close to the observation locations, the observed data govern the forecast field more than the model estimate in proportion to the model-observation mismatch and the chosen covariance scales. The model was then restarted from the beginning of the previous forecast, with the 10 day optimally interpolated gridded data fields for temperature and salinity assimilated over 10 days by melding.

This melding assimilation follows the method of Dombrowsky and Mey (1992), which, in common with optimal interpolation, is based on Gauss Markov theory. At times corresponding to the \( \phi^o \) maps, a weighted sum of observations and forecast is computed as

\[
\phi^o = \mu \phi^o + (1 - \mu)\phi^f
\]  

(4.2)

The weights given by

\[
\mu = \frac{e_f - \gamma e_f^{1/2}e_o^{1/2}}{e_o + e_f - 2\gamma e_f^{1/2}e_o^{1/2}}
\]  

(4.3)

58
are optimal in the sense that the “analysis” estimate $\phi^a$ has the minimum expected mean squared error for the assumed error variance of observations and forecast. The model is reinitialised with $\phi^a$ and the model integration proceeds. We have taken $\gamma = 0.01$ for the correlation between model and observations. (Dombrowsky and Mey (1992) discuss the role of this parameter). At each melding reinitialisation, the forecast error is reset to the analysis error,

$$e_a = \frac{e_f e_o (1 - \gamma^2)}{e_o + e_f - 2\gamma e_f^{1/2} e_o^{1/2}} \quad (4.4)$$

It is assumed that the model error, $e_f$, grew exponentially with a timescale of 10 days.

Each assimilation period forecast has been adjusted to reflect the observations acquired, and this new hindcast state becomes the initial condition for the next ocean forecast cycle forced with the new daily atmospheric prediction.

This sequential suboptimal assimilation has been implemented in various studies (Wilkin, Arango, Haidvogel, Lichtenwalner, Glenn, and Hedstrom (2005), Dombrowsky and Mey (1992)). In Wilkin et al. (2005), temperature and salinity assimilation data were prepared by optimal interpolation of shipboard towed-body data. Surface current observations from coastal radar were projected vertically for assimilation. The results of the simulation were validated with observed features of a coastal ecosystem, and it was shown that intermittent melding provided the forecast system with more significant skill than the nudging assimilation scheme.

### 4.4 Results of the Assimilation Experiment

The assimilation experiment was performed by assimilating objectively analysed sea level anomaly and observed depth profiles of temperature and salinity into a ROMS model of the Indian Ocean. The assimilation model solution is compared to the model solution without assimilation. Both these runs start from the same initial conditions and are not relaxed to climatological fields of SST or salinity as was done in the model runs shown in Chapter 2 and contain the same forcing fields as discussed in the subsection 4.2.1.

A qualitative and quantitative comparison of the model reconstructed ocean fields for the 1 year run done, forcing the model by the flux fields of 2004, with and without assimilation is studied in this chapter. The fields that are compared to study the impact
of assimilation on the ocean dynamics are temperature, salinity, thermocline depth and sea surface height.

4.4.1 Impact of assimilation on the sea surface temperature

The modelled Sea Surface Temperature (SST) is compared with the gridded SST from TMI (TRMM Microwave Imager) (sourced from www.ssmi.com/tmi/tmi_browse.html) satellite data. This provides us with information about the assimilation technique’s ability to reconstruct SST.

The model performs poorly in the simulation of SST if it is not relaxed towards climatological values or if SST is not assimilated into it.

Figure (4.4) shows the monthly average SST field of the Northern Indian Ocean during the months of February on the left panel and April on the right panel. Both the model derived fields, with and without assimilation compare well with observation as it is still early days in the model solution. Though, the model derived SST field for the month of April seems to deteriorate and seems to warm up to more than $2^\circ - 3^\circ$C more than the observed values in this region. The progressive deterioration of the modelled SST as compared to observation can be noticed in the subsequent months seen in the top two panels of Figures (4.5, 4.6), whereas the assimilated run is able to reconstruct the SST field well except that it is about $0.5^\circ$ to $1^\circ$C lower than observed values in certain regions. Also certain features such as the extent of upwelling off the coast of Somalia during the monsoon months of August-September is reconstructed well in the assimilated run. The reasons that the model seems to be failing to simulate the SST well without nudging or assimilation could be either due to a less estimate of the parameterisations involved in the heat budget equation and mixing terms or also due to the non-availability of a good measure of the heat budget for the surface boundary condition of the oceans. Until a good system of measurement is available to estimate the heat budget at the surface of the oceans, we have to rely on modelled values for these quantities whose error in synoptic distribution will entirely depend on the atmospheric models which generate them. Hence the model solutions will be highly sensitive to the surface forcing parameters and also the subgrid scale dynamics parameterisation.

4.4.2 Quantitative Analysis of the Ocean SST Field
Figure 4.4: Comparison of SST reconstructed by the model with and without Data assimilation with TMI SST data.
Figure 4.5: Comparison of SST reconstructed by the model with and without Data assimilation with TMI SST data (Values above 33°C are not plotted)
Figure 4.6: Comparison of SST reconstructed by the model with and without Data assimilation with TMI SST data (Values above 33°C are not plotted)
The Root Mean Square Error in the model reconstructed field of SST compared to the observed TMI SST field were computed and are shown in Figure (4.7). It can be seen that the RMSE over the period of one year has an increasing trend for the model simulated SST without assimilation. The model run with assimilation has lower values of RMSE and does not show an increasing trend of error.

![Figure 4.7: RMSE between SST reconstructed by the model (with and without Assimilation) and TMI SST data](image)

Also, to estimate the correlation between the model reconstructed SST with observations from the TMI satellite data, the correlation coefficients between the model and the observed field were computed and plotted. The bottom panel in figure (4.8) shows the spatial map of correlation between the model derived SST without any assimilation of data and TMI SST and the top panel in figure (4.8) shows the spatial map of correlation between the model derived SST after assimilation and TMI SST. The shaded region in both plots indicate regions with their correlation coefficient above 99% significance level in the red region and above 95% significance in the blue region. It can be seen that the correlation of the data assimilated model SST and observed SST field has a higher correlation almost throughout the model domain compared to the correlation coefficients of the SST from the model run without assimilation and observed field, especially in the Bay of Bengal and southern Arabian Sea region.

### 4.4.3 Impact of assimilation on the Thermocline Depth

The Figure (4.9) shows a plot of the thermocline depth calculated in the Indian Ocean region for the months of January-February-March 2004 of the model run with (right panel) and without (left panel) assimilation. The thermocline depth is the depth of the ocean waters, from the surface, at which the temperature gradient is maximum. The measured and estimated thermocline depths in the Indian Ocean region for these months of the year compare better with the thermocline depths reconstructed by the
Figure 4.8: Correlation coefficient between model SST and TMI SST data with red regions significant above 99% and blue significant above 95%
model run with assimilation. Similarly, Figure (4.10) shows the plot of the calculated thermocline depths in the Indian Ocean region for the months of July-August-September 2004. Again certain measured and estimated values from models match well with the depths reconstructed by the model run with assimilation. The thermocline depths estimated by the model run without assimilation consistently are higher than the values with assimilation. This may be due to several reasons such as fault in the mixing scheme used or heat budget and error in the circulation reconstruction by the model.

In Figure (4.11) the depth profiles of temperature at 65.5E-21N in the Arabian Sea (left panel) and 89.5E-11.5N in the Bay of Bengal (right panel) have been plotted till 500m to compare with the values measured by the Argo Float present at these positions for months of the year 2004. Though the profile in the assimilated run shows a shallower thermocline, the temperature variation with depth has a similar trend as the Argo measured profile, whereas the profile reconstructed by the model run without assimilation deviates in its profile trend as well as estimate of the mixed layer. The thermocline is much more diffuse in the model solution derived without assimilation.

Figure (4.12) shows the depth profiles of salinity at 65.5E-21N in the Arabian Sea and 89.5E-11.5N in the Bay of Bengal till 500m to compare with the values measured by the Argo Float present at these positions throughout the year 2004. The model run with assimilation is able to capture the salinity profile evolution in time better than the model run without assimilation. Though the profiles of the model runs do not match exactly with the actual measured values, the overall trend and values compare well with the measured values.

4.4.4 Impact of assimilation on the sea surface height anomaly

The model simulation of the Sea Surface Height Anomaly(SSHA) is compared with the gridded ERS-TOPEX/Poseidon SSHA every seventh day. Though the model seems to simulate the SSHA well without any assimilation, assimilation of the SSHA data into the model improves the magnitude and position of certain prominent features in the Northern Indian Ocean SSHA such as the Great Whirl off the coast of Somalia during the monsoons. Figure 4.13 shows the SSHA field over the Northern Indian Ocean in the month of September when the Great Whirl is active and has a high SSHA over a prolonged period in this region. The model seems to reconstruct this better in magnitude and position when observed SSHA data is assimilated into the model rather
Figure 4.9: Comparison of Thermocline depth reconstructed by the model with and without Data assimilation for the months of January-February-March 2004.
Figure 4.10: Comparison of Thermocline depth reconstructed by the model with and without Data assimilation for the months of July-August-September 2004.
Figure 4.11: Vertical profiles of Temperature at 65.5E-21N in the Arabian Sea (left panel) and 89.5E-11.5N (right panel) in the Bay of Bengal till 500m.
Figure 4.12: Vertical profiles of Salinity at 65.5E-21N in the Arabian Sea (left panel) and 89.5E-11.5N (right panel) in the Bay of Bengal till 500m
than not. Overall, the model seems to fail in simulating eddies as fine as those observed in the satellite data and this limitation could be only overcome by refining the model to a finer horizontal scale.

4.5 Conclusions

In this chapter, we first described the model configured for the Indian Ocean region which was spun up with climatological surface fluxes and further forced with surface fluxes to prepare for an initial condition to conduct the assimilation experiment run. The model is sensitive to many parameters and also to the surface forcing fluxes. Errors in these parameterisations and forcing fluxes can cause the model to have significant systematic errors. Hence, an experimental run was made of the model configured for the North Indian Ocean with assimilation of subsurface temperature and salinity profile data into the model solution intermittently and sub-optimally. These model solutions were compared to the model solutions obtained without data assimilation into the model run. These comparison results were presented.

These comparisons showed that the model significantly improves in its simulation of the surface temperature evolution and subsurface structure. Also, the models tendency to warm up largely when it is run without any assimilation or relaxation to the observed field is largely curtailed by the assimilation of data into the model. Hence the errors that may be due to errors in the vertical parameterisation or errors in the surface flux data are corrected to some extent by assimilating data into the model.

The reason for performing the data assimilation is to produce the best estimate of the initial conditions for a forecast. Some of the circulation features of the ocean, which may not be studied from direct observations, can be studied better by model simulations. Though the assimilated solutions of the model too did not match exactly with the observed fields, assimilation does significantly improve the model solution both temporally and spatially. Hence, it is to be expected that with better quality and more synoptic observed data for the surface and subsurface fields of the ocean made available, assimilation of these data into the model would significantly improve the model simulations and bring the model solutions closer to reality than before.
Figure 4.13: Comparison of SSHA reconstructed by the model with and without Data assimilation with Topex/Poseidon SSHA data
5 Summary and Conclusions

Abstract

In this concluding chapter we summarize the results of the Dissertation, with an emphasis on novelties, and new problems suggested by this research.

5.1 Objective of the Dissertation

Data assimilation is now well established as an important scientific method and practical tool in the atmospheric sciences (Ghil et al., 1997) but not in oceanography, where adequately sophisticated numerical circulation models and suitable ocean datasets have been essentially lacking. However, with recent substantial improvements in both ocean modeling and observing systems, oceanographers have now begun vigorous exploration of data assimilation in the context of large scale and regional ocean circulation modeling.

At the outset of this Dissertation we noted that data assimilation can be used to obtain a four dimensional realization of the oceanic flow that is simultaneously consistent with the observational evidence and with the dynamical equations of motion and also to provide initial conditions for predictions of oceanic circulation. In the literature ranging from the mathematical framework of data assimilation (Gandin (1965), Sasaki (1958), Weiner (1949), Kolmogorov (1941)) to operational implementation of data assimilation (Bergthorsson and Döös (1955), Panofsky (1949)) and many others, one can find various justifications for the development and study of data assimilation and its impact on model solutions in various regions of the world. In this thesis, we have investigated the impact of data assimilation of subsurface tracer profile data into a terrain-following primitive equation ocean model (ROMS) of the North Indian Ocean.

5.2 Contributions of the Dissertation

In this section we briefly summarize the contributions of this thesis including some problems suggested by this work.
The circulation in the Indian Ocean and various studies and experiments conducted in this region were reviewed to have a good understanding of the complex dynamics of this region. The model set up for numerical experiments had to be configured and evaluated prior to the experimental runs, to evaluate its performance in simulation of the Indian Ocean dynamics. Also, good forcing parameters had to be arrived at to setup the background for the assimilation experimental runs. The model solutions obtained when forced by two different datasets were compared. The two different datasets were COADS and SOC-ERA surface momentum and heat flux datasets, which are described in detail in Section 2.2. The model solution comparisons of these two separate runs showed that the model forced by the COADS fluxes consistently over-estimates the SST, SSS fields which in turn affect the circulation features negatively. The SOC-ERA flux forced model performs better in comparison, in simulating the SST, SSS and surface circulation. ROMS is able to reproduce the known features of the surface circulation as compared to climatology. Though the surface tracers were relaxed to climatology values to make up for the errors in the mixing and frictional parameters and also uncertainty in forcing surface flux data, the model solutions compared well with observed features of the Indian Ocean circulation fields.

The forecast system developed employed the circulation model ROMS in conjunction with simple data assimilation methods that utilized temperature and salinity profile measurements made by Argo floats. Forecast evaluation was carried out in terms of the surface temperature evolution patterns, subsurface structures and also observed main features of the circulation patterns in the Indian Ocean circulation. The mathematical framework for the objective analysis, also known as the optimal interpolation scheme, has been illustrated in 3.3. This scheme was used to interpolate observed data onto the model grid and the corresponding error estimate due to this interpolation is assimilated in the model simulations of the Indian Ocean region using the ROMS model forced by surface fluxes for the year 2004. The data assimilation scheme tested in this dissertation study was the intermittent melding scheme described in (Dombrowsky & Mey, 1992). The assimilation scheme for this procedure is the sequential suboptimal assimilation as illustrated in 3.4. The error in previous literature (Dombrowsky & Mey, 1992) in the error estimate term for this assimilation scheme has been corrected for in the model.

The model, without assimilation, had considerable skill at reproducing the salinity and sea surface heights observed in the Indian Ocean, but this skill was not extended to reproducing a good sea surface temperature field which deteriorated with time. The model simulated SST grew in error with respect to the observed field when tempera-
ture data was not assimilated in the model. Also, the subsurface thermocline structure was considerably more diffuse than what is observed in the Indian Ocean region. With the assimilation of surface and subsurface temperature and salinity profile data, the SST field simulation was significantly improved and also a better structure of the subsurface thermocline was simulated. The assimilated model simulations of SST was coherent with observations at seasonal and monthly time scales. Also, the simulation of the subsurface thermocline structure was improved considerably by the intermittent melding of the temperature profile data. Melding, which reinitialises the model state at intermittent intervals, allows the model to pursue its own trajectory until the next data assimilation step; this simulation is very skillful (as evidenced by the model skill without assimilation). Relaxation to climatological values (as done in the model evaluation experiments described in Chapter 2) or nudging to observations, on the other hand, can unnaturally constrain the model evolution during the forecast interval, denying the model the opportunity to respond accurately to the forcing. This is especially so if the observational system is unable to return data at sufficiently frequent time intervals.

Some aspects of the observational network and model setup were found to be limiting to simulation skill. The model is sensitive to many parameters and also to the surface forcing fluxes. Errors in these parameterisations and forcing fluxes can cause the model to have significant systematic errors. The models tendency to warm up largely when it is run without any assimilation or relaxation to the observed field is largely curtailed by the assimilation of data into the model. Hence the errors that may be due to errors in the vertical parameterisation or errors in the surface flux data are corrected to some extent by assimilating data into the model. The reason for performing the data assimilation is to produce the best estimate of the initial conditions for a forecast. Also, with better model field evolution, we can study some of the circulation features of the ocean, which may not be studied from direct observations. Though the assimilated solutions of the model too didnot match exactly with the observed fields, assimilation does significantly improve the model solution both temporally and spatially. Hence, it is to be expected that with better quality and more synoptic observed data for the surface and subsurface fields of the ocean made available, assimilation of these data into the model would significantly improve the model simulations and bring the model solutions closer to reality than before. This work provides the foundation for continuing modelling efforts in the North Indian Ocean.
5.3 Future Directions

While the model results currently provide good representation of the ocean state, there are undoubtedly improvements that can be made to the model. One area worthy of additional investigation is with regards to the impact of the data assimilation of the complete ARGO profiles dataset that would be available from about 2007-08. The deficiency in the data set used for this research were found to be in the southern Indian Ocean and also partially in the Bay of Bengal too. With an increased number of observations that could be available from these regions in the future, improved results can be expected. Also, another area worthy of investigation is the impact of higher model resolution which would improve the model dynamics but also increase the errors in the objective analysis done to map observed data onto the model grid.

The results that were presented in Chapter 4 were from running the model in a hindcast mode. It is relatively a straight-forward procedure to transition the model into a real-time nowcast/forecast mode. The greatest challenge is to obtain the required forcing data and prepare it for use by the model in a timely fashion.

5.4 Concluding Comments

Data assimilation (DA) has evolved through a number of phases from an era of few available data and poor models using DA methods such as spline interpolation to present day where plentiful data sets and advanced general circulation models are available which make use of high performance DA methods such as the extended Kalman filter and four dimensional variational methods. Communication among those making the observations of oceanic fields, the data managers, modelers and other users who make use of the model forecasts is critical. The ultimate fruits of this labour include, a better understanding of our oceans, the earth, its climate and our role in its beautiful life.
Appendix A  A brief overview of ROMS

In this thesis, the 3-D sigma vertical coordinate ocean model ROMS is used in the study of the impact of data assimilation into an ocean model of the Indian Ocean.

ROMS is a free-surface primitive equation ocean model being used by a broad community for applications ranging from the basin to coastal and estuarine scales (Haidvogel et al. (2000a), Marchesiello et al. (2003), Peliz et al. (2003)). Schepetkin and McWilliams (1998, 2003, 2005) describe in detail the algorithms that comprise the ROMS computational kernel.

The model solves the primitive equations in an earth-centered rotating environment. The model equations are based on the Boussinesq approximation (where density variations are neglected everywhere except in the gravitational force term) and hydrostatic vertical momentum balance (where the buoyant force balances the pressure force) and are expressed in an earth-centered, Cartesian system of coordinates rotating at an angular speed of $\Omega = f y$. The model equations and discretisation methods are described in (Song & Haidvogel, 1994b)

A.1 Equations of the Model

The Primitive equations for momentum balance in x- and y- directions in Cartesian Coordinates are, respectively:

$$\frac{\partial u}{\partial t} + \bar{v} \cdot \nabla u - fv = - \frac{\partial \phi}{\partial x} + F_u + D_u$$

$$\frac{\partial v}{\partial t} + \bar{v} \cdot \nabla v + fu = - \frac{\partial \phi}{\partial y} + F_v + D_v$$

where $f = 2 \sin \phi$, u, v are velocities in the x, y direction and $F_u, F_v$ and $D_u, D_v$ are the corresponding forcing and diffusion terms along the respective directions of u, v.

The advective - diffusive equations are:

$$\frac{\partial T}{\partial t} + \bar{v} \cdot \nabla T = F_T + D_T$$
\[
\frac{\partial S}{\partial t} + \vec{v} \cdot \nabla S = F_S + D_S
\]

where \( T, S \) are temperature and salinity respectively and \( \vec{v} \) is the velocity vector.

The equation of state is:

\[
\rho = \rho(T, S, P)
\]

where \( \rho \) is the density of the ocean water, \( T, S, P \) are temperature, salinity and pressure respectively.

The vertical momentum equation is:

\[
\frac{\partial \phi}{\partial z} = -\frac{\rho g}{\rho_o}
\]

where \( \phi \) is the free surface displacement, \( \rho \) is the density from the equation of state, \( \rho_o \) is the average density along the depth and \( g \) is the acceleration due to gravity.

The model assumes Boussinesq approximation, where the density variations are neglected in the momentum equations except in their contribution to the buoyancy force in the vertical momentum equation. It also assumes hydrostatic approximation i.e., the vertical pressure gradient balances the buoyancy force.

The continuity equation for an incompressible fluid is:

\[
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0
\]

where \( u, v, w \) are the corresponding velocities in \( x, y, z \) directions.

The horizontal viscous and diffusive terms \( D_u \) and \( D_v \) are biharmonic formulations of the form:

\[
D_u = -A_{MHB} \nabla^2 \nabla^2 u
\]

The model uses horizontal curvilinear coordinates and the \( s \)-coordinate in the vertical. The vertical is divided into an equal number of points, according to \( s = z/h(X, Y) \). A nonlinear stretching of the vertical coordinate is applied that depends on the local water depth.

\[
z = h_s + (h - h_s)C(s)
\]
where
\[
C(s) = (1 - \theta_b) \frac{\sinh(\theta s)}{\sinh(\theta)} + \theta_b \frac{\tanh[(\theta(s + 1/2)) - \tanh(\theta/2)]}{2 \tanh(\theta/2)}
\]

with \(-1 \leq s \leq 0, 0 \leq \theta_b \leq 1, 0 \leq \theta \leq 20\). For large \(\theta\), the coordinate lines crowd near the surface; additionally, if \(\theta_b\) approaches 1, resolution at the bottom boundary is enhanced.

The dynamic equations become somewhat more complicated after the coordinate transformation is applied. The momentum equations:
\[
\begin{align*}
\frac{\partial u}{\partial t} + \bar{v} \cdot \nabla u - f v &= - \frac{\partial \phi}{\partial x} - \left( \frac{g \rho}{\rho_0} \right) \frac{\partial z}{\partial x} - g \frac{\partial \zeta}{\partial x} + F_u + D_u \\
\frac{\partial v}{\partial t} + \bar{v} \cdot \nabla v + f u &= - \frac{\partial \phi}{\partial y} - \left( \frac{g \rho}{\rho_0} \right) \frac{\partial z}{\partial y} - g \frac{\partial \zeta}{\partial y} + F_v + D_v
\end{align*}
\]

The advective-diffusive equations are:
\[
\begin{align*}
\frac{\partial T}{\partial t} + \bar{v} \cdot \nabla T &= F_T + D_T \\
\frac{\partial S}{\partial t} + \bar{v} \cdot \nabla S &= F_S + D_S \\
\frac{\partial \phi}{\partial s} &= \left( \frac{-pgH_z}{\rho_0} \right) \\
\frac{\partial H_z}{\partial t} + \frac{\partial H_z u}{\partial x} + \frac{\partial H_z v}{\partial y} + \frac{\partial H_z \Omega}{\partial s} &= 0
\end{align*}
\]

where
\[
\begin{align*}
\bar{v} &= (u, v, \Omega) \\
\bar{v} \cdot \nabla &= u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + \Omega \frac{\partial}{\partial s}
\end{align*}
\]

The vertical velocity in \(s\) coordinates is
\[
\Omega(x, y, s, t) = \frac{1}{H_z} \left[ w - (1 + s) \frac{\partial \zeta}{\partial t} - u \frac{\partial z}{\partial x} - v \frac{\partial z}{\partial y} \right]
\]

and
\[
w = \frac{\partial z}{\partial t} + u \frac{\partial z}{\partial x} + v \frac{\partial z}{\partial y} + \Omega H_z
\]
A.2 Boundary Conditions

The boundary conditions for the above momentum and tracer equations at the sea surface \((z = \zeta)\) are:

\[
K_m \frac{\partial u}{\partial z} = \tau_s^x (x, y, t)
\]

\[
K_m \frac{\partial v}{\partial z} = \tau_s^y (x, y, t)
\]

\[
K_T \frac{\partial T}{\partial z} = \frac{Q_T}{\rho_o c_P}
\]

\[
K_S \frac{\partial S}{\partial z} = (E - P) S
\]

\[
w = \frac{\partial \zeta}{\partial t}
\]

where \(K_m, K_T, K_S\) are the vertical viscosity and diffusivity, \(\tau_s^x, \tau_s^y\) are the components of wind stress acting on the free surface in the \(x\) and \(y\) directions, respectively; \(Q_T\) is the heat flux across the sea surface; \(E\) and \(P\) are the evaporation and precipitation rates, respectively; and \(C_p\) is the heat capacity of sea water. Correspondingly, at the bottom surface, \(z=-h\), the boundary conditions are

\[
K_m \frac{\partial u}{\partial z} = \tau_b^x (x, y, t)
\]

\[
K_m \frac{\partial v}{\partial z} = \tau_b^y (x, y, t)
\]

\[
K_T \frac{\partial T}{\partial z} = 0
\]

\[
K_S \frac{\partial S}{\partial z} = 0
\]

\[-w + \vec{v} \cdot \nabla h = 0\]

where \(K_m, K_T, K_S\) are the vertical viscosity and diffusivity; \(\tau_b^x\) and \(\tau_b^y\) are the bottom surface stresses.

The model equations are solved separately for their external mode representing the depth averaged flow and the internal mode representing the vertically varying component. The depth averaged equations are discretized on a staggered Arakawa C-grid.
A leapfrog-trapezoidal scheme is used for the time integration; this is slightly more stable than the leapfrog itself and strongly suppresses the computational mode. The depth averaged equations are coupled to the vertically varying mode equations through the nonlinear and pressure gradient terms. In order to solve the coupled external and internal mode efficiently, a short time step is used for solving the external mode equations in order to satisfy the CFL condition arising from the fast moving gravity waves. An implicit long time step (about 20 times longer than the short one) is used to obtain the internal mode.

A.3 Model Parameterisation

An important and unresolved issue in ocean circulation modeling is the appropriate parameterisation of subgrid scale processes. One of the purposes of the s-coordinate is to achieve a uniformly high resolution in the surface layer so that mixing parameterisations can be accurately applied.

In the present study, a KPP(K-Profile Parameterisation) scheme of Large et al. (1994) is used to parameterise the vertical eddy viscosity and eddy diffusivity coefficients.


Swallow, J., Molinari, R., Bruce, J., Brown, O., & Evans, R. (1983). Development of near-surface flow pattern and water mass distribution in the somali basin in


