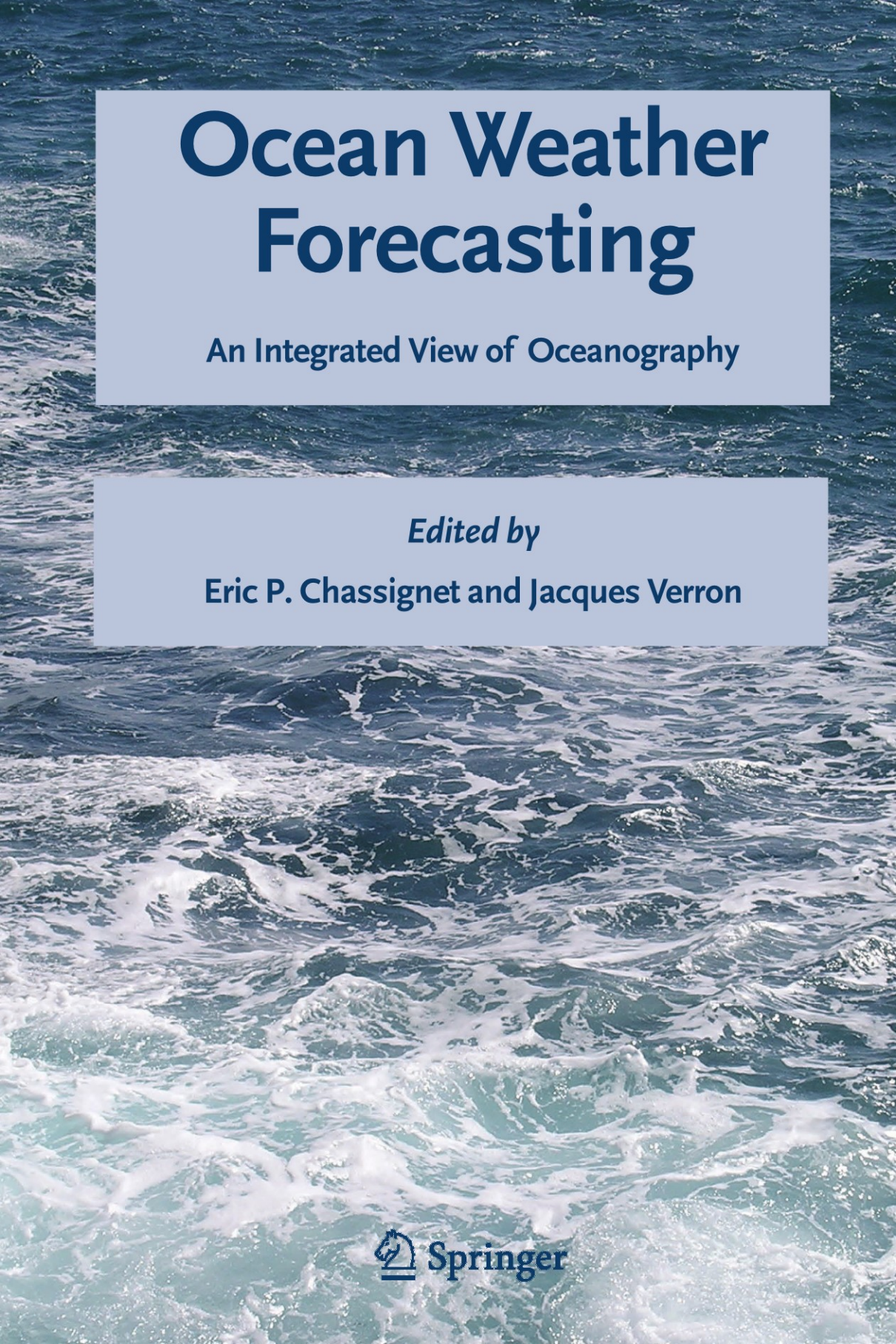

Ocean Weather Forecasting

An Integrated View of Oceanography

Edited by

Eric P. Chassignet and Jacques Verron

Springer

OCEAN WEATHER FORECASTING

Ocean Weather Forecasting

An Integrated View of Oceanography

Edited by

ERIC P. CHASSIGNET
University of Miami, U.S.A.

and

JACQUES VERRON
CNRS, LEGI, Grenoble, France

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3981-6 (HB)
ISBN-13 978-1-4020-3981-2 (HB)
ISBN-10 1-4020-4028-8 (e-book)
ISBN-13 978-1-4020-4028-3 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

Printed in the Netherlands.

This book is dedicated to Christian Le Provost (1943-2004), an eminent scientist in the domains of ocean physics, tides, satellite altimetry, and ocean modeling. He was also a pioneer in the development of operational oceanography.

Contents

Part IV: Data assimilation

Part V: Systems

PREFACE

Progress in a wide range of ocean research and applications depends upon the prompt and dependable availability of ocean information products. The field of physical oceanography has matured to a point where it is now conceivable to combine numerical models and observations via data assimilation in order to provide ocean prediction products on various spatial and time scales. As a result, many nations have begun large-scale efforts to provide routine products to the oceanographic community. The Global Ocean Data Assimilation Experiment (GODAE) provides a framework for these efforts, i.e., a global system of observations, communications, modeling, and assimilation that will deliver regular, comprehensive information on the state of the oceans, in a way that will promote and engender wide utility and availability of this resource for maximum benefit to the community. The societal benefit will be an increased knowledge of the marine environment and ocean climate, predictive skills for societal, industrial, and commercial benefit and tactical and strategic advantage, as well as the provision of a comprehensive and integrated approach to the oceans.

We therefore considered it timely, given the international context, to bring together leading scientists to summarize our present knowledge in ocean modeling, ocean observing systems, and data assimilation to present an integrated view of oceanography and to introduce young scientists to the current state of the field and to a wide range of applications. This book is the end result of an international summer school held in 2004 that aimed, among other things, at forming and motivating the young scientists and professionals that will be the principal movers and users of operational oceanographic outputs in the next 10 years. The chapters collected in this volume cover a wide range of topics and are authored not only by scientists, but also by system developers and application providers.

We would like to thank all the speakers for providing a stimulating series of lectures at this GODAE Summer School. We also express our appreciation to the members of the scientific committee and to the GODAE IGST who contributed in numerous ways to the success of the school. We thank all the attendees (see list in Appendix) for participating actively in the lecture review process and for creating a most cordial atmosphere. We thank Jean-Michel Brankart, Laurence Crosnier, Nicolas Ferry, and David Rozier for preparing and putting together a superb set of student exercises. Finally, our thanks go to Yves Ménard, Joëlle Guinle, Véronique Huix, Nicole Bellefond, and Josiane Brasseur who spent a considerable time with the logistics of the school before and after. A special thank goes to Josiane Brasseur for her help in formatting the manuscripts.

Primary support for this GODAE summer school was provided by the Centre National d'Etudes Spatiales (CNES), the MERSEA EU project, and GODAE. Additional funding was provided by the National Science Foundation (NSF) and by the National Aeronautics and Space Administration (NASA). This support is gratefully acknowledged.

Eric P. Chassignet
Jacques Verron

April 15, 2005

Chapter 1

PERSPECTIVES FROM THE GLOBAL OCEAN DATA ASSIMILATION EXPERIMENT

Neville Smith

Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

Abstract : The Global Ocean Data Assimilation Experiment (GODAE) is introduced, including a discussion of the historical basis and conceptual framework. GODAE aims to develop a global system of observations, communications, modeling and assimilation that will deliver regular, comprehensive information on the state of the oceans in a way that will promote and engender wide utility and availability of this resource for maximum benefit to society. The overall strategy and guiding principles are introduced and the core components discussed. The data and modeling and assimilation systems are intended to provide infrastructure serving a broad range of users and applications. The targeted applications include open-ocean forecasts, coastal and regional prediction, climate assessments and prediction, and reanalyzes for scientific and other purposes. Both internal and external metrics have been developed to assure the quality and reliability of the product streams. The focus at present is on developing an understanding and more intimate relationship with the user community.

Keywords : Ocean, data assimilation, observations, prediction.

1. Introduction

The concept of a Global Ocean Data Assimilation Experiment (GODAE) emerged from the Ocean Observation Panel for Climate (OOPC) in 1997 and derived from concern that attracting the investment for an adequate long-term global ocean observing system depended upon a clear demonstration of the feasibility and value of such a system (Smith and Lefebvre, 1997). Using the First GARP (Global Atmospheric Research Program) Global Experiment (FGGE) as a model, OOPC proposed GODAE as an experiment in which a comprehensive, integrated observing system would be established and held in place for several years and the data

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 1-17.

assimilated into state-of-the art models of the global ocean circulation in near real-time to demonstrate utility.

GODAE recognized the pioneering work in operational oceanography in the U.S. (see Peloquin, 1992, and other papers within that special volume) and the fact that interest in building a broader global capability was emerging in several nations (for example, MERCATOR in France; Bahurel, this volume). This work, among others, guided the development of the concept and ultimately the strategy (International GODAE Steering Team (IGST), 2000) and implementation plan (http://www.bom.gov.au/GODAE/).

As with many international initiatives, GODAE by itself does not provide resources or develop capacity. Rather it relies on the resources and capacity derived from national or regional initiatives and GODAE's role is one of coordination and cooperation and, for example, introducing standards and references for the business of operational oceanography.

This paper recounts the development of GODAE and some perspectives drawn from experience and from those who are thinking of the future of operational ocean analysis and prediction. In order to provide a little context for GODAE in relation to the evolution of ocean science and the development of weather prediction, Section 2 discusses some historical aspects and section 3 some of the lessons learnt from numerical weather prediction. Section 4 discusses the rationale and scope while section 5 introduces the core components. Other chapters of this volume examine these components (e.g., observations, models, assimilation) in more detail. Section 6 discusses applications and the utility of GODAE products and some of the issues surrounding the use of model products. Again, there are several papers in this volume (e.g., Hackett et al.) that go into this area in more detail. Section 7 discusses methods the GODAE community is using to test and validate their products and services. Section 8 discusses the user community and implications for the systems and methods being developed within GODAE. The final section provides some brief conclusions.

2. A little history

Scientific observation of the oceans did not begin in earnest till about the nineteenth century; till this time, exploration and expanding ocean trade routes were the primary concern. Advances in communication technology led to the idea of using under-sea cables to connect the American and European continents. This required knowledge of the sea bed and thus led to exploration of the depth of the ocean; until this point, almost all knowledge of the oceans was derived from surface observation. Along with the improvements in knowledge of the depth of the sea, it was discovered that

life did exist at great depth. Scientific cruises for systematic exploration were born. The British *Challenger* expedition from 1872 to 1876 and German exploration on the *Gazelle* from 1874 to 1876 were two of the early successful deep sea expeditions, taking systematic measurements of ocean currents, temperature, chemistry and biology, as well as sampling bottom sediments.

Valuable trading routes had been started on the open seas and travel time was a critical element of commercial success. M.F. Maury, superintendent of the Depot of Charts and Instruments at Washington, D.C., began to collect and collate information on surface currents and weather conditions leading to the publication of *The Physical Geography of the Sea* (Maury 1859), making it one of the first practical applications of ocean science and ocean observations. If a point in time has to be chosen to mark the beginning of operational oceanography, this time is it. Maury led the organization of an international system for regular observation; sailors on all vessels at sea would regularly record certain observations (e.g., sea state, sea surface temperature, weather, etc.) and, in exchange, they would be provided with charts of ocean currents and weather conditions in order to plan their voyage. The legacy of these early efforts can still be appreciated in the GODAE systems of today.

These scientific endeavors marked the start of what Neumann and Pierson (1966) termed the first era of oceanographic research. The three-dimensional structure of the ocean was being observed for the first time. The second era was born out of the realization that the ocean was not stationary and that its circulation could be partly explained by theoretical relationships (e.g., Ekman, 1905). Exploration of the oceans moved into the four-dimensional era; expeditions of the early twentieth century were making more accurate physical and chemical measurements and the station spacing was closer, driven in part by theoretical revelations. While this era probably marked the first awareness of spatial and temporal sampling problems, it was to be many years later before the ramifications of aliasing and poor spatial resolution were to be fully appreciated.

The third era was characterized by significant technological advances, such as the bathythermograph, and by highly organized, intensive oceanographic surveys which sought quasi-synoptic sampling of large regions. This era also marked the introduction of non-ship instrumentation such as drifting and moored buoys. One of the more imaginative innovations of this period was the neutrally buoyant float (Swallow 1955), a technology that lies at the heart of the Argo campaign of today. This period was also marked by significant advances in theory, not the least being the first theoretical explanations of the gyres and intense western boundary current depicted in Maury's chart (e.g., Stommel, 1948; Sverdrup 1947).

The modern era of oceanography has been shaped by at least three factors. First, costs and logistical considerations have driven the development of mooring and autonomous underwater and surface technology. These advances combined with real-time telemetry not only make synoptic observation of the ocean practical, but allow data to be delivered to models quickly.

A second significant factor is satellites. The vastness of the oceans has, and will forever, preclude near-simultaneous sampling of the oceans by conventional, *in situ* instrumentation, even at the surface. Remote sensing offers the promise of ocean data over all regions of the globe, near-simultaneously, though restricted to a surface view.

A third factor is related to both the previous factors - computing. The growth in computational capacity over the last 50 years has been phenomenal. For observationalists, it has revolutionized instrumentation, allowing more detailed and accurate recording and near-instantaneous processing, both on research ships and on moorings and autonomous devices, and in land-based laboratories. Computing power was the key enabling technology for satellites. Computers have revolutionized the capacity of ocean modelers to represent the circulation of the actual ocean. It is this capacity, as much as any other, which has underpinned the evolution of modern oceanography to the point where routine, operational oceanography is feasible and the concept of GODAE, makes sense.

The legacy from ocean research experiments such as the Tropical Ocean Global Atmosphere Experiment (TOGA; McPhaden et al.,1998) and the World Ocean Circulation Experiment (WOCE; e.g., Smith 2001) is also very important. TOGA developed systematic observation and routine prediction of seasonal-to-interannual climate variations (e.g., El Nino) with requirements closely related to those of GODAE and operational oceanography. WOCE introduced many innovations in observation and developed the models and assimilation methods that are the basis for many GODAE systems.

Perspective #1: Scientific and technical advances over the last century, including accrued knowledge of the dynamics and physics of the ocean, provide the basis for developing the systems of GODAE.

3. Lessons from meteorology

At the First GODAE Symposium, Dr. Tim Palmer delivered a lecture "En Route to GODAE: A brief history of NWP" (see www.bom.gov.au/GODAE) and, within that lecture, he cited from Charney

et al. (1969) concerning US participation in the then Global Atmospheric Research Program (GARP): "It is estimated that the data requirements of computer models are met for only 20 per cent of the earth's surface. Vast oceanic regions remain unobserved… the earth-orbiting satellite affords the opportunity of developing an economically feasible global observing capability." Meteorologists were concerned with their ability to observe the relevant atmospheric variables, at all levels and globally, and to have that data available each day for models and forecasts. Moreover, on the basis of progress made with atmospheric models, they wished to test the hypothesis that models and data assimilation could extend useful predictability and provide useful forecasts, at lead times several days ahead of what was possible at that time.

The goals of GARP were effectively (a) deterministic weather forecasting and (b) understanding climate. The First GARP Global Experiment (FGGE) was conceived to address the challenges above and set down several specific goals:

(i) Development of more realistic models for extended range forecasting, general circulation studies, and climate.

(ii) To assess the ultimate limit of predictability of weather systems.

(iii) To develop more powerful methods for assimilation of meteorological observations and, in particular, for using non-synchronous data...

(iv) To design an optimum composite meteorological observing system for routine numerical weather prediction.

Bengtsson (1981) discusses the impact of FGGE on numerical weather prediction, the meteorological counterpart of the systems GODAE is developing. It is clear that significant progress was made against each of the goals of FGGE and that that experiment was critical in the development of modern weather prediction systems. Palmer also showed the evolution of forecast skill since FGGE, around 2 extra days in lead time in the Northern Hemisphere, and over 3 for the Southern Hemisphere. This progress has been made possible by better observations (particularly remote sensing), better models, faster computers, and most importantly, a vastly improved knowledge of the dynamics and physics of the atmosphere. The improved skill however only tells part of the story. The information content of a modern numerical weather prediction system bears little resemblance to its predecessors during FGGE. Regional models are often operating at scales of 5-10 km or better, and these broad measures of skill do not capture the immense value added through finer resolution (indeed, in some cases, the systems are penalized!). Many forecasts systems are also producing more than one forecast (ensembles) so that the users can now apply forecasts with knowledge of the probability of an event occurring. Assimilation systems are

also being extended, for example to consider ozone, air quality and carbon dioxide. Finally, these same systems are being used to produce consistent (re-)analyses of the atmospheric state.

While there are significant differences between the goals of numerical ocean prediction (the GODAE focus) and numerical weather prediction, it is also clear that our community can benefit from the experiences of that community, including their failures. We will discuss objectives and products that closely parallel those discussed here. It is also likely GODAE systems will utilize and/or share a great deal of the infrastructure developed for weather prediction, including observational networks, data and product communication networks, computers and organizational infrastructure.

One difference that is worth considering is that at this time the numerical ocean prediction community does not have the benefit of a dedicated ocean research program. GARP has morphed into the World Climate Research Program, which does consider climate aspects, but its Programmes do not provide the focus that we need now and in the future.

Perspective #2: We have a good model to follow in the development of numerical weather prediction and we can deliver efficiency and effectiveness by partnering and sharing with this community.

4. The concept of and rationale for GODAE

4.1 The vision

The key to harnessing the powerful resources of the ocean and mitigating and/or exploiting its impact on our environment is knowledge - knowledge of the way the ocean has behaved in the past and of how it is likely to behave in the future. Monitoring and forecasting the behavior of the ocean is one of the major challenges for this century, as a prerequisite to sustained development of the marine environment and the implementation of seasonal prediction and climate forecasting systems.

The vision of GODAE is (IGST, 2000):

> "*A global system of observations, communications, modeling and assimilation, that will deliver regular, comprehensive information on the state of the oceans in a way that will promote and engender wide utility and availability of this resource for maximum benefit to society.*"

Regular depictions of the ocean state will be obtained through synthesis of observations with ocean model estimates. The models will allow us to assimilate and integrate complex information in a way that is consistent with our knowledge of ocean dynamics and physics.

Scientifically, in the totality of its complexity, the problem is enormous. Yet, it is evident that most aspects are now tractable. The benefits of assimilation of ocean observations into ocean and climate models has been demonstrated (e.g., Ji et al., 1998; Giraud et al., 1997; Burnett et al., 2002, and papers within that Volume; Wunsch, 2001). A system of ocean data collection and modeling of the ocean that will allow us to follow the state of the ocean routinely seems in the realm of feasibility (see also Smith and Lefebvre, 1997).

4.2 The rationale

GODAE is inspired by both opportunity and need. There is a genuine user demand for ocean products, for a range of time and space scales (e.g., Flemming 2001, Altalo, this Volume). There is also a concern for future ocean research. A capability for providing regular ocean analyses is required as a framework for scientific research and development. In addition, if we are to build a future with a robust, routine, permanent and well-supported network of ocean observations, then a clear and convincing demonstration of the feasibility, practicality and value of maintaining such a network is required.

The opportunities arise because of the development and maturity of remote and direct observing systems, making global real-time observation feasible; the steady advances in scientific knowledge and our ability to model the global ocean and assimilate data at fine space and time scales; the genuine enthusiasm of the ocean community to promote and implement integrated global observing systems; and the critical advances provided by research programs (see Section 2).

The underlying rationale for the organization of this activity as an international experiment is that achieving the GODAE vision will not happen serendipitously and that the needed capacity will not be realized without a concerted effort to ensure, first, proper integration of the components and, second, the commitment to proving value and viability.

4.3 The approach

Smith (2000) and IGST (2000) introduce the objectives and scope of GODAE and the reader is referred to those publications and the GODAE web site for details.

One premise is that GODAE is not just concerned with prediction in the traditional sense (looking forward in time), but prediction in its most general form, where information is extrapolated and interpolated in space and time, and between fields (Figure 1). The objectives intentionally imply a broad scope in the belief that wide utilization and exploitation of products are essential for cost-efficiency and relevance to society.

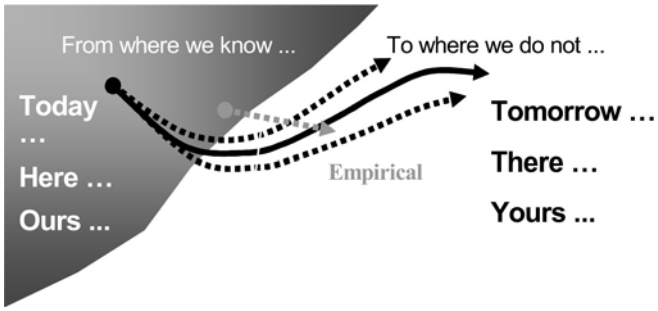

Figure 1. Schematic of the processes used to exploit data. In some cases we use linear, perhaps empirical relationships to relate the current state to, say, a likely future state. In other cases forecasts are produced based on current data ("today"), perhaps at a specific location ("here"), and perhaps for a subset of the total variable space ("ours"), in order to forecast the state in the future ("tomorrow"), at some remote location ("there") or for some variables that are not part of the observables ("yours"). The process involves extrapolation (e.g., a forecast), interpolation (e.g., discrete points to a grid) and interpretation (e.g., inferring winds from sea surface topography).

The strategy for the development of these products is built on the concept of a GODAE "Common" which is shared by all GODAE Partners responsible for realizing the goals and objectives of GODAE. The GODAE Common concept is essential for GODAE, and must also be transported into the "operational" environment, for example through data policies and scientific cooperation.

5. Building the systems

The essential building blocks of GODAE are observations, models and estimation tools (Figure 2). In the GODAE context, these elements are inextricably linked, with obvious two-way interdependencies. The generation of globally consistent fields of ocean temperature, salinity and velocity components through the synthesis of multivariate satellite and in situ data streams into ocean models is a central theme of GODAE.

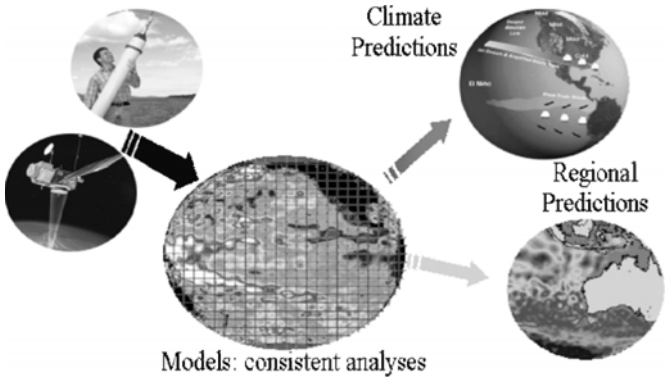

Figure 2. Illustration of the process for taking in situ and remotely sensed data (left) through a model-based assimilation system to produce a self-consistent analysis, which is then used to produce products such as a climate or regional/coastal forecast.

The scope and international nature of GODAE requires distributed data assembly and serving, a multiplicity of assimilation products, distributed product serving and archiving, and a multiplicity of application centers (Figure 3).

5.1 GODAE observational and data needs

Data needed for GODAE model/assimilation systems can be separated into four main classes: atmospheric forcing (wind stress, wind speed, air temperature, specific humidity, precipitation) and sea-ice, data for assimilation (e.g., altimetry, Argo, SST), validation data (e.g., hydrography) and ancillary data (climatologies, bathymetry). Note, however, that the separation into data types is neither definitive nor unique (e.g., forcing data can be used as one of the controls on the assimilation process).

Koblinsky and Smith (2001) discusses the data system and other papers of this Volume discuss details and issues that are of specific concern for GODAE. Remote sensing data is naturally central to the success of GODAE and GODAE has placed particular emphasis on surface topography, surface wind and sea surface temperature data.

GODAE itself has taken two specific initiatives to address specific gaps. In the early stages of GODAE it became clear that the in situ coverage was inadequate for both climate and ocean assimilation purposes. The Argo Pilot Project (Argo Science Team, 1998) was established soon after GODAE was born, and has realized a near-revolution in our capability to observe the ocean in real-time (see papers by Send and by Pouliquen, this Volume). A

second Pilot Project arose in a somewhat unexpected area, sea surface temperature; a field that the community had believed was being estimated well. The GODAE High-Resolution SST Pilot Project (see chapter by I. Robinson in this volume) aims to deliver integrated, high-resolution products, derived from a range of different, but complementary observing systems, that properly respect our understanding of the near-surface temperature structure (e.g., the skin effect) and addresses issues such as the diurnal cycle.

Various data servers will be responsible for maintaining and monitoring the data flow to assimilation groups and to those undertaking validation/evaluations. The GODAE Monterey server and the CORIOLIS Centre (see chapter by S. Pouliquen in this volume) are two examples of this important functionality. One of the tasks is to link the server functions together so that the data users will have a consistent and transparent interface to the variety of data that are available. One of the challenges facing GODAE (and others) is the establishment of adequate metadata to facilitate data tracking, intercomparisons, and distribution of data which may undergo revision through various quality control procedures.

Perspective #3: The real impact of GODAE will come through its ability to bring its complex data and information to applications and users.

5.2 Models and data assimilation

Because of the irregular and incomplete nature of the datasets relative to the scales of interest, a considerable burden in ocean state estimation and forecasting is placed not only on the assimilation components but also on the model. The model provides a capacity to extrapolate information, enabling past data to be used for present analyses, and present data to be used as a basis for predictions of the ocean state at future times (forecasts). Other papers in this Volume discuss approaches to modeling and data assimilation and some of the issues faced by the GODAE community.

Most of the target applications require good representation of, at least, temperature and velocity components and sea level. High resolution operational oceanography requires accurate depiction of mesoscale ocean features such as eddies and the meandering of currents and fronts and of upper ocean structure. Coastal applications require accurate sea level and cross-shelf transport estimates. Seasonal-to-interannual climate forecasts require a good representation of the upper ocean temperature and salinity fields. Decadal climate monitoring and research requires attention to the thermohaline circulation, among other things. Biogeochemical applications

require attention to the upper ocean physical parameterizations and the vertical transports (upwelling). All require considerable computational resources for global simulation and so rely on advanced software developments to take advantage of state of the art computer technology.

Perspective #4: Global high-resolution ocean model assimilation systems are the main focus of GODAE. Regional prototypes have proved critical for development and for regional applications. Sector-specific systems (e.g., for global climate estimates) are also an important aspect. Reanalyses are an important strategy.

An outstanding issue for GODAE, with implications for assimilation and prediction, is the degree to which the key fields mentioned above are predictable and, secondly, the extent to which provided fields (boundary conditions, initial conditions, other inputs) in effect enhance predictability (skill) to the target systems. The applied nature of GODAE only allows it to address these issues in passing, so again it is important that supporting research is fostered to test and understand all aspects of predictability. Note that in the context of GODAE, such research applies not only to temporal predictions (forecasts) but to the more general context (see Fig. 1).

The use of a variety of approaches to modeling and assimilation is regarded as a strength in the strategy of GODAE. Within a framework of intercomparison and progressive evaluations, the diversity of approaches can be used to quantify uncertainties and test reliability of ocean state estimates and initial conditions and forecasts.

Perspective #5: The oceans are predictable ... but when and where, and for how long? What are the dependencies and limitations? Observations? Representation of ocean dynamics and physics? Assimilation? Parameterizations? GODAE will provide only the first installment in our quest to address these issues.

6. The utility of GODAE outputs

The key outcome will be significant improvement in the detail, quality, timeliness, availability and usefulness of ocean analysis and prediction products. The reader is also referred to the GODAE Implementation Plan on http://www.bom.gov.au/GODAE/ for detail of activities by different groups and a more complete description of applications.

Coherent, organized data sets: GODAE aims to develop more coherent, better organized, more widely available and more useful data sets. Such outputs will be realized through:

(a) More effective assembly and availability. From the outset, the GODAE participants recognized that they must work to build coherent data streams that remove the mysteries associated with specific measurement techniques and the confounding problems associated with merging data of different types and formats.

(b) Improving data utility. GODAE places a high-premium on the wide use of data and products to ensure observing efforts realize their full potential in operational systems.

(c) Improving data quality. A sub-project has been launched to coordinate and standardize the GODAE approach (see www.bom.gov.au/GODAE/). As operational oceanography systems mature, they will provide routine, regular and immediate testing of data and thus add value to data sets.

These outputs depend upon adequate devotion of effort to all stages of data handling. Efficiency is realized through rationalization and streamlining of the procedures.

Reanalyses and synoptic ocean analyses: GODAE is most readily associated with products of ocean model assimilation, usually in the form of space-time gridded fields. GODAE includes the continual revision and improvement of analyses, either through re-analysis or through intercomparison activities. The great worth of reanalyzes lies in the fact that they provide dynamically and physically consistent estimates over a period, in a form that is readily used by research, but also by the broader marine community who have interests (dependencies) on knowledge of ocean variability and predictability.

Short-range ocean forecasts: GODAE will have a leading role in short-range ocean prediction and a supporting role in coupled air-sea prediction and surface wind waves via the provision of related ocean fields to application centers. While we might argue that 4-dimensional assimilation is at its roots simply a means for projecting and synthesizing data in space and time, the capacity to extend this projection (initial condition) forward in time to produce forecasts gives the system special value.

Climate applications: The most common application for the GODAE ocean state estimate is as an initial condition for a coupled model forecast (e.g., Ji et al., 1998). One of the primary issues to be faced by this community is how best to use the state estimate; for example, the nature of

the problem might favor an ensemble of initial conditions rather than a single, high-fidelity product. For decadal variability and longer-term change, GODAE focuses on the provision of consistent, high-quality analyses and re-analyses of the ocean.

Coastal applications: Coastal applications will use GODAE ocean state estimates as boundary conditions for coastal/ littoral zone hindcasts and forecasts and analyses (Figure 3).

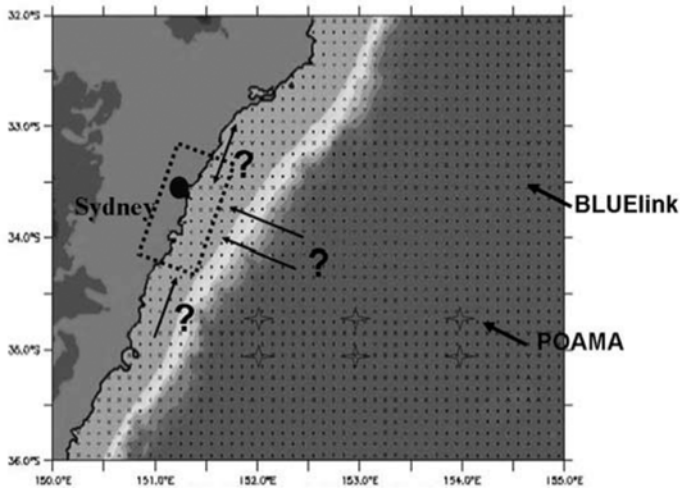

Figure 3. A schematic of the hypothetical nesting of a coastal application near Sydney within the BLUElink ocean forecasting system (indicated by fine dots; see Schiller and Smith, this volume) or a coarser seasonal prediction system "POAMA" (Wang et al., 2001).

It is not yet clear what the accuracy requirements are. Development of GODAE products for these applications will represent a significant research effort within the community. Issues of nesting of models of different resolution, the importance of regional wave and surge models, consistency in bathymetry, forcing, boundary configurations, and input to ecosystem models are critical elements for collaboration. The end users will include regional/local governments responsible for coastal management, as well as coastal industries such as fishing and recreation.

The GODAE approach provides efficiency because the systems can provide information/boundary conditions to multiple users, in a variety of ways. In some prototypes the regional/local modeling is in-built to the modeling system. In other cases the coastal modeling is part of the same project so the interface issues are being solved as part of the project. In yet other cases, boundary conditions are being provided to third parties who

may have knowledge of the source model (and vice versa) but otherwise are running completely independent systems/applications.

There are a large number of issues that impact the utility of GODAE products. We are for the present slave to the errors of our atmospheric partners. Accurate ocean surface current predictions and simulations may prove as elusive as atmospheric fluxes and winds, and we do not yet fully understand the extent to which subsurface currents can be predicted. We do not yet know how well can we "predict" boundary conditions for coastal applications, and how much it matters when we get it "right".

7. How to measure success?

The demonstration of the quality and value of GODAE products for research and operational applications is the central objective of the experiment. We need to set standards for data and products that are testable and defensible. There are two levels of evaluation criteria. Internal (technical) evaluation criteria should measure the performance of the components and functions, effectively within the GODAE Common. External measures and feedback will come from GODAE users and applications.

The scientific rationale for, and a more detailed description of the GODAE metrics are given in Le Provost et al. (2002). The internal metrics will include measures of consistency, quality and performance/efficiency. The so- external evaluation criteria include (a) the impact of GODAE products for the different applications, (b) the utility of GODAE products for the research community, (c) the number of users and their level of satisfaction, (d) the extent of resultant innovation, (e) the utility for observing system evaluation and design, and (f) the extent of uptake by value-adders and other specific users.

Perspective #6: Implementing a rigorous system of internal and external tests and intercomparisons in order to evaluate systems and to set standards is a key task. We need to foster the development of international infrastructure, and national infrastructure, to support and monitor the performance and effectiveness of systems.

8. Users and benefits

At least four types of relationships with end users have been identified.

1. Direct to the Public. This suits the ad hoc and occasional user whose needs are satisfied by directly utilizing the products and services

emanating from GODAE Centers or application centers. There is no intermediary or down-stream value-adding.
2. Via middle-users/value-adders. In this case specialists, varying from private ocean enterprises to sector-specific groups take the output and, perhaps after blending it with other information and/or rendering it in a form that is more useful and "consumable", provide it to their clients. The middle-users have expertise from both the provider and client sides and value-adding is through a partnership.
3. Direct to specific users/sectors. In some cases, specific users may be able to directly exploit GODAE products. The relationship may be commercial. Value-adding is entirely on the user side.
4. Capacity building and education. Here the users do not have access to sophisticated systems or technology and support is needed for the transfer of knowledge.

In simple terms, GODAE faces a challenge of determining the capabilities and product availability relevant to the areas and to build a consolidated view on the requirements for GODAE. Such requirements may not simply be for a particular product but may also involve the timeliness and form of the data and products, or a requirement for GODAE to include certain information in its inputs.

Perspective #7: Determining the utility of products for different users and sectors of the ocean community is the major challenge at this time.

9. Conclusions

The legacy of GODAE will be held in the sustained ocean observing system and in the global and regional operational oceanographic systems that are being developed and tested now and that we envisage being maintained by several nations. GODAE has achieved a level of investment that exceeded its expectations but such investments will only be sustained through proving the utility and use of GODAE deliverables and offset by tangible economic and social returns and outcomes

Like weather prediction, GODAE contains a balance between the practical and applied and the long-term strategic goals. The former represents a commitment to develop practical and useful applications and, through linkages with those able to exploit such products, to promote the development of a rich array of specialist, value-added uses. The latter represents a commitment to provide an appropriate basis for planning and

building future research endeavors and for ensuring that the global ocean observing system, once established, remains effective, efficient and scientifically defensible in the future.

References

Argo Science Team, 1998: On the Design and Implementation of Argo: An Initial plan for a Global Array of profiling Floats. International. CLIVAR Project Office Report Number 2 , GODAE Report No. 5, GODAE International Project Office, Melbourne Australia.

Bengtsson, L., 1981: The impact on FGGE on global medium range forecasts. International Conference on early Results of FGGE and Large-scale Aspects of its Monsoon Experiments. Tallahassee, Jan 1981, ICSU/WMO, Geneva, 1-26.

Burnett, W., J. Harding and G. Heburn, 2002: Overview of operational ocean forecasting in the US Navy: Past, present and future. Oceanography, 15, 4-12.

Charney, J. 1969: Chapter on predictability in Plan for U.S. Participation in the Global Atmospheric Research Program. A report of the U.S. Committee for the Global Atmospheric Research Program to the National Academy of Sciences, Washington, D.C.

Ekman V. W., 1905: On the influence of the earth's rotation on ocean currents. Arkiv. For Matematik, Astronomi och Fysik, 2, 11, 52 pp

Flemming, N., 2001: Dividends from investing in ocean observations: A European perspective. In Ocean Observations for the 21st Century, C. Koblinsky and N Smith (Eds), published by the GODAE Office/BoM, Melbourne.

Giraud, S., S. Baudel, E. Dombrowsky, and P. Bahurel, 1997: The SOPRANE project : Real-time monitoring of the North-East Atlantic – ocean circulation nowcast-forecast for oceanographic scientific campaigns. In Monitoring the oceans in the 2000s: An integrated approach, International Symposium, Biarritz, October 15-17.

International GODAE Steering Team, 2000: The Global Ocean Data Assimilation Experiment Strategic Plan. GODAE Report No. 6, December, 2000.

Ji, M., D.W. Behringer, and A. Leetmaa, 1998: An Improved Coupled Model for ENSO Prediction and Implications for Ocean Initialization. Part II: The coupled model. Mon. Wea. Rev., 126, 1022-1034.

Koblinsky, C. and N Smith, 2001 (Eds): Ocean Observations for the 21st Century, published by the GODAE Office/BoM, Melbourne.

Le Provost, C. et al., 2004: Inter-comparison projects on the North Atlantic and the Med Sea. Version 5, May 3rd 2004, MERSEA Strand-1 - WP4 intercomparison project (available at http://www.mersea.eu.org).

Maury, M.F., 1859: The Physical Geography of the Sea (New York: Harper and Bros.).

McPhaden, M.J., A.J. Busalacchi, R. Cheney, J.-R. Donguy, K.S. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G.T. Mitchum, P.P. Niiler, J. Picaut, R.W. Reynolds, N. Smith, and K. Takeuchi, 1998: The Tropical Ocean Global Atmosphere (TOGA) observing system: a decade of progress. J. Geophys. Res., 103 (C7), 14169-14240.

Neumann, G., and W.J. Pierson, 1966: Principles of Physical Oceanography. Prentice hall, Englewood-Cliffs, N.J.

Peloquin, R., 1992: The Navy ocean modeling and prediction program. Oceanography, 5, 4-8.

Smith, N.R, 2000.: The Global Ocean Data assimilation Experiment, Adv. Space Res., 25, 1089-1098.

Smith, N.R., 2001: Ocean and Climate Prediction – the WOCE Legacy. Chapter 7.4 of "Ocean Circulation and Climate" (Eds. G Siedler, J A Church and J. Gould) , Academic Press, pp 585-602.

Smith, N. and M. Lefebvre, 1997: The Global Ocean Data Assimilation Experiment (GODAE). In "Monitoring the oceans in the 2000s : an integrated approach". International Symposium, Biarritz, October 15-17, 1997.

Stommel, H., 1948: The westward intensification of wind-driven ocean currents. Trans. Amer. Geophys. Union, 29, 202-206.

Sverdrup, H.U., 1947: Wind-driven currents in a baroclinic ocean; with application to the equatorial currents of the eastern Pacific. Proc. Natl. Acad. Sci. USA, 33, 318-326.

Swallow, J.C., 1955: A neutral-buoyancy float for measuring deep currents. Deep-Sea Res., 3, 74-81

Wang, G., R. Kleeman, N. Smith, and F. Tseitkin, 2002: The BMRC coupled general circulation model ENSO forecast system. Mon. Wea. Rev., 130, 975-991.

Wunsch, C., 2001: Global problems and global observations. Chapter 2.1 of "Ocean Circulation and Climate" (Eds. G Siedler, J A Church and J. Gould), Academic Press, pp 47-58.

Chapter 2

SOME OCEAN MODEL FUNDAMENTALS

Stephen M. Griffies
NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA

Abstract The purpose of these lectures is to present elements of the equations and algorithms used in numerical models of the large-scale ocean circulation. Such models generally integrate the ocean's *primitive equations*, which are based on Newton's Laws applied to a continuum fluid under hydrostatic balance in a spherical geometry, along with linear irreversible thermodynamics and subgrid scale (SGS) parameterizations. During formulations of both the kinematics and dynamics, we highlight issues related to the use of a generalized vertical coordinate. The vertical coordinate is arguably the most critical element determining how a model is designed and applications to which a model is of use.

Keywords: Ocean modelling, parameterization, vertical coordinate.

1. Concepts, themes, and questions

Numerical ocean models are computational tools used to understand and predict aspects of the ocean. They are a repository for our best ocean theories, and they provide an essential means to probe a mathematical representation of this very rich and complex geophysical system. That is, models provide an experimental apparatus for the scientific rationalization of ocean phenomena. Indeed, during the past decade, large-scale models have become *the* experimental tool of choice for many oceanographers and climate scientists. The reason for this state of affairs is largely due to improved understanding of both the ocean and ocean models, as well as increased computer power allowing for increasingly realistic representations of ocean fluid dynamics. Without computer models, our ability to develop a robust and testable intellectual basis for ocean and climate dynamics would be severely handicapped.

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 19-73.

The remainder of this section introduces some basic concepts, themes, and questions, some of which are revisited later in the lectures. We present some philosophical notions which motivate a focus on fundamental concepts and notions when designing, constructing, and analyzing ocean models.

1.1 Model environments

The field of ocean model design is presently undergoing a rapid growth phase. It is arguable that the field has reached *adolescence*, with further maturation likely taking another 10-20 years as we take the models to a new level of integrity and innovation. Many applications drive this evolution, such as studies of climate change, operational oceanography, and ultra-refined resolution process studies.

One goal of many developers is that the next decade of model evolution will lead to a reduction in code distinctions which presently hinder the ability of modelers to interchange algorithms, make it difficult to directly compare and reproduce simulations using different codes, and increase the burdens of model maintenance in a world of increasingly complex computational platforms and diverse applications. Notably, the distinctions will not be removed by all modelers using a common algorithm. Such is unreasonable and unwarranted since different scientific problems call for different algorithmic tools. Instead, distinctions may be removed by the development of new codes with general algorithmic structures flexible enough to encompass multiple vertical coordinates, different horizontal grids, various subgrid scale (SGS) parameterizations, and alternate numerical methods.

The word *environment* has recently been proposed to describe these highly flexible and general codes. As yet, no model environment exists to satisfy the needs and desires of most modelers. Yet some models are moving in this direction by providing the ability to choose more than one vertical coordinate. This is a critical first step due to the central importance of vertical coordinates. The present set of lectures formulates the fundamental equations using generalized vertical coordinates, and these equations form the basis for generalized vertical coordinate ocean models. Ideally, the advent of general model environments will allow scientists to use the same code, even though they may use different vertical coordinates, horizontal grids, numerical methods, etc.

Many of the ideas presented here are an outgrowth of research and development with the Modular Ocean Model of Griffies et al., 2004, as well as the MITgcm (Marshall et al., 1997, Adcroft and Campin, 2004). The MITgcm provides for a number of depth-based and pressure-based

vertical coordinates. Another approach, starting from an isopycnal layered model, has been taken by the Hybrid Coordinate Ocean Model (HYCOM) of Bleck, 2002. HYCOM is arguably the most mature of the generalized vertical coordinate models.

From an abstract perspective, it is a minor point that different modelers use the same code, since in principle all that matters should be the continuum equations which are discretized. This perspective has, unfortunately, not been realized in practice. Differences in fundamentals of the formulation and/or numerical methods often serve to make the simulations quite distinct, even when in principle they should be nearly identical. Details do matter, especially when considering long time scale climate studies where small differences have years to magnify.

An argument against merging model development efforts is that there is creative strength in diversity, and so there should remain many ocean codes. A middle ground is argued here, whereby we maintain the framework for independent creative work and innovation, yet little effort is wasted developing redundant software and/or trying to compare different model outputs using disparate conventions. To further emphasize this point, we stress that the problems of ocean climate and operational oceanography are vast and complex, thus requiring tremendous human and computational resources. This situation calls for merging certain efforts to optimize available resources. Furthermore, linking modelers together to use a reduced set of code environments does not squelch creativity nor does it lead to less diversity in algorithmic approaches. Instead, environments ideally can provide modelers with common starting points from which to investigate different methodologies, parameterizations, and the like.

The proposal for model environments is therefore analogous to use of a few spoken/written languages (e.g., english, french) to communicate and formulate arguments, or a few computer languages (e.g., Fortran, C++) to translate numerical equations into computer code. Focusing on a few ocean model environments, rather than many ocean models, can lead to enhanced collaboration by removing awkward and frustrating barriers that exist between the presently wide suite of model codes. Ultimately, such will (it is hoped!) lead to better and more reproducible simulations, thus facilitating the maturation of ocean modelling into a more robust and respectable scientific discipline.

1.2 Some fundamental questions

It is possible to categorize nearly every question about ocean modelling into three classes.

1 Questions of model fundamentals, such as questions raised in this section.

2 Questions of boundary fluxes/forcing, from either the surface air-sea, river-sea, and ice-sea interactions, or forcing from the solid earth boundary. The lectures in this volume from Bill Large touch upon many of the surface flux issues.

3 Questions of analysis, such as how to rationalize the simulation to enhance ones ability to understand, communicate, and conceptualize.

If we ask questions about physical, mathematical, or numerical aspects of an ocean model, then we ask questions about ocean model fundamentals. The subject deals with elements of computational fluid mechanics, geophysical fluid mechanics, oceanography (descriptive and dynamic), and statistical physics. Given the wide scope of the subject, even a monograph such as Griffies, 2004 can only provide partial coverage. We consider even less in these lectures. The hope is that the material will introduce the reader to methods and ideas serving as a foundation for further study.

For the remainder of this section, we summarize a few of the many fundamental questions that designers and users often ask about ocean models. Some of the questions are briefly answered, yet some remain unaswered because they remain part of present day research. It is notable that model users, especially students learning how to use a model, often assume that someone else (e.g., their adviser, the author of a research article, or the author of a book) has devoted a nontrivial level of thought to answering many of the following questions. This is, unfortunately, often an incorrect assumption. The field of ocean modelling is not mature, and there are nearly as many outstanding questions as there are model developers and users. Such hopefully will provide motivation to the student to learn some fundamentals in order to help the field evolve.

Perhaps the most basic question to ask about an ocean model concerns the continuum equations that the model aims to discretize.

- Should the model be based on the non-hydrostatic equations, as relevant for simulations at spatial scales less than 1km, or is the hydrostatic approximation sufficient? Global climate models have all used the hydrostatic approximation, although the model of Marshall et al., 1997 provides an option for using either. Perhaps in 10-20 years, computational power will be sufficient to allow fully non-hydrostatic global climate simulations. Will the simulations

change drastically at scales larger than 1km, or do the hydrostatic models parameterize non-hydrostatic processes sufficiently well for most applications at these scales? Note that the accuracy of the hydrostatic approximation scales as the squared flow aspect ratio (ratio of vertical to horizontal length scales). Atmospheric modelers believe their simulations will be far more realistic with an explicit representation of non-hydrostatic dynamics, such as convection and cloud boundary layer processes. In contrast, it remains unclear how necessary non-hydrostatic simulations are for global ocean climate. Perhaps it will require plenty of experience running non-hydrostatic global models before we have unambiguous answers.

- Should the kinematics be based on incompressible volume conserving fluid parcels, as commonly assumed for ocean models using the Boussinesq approximation, or should the more accurate mass conserving kinematics of the non-Boussinesq fluid be used, as commonly assumed for the more compressible atmosphere. Ocean model designers are moving away from the Boussinesq approximation since only a mass conserving fluid can directly represent sea level changes due to steric effects (see Section 3.4.3 of Griffies, 2004), and because it is simple to use mass conserving kinematics by exploiting the isomorphisms between depth and pressure discussed by DeSzoeke and Samelson, 2002, Marshall et al., 2003, and Losch et al., 2004.

- Can the upper ocean surface be fixed in time with a rigid lid, as proposed decades ago by Bryan, 1969 and used for many years, or should it be allowed to fluctuate with a more realistic free surface so to provide a means to pass fresh water across the ocean surface and to represent tidal fluctuations? Most models today employ a free surface in order to remove the often unacceptable restrictions of the rigid lid. Additionally, many free surface methods remove elliptic problems from hydrostatic models. The absence of elliptic problems from the free surface models greatly enhances their computational efficiency on parallel computers (Griffies et al., 2001).

- Should tracers, such as salt, be passed across the ocean surface via virtual tracer fluxes, as required for rigid lid models, or should the model employ real water fluxes thus allowing for a natural dilution and concentration of tracer upon precipitation and evaporation, respectively? As discussed more fully in Section 3.6, the advent of free surface methods allows for modelers to jettison the unphysical

virtual tracer methods of the rigid lid. Nonetheless, virtual tracer fluxes remain one of the unnecessary legacy approximations plaguing some modern ocean models using free surface methods. The potential problems with virtual tracer fluxes are enhanced as the time scales of the integration go to the decade to century climate scale.

- What is the desired manner to write the discrete momentum equation: advective, as commonly done in B-grid models, or vector invariant, as commonly in C-grid models? The answer to this question may be based more on subjective notions of elegance than clear numerical advantage.

- How accurate should the thermodynamics be, such as the equation of state and the model's "heat" tracer? The work of McDougall and collaborators provides some guidance on these questions (McDougall, 2003, McDougall et al., 2003, Jackett et al., 2004). How important is it to get these things accurate? The perspective taken here is that it is useful to be more accurate and flexible with present day ocean climate models, since the temperature and salinity range over which they are used is quite wide, thus making the older approximations less valid. Additionally, many of the more accurate approaches have been refined to reduce their costs, thus making their use nearly painless.

After deciding on a set of model equations, further questions arise concerning how to cast the continuum partial differential equations onto a finite grid. First, we ask questions about the vertical coordinates. Which one to use?

- Geopotential (z-coordinate): This coordinate is natural for Boussinesq or volume conserving kinematics and is most commonly used in present-day global ocean climate models.

- Pressure: This coordinate is natural for non-Boussinesq or mass conserving kinematics and is commonly used in atmospheric models. As mentioned earlier, the isomorphism between pressure and depth allow for a straightforward transformation of depth coordinates to pressure coordinates, thus removing the Boussinesq approximation from having any practical basis. We return to this point in Section 6.

- Terrain following *sigma* coordinates: This coordinate is commonly used for coastal and estuarine models, with some recent efforts aimed as using it for global modelling (Diansky et al., 2002).

- Potential density or *isopycnal* coordinates: This coordinate is commonly used for idealized adiabatic simulations, with increasing use for operational and global climate simulations, especially when combined with pressure coordinates for the upper ocean in a hybrid context.

- Generalized hybrid vertical coordinates: Models formulated for general vertical coordinates allow for different vertical coordinates depending on the model application and fluid regime. Models with this facility provide an area of focus for the next generation of ocean models.

What about the horizontal grid? Although horizontal grids do not greatly determine the manner that many physical processes are represented or parameterized, they greatly influence on the representation of the solid-earth boundary, and affect details of how numerical schemes are implemented.

- Should we cast the model variables on one of the traditional A through E grids of Arakawa and Lamb, 1977? Which one? The B and C grids are the most common in ocean and atmospheric modelling. Why? Section 3.2 of Griffies et al., 2000a provides some discussion of this question along with references.

- What about spectral methods commonly used in atmospheric models? Can they be used accurately and effectively within the complex geometry of an ocean basin? Haidvogel and Beckmann, 1999 present a summary of these methods with application to the ocean. Typically, spectral methods have not been useful in the horizontal with realistically complex land-sea boundaries, nor in the vertical with realistically sharp pycnoclines. The reason is that a spectral representation of such strong gradients in the ocean can lead to unacceptable Gibbs ripples and unphysically large levels of spurious convective mixing.

- Should the horizontal grid cells be arranged according to spherical coordinates, even when doing so introduces a pesky coordinate singularity at the North Pole? What about generalized orthogonal coordinates such as a bipolar Arctic coupled to a spherical region south of the Arctic (Figure 1)? Such grids are very common today in global modelling, and their use is straightforward in practice since they retain the regular rectangular logic assumed by spherical coordinate models. Or what about strongly curved grid lines that contour the coast, yet remain locally orthogonal? Haidvogel and

Beckmann, 1999 provide some discussion of these grids and their uses.

- What about nested regions of refined resolution where it is critical to explicitly resolve certain flow and/or boundary features? Blayo at this school (see also Blayo and Debreu, 1999) illustrates the potentials for this approach. Can it be successfully employed for long term global climate simulations? What about coastal impacts of climate change? These are important questions at the forefront of ocean climate and regional modelling.

- Can a non-rectangular mesh, such as a cubed sphere, be successfully used to replace all coordinate singularities with milder singularities that allow for both atmosphere and ocean models to jettison polar filtering?[1] The work of Marshall et al., 2003 provide a compelling case for this approach, whereby both the ocean and atmosphere use the same grid and same dynamical core. Figure 2 provides a schematic of a cubed-sphere tiling of the sphere.

- What about icosahedrons, or spherical geodesics as invented by Buckminster Fuller? These grids tile the sphere in a nearly isotropic manner. Work at Colorado State University by David Randall and collaborators has shown some promise for this approach in the atmosphere and ocean.

- What about finite element or triangular meshes popular in engineering, tidal, and coastal applications? These meshes more accurately represent the solid earth boundary. Or what about time dependent adaptive approaches, whereby the grid is refined according to the time dependent flow regimes? Both methods have traditionally failed to perform well for realistic ocean climate simulations due to problems representing stratified and rotating fluids. However, as reported in this volume by Jens Schröter, some important and promising advances have been made by researchers at the University of Reading and Imperial College, both in England, as well as the Alfred-Wegener Institute in Germany. Their efforts have taken strides in overcoming some of the fundamental problems. If this area of research and development is given time to come to fruition, then perhaps in 10 years we will see finite ele-

[1] Polar filtering is a method to reduce the spatial scales of the simulation as one approaches the coordinate singularity at the North Pole. Many computational and numerical problems have been encountered with this approach.

ments commonly used for regional and global models. Such could represent a major advance in ocean modelling.

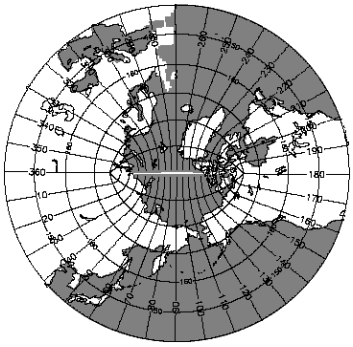

Figure 1. Illustration of the bipolar Arctic as prescribed by Murray, 1996 (see his Figure 7) and realized in the global model discussed in Griffies et al., 2005. A similar grid has also been proposed by Madec and Imbard, 1996. Shown here are grid lines which are labeled with the integers for the grid points. The grid has 360 points in the generalized longitude direction, and 200 points in the generalized latitude direction. This, or similar, bipolar Arctic grids are commonly used in global ocean modelling to overcome problems with the spherical coordinate singularity at the North Pole. Note that the cut across the Arctic is a limitation of the graphics, and does not represent a land-sea boundary in the model domain.

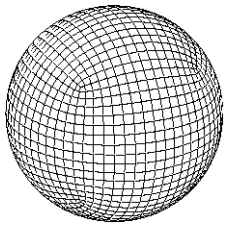

Figure 2. Cubed sphere tiling of the sphere. Note the singularities at the cube corners are much milder than a spherical coordinate singularity found with spherical grids at the poles. The cubed sphere tiling has been implemented in the MITgcm for both the atmosphere and ocean model components. This figure was kindly provided by Alistair Adcroft, a developer of the MITgcm.

What processes are represented explicitly, and what are the important ones to parameterize? This is one of the most critical and difficult questions of ocean model design and use. The lectures by Anne Marie Treguier from this school summarizes many of the issues. She notes that the choice of model resolution and parameterization prejudices the simulation so much so that they effectively determine the "ocean" to be simulated. Discussions in Chassignet and Verron, 1998 thoroughly survey various aspects of the parameterization problem. This book is from a 1998 school on ocean modelling and parameterization. Many of the issues raised there are still unresolved today. Finally, Griffies, 2004 has much to say about some of the common parameterizations used in ocean climate models.

Numerical methods are necessary to transform the continuum equations into accurate and efficient discrete equations for stepping the ocean forward in time. There are many methods of use for doing this task.

- Should they be based on finite volume methods? Such methods are becoming more common in ocean modelling. They provide the numericist with a useful means to take the continuum equations and cast them onto a finite grid.

- What sorts of time stepping schemes are appropriate, and what properties are essential to maintain? Will the ubiquitous leap-frog methods[2] be supplanted by methods that avoid the problematic time splitting mode? Chapter 12 of Griffies, 2004 provides a discussion of these points, and argues for the use of a time staggered method, similar to that discussed by Adcroft and Campin, 2004 and used in the Hallberg Isopycnal Model (Hallberg, 1997) and Modular Ocean Model version 4 (Griffies et al., 2004).

- Should the numerical equations maintain a discrete analog to conservation of energy, tracer, potential vorticity, and potential enstrophy satisfied by the ideal continuum equations? For long term climate simulations, tracer conservation is critical. What about the other conserved quantities?

- What are the essential features needed for the numerical tracer advection operator? Should it maintain positivity of the tracer field? Can such advection operators, which are nonlinear, be easily realized in their adjoint form as required for 4D variational

[2]As noted in Griffies et al., 2000a, the majority of ocean models supported for large-scale oceanography continue to use the leap-frog discretization of the time tendency.

assimilation (see the lectures at this school from Jens Schröter as well as Thuburn and Haine, 2001).

- How should the model treat the Coriolis force? On the B-grid, it is common to do so implicitly or semi-implicitly in time, but this method is not available on the C-grid since the velocity components are not coincident in space. Also, the C-grid spatial averaging of the Coriolis force can lead to problematical null modes (Adcroft et al., 1999).

- What about the pressure gradient calculation? We return to this question in Section 5, where comments are made regarding the difficulties of computing the pressure gradient.

1.3 Two themes

There are two themes emphasized in these lectures.

- How the vertical coordinate is treated is the most fundamental element of an ocean model design.

- The development of ocean model algorithms should be based on rational formulations starting from fundamental principles.

The first theme concerns the central importance of vertical coordinates in ocean model design. Their importance stems from the large distinctions at present between algorithms in models with differing vertical coordinates. Further differences arise in analysis techniques. These fundamental and pervasive distinctions have led to disparate research and development communities oriented around models of a particular class of vertical coordinate. One purpose of these lectures is to describe methods whereby these distinctions at the formulation stage are minimized, thus in principle facilitating the design of a single code capable of employing many vertical coordinates.

The second theme is a "motherhood" statement. What scientist or engineer would disagree? Nonetheless, it remains nontrivial to satisfy for three reasons. First, there are many important elements of the ocean that we do not understand. This ignorance hinders our ability to prescribe rational forms for the very important SGS operators. Second, some approximations (e.g., Boussinesq approximation, rigid lid approximation, virtual tracer fluxes), made years ago for good reasons then, often remain in use today yet need not be made with our present-day modelling capabilities and requirements. These *legacy approximations* often compromise a model's ability to realistically simulate certain aspects of the ocean and/or its interactions with other components of the

climate system. Third, developers are commonly under intense time pressures to "get the model running." These pressures often prompt *ad hoc* measures which, unfortunately, tend to stay around far longer than originally intended.

2. Kinematics of flow through a surface

In our presentation of ocean model fundamentals, we find it useful to start with a discussion of fluid kinematics. Kinematics is that area of mechanics concerned with the intrinsic properties of motion, independent of the dynamical laws governing the motion. In particular, we establish expressions for the transport of fluid through a specified surface. The specification of such transport arises in many areas of oceanography and ocean model design.

There are three surfaces of special interest in this section.

- The lower ocean surface which occurs at the time independent solid earth boundary. This surface is commonly assumed to be impenetrable to fluid.[3] The expression for fluid transport at the lower surface leads to the *solid earth kinematic boundary condition.*

- To formulate budgets for mass, tracer, and momentum in the ocean, we consider the upper ocean surface to be a time dependent permeable membrane through which precipitation, evaporation, ice melt, and river runoff pass. The expression for fluid transport at the upper surface leads to the *upper ocean kinematic boundary condition.*

- A surface of constant generalized vertical coordinate, s, is of importance when establishing the balances of mass, tracer, and momentum within a layer of fluid whose upper and lower bounds are determined by surfaces of constant s. Fluid transport through this surface is said to constitute the *dia-surface* transport.

2.1 Infinitesimal fluid parcels

Mass conservation for an infinitesimal parcel of fluid means that as it moves through the fluid, its mass is constant in time

$$\frac{\mathrm{d}M}{\mathrm{d}t} = 0. \tag{1}$$

[3]This assumption may be broken in some cases. For example, when the lower boundary is a moving sedimentary layer in a coastal estuary, or when there is seeping ground water. We do not consider such cases here.

In this equation, $M = \rho \, dV$ is the parcel's mass, ρ is its *in situ* density, and dV is its infinitesimal volume. The time derivative is taken following the parcel, and is known as a *material* or *Lagrangian* time derivative. Writing $dV = dx \, dy \, dz$, and defining the parcel's velocity as $\mathbf{v} = d\mathbf{x}/dt = (\mathbf{u}, w)$ leads to

$$\frac{d \ln \rho}{dt} = -\nabla \cdot \mathbf{v}. \tag{2}$$

Note that the horizontal coordinates $\mathbf{x}_h = (x, y)$ can generally be spherical coordinates (λ, ϕ), or any other generalized horizontal coordinate appropriate for the sphere, such as those illustrated in Figures 1 and 2 (see chapters 20 and 21 of Griffies, 2004 for a presentation of generalized horizontal coordinates).

For many purposes in fluid mechanics as well as ocean model design, it is useful to transform the frame of reference from the moving parcel to a fixed point in space. This transformation takes us from the material or Lagrangian frame to the *Eulerian* frame. It engenders a difference in how observers measure time changes in a fluid parcel's properties. In particular, the material time derivative picks up a *transport* or *advective* term associated with motion of the parcel

$$\boxed{\frac{d}{dt} = \partial_t + \mathbf{v} \cdot \nabla.} \tag{3}$$

This relation allows us to write the Lagrangian expression (2) for mass conservation in an Eulerian conservation form[4]

$$\rho_{,t} + \nabla \cdot (\rho \, \mathbf{v}) = 0. \tag{4}$$

Fluids that conserve mass are said to be *compressible* since the volume of a mass conserving fluid parcel can expand or contract based on pressure forces acting on the parcel, or properties such as temperature and salinity. However, in many circumstances, it is useful to consider the kinematics of a parcel that conserves its volume, in which case

$$\frac{1}{dV} \frac{dV}{dt} = -\nabla \cdot \mathbf{v} = 0. \tag{5}$$

The non-divergence condition $\nabla \cdot \mathbf{v} = 0$ provides a constraint on the parcel's velocity that must be satisfied at each point of the fluid. Fluid

[4]Throughout these lectures, a comma is used as a shorthand for partial derivative. Hence, $\rho_{,t} = \partial \rho / \partial t$. This notation follows Griffies, 2004, and is commonly used in mathematical physics. It is a useful means to distinguish a derivative from some of the many other uses of subscripts, such as a tensor component or as part of the name of a variable such as the fresh water flux q_{w} introduced in equation (27).

parcels that conserve their volume are known as *Boussinesq* parcels, whereas mass conserving parcels are *non-Boussinesq*. Non-Boussinesq parcels are generally considered in atmospheric dynamics, since the atmosphere is far more compressible than the ocean. However, most new ocean models are removing the Boussinesq approximation since straightforward means are known to solve the more general non-Boussinesq evolution using pressure-based coordinates.

2.2 Solid earth kinematic boundary condition

To begin our discussion of fluid flow through a surface, we start with the simplest surface: the time independent solid earth boundary. As mentioned earlier, one typically assumes in ocean modelling that there is no fluid crossing the solid earth lower boundary. In this case, a no-normal flow condition is imposed at the solid earth boundary at the depth

$$z = -H(x, y). \tag{6}$$

To develop a mathematical expression for the boundary condition, we note that the outward unit normal pointing from the ocean into the underlying rock is given by[5] (see Figure 3)

$$\hat{\mathbf{n}}_{\rm H} = -\frac{\nabla(z + H)}{|\nabla(z + H)|}. \tag{7}$$

Furthermore, we assume that the bottom topography can be represented as a continuous function $H(x, y)$ that does not possess "overturns." That is, we do not consider caves or overhangs in the bottom boundary where the topographic slope becomes infinite. Such would make it difficult to consider the slope of the bottom in our formulations. This limitation is common for ocean models.[6]

A no-normal flow condition on fluid flow at the ocean bottom implies

$$\mathbf{v} \cdot \hat{\mathbf{n}}_{\rm H} = 0 \qquad \text{at} \quad z = -H(x, y). \tag{8}$$

Expanding this constraint into its horizontal and vertical components leads to

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at} \quad z = -H(x, y), \tag{9}$$

[5]The three dimensional gradient operator $\nabla = (\partial_x, \partial_y, \partial_z)$ reduces to the two dimensional horizontal operator $\nabla_z = (\partial_x, \partial_y, 0)$ when acting on functions that depend only on the horizontal directions. To reduce notation clutter, we do not expose the z subscript in cases where it is clear that the horizontal gradient is all that is relevant.

[6]For hydrostatic models, the solution algorithms rely on the ability to integrate vertically from the ocean bottom to the top, uninterrupted by rock in between. Non-hydrostatic models do not employ such algorithms, and so may in principle allow for arbitrary bottom topography, including overhangs.

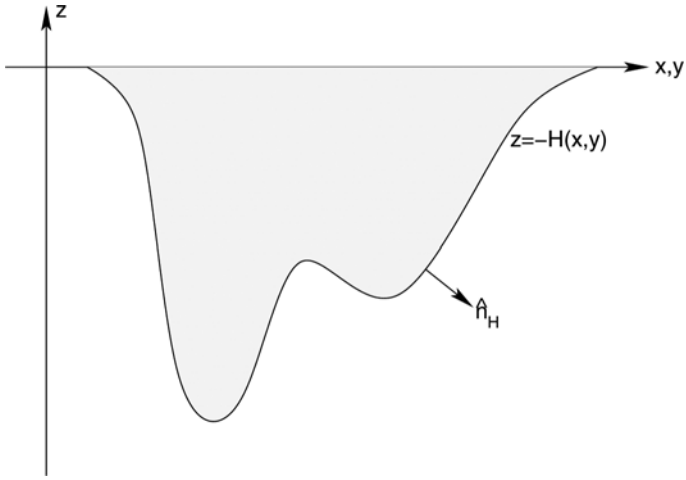

Figure 3. Schematic of the ocean's bottom surface with a smoothed undulating solid earth topography at $z = -H(x, y)$ and outward normal direction $\hat{\mathbf{n}}_{\mathrm{H}}$. Undulations of the bottom are far greater than the surface height (see Figure 4), as they can reach from the ocean bottom at 5000m-6000m to the surface over the course of a few kilometers (slopes on the order of 0.1 to 1.0). It is important for simulations to employ numerics that facilitate an accurate representation of the ocean bottom.

which can be written in the material derivative form

$$\frac{\mathrm{d}(z+H)}{\mathrm{d}t} = 0 \qquad \text{at} \quad z = -H(x,y). \tag{10}$$

Equation (10) expresses in a material or Lagrangian form the impenetrable nature of the solid earth lower surface, whereas equation (9) expresses the same constraint in an Eulerian form.

2.3 Generalized vertical coordinates

We now consider the form of the bottom kinematic boundary condition in generalized vertical coordinates. Generalized vertical coordinates provide the ocean theorist and modeler with a powerful set of tools to describe ocean flow, which in many situations is far more natural than the more traditional geopotential coordinates (x, y, z) that we have been using thus far. Therefore, it is important for the student to gain some exposure to the fundamentals of these coordinates, as they are ubiquitous in ocean modelling today.

Chapter 6 of Griffies, 2004 develops a calculus for generalized vertical coordinates. Some experience with these equations is useful to nurture an intuition for ocean modelling in generalized vertical coordinates.

Most notably, these coordinates, when used with the familiar horizontal coordinates (x, y), form a non-orthogonal triad, and thus lead to some unfamiliar relationships. To proceed in this section, we present some salient results of the mathematics of generalized vertical coordinates, and reserve many of the derivations for Griffies, 2004.

When considering generalized vertical coordinates in oceanography, we always assume that the surfaces cannot overturn on themselves. This constraint means that the Jacobian of transformation between the generalized vertical coordinate

$$s = s(x, y, z, t) \tag{11}$$

and the geopotential coordinate z, must be one signed. That is, the specific thickness

$$\frac{\partial z}{\partial s} = z_{,s} \tag{12}$$

is of the same sign throughout the ocean fluid. The name *specific thickness* arises from the property that

$$\mathrm{d}z = z_{,s}\, \mathrm{d}s \tag{13}$$

is an expression for the thickness of an infinitesimal layer of fluid bounded by two constant s surfaces.

Deriving the bottom kinematic boundary condition in s-coordinates requires a relation between the vertical velocity component used in geopotential coordinates, $w = \mathrm{d}z/\mathrm{d}t$, and the pseudo-velocity component $\mathrm{d}s/\mathrm{d}t$. For this purpose, we refer to some results from Section 6.5.5 of Griffies, 2004. As in that discussion, we note isomorphic relations

$$\mathrm{d}z/\mathrm{d}t = z_{,t} + \mathbf{u} \cdot \nabla_s z + z_{,s}\, \mathrm{d}s/\mathrm{d}t \tag{14}$$

$$\mathrm{d}s/\mathrm{d}t = s_{,t} + \mathbf{u} \cdot \nabla_z s + s_{,z}\, \mathrm{d}z/\mathrm{d}t, \tag{15}$$

with rearrangement leading to

$$\mathrm{d}z/\mathrm{d}t = z_{,s}\, (\mathrm{d}/\mathrm{d}t - \partial_t - \mathbf{u} \cdot \nabla_z)\, s. \tag{16}$$

This expression is relevant when measurements are taken on surfaces of constant geopotential, or depth. To apply this relation to the ocean bottom, which is generally not a surface of constant depth, it is necessary to transform the constant depth gradient ∇_z to a horizontal gradient taken along the bottom. We thus proceed as in Section 6.5.3 of Griffies, 2004 and consider the time-independent coordinate transformation

$$(\overline{x}, \overline{y}, \overline{z}, \overline{t}) = (x, y, -H(x, y), t). \tag{17}$$

The horizontal gradient taken on constant depth surfaces, ∇_z, and the horizontal gradient along the bottom, $\nabla_{\overline{z}}$, are thus related by

$$\nabla_{\overline{z}} = \nabla_z - (\nabla H)\, \partial_z. \tag{18}$$

Using this result in equation (16) yields

$$s_{,z}\,(w + \mathbf{u}\cdot\nabla H) = (\mathrm{d}/\mathrm{d}t - \partial_t - \mathbf{u}\cdot\nabla_{\overline{z}})\, s \qquad \text{at} \quad z = -H. \tag{19}$$

The left hand side vanishes due to the kinematic boundary condition (9), which then leads to

$$\mathrm{d}s/\mathrm{d}t = (\partial_t + \mathbf{u}\cdot\nabla_{\overline{z}})\, s \qquad \text{at} \quad s = s(x, y, z = -H(x, y), t). \tag{20}$$

The value of the generalized coordinate at the ocean bottom can be written in the shorthand form

$$s_{\mathrm{bot}}(x, y, t) = s(x, y, z = -H, t) \tag{21}$$

which leads to

$$\boxed{\frac{\mathrm{d}\,(s - s_{\mathrm{bot}})}{\mathrm{d}t} = 0 \qquad \text{at} \quad s = s_{\mathrm{bot}}.} \tag{22}$$

This relation is analogous to equation (10) appropriate to z-coordinates. Indeed, it is actually a basic statement of the impenetrable nature of the solid earth lower boundary, which is true regardless the vertical coordinates.

2.4 Upper surface kinematic condition

The upper ocean surface is penetrable and time dependent and full of breaking waves. Changes in ocean tracer concentration arise from precipitation, evaporation, river runoff,[7] and ice melt. These fluxes are critical agents in forcing the large scale ocean circulation via changes in ocean density and hence the water mass characteristics.

To describe the kinematics of water transport into the ocean, it is useful to introduce an effective transport through a smoothed ocean surface, where smoothing is performed via an ensemble average. We assume that this averaging leads to a surface absent overturns or breaking waves, thus

[7]River runoff generally enters the ocean at a nonzero depth rather than through the surface. Many global models, however, have traditionally inserted river runoff to the top model cell. Such can become problematic numerically and physically when the top grid cells are refined to levels common in coastal modelling. Hence, more applications are now considering the input of runoff throughout a nonzero depth.

facilitating a mathematical description analogous to the ocean bottom just considered. The vertical coordinate takes on the value

$$z = \eta(x, y, t) \tag{23}$$

at this idealized ocean surface.

We furthermore assume that density of the water crossing the ocean surface is ρ_w, which is a function of the temperature, salinity, and pressure. Different water densities can be considered for precipitation, evaporation, runoff, and ice melt, but this level of detail is not warranted for present purposes. The mass transport crossing the ocean surface can be written

$$(\text{MASS/TIME}) \text{ THROUGH SURFACE} = \hat{\mathbf{n}}_\eta \cdot \hat{\mathbf{n}}_w \, (P - E + R) \, \rho_w \, dA_\eta. \tag{24}$$

In this expression, $P > 0$ is the volume per time per area of precipitation entering the ocean, $E > 0$ is the evaporation leaving the ocean, and $R > 0$ is the river runoff and ice melt entering the ocean. The unit normal

$$\hat{\mathbf{n}}_\eta = \frac{\nabla (z - \eta)}{|\nabla (z - \eta)|} \tag{25}$$

points from the ocean surface at $z = \eta$ into the overlying atmosphere, whereas the unit normal $\hat{\mathbf{n}}_w$ orients the flow of the water mass transported across the ocean surface (see Figure 4). Finally, the area element dA_η measures the infinitesimal area on the ocean surface $z = \eta$, and it is given by (see Section 20.13.2 of Griffies, 2004)

$$dA_\eta = |\nabla(z - \eta)| \, dx \, dy. \tag{26}$$

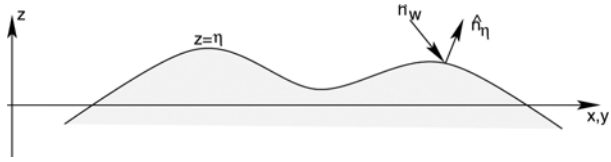

Figure 4. Schematic of the ocean's upper surface with a smoothed undulating surface height at $z = \eta(x, y, t)$, outward normal direction $\hat{\mathbf{n}}_\eta$, and freshwater normal direction $\hat{\mathbf{n}}_w$. Undulations of the surface height are on the order of a few meters due to tidal fluctuations in the open ocean, and order 10m-20m in certain embayments (e.g., Bay of Fundy in Nova Scotia). When imposing the weight of sea ice onto the ocean surface, the surface height can depress even further, on the order of 5m-10m, with larger values possible in some cases. It is important for simulations to employ numerical schemes facilitating such wide surface height undulations.

We now introduce a more convenient expression for the mass transport across the surface by exploiting our assumption that the ocean surface has no overturns. For this purpose, define

$$q_{\rm w}\, {\rm d}A = \hat{\mathbf{n}}_\eta \cdot \hat{\mathbf{n}}_{\rm w}\, (P - E + R)\, {\rm d}A_\eta, \tag{27}$$

where

$${\rm d}A = {\rm d}x\, {\rm d}y \tag{28}$$

is the horizontal projection of the surface area element ${\rm d}A_\eta$. The volume per time per horizontal area of fluid crossing the ocean surface is therefore defined by $q_{\rm w}$

$$\boxed{\begin{aligned} q_{\rm w} &= \frac{\hat{\mathbf{n}}_\eta \cdot \hat{\mathbf{n}}_{\rm w}\, (P - E + R)\, {\rm d}A_\eta}{{\rm d}A} \\ &= \frac{\text{(VOLUME/TIME) THROUGH FREE SURFACE}}{\text{HORIZONTAL AREA UNDER FREE SURFACE}}. \end{aligned}} \tag{29}$$

This is the surface water flux that appears in ocean model budgets for mass, tracer, and momentum.

As discussed in Section 3.4.3 of Griffies, 2004, the mass budget per horizontal area of a column of fluid extending from the ocean surface to its bottom is given by

$$\partial_t \left(\int_{-H}^{\eta} {\rm d}z\, \rho \right) = -\nabla \cdot \left(\int_{-H}^{\eta} {\rm d}z\, \rho\, \mathbf{u} \right) + q_{\rm w}\, \rho_{\rm w}. \tag{30}$$

This budget says that the time tendency of the total fluid mass per unit horizontal area within a column (left hand side) is balanced by the convergence of mass into the column (first term on the right hand side) and transport across the upper ocean surface (second term on the right hand side). To develop the upper ocean kinematic boundary condition, perform the derivatives in equation (30), keeping in mind Leibnitz's Rule when differentiating an integral. This step then leads to

$$[\rho\, (\partial_t + \mathbf{u} \cdot \nabla)\, \eta]_{z=\eta} + [\rho\, \nabla H \cdot \mathbf{u}]_{z=-H} + \int_{-H}^{\eta} {\rm d}z\, [\rho_{,t} + \nabla \cdot (\rho\, \mathbf{u})] = \rho_{\rm w}\, q_{\rm w}. \tag{31}$$

Use of the mass conservation equation (4) yields

$$[\rho\, (\eta_{,t} + \mathbf{u} \cdot \nabla \eta - w)]_{z=\eta} + [\rho\, (w + \nabla H \cdot \mathbf{u})]_{z=-H} = \rho_{\rm w}\, q_{\rm w}. \tag{32}$$

The solid earth kinematic boundary condition (9) allows us to cancel the second term on the left hand side, thus leading to the surface ocean

kinematic boundary condition

$$\rho\,(\partial_t + \mathbf{u}\cdot\nabla)\,\eta = \rho_{\rm w}\,q_{\rm w} + \rho\,w \qquad \text{at} \quad z=\eta \tag{33}$$

which can be written in the material form

$$\rho\left(\frac{\mathrm{d}(z-\eta)}{\mathrm{d}t}\right) = -\rho_{\rm w}\,q_{\rm w} \qquad \text{at} \quad z=\eta. \tag{34}$$

Contrary to the solid earth condition (10), where $z+H$ is materially constant, permeability of the ocean surface leads to a nontrivial material evolution of $z-\eta$.

To derive the analogous s-coordinate boundary condition, we proceed as for the bottom. Here, the coordinate transformation is time dependent

$$(\overline{x},\overline{y},\overline{z},\overline{t}) = (x,y,\eta(x,y,t),t). \tag{35}$$

The horizontal gradient and time derivative operators are therefore related by

$$\nabla_{\overline{z}} = \nabla_z + (\nabla\,\eta)\,\partial_z \tag{36}$$

$$\partial_{\overline{t}} = \partial_t + \eta_{,t}\,\partial_z. \tag{37}$$

Hence, the relation (16) between vertical velocity components takes the following form at the ocean surface

$$w = z_{,s}\,(\mathrm{d}/\mathrm{d}t - \partial_{\overline{t}} - \mathbf{u}\cdot\nabla_{\overline{z}})\,s + (\partial_t + \mathbf{u}\cdot\nabla)\eta \qquad \text{at} \quad z=\eta. \tag{38}$$

Substitution of the z-coordinate kinematic boundary condition (33) leads to

$$\rho\,z_{,s}\;(\mathrm{d}/\mathrm{d}t - \partial_{\overline{t}} - \mathbf{u}\cdot\nabla_{\overline{z}})\;s = -\rho_{\rm w}\,q_{\rm w} \qquad \text{at} \quad s=s_{\rm top} \tag{39}$$

where $s_{\rm top} = s(x,y,z=\eta,t)$ is the value of the generalized vertical coordinate at the ocean surface. Reorganizing the result (39) leads to the material time derivative form

$$\boxed{\rho\,z_{,s}\left(\frac{\mathrm{d}(s-s_{\rm top})}{\mathrm{d}t}\right) = -\rho_{\rm w}\,q_{\rm w} \qquad \text{at} \quad s=s_{\rm top}} \tag{40}$$

which is analogous to the z-coordinate result (34). Indeed, it can be derived trivially by noting that $\mathrm{d}z/\mathrm{d}t = z_{,s}\,\mathrm{d}s/\mathrm{d}t$. Even so, it is useful to have gone through the previous manipulations in order to garner experience and confidence with the formalism. Such confidence becomes of particular use in the next section focusing on the dia-surface flux.

2.5 Dia-surface transport

We seek an expression for the flux of fluid passing through a surface of constant generalized vertical coordinate. The result will be an expression for the *dia-surface* transport. It plays a fundamental role in generalized vertical coordinate modelling. Our derivation here follows that given in Section 6.7 of Griffies, 2004.

At an arbitrary point on a surface of constant generalized vertical coordinate (see Figure 5), the flux of fluid in the direction normal to the surface is given by

$$\text{SEAWATER FLUX IN DIRECTION } \hat{\mathbf{n}} = \mathbf{v} \cdot \hat{\mathbf{n}}, \tag{41}$$

with

$$\hat{\mathbf{n}} = \nabla s \, |\nabla s|^{-1} \tag{42}$$

the surface unit normal direction. Introducing the material time derivative $\mathrm{d}s/\mathrm{d}t = s_{,t} + \mathbf{v} \cdot \nabla s$ leads to the equivalent expression

$$\mathbf{v} \cdot \hat{\mathbf{n}} = |\nabla s|^{-1} \, (\mathrm{d}/\mathrm{d}t - \partial_t) \, s. \tag{43}$$

That is, the normal component to a fluid parcel's velocity is proportional to the difference between the material time derivative of the surface and its partial time derivative.

Since the surface is generally moving, the net flux of seawater penetrating the surface is obtained by subtracting the velocity of the surface $\mathbf{v}^{(\mathrm{ref})}$ in the $\hat{\mathbf{n}}$ direction from the velocity component $\mathbf{v} \cdot \hat{\mathbf{n}}$ of the fluid parcels

$$\text{FLUX OF SEAWATER THROUGH SURFACE} = \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\mathrm{ref})}). \tag{44}$$

The velocity $\mathbf{v}^{(\mathrm{ref})}$ is the velocity of a reference point fixed on the surface, and it is written

$$\mathbf{v}^{(\mathrm{ref})} = \mathbf{u}^{(\mathrm{ref})} + w^{(\mathrm{ref})} \, \hat{\mathbf{z}}. \tag{45}$$

Since the reference point remains on the same $s = \text{const}$ surface, $\mathrm{d}s/\mathrm{d}t = 0$ for the reference point. Consequently, we can write the vertical velocity component $w^{(\mathrm{ref})}$ as

$$w^{(\mathrm{ref})} = -z_{,s} \, (\partial_t + \mathbf{u}^{(\mathrm{ref})} \cdot \nabla_z) \, s, \tag{46}$$

where equation (16) was used with $\mathrm{d}s/\mathrm{d}t = 0$. This result then leads to

$$\begin{aligned} \hat{\mathbf{n}} \cdot \mathbf{v}^{(\mathrm{ref})} &= \hat{\mathbf{n}} \cdot \mathbf{u}^{(\mathrm{ref})} + \hat{\mathbf{n}} \cdot \hat{\mathbf{z}} \, w^{(\mathrm{ref})} \\ &= -s_{,t} \, |\nabla s|^{-1}, \end{aligned} \tag{47}$$

which says that the normal component of the surface's velocity vanishes when the surface is static, as may be expected. It then leads to the following expression for the net flux of seawater crossing the surface

$$\boxed{\begin{aligned}\hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\text{ref})}) &= |\nabla s|^{-1} \, (\partial_t + \mathbf{v} \cdot \nabla) \, s \\ &= |\nabla s|^{-1} \, \mathrm{d}s/\mathrm{d}t.\end{aligned}} \tag{48}$$

Hence, the material time derivative of the generalized surface vanishes if and only if no water parcels cross it. This important result is used throughout ocean theory and modelling.

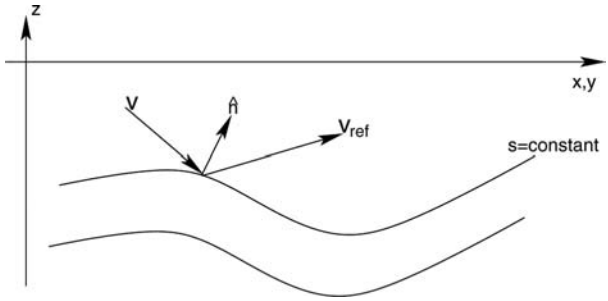

Figure 5. Surfaces of constant generalized vertical coordinate living interior to the ocean. An upward normal direction $\hat{\mathbf{n}}$ is indicated on one of the surfaces. Also shown is the orientation of a fluid parcel's velocity $\mathbf{v}$ and the velocity $\mathbf{v}^{(\text{ref})}$ of a reference point living on the surface.

Expression (48) gives the volume of seawater crossing a generalized surface, per time, per area. The area normalizing the volume flux is that area $\mathrm{d}A_{(\hat{\mathbf{n}})}$ of an infinitesimal patch on the surface of constant generalized vertical coordinate with outward unit normal $\hat{\mathbf{n}}$. This area can be written (see equation (6.58) of Griffies, 2004)

$$\mathrm{d}A_{(\hat{\mathbf{n}})} = |z_{,s} \, \nabla s| \, \mathrm{d}x \, \mathrm{d}y. \tag{49}$$

Hence, the volume per time of fluid passing through the generalized surface is

$$\begin{aligned}(\text{VOLUME/TIME}) \text{ THROUGH SURFACE} &= \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\text{ref})}) \, \mathrm{d}A_{(\hat{\mathbf{n}})} \\ &= |z_{,s}| \, (\mathrm{d}s/\mathrm{d}t) \, \mathrm{d}x \, \mathrm{d}y,\end{aligned} \tag{50}$$

and the magnitude of this flux is

$$|\hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\text{ref})}) \, \mathrm{d}A_{(\hat{\mathbf{n}})}| \equiv |w^{(s)}| \, \mathrm{d}x \, \mathrm{d}y. \tag{51}$$

We introduced the expression

$$w^{(s)} = z_{,s}\, \mathrm{d}s/\mathrm{d}t, \tag{52}$$

which measures the volume of fluid passing through the surface, per unit area $\mathrm{d}A = \mathrm{d}x\, \mathrm{d}y$ of the horizontal projection of the surface, per unit time. That is,

$$\boxed{\begin{aligned} w^{(s)} &\equiv \frac{\hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(\mathrm{ref})})\, \mathrm{d}A_{(\hat{\mathbf{n}})}}{\mathrm{d}A} \\ &= \frac{\text{(VOLUME/TIME) OF FLUID THROUGH SURFACE}}{\text{AREA OF HORIZONTAL PROJECTION OF SURFACE}}. \end{aligned}} \tag{53}$$

The quantity $w^{(s)}$ is called the dia-surface velocity component. It is directly analogous to the fresh water flux q_{w} defined in equation (27), which measures the volume of freshwater crossing the ocean surface, per unit time per horizontal area. To gain some experience with the dia-surface velocity component, it is useful to write it in the equivalent forms

$$\boxed{\begin{aligned} w^{(s)} &= z_{,s}\, \mathrm{d}s/\mathrm{d}t \\ &= z_{,s}\, \nabla s \cdot (\mathbf{v} - \mathbf{v}^{(\mathrm{ref})}) \\ &= (\hat{\mathbf{z}} - \mathbf{S}) \cdot (\mathbf{v} - \mathbf{v}^{(\mathrm{ref})}) \end{aligned}} \tag{54}$$

where

$$\begin{aligned} \mathbf{S} &= \nabla_s z \\ &= -z_{,s}\, \nabla_z s \end{aligned} \tag{55}$$

is the slope of the s surface as projected onto the horizontal directions. For example, if the slope vanishes, then the dia-surface velocity component measures the flux of fluid moving vertically relative to the motion of the generalized surface. When the surface is static and flat, then the dia-surface velocity component is simply the vertical velocity component $w = \mathrm{d}z/\mathrm{d}t$.

The expression (52) for $w^{(s)}$ allows one to write the material time derivative in one of the following equivalent manners

$$\boxed{\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} &= \left(\frac{\partial}{\partial t}\right)_z + \mathbf{u} \cdot \nabla_z + w \left(\frac{\partial}{\partial z}\right) \\ &= \left(\frac{\partial}{\partial t}\right)_s + \mathbf{u} \cdot \nabla_s + \frac{\mathrm{d}s}{\mathrm{d}t} \left(\frac{\partial}{\partial s}\right) \\ &= \left(\frac{\partial}{\partial t}\right)_s + \mathbf{u} \cdot \nabla_s + w^{(s)} \left(\frac{\partial}{\partial z}\right), \end{aligned}} \tag{56}$$

where $\partial_s = z_{,s}\, \partial_z$. The last form motivates some to consider $w^{(s)}$ as a vertical velocity component that measures the rate at which fluid parcels penetrate the surface of constant generalized coordinate (see Appendix A to McDougall, 1995). One should be mindful, however, to distinguish $w^{(s)}$ from the generally different vertical velocity component $w = \mathrm{d}z/\mathrm{d}t$, which measures the water flux crossing constant geopotential surfaces.

We close with a few points of clarification for the case where no fluid parcels cross the generalized surface. Such occurs, in particular, in the case of adiabatic flows with $s = \rho$ an isopycnal coordinate. In this case, the material time derivative (56) only has a horizontal two-dimensional advective component $\mathbf{u} \cdot \nabla_s$. This result *should not* be interpreted to mean that the velocity of a fluid parcel is strictly horizontal. Indeed, it generally is not, as the previous derivation should make clear. Rather, it means that the transport of fluid properties occurs along surfaces of constant s, and such transport is measured by the convergence of horizontal advective fluxes as measured along surfaces of constant s. We revisit this point in Section 3.2 when discussing tracer transport (see in particular Figure 7).

3. Mass and tracer budgets

The purpose of this section is to extend the kinematics discussed in the previous section to the case of mass and tracer budgets for finite domains within the ocean fluid. In the formulation of ocean models, these domains are thought of as discrete model grid cells.

3.1 General formulation

Assume that mass and tracer are altered within a finite region by transport across boundaries of the region and by sources within the region. Hence, the tracer mass within an arbitrary fluid region evolves according to

$$\partial_t \left(\iiint C\, \rho\, \mathrm{d}V \right) = \iiint \mathcal{S}^{(\mathrm{C})}\, \rho\, \mathrm{d}V - \iint \mathrm{d}A_{(\hat{\mathbf{n}})}\, \hat{\mathbf{n}} \cdot [(\mathbf{v} - \mathbf{v}^{\mathrm{ref}})\, \rho\, C + \mathbf{J}]. \tag{57}$$

The left hand side of this equation is the time tendency for the tracer mass within the region, where C is the tracer concentration and ρ is the *in situ* fluid density (mass of seawater per volume). As discussed in Sections 5.1 and 5.6 of Griffies, 2004, C represents a mass of tracer per mass of seawater for non-thermodynamic tracers such as salt or biogeochemical tracers, whereas C represents the potential temperature or the conservative temperature (McDougall, 2003) for the "heat" tracer

used in the model. On the right hand side, $\mathcal{S}^{(C)}$ represents a tracer source with units of tracer concentration per time. As seen in Section 2.5, $\mathrm{d}A_{(\hat{\mathbf{n}})}\, \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{\mathrm{ref}})$ measures the volume per time of fluid penetrating the domain boundary at a point.

The tracer flux $\mathbf{J}$ arises from subgrid scale transport, such as diffusion and/or unresolved advection. This flux is assumed to vanish when the tracer concentration is uniform, in which case the tracer budget (57) reduces to a mass budget. In addition to the tracer flux, it is convenient to define the *tracer concentration flux* $\mathbf{F}$ via

$$\mathbf{J} = \rho\, \mathbf{F}, \tag{58}$$

where the dimensions of $\mathbf{F}$ are velocity $\times$ tracer concentration.

3.2 Budget for an interior grid cell

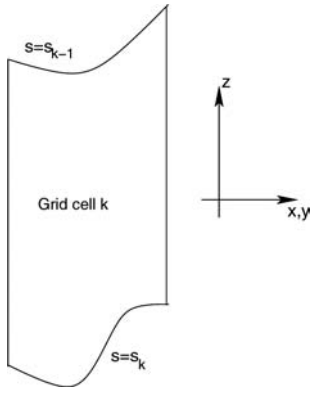

Figure 6. Schematic of an ocean grid cell labeled by the vertical integer k. Its sides are vertical and oriented according to $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, and its horizontal position is fixed in time. The top and bottom surfaces are determined by constant generalized vertical coordinates s_{k-1} and s_k, respectively. Furthermore, the top and bottom are assumed to always have an outward normal with a nonzero component in the vertical direction $\hat{\mathbf{z}}$. That is, the top and bottom are never vertical. Note that we take the convention that the discrete vertical label k increases as moving downward in the column, and grid cell k is bounded at its upper face by $s = s_{k-1}$ and lower face by $s = s_k$.

Consider the budget for a region bounded away from the ocean surface and bottom, such as that shown in Figure 6. There are two assumptions which define a grid cell region in this case.

- The sides of the cell are vertical, and so they are parallel to $\hat{\mathbf{z}}$ and aligned with the horizontal coordinate directions $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$. Their horizontal positions are fixed in time.

- The top and bottom of the cell are defined by surfaces of constant generalized vertical coordinate $s = s(x, y, z, t)$. The generalized surfaces do not overturn, which means that $s_{,z}$ is single signed throughout the ocean.

These assumptions lead to the following results for the sides of the grid cell

$$\text{TRACER MASS ENTERING CELL WEST FACE} = \iint_{x=x_1} \mathrm{d}y\,\mathrm{d}z\,(u\,\rho\,C + \rho\,F^x) \tag{59}$$

$$\text{TRACER MASS LEAVING CELL EAST FACE} = -\iint_{x=x_2} \mathrm{d}y\,\mathrm{d}z\,(u\,\rho\,C + \rho\,F^x) \tag{60}$$

where $x_1 \leq x \leq x_2$ defines the domain boundaries for the east-west coordinates.[8] Similar results hold for the tracer mass crossing the cell in the north-south directions. At the top and bottom of the grid cell[9]

$$\text{TRACER MASS ENTERING CELL BOTTOM FACE} = \iint_{s=s_k} \mathrm{d}x\,\mathrm{d}y\;\rho\,(w^{(s)}\,C + F^{(s)}) \tag{61}$$

$$\text{TRACER MASS LEAVING CELL TOP FACE} = -\iint_{s=s_{k-1}} \mathrm{d}x\,\mathrm{d}y\;\rho\,(w^{(s)}\,C + F^{(s)}). \tag{62}$$

To reach this result, we used a result from Section 2.5 to write the volume flux passing through the top face of the grid cell

$$\mathrm{d}A_{(\hat{\mathbf{n}})}\,\hat{\mathbf{n}}\cdot(\mathbf{v} - \mathbf{v}^{\mathrm{ref}}) = w^{(s)}\,\mathrm{d}x\,\mathrm{d}y, \tag{63}$$

with $w^{(s)} = z_{,s}\,\mathrm{d}s/\mathrm{d}t$ the dia-surface velocity component. A similar relation holds for the bottom face of the cell. The form of the SGS flux passing across the top and bottom is correspondingly given by

$$\mathrm{d}A_{(\hat{\mathbf{n}})}\,\hat{\mathbf{n}}\cdot\mathbf{J} = J^{(s)}\,\mathrm{d}x\,\mathrm{d}y. \tag{64}$$

Since the model is using the generalized coordinate s for the vertical, it is convenient to do the vertical integrals over s instead of z. For this

[8] We use generalized horizontal coordinates, such as those discussed in Griffies, 2004. Hence, the directions east, west, north, and south may not correspond to the usual geographic directions. Nonetheless, this terminology is useful for establishing the budgets, whose validity is general.

[9] As seen in Section 6, for pressure-like vertical coordinates, s increases with depth. For depth-like vertical coordinates, s decreases with depth. It is important to keep this sign difference in mind when formulating the budgets in the various coordinates. Notably, the specific thickness $z_{,s}$ carries the sign.

purpose, recall that with $z_{,s}$ single signed, the vertical thickness of a grid cell is

$$\mathrm{d}z = z_{,s}\,\mathrm{d}s. \tag{65}$$

Bringing these results together, and taking the limit as the volume of the cell in (x, y, s) space goes to zero (i.e., $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \to 0$) leads to

$$\partial_t(z_{,s}\,\rho\,C) = z_{,s}\,\rho\,\mathcal{S}^{(\mathrm{C})} - \nabla_s\cdot[z_{,s}\,\rho\,(\mathbf{u}\,C + \mathbf{F})] - \partial_s\,[\,\rho\,(w^{(s)}\,C + F^{(s)})] \tag{66}$$

Notably, the horizontal gradient operator ∇_s is computed on surfaces of constant s, and so it is distinct generally from the horizontal gradient ∇_z taken on surfaces of constant z. Instead of taking the limit as $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \to 0$, it convenient for discretization purposes to take the limit as the time independent horizontal area $\mathrm{d}x\,\mathrm{d}y$ goes to zero, thus maintaining the time dependent thickness $\mathrm{d}z = z_{,s}\,\mathrm{d}s$ inside the derivative operators. In this case, the thickness weighted tracer mass budget takes the form

$$\boxed{\begin{aligned}\partial_t(\mathrm{d}z\,\rho\,C) = \mathrm{d}z\,\rho\,\mathcal{S}^{(\mathrm{C})} - \nabla_s\cdot[\mathrm{d}z\,\rho\,(\mathbf{u}\,C + \mathbf{F})] \\ - [\rho\,(w^{(s)}\,C + F^{(s)})]_{s=s_{k-1}} + [\rho\,(w^{(s)}\,C + F^{(s)})]_{s=s_k}.\end{aligned}} \tag{67}$$

Similarly, the thickness weighted mass budget is

$$\boxed{\begin{aligned}\partial_t(\mathrm{d}z\,\rho) = \mathrm{d}z\,\rho\,\mathcal{S}^{(\mathrm{M})} - \nabla_s\cdot(\mathrm{d}z\,\rho\,\mathbf{u}) \\ - (\rho\,w^{(s)})_{s=s_{k-1}} + (\rho\,w^{(s)})_{s=s_k}.\end{aligned}} \tag{68}$$

where $\mathcal{S}^{(\mathrm{M})}$ is a mass source with units of inverse time that is related to the tracer source via

$$\mathcal{S}^{(\mathrm{M})} = \mathcal{S}^{(\mathrm{C})}(C = 1), \tag{69}$$

and the SGS flux vanishes with a uniform tracer

$$\mathbf{F}(C = 1) = 0. \tag{70}$$

3.3 Budgets without dia-surface fluxes

To garner some experience with these budgets, it is useful to consider the special case of zero dia-surface transport, either via advection or SGS fluxes, and zero tracer/mass sources. In this case, the thickness weighted mass and tracer mass budgets take the simplified form

$$\partial_t(\mathrm{d}z\,\rho) = -\nabla_s\cdot(\mathrm{d}z\,\rho\,\mathbf{u}) \tag{71}$$

$$\partial_t(\mathrm{d}z\,\rho\,C) = -\nabla_s\cdot[\mathrm{d}z\,\rho\,(\mathbf{u}\,C + \mathbf{F})]. \tag{72}$$

The first equation says that the time tendency of the thickness weighted density (mass per area) at a point between two surfaces of constant

generalized vertical coordinate is given by the horizontal convergence of mass per area onto that point. The transport is quasi-two-dimensional in the sense that it is only a two-dimensional convergence that determines the evolution. The tracer equation has an analogous interpretation. We illustrate this situation in Figure 7. As emphasized in our discussion of the material time derivative (56), this simplification of the transport equation does not mean that fluid parcels are strictly horizontal. Indeed, such is distinctly not the case when the surfaces are moving.

A further simplification of the mass and tracer mass budgets ensues when considering adiabatic and Boussinesq flow in isopycnal coordinates. We consider ρ now to represent the constant potential density of the finitely thick fluid layer. In this case, the mass and tracer budgets reduce to

$$\partial_t(\mathrm{d}z) = -\nabla_\rho \cdot (\mathrm{d}z\,\mathbf{u}) \tag{73}$$

$$\partial_t(\mathrm{d}z\,C) = -\nabla_\rho \cdot [\mathrm{d}z\,(\mathbf{u}\,C + \mathbf{F})]. \tag{74}$$

Equation (73) provides a relation for the thickness of the density layers, and equation (74) is the analogous relation for the tracer within the layer. These expressions are commonly used in the construction of adiabatic isopycnal models, which are often used in the study of geophysical fluid mechanics of the ocean.

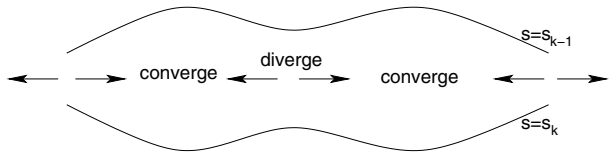

Figure 7. Schematic of the horizontal convergence of mass between two surfaces of constant generalized vertical coordinates. As indicated by equation (71), when there is zero dia-surface transport, it is just the horizontal convergence that determines the time evolution of mass between the layers. Evolution of thickness weighted tracer concentration in between the layers is likewise evolved just by the horizontal convergence of the thickness weighted advective and diffusive tracer fluxes (equation (72)). In this way, the transport is quasi-two-dimensional when the dia-surface transports vanish. A common example of this special system is an adiabatic ocean where the generalized surfaces are defined by isopycnals.

3.4 Cells adjacent to the ocean bottom

For a grid cell adjacent to the ocean bottom (Figure 8), we assume that just the bottom face of this cell abuts the solid earth boundary. The outward normal $\hat{\mathbf{n}}_{\mathrm{H}}$ to the bottom is given by equation (7), and the

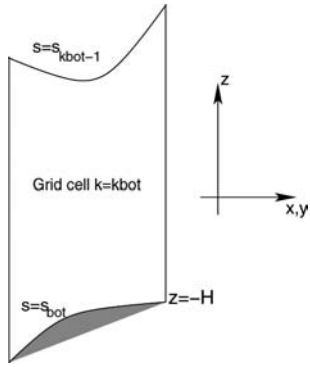

Figure 8. Schematic of an ocean grid cell next to the ocean bottom labeled by $k = k_{\rm bot}$. Its top face is a surface of constant generalized vertical coordinate $s = s_{\rm kbot-1}$, and the bottom face is determined by the ocean bottom topography at $z = -H$ where $s_{\rm bot}(x, y, t) = s(x, y, z = -H, t)$.

area element along the bottom is

$$\mathrm{d}A_{\rm H} = |\nabla(z + H)|\, \mathrm{d}x\, \mathrm{d}y. \tag{75}$$

Hence, the transport across the solid earth boundary is

$$-\iint \mathrm{d}A_{\rm H}\, \hat{\mathbf{n}}_{\rm H} \cdot (\mathbf{v}\, \rho\, C + \mathbf{J}) = \iint \mathrm{d}x\, \mathrm{d}y\, (\nabla H + \hat{\mathbf{z}}) \cdot (\mathbf{v}\, \rho\, C + \mathbf{J}). \tag{76}$$

We assume that there is zero mass flux across the bottom, in which case the advective flux drops out since $\mathbf{v} \cdot (\nabla H + \hat{\mathbf{z}}) = 0$ (equation (9)). However, the possibility of a nonzero geothermal tracer transport warrants a nonzero SGS tracer flux at the bottom, in which case the bottom tracer flux is written

$$Q^{(\rm C)}_{(\rm bot)} = (\nabla H + \hat{\mathbf{z}}) \cdot \mathbf{J}. \tag{77}$$

The corresponding thickness weighted budget is given by

$$\boxed{\begin{aligned} \partial_t\, (\mathrm{d}z\, \rho\, C) = {} & \mathrm{d}z\, \rho\, \mathcal{S}^{(\rm C)} - \nabla_s \cdot [\mathrm{d}z\, \rho\, (\mathbf{u}\, C + \mathbf{F})] \\ & - \Big[\rho\, (w^{(s)}\, C + z_{,s}\, \nabla s \cdot \mathbf{F})\Big]_{s=s_{kbot-1}} \\ & + Q^{(\rm C)}_{(\rm bot)}, \end{aligned}} \tag{78}$$

and the corresponding mass budget is

$$\boxed{\partial_t\, (\mathrm{d}z\, \rho) = \mathrm{d}z\, \rho\, \mathcal{S}^{(\rm M)} - \nabla_s \cdot (\mathrm{d}z\, \rho\, \mathbf{u}) - (\rho\, w^{s)})_{s=s_{kbot-1}}.} \tag{79}$$

3.5 Cells adjacent to the ocean surface

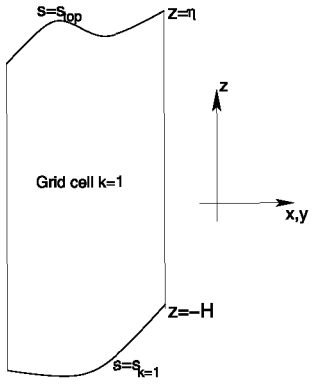

Figure 9. Schematic of an ocean grid cell next to the ocean surface labeled by $k = 1$. Its top face is at $z = \eta$, and the bottom is a surface of constant generalized vertical coordinate $s = s_{k=1}$.

For a grid cell adjacent to the ocean surface (Figure 9), we assume that just the upper face of this cell abuts the boundary between the ocean and the atmosphere or sea ice. The ocean surface is a time dependent boundary with $z = \eta(x, y, t)$. The outward normal $\hat{\mathbf{n}}_\eta$ is given by equation (25), and its area element $\mathrm{d}A_\eta$ is given by equation (26).

As the surface can move, we must measure the advective transport with respect to the moving surface. Just as in the dia-surface transport discussed in Section 2.5, we consider the velocity of a reference point on the surface

$$\mathbf{v}^{\mathrm{ref}} = \mathbf{u}^{\mathrm{ref}} + \hat{\mathbf{z}}\, w^{\mathrm{ref}}. \tag{80}$$

Since $z = \eta$ represents the vertical position of the reference point, the vertical component of the velocity for this point is given by

$$w^{\mathrm{ref}} = (\partial_t + \mathbf{u}^{\mathrm{ref}} \cdot \nabla)\, \eta \tag{81}$$

which then leads to

$$\mathbf{v}^{\mathrm{ref}} \cdot \nabla\, (z - \eta) = \eta_{,t}. \tag{82}$$

Hence, the advective transport leaving the ocean surface is

$$\begin{aligned} \iint_{z=\eta} \mathrm{d}A_{(\hat{\mathbf{n}})}\, \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{\mathrm{ref}})\, \rho\, C &= \iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, (-\eta_{,t} + w - \mathbf{u} \cdot \nabla\eta)\, \rho\, C \\ &= -\iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, \rho_{\mathrm{w}}\, q_{\mathrm{w}}\, C, \end{aligned} \tag{83}$$

where the surface kinematic boundary condition (33) was used. The negative sign on the right hand side arises from our convention that $q_w > 0$ represents an input of water to the ocean domain. In summary, the tracer flux leaving the ocean free surface is given by

$$\iint_{z=\eta} \mathrm{d}A_{(\hat{\mathbf{n}})}\, \hat{\mathbf{n}} \cdot [(\mathbf{v} - \mathbf{v}^{\mathrm{ref}})\, \rho\, C + \mathbf{J}] = \iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, (-\rho_w\, q_w\, C + \nabla\,(z-\eta) \cdot \mathbf{J}). \tag{84}$$

In the above, we formally require the tracer concentration precisely at the ocean surface $z = \eta$. However, as mentioned at the start of Section 2.4, it is actually a fiction that the ocean surface is a smooth mathematical function. Furthermore, seawater properties precisely at the ocean surface, known generally as *skin properties*, are generally not what an ocean model carries as its prognostic variable in its top grid cell. Instead, the model carries a bulk property averaged over roughly the upper few tens of centimeters. The lectures at this school by Professor Ian Robinson discuss these important points in the context of measuring sea surface temperature from a satellite, where the satellite measures the skin temperature, not the *foundational* or bulk temperature carried by large-scale ocean models.

To proceed in formulating the boundary condition for an ocean climate model, whose grid cells we assume to be at least a meter in thickness, we consider there to be a boundary layer model that provides us with the total tracer flux passing through the ocean surface. Developing such a model is a nontrivial problem in air-sea and ice-sea interaction theory and phenomenology. For present purposes, we do not focus on these details, and instead just introduce this flux in the form

$$Q^{(\mathrm{C})} = -\rho_w\, q_w\, C_w + Q^{(\mathrm{C})}_{(\mathrm{turb})} \tag{85}$$

where C_w is the tracer concentration in fresh water. The first term represents the advective transport of tracer through the surface with the fresh water (i.e., ice melt, rivers, precipitation, evaporation). The term $Q^{(\mathrm{C})}_{(\mathrm{turb})}$ arises from parameterized turbulence and/or radiative fluxes, such as sensible, latent, shortwave, and longwave heating appropriate for the temperature equation. A positive value for $Q^{(\mathrm{C})}_{(\mathrm{turb})}$ signals tracer leaving the ocean through its surface. In the special case of zero fresh water flux, then

$$\nabla\,(z-\eta) \cdot \mathbf{J} = Q^{(\mathrm{C})}_{(\mathrm{turb})} \qquad \text{if} \quad q_w = 0. \tag{86}$$

In general, it is not possible to make this identification. Instead, we must settle for the general expression

$$\iint_{z=\eta} \mathrm{d}A_{(\hat{\mathbf{n}})}\, \hat{\mathbf{n}} \cdot [(\mathbf{v} - \mathbf{v}^{\mathrm{ref}})\, \rho\, C + \mathbf{J}] = \iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, (-\rho_{\mathrm{w}}\, q_{\mathrm{w}}\, C_{\mathrm{w}} + Q^{(\mathrm{C})}_{(\mathrm{turb})}). \quad (87)$$

The above results lead to the thickness weighted tracer budget for the ocean surface grid cell

$$\boxed{\begin{aligned} \partial_t\, (\mathrm{d}z\, \rho\, C) = \mathrm{d}z\, \rho\, \mathcal{S}^{(\mathrm{C})} - \nabla_s \cdot [\mathrm{d}z\, \rho\, (\mathbf{u}\, C + \mathbf{F})] \\ + \Big[\rho\, (w^{(s)}\, C + z_{,s}\, \nabla s \cdot \mathbf{F})\Big]_{s=s_{k=1}} \\ + (\rho_{\mathrm{w}}\, q_{\mathrm{w}}\, C_{\mathrm{w}} - Q^{(\mathrm{turb})}_{(\mathrm{C})}), \end{aligned}} \quad (88)$$

and the corresponding mass budget

$$\boxed{\partial_t\, (\mathrm{d}z\, \rho) = \mathrm{d}z\, \rho\, \mathcal{S}^{(\mathrm{M})} - \nabla_s \cdot (\mathrm{d}z\, \rho\, \mathbf{u}) + (\rho\, w^{(s)})_{s=s_{k=1}} + \rho_{\mathrm{w}}\, q_{\mathrm{w}}.} \quad (89)$$

3.6 Surface boundary condition for salt

We close this section by mentioning the free ocean surface boundary condition for salt and other material tracers. Salt is transferred into the ocean with brackish river water and ice melt of nonzero salinity. Yet evaporation and precipitation generally leave the salt content of the ocean unchanged. In these latter cases, the boundary layer tracer flux (85) vanishes

$$Q^{(\mathrm{salt})} = 0. \quad (90)$$

This trivial boundary condition is also appropriate for many other material tracers, such as those encountered with ocean biogeochemical processes. In these cases, the tracer concentration *is not* altered via the passage of tracer across the surface. Instead, it is altered via the transport of fresh water across the ocean free surface which acts to dilute or concentrate the tracer.

The boundary condition (90) is often replaced in ocean models by a *virtual tracer flux* condition, whereby tracer is transferred into the model in lieu of altering the ocean water mass via the transport of fresh water. Virtual tracer flux boundary conditions are required for rigid lid models (Bryan, 1969) that maintain a constant volume and so cannot incorporate surface fresh water fluxes. However, there remain few rigid lid models in use today, and there is no reason to maintain the virtual tracer flux in the more commonly used free surface models. The differences in solution may be minor for many purposes, especially short

integrations (e.g., less than a year). However, the feedbacks related to climate and climate change may be nontrivial. Furthermore, the changes in model formulation are minor once a free surface algorithm has been implemented. Thus, it is prudent and straightforward to jettison the virtual tracer flux in favor of the physically motivated boundary condition (90) (Huang, 1993 and Griffies et al., 2001).

4. Linear momentum budget

The purpose of this section is to formulate the budget for linear momentum over a finite region of the ocean, with specific application to ocean model grid cells. The material here requires many of the same elements as in Section 3, but with added complexity arising from the vector nature of momentum, and the additional considerations of forces from pressure, friction, gravity, and planetary rotation.

4.1 General formulation

The budget of linear momentum for a finite region of fluid is given by the following relation based on Newton's second and third laws

$$\begin{aligned} \partial_t \left(\iiint \mathrm{d}V \, \rho \, \mathbf{v} \right) = & - \iint \mathrm{d}A_{(\hat{\mathbf{n}})} \, [\, \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{\mathrm{ref}})] \, \rho \, \mathbf{v} \\ & + \iint \mathrm{d}A_{(\hat{\mathbf{n}})} \, (\hat{\mathbf{n}} \cdot \boldsymbol{\tau} - \hat{\mathbf{n}} \, p) \\ & - \iiint \mathrm{d}V \, \rho \, [\, g \, \hat{\mathbf{z}} + (f + \mathcal{M}) \, \hat{\mathbf{z}} \wedge \mathbf{v}]. \end{aligned} \tag{91}$$

The left hand side is the time tendency of the region's linear momentum. The first term on the right hand side is the advective transport of linear momentum across the boundary of the region, with recognition that the region's boundaries are generally moving with velocity $\mathbf{v}^{\mathrm{ref}}$. The second term is the integral of the contact stresses due to friction and pressure. These stresses act on the boundary of the fluid domain. The stress tensor $\boldsymbol{\tau}$ is a symmetric second order tensor that parameterizes subgrid scale transport of momentum. The final term on the right hand side is the volume integral of body forces due to gravity and the Coriolis force.[10] In addition, there is a body force arising from the nonzero curvature of the spherical space. This curvature leads to the advection metric frequency (see equation (4.49) of Griffies, 2004) $\mathcal{M} = v \, \partial_x \ln \mathrm{d}y - u \, \partial_y \ln \mathrm{d}x$. The

[10] The wedge symbol $\wedge$ represents a vector cross product, also commonly written as $\times$. The wedge is typically used in the physics literature, and is preferred here to avoid confusion with the horizontal coordinate x.

advection metric frequency arises since linear momentum is not conserved on the sphere.[11] Hence, the linear momentum budget picks up this extra term that is a function of the chosen lateral coordinates. The advection metric frequency is analogous to, but far smaller than, the Coriolis frequency.

Unlike the case of the tracer and mass balances considered in Section 3, we do not consider momentum sources interior to the fluid domain. Such may be of interest and can be introduced without difficulty. The goal of the remainder of this section is to consider the linear momentum balance for finite grid cells in an ocean model.

4.2 An interior grid cell

At the west side of a grid cell, $\hat{\mathbf{n}} = -\hat{\mathbf{x}}$ whereas $\hat{\mathbf{n}} = \hat{\mathbf{x}}$ on the east side. Hence, the advective transport of linear momentum entering through the west side of the grid cell and that which is leaving through the east side are given by

$$\text{TRANSPORT ENTERING FROM WEST} = \iint_{x=x_1} \mathrm{d}y\, \mathrm{d}s\; z_{,s}\, u\, (\rho\, \mathbf{v}) \tag{92}$$

$$\text{TRANSPORT LEAVING THROUGH EAST} = -\iint_{x=x_2} \mathrm{d}y\, \mathrm{d}s\; z_{,s}\, u\, (\rho\, \mathbf{v}). \tag{93}$$

Similar results hold for momentum crossing the cell boundaries in the north and south directions. Momentum crossing the top and bottom surfaces of an interior cell is given by

$$\text{TRANSPORT ENTERING FROM THE BOTTOM} = \iint_{s=s_2} \mathrm{d}x\, \mathrm{d}y\; w^{(s)}\, (\rho\, \mathbf{v}) \tag{94}$$

$$\text{TRANSPORT LEAVING FROM THE TOP} = -\iint_{s=s_1} \mathrm{d}x\, \mathrm{d}y\; w^{(s)}\, (\rho\, \mathbf{v}). \tag{95}$$

[11] Angular momentum is conserved for frictionless flow on the sphere in the absence of horizontal boundaries (see Section 4.11.2 of Griffies, 2004).

Forces due to the contact stresses at the west and east sides are given by

$$\text{CONTACT FORCE ON WEST SIDE} = -\iint_{x=x_1} \mathrm{d}y\,\mathrm{d}s\; z_{,s}\,(\hat{\mathbf{x}}\cdot\boldsymbol{\tau} - \hat{\mathbf{x}}\,p) \tag{96}$$

$$\text{CONTACT FORCE ON EAST SIDE} = \iint_{x=x_2} \mathrm{d}y\,\mathrm{d}s\; z_{,s}\,(\hat{\mathbf{x}}\cdot\boldsymbol{\tau} - \hat{\mathbf{x}}\,p) \tag{97}$$

with similar results at the north and south sides. At the top of the cell, $\mathrm{d}A_{(\hat{\mathbf{n}})}\,\hat{\mathbf{n}} = \nabla s\,\mathrm{d}x\,\mathrm{d}y$ whereas $\mathrm{d}A_{(\hat{\mathbf{n}})}\,\hat{\mathbf{n}} = -\nabla s\,\mathrm{d}x\,\mathrm{d}y$ at the bottom. Hence,

$$\text{CONTACT FORCE ON CELL TOP} = \iint_{s=s_{k-1}} \mathrm{d}x\,\mathrm{d}y\; z_{,s}\,(\nabla s\cdot\boldsymbol{\tau} - p\,\nabla s) \tag{98}$$

$$\text{CONTACT FORCE ON CELL BOTTOM} = -\iint_{s=s_k} \mathrm{d}y\,\mathrm{d}s\; z_{,s}\,(\nabla s\cdot\boldsymbol{\tau} - p\,\nabla s). \tag{99}$$

Bringing these results together, and taking limit as the time independent horizontal area $\mathrm{d}x\,\mathrm{d}y \to 0$, leads to the thickness weighted budget for the momentum per horizontal area of an interior grid cell

$$\begin{aligned}
\partial_t\,(\mathrm{d}z\,\rho\,\mathbf{v}) = &-\nabla_s\cdot[\,\mathrm{d}z\,\mathbf{u}\,(\rho\,\mathbf{v})] \\
&+ (w^{(s)}\,\rho\,\mathbf{v})_{s=s_k} - (w^{(s)}\,\rho\,\mathbf{v})_{s=s_{k-1}} \\
&+ \partial_x\,[\,\mathrm{d}z\,(\hat{\mathbf{x}}\cdot\boldsymbol{\tau} - \hat{\mathbf{x}}\,p)] \\
&+ \partial_y\,[\,\mathrm{d}z\,(\hat{\mathbf{y}}\cdot\boldsymbol{\tau} - \hat{\mathbf{y}}\,p)] \\
&+ [z_{,s}\,(\nabla s\cdot\boldsymbol{\tau} - p\,\nabla s)]_{s=s_{k-1}} \\
&- [z_{,s}\,(\nabla s\cdot\boldsymbol{\tau} - p\,\nabla s)]_{s=s_k} \\
&- \rho\,\mathrm{d}z\,[\,g\,\hat{\mathbf{z}} + (f+\mathcal{M})\,\hat{\mathbf{z}}\,\wedge\,\mathbf{v}].
\end{aligned} \tag{100}$$

Note that both the time and horizontal partial derivatives are for positions fixed on a constant generalized vertical coordinate surface. Additionally, we have yet to take the hydrostatic approximation, so these equations are written for the three components of the vertical velocity.

The first term on the right hand side of the thickness weighted momentum budget (100) is the convergence of advective momentum fluxes occurring within the layer. We discussed the analogous flux convergence for the tracer and mass budgets in Section 3.3. The second and third terms arise from the transport of momentum across the upper and lower constant s interfaces. The fourth and fifth terms arise from the

horizontal convergence of pressure and viscous stresses. The sixth and seventh terms arise from the frictional and pressure stresses acting on the constant generalized surfaces. These forces provide an interfacial stress between layers of constant s. Note that even in the absence of frictional stresses, interfacial stresses from pressure acting on the generally curved s surface can transmit momentum between vertically stacked layers. The final term arises from the gravitational force, the Coriolis force, and the advective frequency.

4.3 Cell adjacent to the ocean bottom

As for the tracer and mass budgets, we assume zero mass flux through the ocean bottom at $z = -H(x, y)$. However, there is generally a nonzero stress at the bottom due to both the pressure between the fluid and the bottom, and unresolved features in the flow which can correlate or anti-correlate with bottom topographic features (Holloway, 1999). The area integral of the stresses lead to a force on the fluid at the bottom

$$\mathbf{F}_{\text{bottom}} = - \iint_{z=-H} \mathrm{d}x\, \mathrm{d}y\, [\nabla(z+H) \cdot \boldsymbol{\tau} - p\, \nabla(z+H)]. \tag{101}$$

Details of the stress term requires fine scale information that is generally unavailable. For present purposes we assume that some boundary layer model provides information that is schematically written

$$\boldsymbol{\tau}^{\text{bot}} = \nabla(z+H) \cdot \boldsymbol{\tau} \tag{102}$$

where $\boldsymbol{\tau}^{\text{bot}}$ is a vector bottom stress. Taking the limit as the horizontal area vanishes leads to the thickness weighted budget for momentum per horizontal area of a grid cell next to the ocean bottom

$$\begin{aligned} \partial_t\, (\mathrm{d}z\, \rho\, \mathbf{v}) = & - \nabla_s \cdot [\, \mathrm{d}z\, \mathbf{u}\, (\rho\, \mathbf{v})] - (w^{(s)}\, \rho\, \mathbf{v})_{s=s_{kbot-1}} \\ & + \partial_x\, [\, \mathrm{d}z\, (\hat{\mathbf{x}} \cdot \boldsymbol{\tau} - \hat{\mathbf{x}}\, p)] \\ & + \partial_y\, [\, \mathrm{d}z\, (\hat{\mathbf{y}} \cdot \boldsymbol{\tau} - \hat{\mathbf{y}}\, p)] \\ & + [z_{,s}\, (\nabla s \cdot \boldsymbol{\tau} - p\, \nabla s)]_{s=s_{kbot-1}} \\ & - \boldsymbol{\tau}^{\text{bot}} + p_{\text{b}}\, \nabla(z+H) \\ & - \rho\, \mathrm{d}z\, [\, g\, \hat{\mathbf{z}} + (f + \mathcal{M})\, \hat{\mathbf{z}} \wedge \mathbf{v}]. \end{aligned} \tag{103}$$

4.4 Cell adjacent to the ocean surface

There is a nonzero mass and momentum flux through the upper ocean surface at $z = \eta(x, y, t)$, and contact stresses are applied from resolved and unresolved processes involving interactions with the atmosphere and

sea ice. Following the discussion of the tracer budget at the ocean surface in Section 3.5 leads to the expression for the transport of momentum into the ocean due to mass transport at the surface

$$- \iint \mathrm{d}A_{(\hat{\mathbf{n}})}\, \hat{\mathbf{n}} \cdot [(\mathbf{v} - \mathbf{v}^{\mathrm{ref}})\, \rho\, \mathbf{v} = \iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, \rho_{\mathrm{w}}\, q_{\mathrm{w}}\, \mathbf{v}. \tag{104}$$

The force arising from the contact stresses at the surface is written

$$\mathbf{F}_{\mathrm{contact}} = \iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, [\nabla\, (z - \eta) \cdot \boldsymbol{\tau} - p\, \nabla\, (z - \eta)]. \tag{105}$$

Bringing these results together leads to the force acting at the ocean surface

$$\mathbf{F}_{\mathrm{surface}} = \iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, [\nabla\, (z - \eta) \cdot \boldsymbol{\tau} - p\, \nabla\, (z - \eta) + \rho_{\mathrm{w}}\, q_{\mathrm{w}}\, \mathbf{v}]. \tag{106}$$

Details of the various terms in this force are generally unknown. Therefore, just as for the tracer at $z = \eta$ in Section 3.5, we assume that a boundary layer model provides information about the total force, and that this force is written

$$\mathbf{F}_{\mathrm{surface}} = \iint_{z=\eta} \mathrm{d}x\, \mathrm{d}y\, [\boldsymbol{\tau}^{\mathrm{top}} - p_{\mathrm{a}}\, \nabla\, (z - \eta) + \rho_{\mathrm{w}}\, q_{\mathrm{w}}\, \mathbf{v}_{\mathrm{w}}], \tag{107}$$

where $\mathbf{v}_{\mathrm{w}}$ is the velocity of the fresh water. This velocity is typically taken to be equal to the velocity of the ocean currents in the top cells of the ocean model, but such is not necessarily the case when considering the different velocities of, say, river water and precipitation. The stress $\boldsymbol{\tau}^{\mathrm{top}}$ is that arising from the wind, as well as interactions between the ocean and sea ice. Letting the horizontal area vanish leads to the thickness weighted budget for a grid cell next to the ocean surface

$$\begin{aligned} \partial_t\, (\mathrm{d}z\, \rho\, \mathbf{v}) = & - \nabla_s \cdot [\, \mathrm{d}z\, \mathbf{u}\, (\rho\, \mathbf{v})] + (w^{(s)}\, \rho\, \mathbf{v})_{s=s_{k=1}} \\ & + \partial_x\, [\, \mathrm{d}z\, (\hat{\mathbf{x}} \cdot \boldsymbol{\tau} - \hat{\mathbf{x}}\, p)] \\ & + \partial_y\, [\, \mathrm{d}z\, (\hat{\mathbf{y}} \cdot \boldsymbol{\tau} - \hat{\mathbf{y}}\, p)] \\ & - [z_{,s}\, (\nabla s \cdot \boldsymbol{\tau} - p\, \nabla s)]_{s=s_{k=1}} \\ & + [\, \boldsymbol{\tau}^{\mathrm{top}} - p_{\mathrm{a}}\, \nabla\, (z - \eta) + \rho_{\mathrm{w}}\, q_{\mathrm{w}}\, \mathbf{v}_{\mathrm{w}}] \\ & - \rho\, \mathrm{d}z\, [\, g\, \hat{\mathbf{z}} + (f + \mathcal{M})\, \hat{\mathbf{z}} \wedge \mathbf{v}]. \end{aligned} \tag{108}$$

5. The pressure force

A hydrostatic fluid maintains the balance $p_{,z} = -\rho\, g$. This balance means that the pressure at a point in a hydrostatic fluid is determined by the weight of fluid above this point. This relation is maintained quite well in the ocean on spatial scales larger than roughly 1km. Precisely, when the squared ratio of the vertical to horizontal scales of motion is small, then the hydrostatic approximation is well maintained. In this case, the vertical momentum budget reduces to the hydrostatic balance, in which case vertical acceleration and friction are neglected. If we are interested in explicitly representing such motions as Kelvin-Helmholtz billows and flow within a convective chimney, vertical accelerations are nontrivial and so the non-hydrostatic momentum budget must be used.

The hydrostatic balance greatly affects the algorithms used to numerically solve the equations of motion. The paper by Marshall et al., 1997 highlights these points in the context of developing an algorithm suited for both hydrostatic and non-hydrostatic simulations. However, so far in ocean modelling, no global simulations have been run at resolutions sufficiently refined to require the non-hydrostatic equations. Additionally, many regional and coastal models, even some with resolutions refined smaller than 1km, still maintain the hydrostatic approximation, and thus they must parameterize the unrepresented non-hydrostatic motions.

At a point in the continuum, the horizontal pressure gradient force for the hydrostatic and non-Boussinesq set of equations can be written[12]

$$\boxed{\begin{aligned} \rho^{-1}\, \nabla_z p &= \rho^{-1}\, (\nabla_s - \nabla_s\, z\, \partial_z)\, p \\ &= \rho^{-1}\, \nabla_s\, p + g\, \nabla_s\, z, \\ &= \nabla_s\, (p/\rho + g\, z) - p\, \nabla_s \rho^{-1} \end{aligned}} \tag{109}$$

where the hydrostatic relation $p_{,z} = -\rho\, g$ was used to reach the second equality. The term $p/\rho + g\, z$ is known as the Montgomery potential. For cases where the density term $\nabla_s \rho$ vanishes (such as when s is proportional to density), the pressure gradient force takes the form of a total gradient, and so it has a zero curl thus facilitating the formulation of vorticity budgets.

In general, the difficulty of numerically realizing the pressure gradient force arises when there are contributions from *both* the Montgomery potential and the density gradient terms in equation (109). Naive discretizations result in both terms being large and of opposite sign in

[12] For a Boussinesq fluid, equation (109) is modified by a factor of ρ/ρ_o. Hence, the same issues arise when numerically implementing the pressure gradient force with generalized vertical coordinates in either the Boussinesq or non-Boussinesq fluids.

many regions. Hence, they expose the calculation to nontrivial numerical truncation errors which can lead to spurious pressure gradients that spin up an unforced fluid with initially flat isopycnals. Significant effort has gone into reducing such *pressure gradient errors*, especially in terrain following models where undulations of the coordinate surfaces can be large with realistic bottom topography (e.g., see Figure 12). Some of these issues are summarized, with references, in Section 2 of Griffies et al., 2000a. Perhaps the most promising approach is that proposed by Shchepetkin and McWilliams, 2002. It is notable that difficulties with pressure gradient errors have largely been responsible for the near absence of sigma models being used for long term global ocean climate simulations.[13]

6. Elements of vertical coordinates

As discussed in Griffies et al., 2000a, there are broadly three regimes of the ocean germane to the considerations of a vertical coordinate.

- Upper ocean mixed layer: This is a generally turbulent region dominated by transfers of momentum, heat, freshwater, and tracers with the overlying atmosphere, sea ice, rivers, etc. It is of prime importance for climate system modelling and operational oceanography. It is typically very well mixed in the vertical through three-dimensional convective/turbulent processes. These processes involve non-hydrostatic physics which requires very high horizontal and vertical resolution (i.e., a vertical to horizontal grid aspect ratio near unity) to explicitly represent. A parameterization of these processes is therefore necessary in primitive equation ocean models. In this region, it is essential to employ a vertical coordinate that facilitates the representation and parameterization of these highly turbulent processes. Geopotential and pressure coordinates, or their relatives, are the most commonly used coordinates as they facilitate the use of very refined vertical grid spacing, which can be essential to simulate the strong exchanges between the ocean and atmosphere, rivers, and ice.

- Ocean interior: Tracer transport processes in the ocean interior predominantly occur along neutral directions (McDougall, 1987). The transport is dominated by large scale currents and mesoscale eddy fluctuations. Water mass properties in the interior thus tend

[13]The work of Diansky et al., 2002 is the only case known by the author of a global sigma model used for climate purposes.

to be preserved over large space and time scales (e.g., basin and decade scales). This property of the ocean interior is critical to represent in a numerical simulation of ocean climate. An isopycnal coordinate framework is well suited to this task, whereas geopotential and sigma models have problems associated with numerical truncation errors. As discussed by Griffies et al., 2000b, the problem becomes more egregious as the model resolution is refined, due to the enhanced levels of eddy activity that pumps tracer variance to the grid scale. Quasi-adiabatic dissipation of this variance is difficult to maintain in non-isopycnal models.

- Ocean bottom: The solid earth bottom topography directly influences the overlying currents. In an unstratified ocean, the balanced flow generally follows lines of constant f/H, where f is the Coriolis parameter and H ocean depth. Additionally, there are several regions where density driven currents (overflows) and turbulent bottom boundary layer (BBL) processes act as a strong determinant of water mass characteristics. Many such processes are crucial for the formation of deep water properties in the World Ocean, and for representing coastal processes in regional models. It is for this reason that sigma models have been developed over the past few decades, with their dominant application focused on the coastal and estuarine problem.

These three regimes impact on the design of vertical coordinates for ocean models. In this section, we detail some vertical coordinates and summarize their strengths and weaknesses, keeping in mind the above physical considerations.

6.1 Depth based vertical coordinates

We use depth based vertical coordinates in this section to discretize the Boussinesq equations.[14] Depth based coordinates are also known as *volume based* coordinates, since for a Boussinesq model which uses depth as the vertical coordinate, the volume of interior grid cells is constant in the absence of sources. Correspondingly, depth based coordinates are naturally suited for Boussinesq fluids.

[14] Greatbatch and McDougall, 2003 discuss an algorithm for non-Boussinesq dynamics in a z-model. Their methods are implemented in the MOM4 code of Griffies et al., 2004. This approach may be of special use for non-Boussinesq non-hydrostatic z-models. However, when focusing on hydrostatic models as we do here, pressure based vertical coordinates discussed in Section 6.2 are more convenient.

The equations describing a Boussinesq fluid are derived from the non-Boussinesq set derived in Sections 3 and 4 by replacing all appearances of *in situ* density ρ by a constant density ρ_o, *except* when density is used to compute the buoyancy forces arising from gravity. The density ρ_o is a representative density of the ocean fluid, such as $\rho_o = 1035\,\mathrm{kg\,m^{-3}}$. For much of the ocean, the *in situ* density varies less than 2% from this value (see page 47 of Gill, 1982).

Depth coordinate. With a free surface, the vertical domain over which the z-coordinate $s = z$ ranges is given by the time dependent interval $-H \leq z \leq \eta$. Consequently, the sum of the vertical grid cell increments equals to the total depth of the column $\sum_k \mathrm{d}z = H + \eta$. The trivial specific thickness $z_{,s} = 1$ simplifies the Boussinesq budgets.

The depth coordinate is useful for many purposes in global climate modelling, and models based on depth are the most popular ocean climate models. Their advantages include the following.

- Simple numerical methods have been successfully used in this framework.
- The horizontal pressure gradient can be easily represented in an accurate manner.
- The equation of state for ocean water can be accurately represented in a straightforward manner (e.g., McDougall et al., 2003).
- The upper ocean mixed layer is well parameterized using a z-coordinate.

Unfortunately, these models have some well known disadvantages, which include the following.

- Representation of tracer transport within the quasi-adiabatic interior is cumbersome, with problems becoming more egregious as mesoscale eddies are admitted (Griffies et al., 2000b).
- Representation and parameterization of bottom boundary layer processes and flow are unnatural.

Grid cells have static vertical increments $\mathrm{d}s = \mathrm{d}z$ when $s = z$, except for the top. At the top, $\partial_t(\mathrm{d}z) = \eta_{,t}$. The time dependent vertical range of the coordinate slightly complicates a numerical treatment of the surface cell in z-models (see Griffies et al., 2001 for details of one such treatment). More problematic, however, is the possibility of a vanishing top grid cell. That is, the surface cell can be lost (i.e., can become dry) if the free surface depresses below the depth of the top grid cell's

bottom face. This is a very inconvenient feature that limits the use of z-coordinates.[15] In particular, the following studies may require very refined vertical resolution and/or large undulations of the surface height, and so would not be accessible with a conventional free surface z-model.

- Process studies of surface mixing and biological cycling may warrant very refined upper ocean grid cell thickness, some as refined as 1m.
- Realistic tidal fluctuations in some parts of the World Ocean can reach 10m-20m.
- Coastal models tend to require refined vertical resolution to represent shallow coastal processes along the continental shelves and near-shore.
- When coupled to a sea ice model, the weight of the ice will depress the ocean free surface.

An example of depth coordinates. In some of the following discussion, we illustrate aspects of vertical coordinates by diagnosing values for the coordinates from a realistic z-model run with partial step thicknesses. *Partial steps* have arbitrary thickness which are set to accurately represent the bottom topography. The partial step technology was introduced by Adcroft et al., 1997 in the C-grid MITgcm, and further discussed by Pacanowski and Gnanadesikan, 1998 for the B-grid Modular Ocean Model (MOM). Figure 10 compares the representation of topography in a z-model using partial steps as realized in the MOM code of Griffies et al., 2004. Many z-models have incorporated the partial step technology as it provides an important facility to accurately represent flow and waves near topography.

In the representation of bottom topography, there is an artificial distinction between a vertical face of a cell and its horizontal top and bottom faces. There is no such distinction in the real ocean. As noted in Anne Marie Treguier's lectures at this school, the block structure of topography in z-models has the potential to affect the level of bottom friction. The effects on bottom friction come in by noting that for a C-grid, it is straightforward to run with free-slip side walls as well as bottom faces. In contrast, B-grids use a no-slip side wall and free slip

[15] Linearized free surfaces, in which the budgets for tracer and momentum are formulated assuming a constant top cell thickness, avoid problems with vanishing top cells. However, such models do not conserve total tracer or volume in the presence of a surface fresh water flux (see Griffies et al., 2001, Campin et al., 2004 for discussion).

bottom face. Hence, depending on the interior viscosity and bottom stress parameterization, B-grid models will generally have more bottom friction than C-grid models. With partial steps, the area of the side walls are reduced, thus reducing the area of no-slip side walls in the B-grid. The effective bottom friction in the B-grid is therefore less with partial step topography.

Because of partial steps, the level next to the ocean bottom has grid cell centers that are generally at different depths. That is, the bottom cell in a partial step z-model is likened to a sigma-layer. All other cells, including the surface, have grid cell centers that are at fixed depths. Figure 11 illustrates the lines of constant partial step depth for this model.

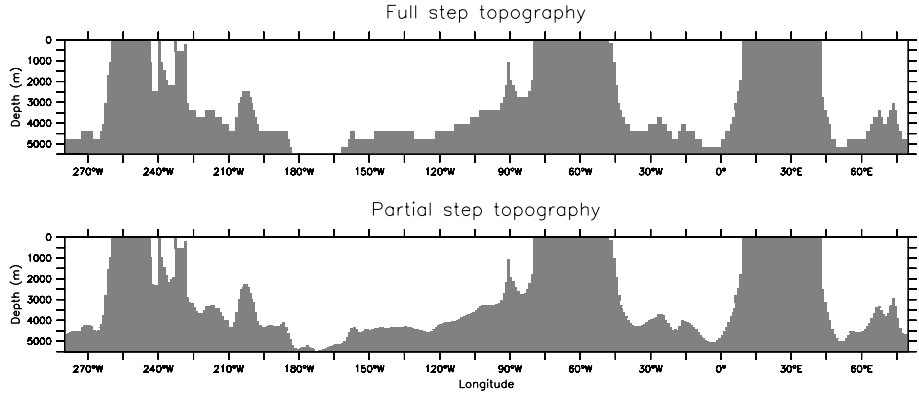

Figure 10. Comparison of the partial step versus full step representation of topography along the equator as realized in the z-model discussed by Griffies et al., 2005. The model horizontal grid has one degree latitudinal resolution. The main differences are in the deep ocean in regions where the topographic slope is gradual. Steep sloped regions, and those in the upper ocean with refined vertical resolution, show less distinctions.

Depth deviation coordinate. The depth deviation coordinate $s = z - \eta$ removes the restriction on upper ocean grid cell resolution present with $s = z$. That is, $s = 0$ is the time independent coordinate value of the ocean surface, no matter how much the free surface depresses or grows. Hence, no surface cells vanish so long as $\eta > -H$. However, $-(H + \eta) \leq s \leq 0$, and so the bottom of a column is a time dependent surface. Consequently, by solving the problem at the ocean surface, the deviation coordinate introduces a problem to the ocean bottom where bottom cells can now vanish.

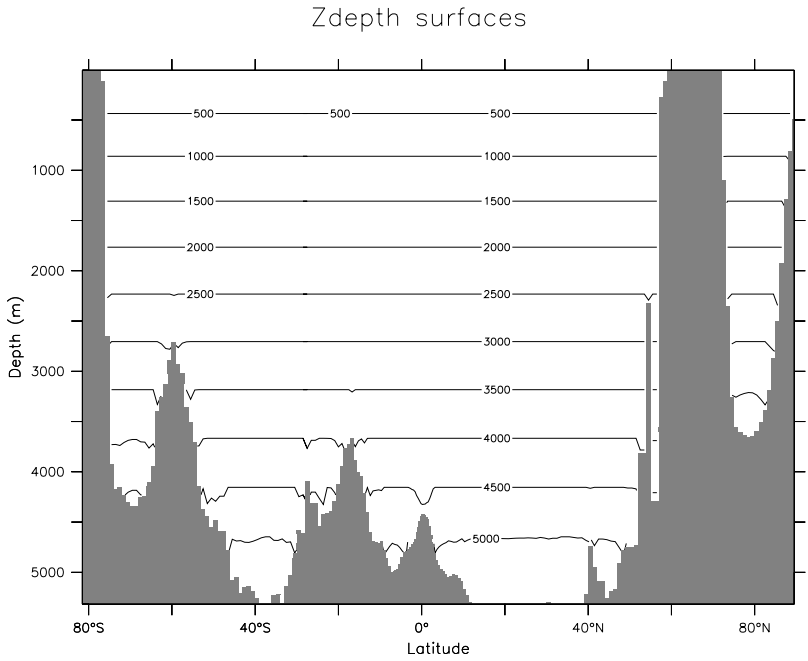

Figure 11. Constant depth surfaces in a realistic ocean model. Deviations from horizontal next to the bottom arise from the use of partial bottom cell thicknesses, as illustrated in Figure 10. Shown here is a section along $150^\circ W$.

Zstar coordinate. To overcome problems with vanishing surface and/or bottom cells, we consider the *zstar* coordinate

$$z^* = H\,(z-\eta)/(H+\eta). \tag{110}$$

This coordinate is closely related to the "eta" coordinate used in many atmospheric models (see Black, 1994 for a review). It was originally used in ocean models by Stacey et al., 1995 for studies of tides next to shelves, and it has been recently promoted by Adcroft and Campin, 2004 for global climate modelling.

The surfaces of constant z^* are quasi-horizontal. Indeed, the z^* coordinate reduces to z when η is zero. In general, when noting the large differences between undulations of the bottom topography versus undulations in the surface height, it is clear that surfaces constant z^* are

very similar to the depth surfaces shown in Figure 11. These properties greatly reduce difficulties of computing the horizontal pressure gradient relative to terrain following sigma models discussed next. Additionally, since $z^* = z$ when $\eta = 0$, no flow is spontaneously generated in an unforced ocean starting from rest, regardless the bottom topography. This behavior is in contrast to the case with sigma models, where pressure gradient errors in the presence of nontrivial topographic variations can generate spontaneous flow from a resting state. The quasi-horizontal nature of the coordinate surfaces also facilitates the implementation of neutral physics parameterizations in z^* models using the same techniques as in z-models (see Chapters 13-16 of Griffies, 2004 for a discussion of neutral physics in z-models).

The range over which z^* varies is time independent $-H \leq z^* \leq 0$. Hence, all cells remain nonvanishing, so long as the surface height maintains $\eta > -H$. This is a minor constraint relative to that encountered on the surface height when using $s = z$ or $s = z - \eta$.

Because z^* has a time independent range, all grid cells have static increments $\mathrm{d}s$, and the sum of the vertical increments yields the time independent ocean depth $\sum_k \mathrm{d}s = H$. The z^* coordinate is therefore invisible to undulations of the free surface, since it moves along with the free surface. This property means that no spurious vertical transport is induced across surfaces of constant z^* by motion of external gravity waves. Such spurious transport can be a problem in z-models, especially those with tidal forcing. Quite generally, the time independent range for the z^* coordinate is a very convenient property that allows for a nearly arbitrary vertical resolution even in the presence of large amplitude fluctuations of the surface height.

Depth sigma coordinate. The depth-sigma coordinate

$$\sigma = (z - \eta)/(H + \eta) \tag{111}$$

is the canonical *terrain following* coordinate. Figure 12 illustrates this coordinate in a realistic model. The sigma coordinate has a long history of use in coastal modelling. For reviews, see Greatbatch and Mellor, 1999 and Ezer et al., 2002. Models based on the sigma coordinate have also been successfully extended to basinwide studies, as well as recent global work by Diansky et al., 2002.

Just as for $z^* = H\,\sigma$, the range over which the sigma coordinate varies is time independent and given by $-1 \leq \sigma \leq 0$. Hence, all cells have static grid increments $\mathrm{d}s$, and the sum of the vertical increments yields unity $\sum_k \mathrm{d}s = 1$. So long as the surface height is not depressed

deeper than the ocean bottom (i.e., so long as $\eta > -H$), then all cells remain nonvanishing.[16]

In addition to not worrying about vanishing grid cells, some key advantages of sigma models are the following.

- They provide a natural framework to represent bottom influenced flow and to parameterize bottom boundary layer processes.
- Thermodynamic effects associated with the equation of state are well represented.

However, some of the disadvantages are the following:

- As with the z-models, the representation of the quasi-adiabatic interior is cumbersome due to numerical truncation errors inducing unphysically large levels of spurious mixing, especially in the presence of vigorous mesoscale eddies. Parameterization of these processes using neutral physics schemes may be more difficult numerically than in the z-models. The reason is that neutral directions generally have slopes less than 1/100 relative to the horizontal, but can have order unity slopes relative to sigma surfaces. The larger relative slopes precludes the *small slope approximation* commonly made with z-model implementations of neutral physics. The small slope approximation provides for simplification of the schemes, and improves computational efficiency.
- Sigma models have difficulty accurately representing the horizontal pressure gradient in the presence of realistic topography, where slopes are commonly larger than 1/100.

Although there are regional simulations using terrain following models, Griffies et al., 2000a notes that there are few examples of global climate models running with this vertical coordinate. Diansky et al., 2002 is the only exception known to the author. This situation is largely due to problems representing realistic topography without incurring unacceptable pressure gradient errors, as well as difficulties implementing parameterizations of neutral physical processes. There are notable efforts to resolve these problems, such as the pressure gradient work of Shchepetkin and McWilliams, 2002. Continued efforts along these lines may soon facilitate the more common use of terrain following coordinates for global ocean climate modelling.

[16] If $\eta < -H$, besides drying up a region of ocean, the specific thickness $z_{,s} = H + \eta$ changes sign, which signals a singularity in the vertical grid definition. The same problem occurs for the z^* coordinate.

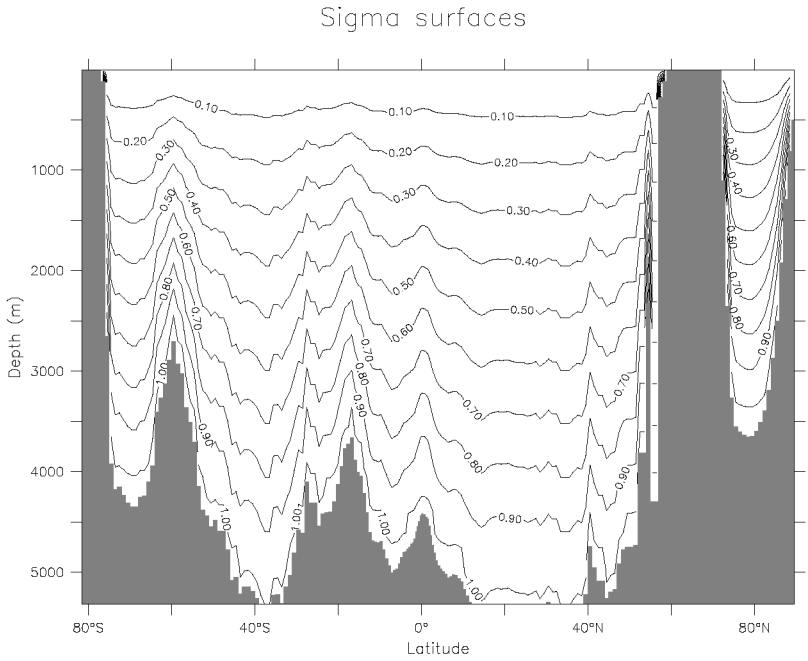

Figure 12. Constant sigma surfaces as diagnosed in a z-model. Shown here is a section along $150°W$, as in Figure 11. Note the strong variations in the contours, as determined by changes in the bottom topography.

Summary of the depth based vertical coordinates. Depth based vertical coordinates are naturally used for Boussinesq equations. These coordinates and their specific thicknesses $z_{,s}$ are summarized in Table 2.1. Notably, both the sigma and zstar coordinates have time independent ranges, but time dependent specific thicknesses. In contrast, the depth and depth deviation coordinates have time dependent depth ranges and time independent specific thicknesses. If plotted with the same range as those given in Figure 11, surfaces of constant depth deviation and constant zstar are indistinguishable from surfaces of constant depth. This result follows since the surface height undulations are so much smaller than undulations in the bottom topography, thus making the depth deviation and zstar coordinates very close to horizontal in most parts of the ocean.

Coordinate	Definition	Range	$z_{,s}$
geopotential	z	$-H \leq z \leq \eta$	1
z-deviation	$z' = z - \eta$	$-(H+\eta) \leq z' \leq 0$	1
z-star	$z^* = H\,(z-\eta)/(H+\eta)$	$-H \leq z^* \leq 0$	$1+\eta/H$
z-sigma	$\sigma = (z-\eta)/(H+\eta)$	$-1 \leq \sigma \leq 0$	$H+\eta$

Table 2.1. Table of vertical coordinates based on depth. These coordinates are naturally used for discretizing the Boussinesq equations.

6.2 Pressure based coordinates

The second class of vertical coordinates that we discuss is based on pressure. Pressure based coordinates provide a straightforward way to generalize Boussinesq depth based models to non-Boussinesq pressure models (Huang et al., 2001, DeSzoeke and Samelson, 2002, Marshall et al., 2003, Losch et al., 2004). The reason is that there is an isomorphism between the Boussinesq equations written in depth based coordinates and non-Boussinesq equations written in pressure based coordinates.

Pressure based vertical coordinates of interest include the following:

$$s = p \qquad \text{pressure} \tag{112}$$

$$s = \left(\frac{p - p_{\rm a}}{p_{\rm b} - p_{\rm a}}\right) \qquad \text{pressure-sigma} \tag{113}$$

$$s = p^o_{\rm b} \left(\frac{p - p_{\rm a}}{p_{\rm b} - p_{\rm a}}\right) \qquad \text{pressure-star.} \tag{114}$$

In these equations, p is the hydrostatic pressure, $p_{\rm a}$ is the pressure applied at the ocean surface from any media above the ocean, such as the atmosphere and sea ice, $p_{\rm b}$ is the hydrostatic pressure at the solid-earth lower boundary, and $p^o_{\rm b}$ is a time independent reference pressure, usually taken to be the bottom pressure in a resting ocean.[17] Since $p_{,z} = -\rho\, g < 0$ is single signed for the hydrostatic fluid, pressure provides a well defined vertical coordinate. Strengths and weaknesses of the corresponding depth based coordinates also hold for the pressure based coordinates, with the main difference being that pressure based models are non-Boussinesq.

A technical reason that the pressure based coordinates considered here are so useful for non-Boussinesq hydrostatic modelling is that $\rho\, z_{,s}$

[17] Note that equation (11.64) of Griffies, 2004 used the time dependent $p_{\rm b}$ rather than the time independent reference pressure $p^o_{\rm b}$. The former vertical coordinate has not been used in practice, and so we focus here on that coordinate defined with the reference pressure $p^o_{\rm b}$.

Coordinate	**Definition**	**Range**	$g\,\rho\,z_{,s}$
pressure	p	$p_\mathrm{a} \le p \le p_\mathrm{b}$	-1
p-deviation	$p' = p - p_\mathrm{a}$	$0 \le p' \le p_\mathrm{b} - p_\mathrm{a}$	-1
pstar	$p^* = p^o_\mathrm{b}\,(p - p_\mathrm{a})/(p_\mathrm{b} - p_\mathrm{a})$	$0 \le p^* \le p^o_\mathrm{b}$	$-(p_\mathrm{b} - p_\mathrm{a})/p^o_\mathrm{b}$
p-sigma	$\sigma = (p - p_\mathrm{a})/(p_\mathrm{b} - p_\mathrm{a})$	$0 \le \sigma \le 1$	$-(p_\mathrm{b} - p_\mathrm{a})$

Table 2.2. Table of vertical coordinates based on pressure. These coordinates are naturally used for non-Boussinesq dynamics.

is either a constant or a two-dimensional field. In contrast, for depth based models $\rho\,z_{,s}$ is proportional to the three-dimensional *in situ* density ρ, thus necessitating special algorithmic treatment for non-Boussinesq z-models (see the discussions in Greatbatch and McDougall, 2003 and Griffies, 2004). Table 2.2 summarizes some pressure based coordinates.

As Table 2.2 reveals, the specific thickness $z_{,s}$ is negative for the pressure-based coordinates, whereas it is positive for the depth-based coordinate (Table 2.1). The sign change arises since upward motion in a fluid column increases the geopotential coordinate z yet decreases the hydrostatic pressure p. To establish a convention, we assume that the thickness of a grid cell in z space is always positive

$$\mathrm{d}z = z_{,s}\,\mathrm{d}s > 0 \tag{115}$$

as is the case in the conventional z-models. With $z_{,s} < 0$ for the pressure-based vertical coordinates, the thickness of grid cells in s space is negative

$$\mathrm{d}s < 0 \qquad \text{for pressure-based coordinates with } z_{,s} < 0. \tag{116}$$

6.3 Isopycnal coordinates

Isopycnal models discretize the vertical into potential density classes. Some key advantages of isopycnal models are the following:

- Tracer transport in the ocean interior is well represented due to the natural ability of these models to maintain water mass properties.
- The bottom topography is represented in a piecewise linear fashion, hence avoiding the need to distinguish bottom from side as traditionally done with z-models.
- In some cases, flow near topographically critical regions, such as overflows, can be well resolved by isopycnal models due to the natural tendency of the coordinate surfaces to become refined in these regions.

- For a fluid with a linear equation of state, the horizontal pressure gradient can be easily represented.
- For an adiabatic fluid, the volume (for a Boussinesq fluid) or mass (for a non-Boussinesq fluid) between isopycnals is conserved.

Some of the disadvantages are the following:

- Representing the effects of a realistic (nonlinear) equation of state is cumbersome.
- The thermal wind balance is based on *in situ* density, not potential density. Hence, the further away from the reference pressure, the less accurate the pressure gradient force can be represented solely by the isopycnal gradient of the Montgomery.
- An isopycnal coordinate is inappropriate for regions where density becomes unstratified, such as mixed layers or deep convection regions.

Figure 13 illustrates isopycnal surfaces for a section in the model used to generate Figures 11 and 12.

6.4 Two algorithms

Adcroft and Hallberg, 2004 distinguish two classes of algorithms used to update the model state: quasi-Eulerian and quasi-Lagrangian. The main distinguishing characteristic of these algorithms is how they compute the dia-surface velocity component (Section 2.5). The two algorithm classes have traditionally been associated with two classes of vertical coordinates.

- Quasi-Eulerian algorithms diagnose their vertical velocity component from the *continuity equation*. Geopotential and sigma models have traditionally employed this approach.
- Quasi-Lagrangian algorithms set the vertical velocity component based on specified constraints, and they update the thickness betwen layers via the continuity equation to prognostically move layers around. Isopycnal vertical coordinate models typically use this approach. For example, adiabatic simulations with isopycnal coordinates set the diapycnal velocity component to zero, thus exactly preserving the integrity of the chosen density classes. For non-adiabatic simulations, the diapycnal flux is based on parameterizations of diabatic processes such as arise from the nonlinear equation of state or small scale mixing. A summary of these ideas

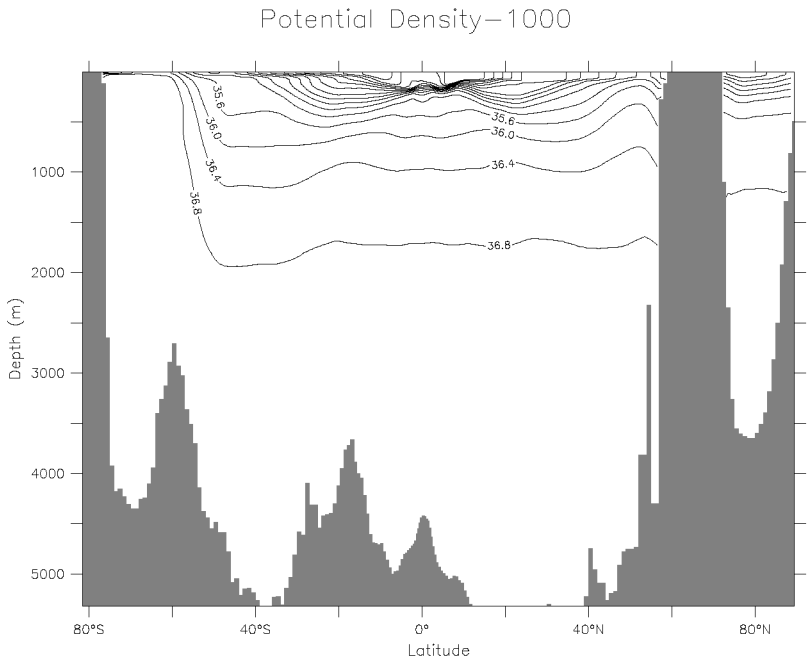

Figure 13. Constant potential density surfaces (minus 1000) in units of $\mathrm{kg\,m^{-3}}$. Potential density is here referenced to 2000db, which is the common reference for isopycnal models based on the work of Sun et al., 1999. Shown here is a section along $150°W$, as in Figure 11. Note the weak stratification in the deep, which is spanned by only one density layer. However, in a realistic isopycnal model, the choice of density classes used to partition the ocean would be non-uniform, in contrast to that used here. In that way, the model will have more layers in the deep and so will better represent interactions with the bottom topography than suggested by this figure.

can be found in Chassignet and Bleck's lectures in this volume, as well as Bleck's lectures in Chassignet and Verron, 1998.

Notably, it is possible to, say, design a z-coordinate model based on quasi-Lagrangian methods, or isopycnal models based on quasi-Eulerian methods. However, such has traditionally not been the case, with the distinctions mentioned above the usual situation.

There is presently no general consensus in the ocean modelling community regarding the best choice of vertical coordinate or the best al-

gorithm methodology. For example, those aiming to faithfully represent the ocean's quasi-adiabatic interior generally prefer an isopycnal layered model using quasi-Lagrangian methods over either terrain following or geopotential models using quasi-Eulerian methods. However, there have been decades of experience with z-models for global climate modelling, largely due to the simplicity of representing and parameterizing air-sea and ice-sea interactions as well as the ocean mixed layer. Hence, these models remain the dominant tool for global climate modelers, even given their well known problems with spurious mixing and difficulties handling overflow processes (see the discussion in Griffies et al., 2000a). Additionally, non-hydrostatic models, such as that from Marshall et al., 1997, have traditionally used geopotential coordinates. Indeed, there is presently no non-hydrostatic algorithm for use in the ocean that is based on a quasi-Lagrangian algorithm. That is, all layered models for the ocean are hydrostatic. Finally, those focusing on shallow ocean dynamics and estuaries have traditionally chosen terrain following coordinates due to their fidelity with bottom boundary layer processes. However, such models have only recently been employed for global climate studies, largely due to difficulties with pressure gradient errors (see Section 2 of Griffies et al., 2000a).

In summary, it is unlikely that modelers will arrive at one universally best vertical coordinate. Instead, vertical coordinates will remain chosen for the particular model application in mind. Modelers may, however, converge on an optimal algorithm methodology, especially if quasi-Lagrangian methods can be extended to non-hydrostatic models. In general, it is useful for model designs to evolve from being based on a single vertical coordinate, to model environments mentioned in Section 1.3 that are flexible enough to include many vertical coordinate algorithms.

7. Closing remarks

It is incumbent on ocean model designers and developers to provide a thorough and pedagogical rationalization of their codes, from the basic equations that the model aims to integrate, to the limitations of their subgrid scale (SGS) parameterizations. Likewise, it is essential that model users understand elements of the model algorithms and SGS parameterizations. The sophisticated and productive use of ocean models comes from a firm understanding of model fundamentals. It is hoped that through more schools such as this one or those documented by O'Brien, 1986, Chassignet and Verron, 1998, and others, as well as books on the subject of geophysical fluid modelling such as Haltiner and

Williams, 1980, Durran, 1999, Haidvogel and Beckmann, 1999, Kantha and Clayson, 2000a, Kantha and Clayson, 2000b, and Griffies, 2004, students aiming to use ocean models will readily learn to scrutinize the simulation's output in a scientifically sound and rational manner so as to improve the models, and ultimately to better understand and predict the ocean.

Acknowledgements

These lectures were prepared for the GODAE Summer School on Operational Oceanography held at La Londe Les Maures, France from September 20 to October 1, 2004. I thank the organizers Eric Chassignet and Jacques Verron for inviting me to give these lectures, and for developing a fine program in a wonderful place. Many students provided important feedback to help develop these lectures, as did Anne Marie Treguier. I also thank Alistair Adcroft, Bob Hallberg, and Trevor McDougall for their lessons over the years about generalized vertical coordinates.

References

Adcroft, A. and Hallberg, R. W. (2004). On methods for solving the oceanic equations of motion in generalized vertical coordinates. *Ocean Modelling*, page in press.

Adcroft, A., Hill, C., and Marshall, J. (1997). Representation of topography by shaved cells in a height coordinate ocean model. *Monthly Weather Review*, 125:2293–2315.

Adcroft, A., Hill, C., and Marshall, J. (1999). A new treatment of the coriolis terms in c-grid models at both high and low resolutions. *Monthly Weather Review*, 127:1928–1936.

Adcroft, Alistair and Campin, Jean-Michel (2004). Rescaled height coordinates for accurate representation of free-surface flows in ocean circulation models. *Ocean Modelling*, 7:269–284.

Arakawa, Akio and Lamb, Vivian R. (1977). The UCLA general circulation model. In Chang, Julius, editor, *Methods in Computational Physics: General Circulation Models of the Atmosphere*, volume 17, pages 174–265. Academic Press.

Black, Thomas L. (1994). The new NMC mesoscale eta model: description and forecast examples. *Weather and Forecasting*, 9:265–278.

Blayo, Eric and Debreu, L. (1999). Adaptive mesh refinement for finite difference ocean model: some first experiments. *Journal of Physical Oceanography*, 29:1239–1250.

Bleck, Rainer (2002). An oceanic general circulation model frame in hybrid isopycnic-cartesian coordinates. *Ocean Modelling*, 4:55–88.

Bryan, K. (1969). A numerical method for the study of the circulation of the world ocean. *Journal of Computational Physics*, 4:347–376.

Campin, Jean-Michel, Adcroft, Alistair, Hill, Chris, and Marshall, John (2004). Conservation of properties in a free-surface model. *Ocean Modelling*, 6:221–244.

Chassignet, Eric P. and Verron, J. (1998). *Ocean Modeling and Parameterization*, volume 516 of *NATO ASI Mathematical and Physical Sciences Series*. Kluwer Academic Publishers.

DeSzoeke, R. A. and Samelson, R. M. (2002). The duality between the Boussinesq and non-Boussinesq hydrostatic equations of motion. *Journal of Physical Oceanography*, 32:2194–2203.

Diansky, N. A., Bagno, A. V., and Zalensy, V. B. (2002). Global ocean circulation sigma-model and its sensitivity to the wind stress forcing. *Izvestia, Atmospheric and Oceanic Physics*, 38:477–494.

Durran, D. R. (1999). *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*. Springer Verlag, Berlin. 470 pp.

Ezer, T., Arango, H., and Shchepetkin, A. F. (2002). Developments in terrain-following ocean models: Intercomparisons of numerical aspects. *Ocean Modelling*, 4:249–267.

Gill, A. (1982). *Atmosphere-Ocean Dynamics*, volume 30 of *International Geophysics Series*. Academic Press, London. 662 + xv pp.

Greatbatch, R. J. and McDougall, Trevor J. (2003). The non-Boussinesq temporal-residual-mean. *Journal of Physical Oceanography*, 33:1231–1239.

Greatbatch, Richard J. and Mellor, G. L. (1999). An overview of coastal ocean models. In Mooers, C. N. K., editor, *Coastal Ocean Prediction*, volume 56 of *Coastal and Estuarine Studies*, pages 31–57. American Geophysical Union.

Griffies, Stephen M. (2004). *Fundamentals of ocean climate models*. Princeton University Press, Princeton, USA. 496 pages.

Griffies, Stephen M., Böning, C., Bryan, F. O., Chassignet, E. P., Gerdes, R., Hasumi, H., Hirst, A., Treguier, A.-M., and Webb, D. (2000a). Developments in ocean climate modelling. *Ocean Modelling*, 2:123–192.

Griffies, Stephen M. Gnanadesikan, Anand, Dixon, Keith W. Dunne, John P. Gerdes, Rüdiger, Harrison, Matthew J. Held, Isaac M. Pacanowski, Ronald C. Rosati, Anthony, Samuels, Bonita L. Spelman, Michael J. Winton, Michael, and Zhang, Rong (2005). Formulation of an ocean model for use in global climate simulations. *Ocean Modelling*, page in prep.

Griffies, Stephen M., Harrison, Matthew J., Pacanowski, Ronald C., and Rosati, Anthony (2004). *A Technical Guide to MOM4*. NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, USA. 337 pp.

Griffies, Stephen M., Pacanowski, R. C., and Hallberg, Robert W. (2000b). Spurious diapycnal mixing associated with advection in a z-coordinate ocean model. *Monthly Weather Review*, 128:538–564.

Griffies, Stephen M., Pacanowski, R. C., Schmidt, R. M., and Balaji, V. (2001). Tracer conservation with an explicit free surface method for z-coordinate ocean models. *Monthly Weather Review*, 129:1081–1098.

Haidvogel, D. B. and Beckmann, A. (1999). *Numerical Ocean Circulation Modeling*. Imperial College Press, London.

Hallberg, Robert W. (1997). Stable split time stepping schemes for large-scale ocean modeling. *Journal of Computational Physics*, 135:54–65.

Haltiner, G. T. and Williams, R. T. (1980). *Numerical Prediction and Dynamic Meteorology*. John Wiley and Sons, New York, USA.

Holloway, Greg (1999). Moments of probable seas: statistical dynamics of Planet Ocean. *Physica D*, 133:199–214.

Huang, R. X. (1993). Real freshwater flux as a natural boundary condition for the salinity balance and thermohaline circulation forced by evaporation and precipitation. *Journal of Physical Oceanography*, 23:2428–2446.

Huang, R. X., Jin, Xiangze, and Zhang, Xuehong (2001). An oceanic general circulation model in pressure coordinates. *Advances in Atmospheric Physics*, 18:1–22.

Jackett, D. R., McDougall, T. J., Feistel, R., Wright, D. G., and Griffies, Stephen M. (2004). Updated algorithms for density, potential temperature, conservative temperature, and freezing temperature of seawater. *Journal of Atmospheric and Oceanic Technology*, page submitted.

Kantha, L. H. and Clayson, C. A. (2000a). *Numerical Models of Oceans and Oceanic Processes*. Academic Press, New York, USA. 936 pp.

Kantha, L. H. and Clayson, C. A. (2000b). *Small Scale Processes in Geophysical Fluid Flows*. Academic Press, New York, USA. 883 pp.

Losch, M., Adcroft, A., and Campin, J.-M. (2004). How sensitive are coarse general circulation models to fundamental approximations in the equations of motion? *Journal of Physical Oceanography*, 34:306–319.

Madec, G. and Imbard, M. (1996). A global ocean mesh to overcome the North Pole singularity. *CD*, 12:381–388.

Marshall, J., Adcroft, A., Campin, J.-M., and Hill, C. (2003). Atmosphere-ocean modeling exploiting fluid isomorphisms. *Journal of Physical Oceanography*. in press.

Marshall, J., Hill, C., Perelman, L., and Adcroft, A. (1997). Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling. *Journal of Geophysical Research*, 102:5733–5752.

McDougall, T. J. (1987). Neutral surfaces. *Journal of Physical Oceanography*, 17:1950–1967.

McDougall, T. J. (1995). The influence of ocean mixing on the absolute velocity vector. *Journal of Physical Oceanography*, 25:705–725.

McDougall, T. J. (2003). Potential enthalpy: a conservative oceanic variable for evaluating heat content and heat fluxes. *Journal of Physical Oceanography*, 33:945–963.

McDougall, T. J., Jackett, D. R., Wright, D. G., and Feistel, R. (2003). Accurate and computationally efficient algorithms for potential temperature and density of seawater. *Journal of Atmospheric and Oceanic Technology*, 20:730–741.

Murray, R. J. (1996). Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics*, 126:251–273.

O'Brien, James J. (1986). *Advanced Physical Oceanographic Numerical Modelling*. D. Reidel Publishing Company.

Pacanowski, Ronald C. and Gnanadesikan, A. (1998). Transient response in a z-level ocean model that resolves topography with partial-cells. *Monthly Weather Review*, 126:3248–3270.

Shchepetkin, A. F. and McWilliams, J. C. (2002). A method for computing horizontal pressure-gradient force in an ocean model with a non-aligned vertical coordinate. *Journal of Geophysical Research*, 108:35.1–35.34.

Stacey, Michael W., Pond, Stephen, and Nowak, Zenon P. (1995). A numerical model of the circulation in Knight Inlet, British Columbia, Canada. *Journal of Physical Oceanography*, 25:1037–1062.

Sun, S., Bleck, R., Rooth, C., Dukowicz, J., Chassignet, E., and Killworth, P. D. (1999). Inclusion of thermobaricity in isopycnic-coordinate ocean models. *Journal of Physical Oceanography*, 29:2719–2729.

Thuburn, John and Haine, Thomas W. N. (2001). Adjoints of nonoscillatory advection schemes. *Journal of Computational Physics*, 171:616–631.

Chapter 3

MODELS OF THE OCEAN: WHICH OCEAN?

Anne Marie Treguier
CNRS, LPO, Plouzané, France

Abstract Physics actually represented in an ocean model depend on each model's resolution and its parameterization of subgridscale effects. This chapter is a review of parameterizations used in ocean models, focussing on operational ocean forecasting systems for the North Atlantic and Mediterranean Sea. This review is limited to z-coordinate models. A detailed presentation of the physics underlying each parameterization is out of the scope of this short chapter, but we try to discuss some uncertainties of the physical basis of current parameterizations. The concept of subgrid scale effects and some interesting properties of the diffusion equation are presented first. Because ocean turbulence is strongly anisotropic, parameterization in the vertical and horizontal (or isopycnal) directions differ and are presented separately. Special sections are devoted to bottom boundary layers, flow topography interactions, and the dynamical effects of mesoscale eddies.

Keywords: Parameterizations, ocean modelling, numerical models, subgrid scale physics, diffusion, viscosity.

A simplistic view of the surface ocean circulation is often found in geographical maps, with arrows displaying the direction and location of the main surface currents. Those are the large scale, wind driven currents which Sverdrup and Stommel, among others pioneers, have tried to understand. The first models of the wind-forced ocean circulation that they built were two-dimensional, used the simplified quasi-geostrophic equations, and the western boundary currents were viscous boundary layers.

This linear, viscous ocean is still what is represented in most climate models today. The sea surface height distribution in the South Atlantic

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 75-108.

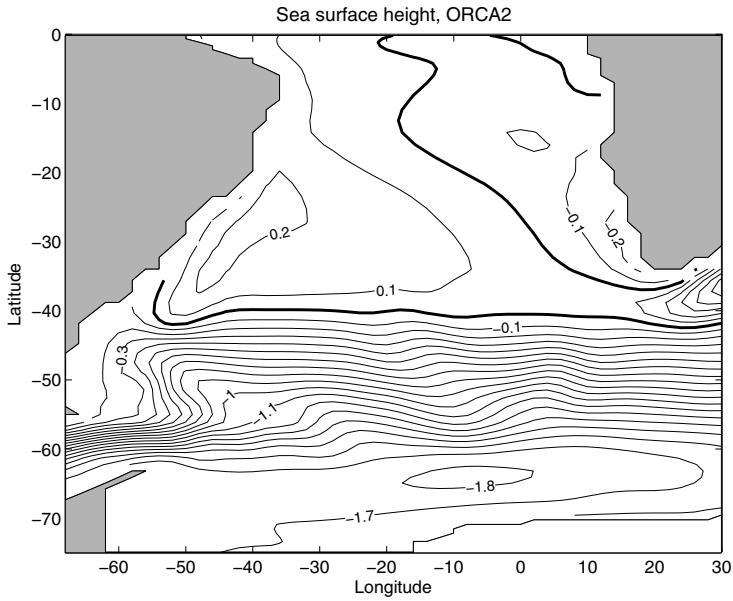

Figure 1. Sea surface height (ssh) in the south Atlantic in the global ORCA2 model (G. Madec). The ssh is averaged over one year, at the end of a 100 year experiment. Contour interval is 0.1 m, the zero contour is indicated in dark.

from the global 2° model ORCA2 (Fig. 1) is completely in agreement with the image of the ocean conveyed by simplified maps. Of course, the climate models are three dimensional, so that they are able to represent the global overturning circulation. In fact, the latter is often represented by a diagram of the "conveyor belt", similar to the sketchy geographical maps of surface currents.

Observations show that the real ocean is turbulent over a wide spectrum of spatial and temporal scales, and non-viscous. Today high resolution models begin to represent realistically the ocean we observe, and provide pictures in stark contrast with Fig. 1. One example is the POP 1/10° global model represented in Fig. 2 (Maltrud and McClean, 2004).

ORCA2 and POP 1/10 are so different that one may argue they do not represent the same ocean. A similar contrast exists between the ATL1 and ATL6 models of the CLIPPER group (Treguier et al., 2001). ATL1 and ATL6 are Atlantic models with 1° and 1/6° spatial resolution, respectively. Fig. 3 represents float trajectories during 5 years in the deep western boundary current of the South Atlantic, at 1800m. The coarse resolution ATL1 model depicts a sluggish western boundary current, with a well-defined southward velocity. Most of the floats reach

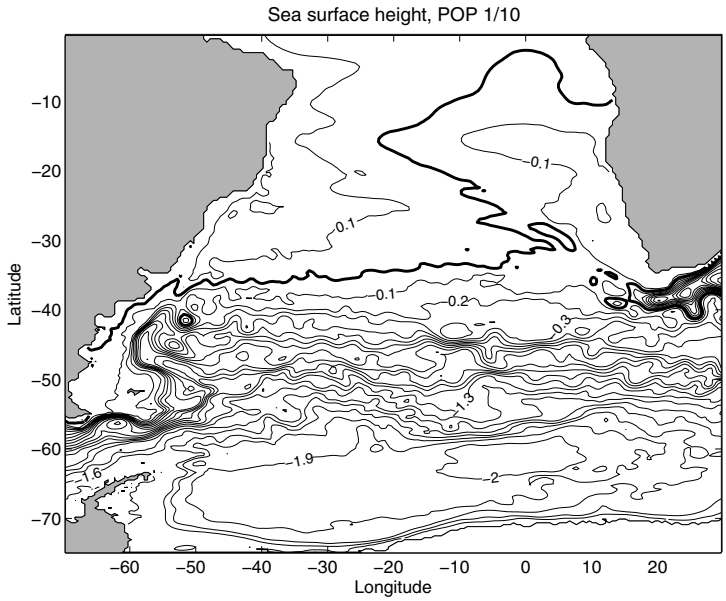

Figure 2. Sea surface height (ssh) in the south Atlantic in the global POP 1/10° model. The ssh is averaged over one year (fifth year of the experiment). Contour interval is 0.1 m, the zero contour is indicated in dark.

30°S after 5 years. In the ATL6 model, some trajectories go north instead of south because the western boundary current at that latitude often breaks down in a series of eddies (this has recently been observed by Dengler et al. (2004). Only one ATL6 float goes farther south than 30°S, but it gets there faster than the ATL1 floats, and many ATL6 floats escape into the interior of the ocean: this behavior is illustrative of chaotic mixing. The flow of the deep water in models like ATL1 is consistent with our simplified picture of the "conveyor belt"; high resolution models like ATL6 provide a picture much more difficult to interpret. They are closer to the real ocean, and yet too far from it to give us confidence in quantitative estimates. For example, the eddy kinetic energy at 2400m near 34°W, 22°S is 7.1 cm^2s^{-2} in ATL6 while a value of 62.1 cm^2s^{-2} has been measured there (Treguier et al., 2001). Underestimation of eddy kinetic energy at depths is very common in ocean models (Penduff et al., 2002).

It is important to realize that ATL1 and ATL6 are the same model, from the numerical and computational point of view. Both solve the same primitive equations, with the same code (OPA8.1, Madec et al. 1998). However, they are not models of the "same ocean", because they

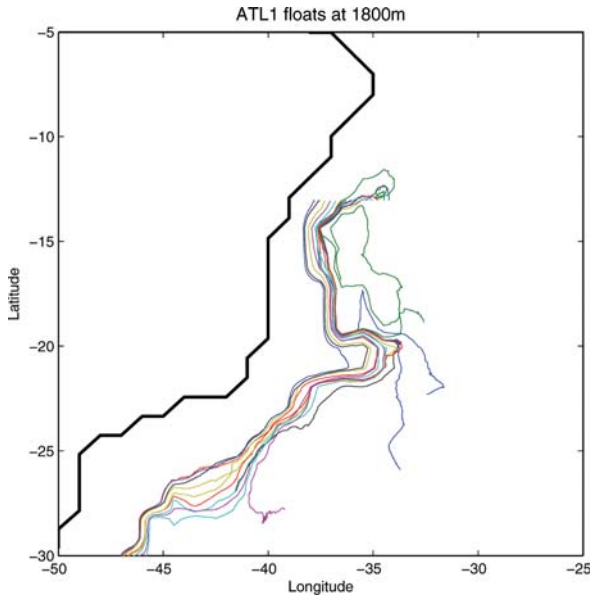

Figure 3. Trajectories of 15 numerical floats seeded at 1800 m, at 13°S every 1/4° from 37.75°W to 34.25°W. The trajectories are integrated during 5 years in the ATL1 1° model.

do not use the same parameterizations. In choosing the resolved spatial and temporal scale, and the parameterizations of the subgrid scales, the modeller effectively chooses the ocean he (or she) wishes to model. The present chapter discusses these choices.

When setting up an ocean model configuration, we have to ask ourselves which parameterizations are the most suitable, which coefficients to use for those parameterizations, and how changes in those coefficients would affect our solutions. We can answer surprisingly few of those questions for realistic ocean models, due to the complex interaction between different parameterizations (not to mention the interplay between physical parameterizations and numerical schemes). This is especially true for eddy permitting models for which extensive parameterization studies are not yet feasible due to the computational cost and the long time scales involved.

This chapter includes a discussion of sub-grid scale effects and general properties of the diffusion equation (part 2), and a discussion about parameterizations used in ocean models (parts 3 to 6). The parameterization issue has recently been the subject of a whole set of courses (Chassignet and Verron, 1998) It is discussed in the context of climate models in a paper by Griffies et al. (2000a), and also in Griffies' book

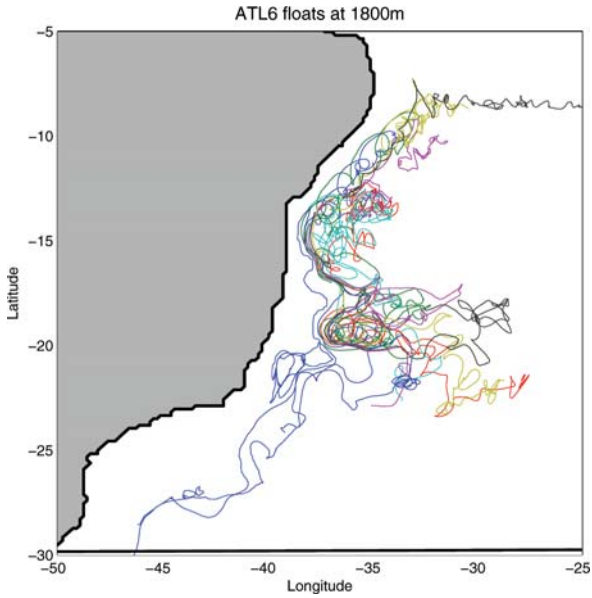

Figure 4. Same as Fig. 3, for the ATL6 1/6° model.

(Griffies, 2004) where the relationship between numerical schemes and parameterizations is analyzed in depth. In this short course I will survey parameterizations, hopefully providing a useful (albeit superficial) introduction to the more exhaustive material. Focus is on current practice, with little discussion of the underlying physical processes. The interested reader is referred to Chassignet and Verron (1998), or references therein.

1. Sub-grid scale effects in ocean models

1.1 Convergence of numerical solutions

We can write the prognostic equations of an ocean model in the general form:

$$\frac{\partial \mathbf{Y}}{\partial t} + \mathbf{V}.\nabla \mathbf{Y} + F(\mathbf{Y}) = 0 \tag{1}$$

where $\mathbf{Y} = (\mathbf{V}, T, S)$ is the vector of prognostic variables with $\mathbf{V}$ the velocity vector, T the potential temperature, S the salinity. The second term is nonlinear advection and the third term F represent all other terms, including external forcings. The equations must be discretized in order to be solved numerically: in ocean models this is usually done by choosing a mesh of grid points and using finite difference formulae.

The model solves for $\mathbf{Y_R}$ which is the resolved state of the ocean, on spatial scales of a grid cell, at discrete times. Using the terminology of Boer and Denis (1997), $\mathbf{Y_R}$ results from applying a "numeric resolution" operator $()_R$ to the state vector $\mathbf{Y}$. The definition of the "resolved" scales involves some kind of averaging: appropriate averaging operators for ocean dynamics are discussed extensively by Griffies (2004). Let us apply the operator $()_R$ to (1):

$$\begin{aligned} \frac{\partial \mathbf{Y_R}}{\partial t} &+ \mathbf{V_R}.\nabla \mathbf{Y_R} + F_R(\mathbf{Y_R}) = \\ &- ((\mathbf{V}.\nabla \mathbf{Y})_R - \mathbf{V_R}.\nabla \mathbf{Y_R}) - (F(\mathbf{Y})_R - F_R(\mathbf{Y}_R)) \,. \end{aligned} \quad (2)$$

We have to account for the effect of unresolved scales on the evolution of $\mathbf{Y_R}$ (the right hand side of (2)). When this effect is not represented correctly we make a *parameterization error*, which is different from the *numerical error* made by using a finite difference approximation in solving the left hand side of (2). Note that the advective contribution to subgrid scale effects (first term on the right hand side of (2)) does not vanish even in an inviscid fluid. This happens because turbulent motions usually generate a cascade of variance of the resolved quantity towards small scales (say, for a tracer, as discussed for example in Dubos and Babiano (2002). It is necessary for the parameterization to dissipate tracer variance to represent this cascade in the limited spectral space of a numerical model, even when the physical processes involved are related to stirring rather than mixing.

From a mathematical point of view, one would like to see the solution of an ocean model to converge as the resolution is increased (that is, progressive refinements of the resolution should bring smaller and smaller changes in the solution). However, when we refine the grid (as between Fig. 1 and Fig. 2) we also change the parameterizations on the right hand side, and thus we solve different equations. The huge differences in the solutions of ORCA2 and POP 1/10 do not come from a faulty numerical scheme; rather they come from the fact that the parameterizations differ. Let us note, however, that even the numerical (mathematical) convergence of z coordinates model solutions is not demonstrated, and indeed there are examples of non-convergence (Gerdes, 1993) due to the staircase representation of the topography.

Taking a physical point of view, convergence can be expected only over a range of scales where the dynamics of the flow remains qualitatively the same, so that the same parameterizations can be consistently applied. Atmospheric scientists have been able to set up test problems to look for the convergence of the dynamical core of their climate models (allowing representation of synoptic scale turbulence). Boer and Denis

(1997) present such a setting: an aquaplanet (no topography), a dry atmosphere, with a large scale forcing including prescribed heating and weak relaxation to a temperature profile. In the ocean it is much more difficult to find test problems that are relevant to climate. Two similar problems have been submitted to a convergence test. The first one is the flat-bottom, quasigeostrophic basin, relevant to the study of the upper ocean wind forced response (Siegel et al., 2001). The domain had a width of 3500 km with six layers in the vertical; the smallest dynamical spatial scale, the sixth internal Rossby radius, was close to 10 km. The second test case is a layered model of the North Atlantic (Hurlburt and Hogan, 2000) with 6 layers in the vertical, realistic coastline and topography restricted to the bottom layer. In both studies the authors still found significant differences between horizontal resolutions of 3 and 1.5 km (1/32°and 1/64°), either in energy and potential vorticity fluxes or in local aspects of the circulation. However, the differences were smaller than between lower resolution cases (say, between 1/8°and 1/16°), suggesting that the highest resolution cases approached convergence.

It is possible to relate the oceanic case to the atmospheric case considering the different dynamical scales (Rossby radii) in the two fluids. Boer and Denis (1997) consider in their test case that the dynamics have converged at T63, that is, a resolution of 1.87° (about 150 km at mid latitudes). This is 18% of the first internal Rossby radius R_o which is about 800 km in the atmosphere. An equivalent resolution in the ocean would be 7 km in the subtropics (R_o = 40 km) and 2 km in subpolar regions (R_o = 12 km). Those results suggest that none of today's basin scale models can be called "eddy resolving" in the subpolar regions, and that parameterizations should take into account the part of the mesoscale spectrum that is not resolved.

1.2 Subgrid scale turbulence

The first subgrid scale effects that usually come to mind are those related to the nonlinear advection terms, that is, the first term on the rhs of (2). To develop parameterizations one further assumes that

$$(V.\nabla \mathbf{Y})_R - V_R.\nabla \mathbf{Y_R} = (V'.\nabla \mathbf{Y}')_R, \tag{3}$$

where $\mathbf{Y}' = \mathbf{Y} - \mathbf{Y_R}$ is the subgrid scale part of $\mathbf{Y}$. This is true only if the "resolution operator" has the properties of a Reynolds average, which is not the case for a spatial truncation (among other properties, a Reynolds average commutes with spatial and temporal derivatives, and the average of the deviation $\mathbf{Y}'$ is zero). Assuming a Reynolds decomposition (for lack of something more accurate) the subgrid scale effects appear as the divergence of eddy fluxes. Equations can be written for

those eddy fluxes, such as the Turbulent Kinetic Energy (TKE) equation. They involve higher order moments of the turbulent variables so that a "closure hypothesis" is required to solve them: this consists in using an empirical relationship to express higher moments in term of the lower-order moments. A classical example of closure model for vertical mixing in the ocean is provided by Mellor and Yamada (1982). The simplest closure applies to the advection of a passive tracer by homogeneous and isotropic turbulence. Eddy fluxes in that case can be modelled by analogy with the molecular diffusivity (Fickian hypothesis): for example

$$(w'T')_R = -\kappa \frac{\partial T_R}{\partial z}. \tag{4}$$

The vertical eddy temperature flux is down the gradient of resolved temperature.

It is usually assumed that the ocean turbulence is isotropic at the centimeter scale, so this simple parameterization would apply. At larger scales, the physical processes that one needs to parameterize are more complex and no longer isotropic. The first ingredients that break isotropy are the effects of gravity and stratification. Stratified fluid supports internal waves, which can carry energy far from their generation site; stratification inhibits cross-isopycnal motion and cross-isopycnal mixing. Furthermore, when it is unstable, stratification generates convective instabilities. These must be parameterized regardless of the grid scale and time step of the model when the hydrostatic approximation is made (it is the case in primitive equation models). An additional physical process in the ocean is the double diffusive convection arising from the different molecular diffusivities of heat and salt. Going to larger scales, the earth rotation comes into play (time scale of one day, horizontal scale of hundreds of meters). It creates the possibility of resonant inertial motions, and further inhibits vertical motion. Finally, at the mesoscale, the variation of the Coriolis parameter with latitude is important. The vanishing of the Coriolis force at the equator makes it a waveguide and allows inertial instability. The β effect at mid latitudes tends to favor zonal motions and inhibit meridional mixing. All those physical processes are reviewed in detail in the book edited by Chassignet and Verron (1998).

In three dimensions, a linear relationship as (4) between local eddy fluxes and local mean gradient components can be expressed as the product of the gradient vector by a matrix:

$$(v_i'T')_R = -T_{ij} \frac{\partial T_R}{\partial x_j}, \tag{5}$$

where v_i are the velocity components and T_{ij} is the mixing tensor. Mathematically, the tensor can be decomposed as the sum of a symmetric part

K_{ij} and an antisymmetric S_{ij} part. With classical isotropic diffusion, K_{ij} is diagonal with mixing coefficient κ along the diagonal, and S_{ij} is zero. Taking into account the anisotropy of ocean motions requires a different coefficient for horizontal and vertical mixing. More generally, the symmetric tensor K_{ij} can be diagonalized along principal mixing directions; in the ocean those are assumed to be along and across isopycnal (isoneutral) directions respectively (see section 5.2). The corresponding paramerization in the temperature and salinity equations for an ocean model is called "isopycnal laplacian diffusion", by contrast with horizontal diffusion.

One way to understand the antisymmetric part S_{ij} is the following. With eddy fluxes defined by (5) the equation for the resolved temperature T_R includes the divergence of the eddy fluxes, with a contribution from the antisymmetric tensor written as:

$$\nabla(-S_{ij}\frac{\partial T_R}{\partial x_j}). \tag{6}$$

It is easily demonstrated that this term is identical to an advection of T_R by a velocity V^* with components defined by:

$$v_i^* = \frac{\partial S_{ij}}{\partial x_j}. \tag{7}$$

As a consequense, in our idealized framework of a linear relationship between eddy fluxes and mean gradients, parameterizations can be classified in three components: the vertical (cross isopycnal) and the lateral (isopycnal) mixing associated with the symmetric tensor K_{ij}, and the advective eddy effect associated with S_{ij}. Those three components will be considered in turn in sections 3, 5 and 6 of this chapter. The interested reader will find a complete discussion of the mixing tensor (as well as an alternative presentation using the notion of skew flux) in Griffies (2004).

1.3 The diffusion equation

Parameterizations often assume a flux gradient relationship like (4), and look like a diffusion. It is important to realize that the diffusion equation has some unexpected properties when the mixing coefficient is allowed to vary. Let us consider for example the evolution of a vertical profile of potential temperature, when the vertical eddy flux is parameterized by (4). The initial temperature perturbation T is sinusoidal over a depth H and evolves according to the equation:

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(\kappa\frac{\partial T}{\partial z}\right) = \kappa\frac{\partial^2 T}{\partial z^2} + \frac{\partial \kappa}{\partial z}\frac{\partial T}{\partial z}. \tag{8}$$

With constant κ this is a classical diffusion equation and the perturbation decays with a characteristic time $\tau = H^2/(\pi^2\kappa)$.

When κ varies vertically, the second term on the right-hand side of (8) is non zero. It is similar to a vertical advection with velocity

$$w_\kappa = -\partial\kappa/\partial z.$$

This term can lead to a sharpening of the large scale gradients (P. Klein, personal communication). To see this, let us consider the equation for the temperature gradient T_z, obtained by taking the vertical derivative of (8):

$$\frac{\partial T_z}{\partial t} = \kappa\frac{\partial^2 T_z}{\partial z^2} + 2\frac{\partial\kappa}{\partial z}\frac{\partial T_z}{\partial z} + \frac{\partial^2\kappa}{\partial z^2}T_z. \tag{9}$$

The first term is the diffusion, the second the advective contribution, and the third term can cause an exponential growth of the temperature gradient when $\partial^2\kappa/\partial z^2$ is large enough. This happens if κ varies more rapidly in space than T. Let us assume, for instance, that the initial T profile has some small scale variations superimposed on it, and that the physical processes generating mixing are very sensitive to the presence of those small scales. This situation is displayed in Fig. 5. The mixing cofficient has large values where the small scales are present, in the upper third part of the water column. Instead of decaying, the profile of temperature after 120 days has a much stronger gradient. The T profile, initially of typical scale H, varies now with the typical spatial scale of the κ profile. The temperature perturbation has also been advected downwards. Note that although a sharpening of the gradient has occurred, diffusion has smoothed the local extrema in the initial profile as expected (this property of the diffusion equation is independent of the structure of the diffusion coefficient).

Even more interesting things happen when κ is a nonlinear decreasing function of $\partial T/\partial z$. When κ is nonlinear enough, the parameterization can generate discontinuities and staircases in the temperature profile. Fig. 6 shows the final state of the evolution of eq.(8) with κ proportional to $\exp(-(dT/dz)^2)$. This effect was noted by Phillips (1972) and more recently by Ruddick et al. (1989)

Letting κ be a decreasing function of the vertical temperature gradient is precisely what parametrizations of vertical mixing do: stratification inhibit vertical mixing by providing a strong restoring force (buoyancy force), thus limiting vertical displacements. Most parameterizations of vertical mixing are based on the Richardson number of the large scale flow. Let us define first the Vaisala frequency N:

$$N^2 = \frac{-g\partial\rho/\partial z}{\rho_0},$$

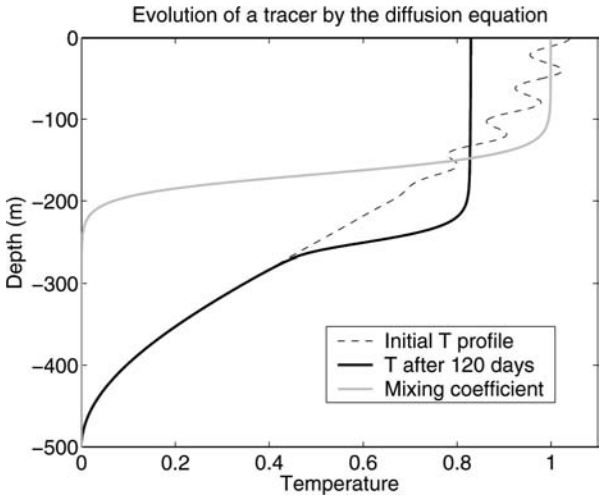

Figure 5. Initial and final solution of the diffusion of a tracer according to (8) with no flux boundary conditions, when $\kappa = 0.005 \tanh(\alpha(z - H/3)) + 1)$ $\mathrm{m}^2.\mathrm{s}^{-1}$, with $H = 500$ m and $\alpha = (0.05H)^{-1}$. The profile of the mixing coefficient κ is multiplied by 100 to be displayed on the same scale as T.

where ρ is density and g gravity. The Richardson number is:

$$Ri = \frac{N^2}{(\partial u/\partial z)^2 + (\partial v/\partial z)^2},$$

with u and v the horizontal components of velocity. This dimensionless number expresses the competition between the stabilizing effect of stratification and the destabilizing effect of the shear. Parameterizations of vertical mixing always produce mixing coefficients that are strongly nonlinear functions of the Richardson number, displaying an almost "step-like" behavior with strong mixing at low Richardson numbers and little mixing for Richardson numbers above critical (see for example fig 23 of Blanke and Delecluse, 1993, or Fig. 5 of Large, 1998). This behavior is sound physically and is observed in the ocean, but may create numerical problems. Modellers need to be aware of the profound implications of spatially variable mixing coefficients.

1.4 Subgrid scale effects of external forcings and boundary conditions

Besides the nonlinear interactions inside the fluid itself, external forcings also generate subgrid scale effects: it is the case for ocean-atmosphere interactions. For example, heat fluxes and evaporation depend on the

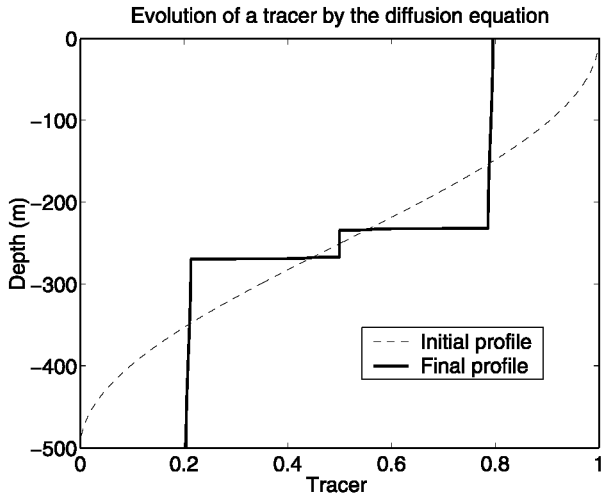

Figure 6. Initial and final solution of the diffusion of a tracer according to (8) with no flux boundary conditions, when $\kappa = 0.01 \exp(-(0.59 H dT/dz)^2)$, with $H = 500$ m.

sea surface temperature, and can be very different above mesoscale eddies. Arhan et al. (1999) estimate an average heat loss of 620 W.m^{-2} above an Agulhas eddy in 6 months, much higher than climatological values in the area. No attempts have been made yet to parameterize this effect in climate models where eddies are absent. In eddy resolving ocean models, subgrid scale effects arise because of the low resolution of the forcing fields or the atmospheric models used for coupling. The mesoscale response of the atmosphere to SST perturbations is ignored in such models. Finally, according to the temporal resolution of the forcing fields, there may be non-resolved time scales as well: for exemple the effect of wind bursts, or the diurnal cycle of radiative forcing. These sub-grid scale effects will not be discussed further here but should be kept in mind.

Perhaps the most important and complex sub-grid scale effect arises through the boundary conditions, namely the shape of the ocean basins. The first example is the communication between ocean basins and semi-enclosed seas: according to the spatial resolution of the model, it is possible or not to represent some straits. Some aspects of the parameterizations of subgrid scale topography are presented in section 4.

2. Parameterizations in the vertical

After introducing sub-gridscale effects, let us now review parameterizations, considering in turn the vertical direction (this section), bottom

Parameters	PSY2	FOAM	MFS-MOM	MFS-OPA
Code	OPA	MOM	MOM	OPA
Time step	δt = 800s	δt =1200s	δt =900s	δt =600s
Domain	N. Atlantic+Med	N. Atlantic	Mediterranean	
Horizontal grid:				
Max δx	7 km	12 km	12 km	6 km
Min δx	3 km	12 km	9.8 km	4.9 km
Vertical grid:				
levels	43	20	31	72
Max δz	300 m	615 m	300 m	300 m
Min δz	6 m	10 m	10 m	3 m
Lateral boundary condition:	partial-slip	no-slip	no-slip	no-slip

Table 1. Grid and domain for four z-coordinate models used in forecasting systems: PSY2 (MERCATOR, France), FOAM (U.K. Met Office) and two MFS systems (Italy)

and topographic effects (section 4), mixing laterally or along isopycnals (section 5) and dynamical effects of mesoscale eddies (section 6). Because this book is about operational oceanography, I will consider as examples three z-coordinates ocean models that are currently part of an operational forecasting system (table 1).

The parameterizations used in those models are listed in table 2. Three of the models, PSY2, MFS-OPA and MFS-MOM have quite similar, rather simple parameterizations. FOAM has more complex parameterizations, mainly because this eddy permitting model has been derived from a lower resolution climate model (Gordon et al., 2000). It would be interesting to know the impact that those more elaborate parameterizations have on the results of the forecasting system.

In this section we discuss the parameterization of processes that cause vertical mixing in the surface boundary layer or in the interior: convective mixing, shear instabilities, inertial waves breaking, double diffusion. It is important to note that the parameterizations of those processes are often lumped together in one package. This is the case of the "KPP" parameterization (Large et al., 1994). Its originality lies in the parameterization of the surface mixed layer, using a prescribed vertical profile of fluxes adjusted on a mixed layer depth (K-profile), but Large et al. (1994) also propose a parameterization of convection, interior and double diffusive mixing. This is why the parametrizations for vertical diffusivity and viscosity in the interior and in the upper mixed layer are lumped together in table 2.

	PSY2	FOAM	MFS-MOM	MFS-OPA
Vertical diffusivity				
Background	1. 10^{-5}m^{2}s^{-1}		3. 10^{-5}m^{2}s^{-1}	3. 10^{-5}m^{2}s^{-1}
Ri dependent	TKE	PP81	none	none
Other	Surf.E	KPP+KT67	none	none
Vertical viscosity				
Background	1. 10^{-4}m^{2}s^{-1}		1.5 10^{-4}m^{2}s^{-1}	1.5 10^{-4}m^{2}s^{-1}
Ri dependent	TKE	PP81	none	none
Other	Surf.E	KPP+KT67		
Convection				
Adjustment	no	yes	yes	no
Enhanced mixing	1 m^{2}s^{-1}			1 m^{2}s^{-1}
Lateral diffusivity				
Hor. Biharmonic	3 10^9m^{4}s^{-1}	none	1.5 10^{10}m^{4}s^{-1}	3 10^9m^{4}s^{-1}
Spatial variation	$\propto \delta x^3$		Constant	Constant
Laplacian	none	100 m^2.s^{-1}	none	none
Orientation	hor	iso	hor	hor
slope limitation	no	GE91	no	no
Horizontal background	no	10 m^2.s^{-1}	no	no
Lateral viscosity				
Biharmonic	9 10^9m^{4}s^{-1}	2.6 10^9m^{4}s^{-1}	5 10^9m^{4}s^{-1}	5 10^9m^{4}s^{-1}
Laplacian	none	30 m^2.s^{-1}	none	none
Spatial variation	$\propto \delta x^3$	no	no	no
Bottom friction				
Type	Quadratic	Quadratic	none	quadratic
Coefficient	C_d =1.3 10^{-3}	C_d =1.225 10^{-3}		C_d =10^{-3}
Bottom boundary layer				
Type	none	GO00	none	none

Table 2. Parameterizations for four z-coordinate models used in forecasting systems. Schemes are TKE (Blanke and Delecluse, 1993), PP81 (Pacanowski and Philander, 1981), KPP Large et al., 1994, KT67 (Kraus and Turner, 1967). Note that in FOAM KPP is a modified version (Gordon et al., 2000). Surf.E is an enhancement of background coefficients near the surface. The orientation is either horizontal (hor) or iso-neutral (iso). Note that locally, neutral and isopycnal directions are identical. In the case of isopycnal mixing, modellers have to modify the algorithm whe slopes are too steep. In FOAM this is done using the scheme GE91 of Gerdes (1993). The Bottom boundary layer (GO00) used in FOAM is documented in Gordon et al. (2000).

2.1 Local and non-local parameterizations

As reviewed in detail by Large (1998) parameterizations can be classified into local and nonloncal. Local parameterizations following (4) assume that the eddy fluxes depend on the *local* properties of the large scale flow. They are often based on one-dimensional turbulence closure models, like the TKE model of Blanke and Delecluse (1993). The one-dimensional "stand-alone" models are implemented with grid spacing of order one meter in the vertical; it is unclear how they perform with the typical grid spacing of ocean models (5-10 m at the surface, quickly increasing to 20-50 m at 100 m depth). It is important to keep in mind that the classical Ekman layer depth is $h_e = \sqrt{2\nu/f}$. At mid latitudes, the vertical viscosity ν has to be larger than 5. $10^{-3}\mathrm{m}^2.\mathrm{s}^{-1}$ for h_e to be larger than 10 m (that is, for the Ekman depth to be larger than the first model layer thickness). In the absence of high frequency forcing, in the absence of night time convection, and with low vertical resolutions, turbulent closures cannot produce high enough mixing at the top layer interface. This explains why non-local parameterizations are attractive, like the old Kraus-Turner (Krauss and Turner, 1967) parameterization and the new KPP scheme (Large et al., 1994) used in the FOAM model (Table 2). This is also the rationale for "ad-hoc" fixes like the increase of the background coefficient in the upper layers found in the PSY2 model.

The present versions of the MFS models do not use any parameterization of the surface mixed layer (table 2). In that case, convection is the only source of enhanced mixing. Convection driven by surface cooling allows to reach realistic mixed layer depths in winter in the Mediterranean sea (see Crosnier and Le Provost in this volume). In summer, a shallow convection is driven by the penetration of incoming short wave radiation which warms the water down to a depth of about 15 m, while the outgoing longwave radiation cools the top layer only. Without penetrative solar radiation the mixed layer depth in MFS in summer would be restricted to the first model layer.

The performance of different vertical mixing schemes in realistic ocean models is not well documented, so that it is very difficult to make an objective choice among the different parameterizations. The use of KPP instead of crude parameterizations, like an imposed uniform mixed layer depth, clearly brings an improvement (Large, 1998). An improvement was also found by Blanke and Delecluse (1993) when using TKE instead of the Richardson-dependent scheme of Pacanowski and Philander (1981). On the other hand, a recent comparison of KPP with the TKE scheme (Chanut and Molines, 2004) shows little difference in the 1° CLIPPER Atlantic model. What makes the picture even fuzzier is

the fact that changing tunable constants within one parameterization package has significant effects (Matteoli, 2003). This lack of thorough sensitivity studies, especially for eddy permitting models, will stand out as we discuss the different processes leading to vertical mixing.

2.2 Convection

Early primitive equation models (Cox, 1984) represented convection by an iterative adjustment, which modified temperature and salinity in a water column. Adjustment schemes have convergence problems in some cases and tend to be costly; moreover the time scale of adjustment is one time step, which is too short in high resolution configurations. Despite these shortcomings, they are still in use in some forecasting models (Table 2).

Nowadays convective adjustment is more frequently represented by increasing the vertical mixing coefficient to a very large value in the case of convection. This procedure has been found to be a satisfactory parameterization of the effect of convective plumes by Klinger et al. (1996), with a mixing coefficient of 10 $m^2.s^{-1}$. Scalings suggest values up to 50 $m^2.s^{-1}$ (Send and Käse, 1998). The PSY2 model uses a coefficient of 1 $m^2.s^{-1}$ (table 2); the ORCA2 model uses 100 $m^2.s^{-1}$. Users of the KPP scheme take values from 0.1 to 10 $m^2.s^{-1}$.

The criterion for the onset of convection varies among models; convection is active as soon as the Vaisala frequency N^2 is negative in some models (PSY2) while the criterion in KPP is $N^2 < -0.2\ 10^{-4} s^{-2}$. The relative mixing of momentum and tracers also varies between models. Momentum should be mixed like tracers in convective plumes if the time scale t_{mix} for a parcel to move down the plume is shorter than the $1/f$, the time for geostrophic adjustment. With plume vertical velocities w of order 3 to 10 cm/s (Klinger et al., 1996), $t_{mix} = h/w$ reaches 12 h for deep convection, thus comparable to $1/f$. Tests performed with the ORCA2 model (Matteoli, 2003) show important differences in mean surface velocities (up to 10 $cm.s^{-1}$) in the Antarctic circumpolar current, with and without momentum mixing in the case of convection.

2.3 Interior mixing

As emphasized in the review by J. Toole (Toole, 1998), observations have shown increased levels of mixing in the abyss and over rough topography, compared with the low values found in the thermocline by microstructure measurements and tracer releases. More recently, the role of internal tides as an energy source for mixing has been empha-

sized. Maps of energy flux have been derived from tidal models leading to parameterizations of vertical mixing (Laurent et al., 2002)

Spatially variable vertical mixing coefficients based on topographic roughness or tidal mixing have yet to be tested extensively in models. Studies by Hasumi and Suginohara (1999) and Simmons et al. (2004) show modest improvements in coarse resolution models integrated to equilibrium. The effect of such parameterizations over shorter time scales in higher resolution models needs to be assessed, since uniform mixing coefficients are clearly not acceptable based on the observations.

One has to be aware that z coordinate models have difficulty achieving the vertical mixing coefficients of order $10^{-5}\mathrm{m}^2.\mathrm{s}^{-1}$ in the thermocline, depending on their advections scheme (Griffies et al., 2000b). Since the last decade modellers tend to abandon centered advection schemes, which lead to the generation of unphysical temperatures and salinities near fronts (note, for example, that the ATL6 model of Fig. 4 has one grid point with temperatures lower than -3°C at the bottom of the Faroe Bank channel outflow; salinities close to 42 PSU are found downstream of Gibraltar in the 1/10° model of Smith et al., 2000). Diffusive schemes like FCT (Flux Corrected Transport) avoid such problems, but in eddy permitting z-models they can cause a large amount of diapycnal mixing (Griffies et al., 2000b). The same is true for non-eddy resolving models when the western boundary current is marginally resolved: Griffies et al. (2000b) find spurious diapycnal mixing of 3 $10^{-4}\mathrm{m}^2.\mathrm{s}^{-1}$ due to the advection scheme in that case.

2.4 Double diffusive mixing

Double diffusion occurs in stably stratified situations, either when warm and salty water overlies cold, fresh water (salt fingering) or when cold, fresh water overlies warm and salty water (diffusive convection). Those processes generate turbulent mixing of heat and salt with different coefficients, dependent on the density ratio $R_\rho = \alpha\partial_z T/\beta\partial_z S$, where α and β are coefficients of thermal expansion and saline contraction. A parameterization has been proposed by Large et al. (1994) but it has not been thoroughly tested. A slightly different one has been proposed by Merryfield et al. (1999) and tested in a coarse resolution model, showing and improvement in the representation of water mass temperature and salinity although the effect on the circulation was small. None of the forecasting models listed in Table 2 uses a parameterization for double diffusion, even though this dynamical process is important in the ocean (Schmitt, 1998). A better and cleaner representation of "background" interior vertical mixing may be needed in z-coordinate models before

adding a parameterization of double diffusion can have a demonstrable positive impact on the solutions.

3. Bottom boundary layer and topographic effects

3.1 Bottom friction

Two-dimensional geostrophic turbulence has the property that energy cascades towards large scales, and enstrophy (the relative vorticity squared) cascades towards small scales (Batchelor, 1969). Thus, if eddies have their energy source at a scale close to the internal Rossby radius, nonlinear interactions tend to transfer this energy to larger scales where it must be dissipated. Viscous bottom drag can provide the energy sink which is required to equilibrate the flow. It is thus necessary to include a parameterization of bottom drag in eddy-resolving models. The strength of the bottom drag can have an influence on the spatial organisation of the flow because it affects the baroclinic instability of eastwards jets (Riviere et al., 2004).

Another interesting effect of bottom drag happens in overflow regions. In a rotating fluid, and under the hydrostatic approximation, dense plumes have a strong tendency to follow isobaths rather than plunging. A high bottom friction makes the flow less geostrophic and increases the rate of descent of the plumes (Stratford and Haines, 2000).

Despite such important dynamical effects, there does not seem to be any study documenting the effect of bottom drag in high resolution basin flows, especially in the presence of bottom topography. Most models use a quadratic bottom drag with constant coefficient (Table. 2).

3.2 Effects of overflows

Overflows are currents from marginal or semi-enclosed seas into the main ocean basins, through sills or along continental slopes. They set the properties of many water masses (Price and Yang, 1998). A major difficulty in modelling overflows is that many of them are subgrid scale physics for a given choice of resolution: the width of the strait or channel is narrower than the grid size. Note that straits can be one grid point-wide on a staggered "C" grid such as used in the OPA code, but two grid points are necessary to allow throughflow on a "B" grid with no-slip boundary condition (see Haidvogel and Beckmann, 1999, for a definition of staggered grids). The modeller may decide to open too wide a strait. In that case, the transport may be too large due to the exaggerated cross-section. G. Madec (personal communication) decreases the grid

size locally at a strait: but this trick is possible only when the strait is between land points because the grid size $\delta x, \delta y$ does not depend on z. It cannot be used for deep fracture zones in the middle of ocean basins.

The influence of overflow waters is especially important in the North Atlantic, with North Atlantic Deep water (at depths of 2000-3000 m) coming over Denmark Straits and the Faroe-Scotland ridge, Mediterranean water (1000 m depth) coming through Gibraltar, and Antarctic bottom water (4000 m depth) spreading over sills in the mid-Atlantic ridge (Romanche and Vema fracture zones, for example). The overflows are very badly represented in z coordinate models with staircase topography, generally leading to excessive mixing (Willebrand et al., 2001). There are simple models (such as streamtubes) to calculate exchange between basins in simplified cases, that could be the basis for parameterizations (Price and Yang, 1998). The problem with this method is that each overflow must be specified at a given grid cell or set of grid cells, which is quite cumbersome in a world ocean. Another drawback is that such a parameterization introduces grid scale sources and sinks that may not be handled well by the numerics.

Modellers look for parameterizations valid everywhere in the domain, such as the "Bottom boundary layer" (BBL) parameterization (Beckmann and Döscher, 1997). Since this pioneering work, BBL parameterizations have been implemented in many ocean models. However, we don't have yet a complete picture of their efficiency, depending on model characteristics. Dengg et al. (1999) find that a BBL parameterization induces a large improvement in a 1/3° model of the Atlantic using a centered advection scheme. However in the 1/6°model of the Atlantic we find that the improvement is modest, although the BBL parameterization is similar and both models use the same isopycnal mixing of tracers (Fig. 7). What causes the different performance of the BBL parameterization in those two models is unclear; numerical details may matter.

The FOAM model was developped from the HadCM3 ocean component at 1.8° resolution. In the latter, a variant of a diffusive BBL scheme was implemented that dramatically improved the representation of the Nordic seas overflows (Gordon et al., 2000). In this scheme, when bottom water at a grid cell is denser than the deeper water colums around, the algorithm looks for the level of neutral buoyancy of that bottom water and mixes the dense water into that model level. Lighter water is then moved up in the water column to replace the dense water. The behavior of this scheme in the 1/9° version of FOAM has no yet been evaluated in detail.

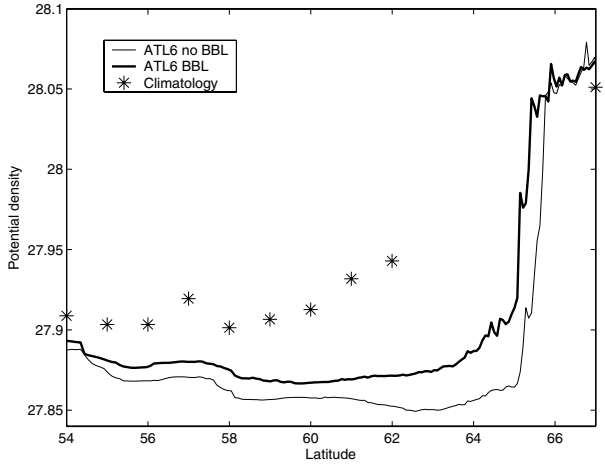

Figure 7. Zonal maximum of bottom density in the Irminger Sea (45°W-25°W), in the climatology and in two ATL6 experiments with and without BBL (the average of the 13rd model year is used).

A new difficulty has appeared with the generalization of the "partial cell" representation of bottom topography in z coordinate models. BBL parameterizations generate fluxes of tracers between bottom cells situated in neighbouring fluid columns. With partial cells the thickness of those bottom cell can vary widely, which may introduce spurious noise in the BBL fluxes.

In this chapter we have chosen examples from z-coordinate models only, but the choice of vertical coordinate is very important for the representation of overflows (Griffies 2005, this book). σ coordinate models handle overflows very well provided their vertical resolution near the bottom is good enough. This was not the case in the DYNAMO σ model, (Willebrand et al., 2001), but one example is the model of the Mediterranean outflow by Jungclaus and Mellor (2000). It is possible in σ models to retain a spatially homogeneous vertical resolution in the bottom boundary layer, which seems ideal for overflow representation, but does have an extra numerical cost.

3.3 Other flow-topography interaction

The interaction of flow with subgrid scale topography can generate internal waves, which can propagate in the water column and increase vertical mixing if they break. This process is generally parameterized as part of the vertical mixing due to internal waves, which has been discussed in section 3.4.

Regarding low frequency motions, the statistical effect of unresolved topographic roughness has been explored in the framework of the quasi-geostrophic (QG) equations, starting with Rhines (1977). The main effect of bottom roughness is to scatter the barotropic energy into baroclinic modes, and decrease the energy of mesoscale motions in the deep layers. There has been no attempt to parameterize this effect in ocean models. On the contrary, the effect of bottom roughness is probably overestimated already in standard z-coordinate models with unsmoothed staircase topography. Penduff et al. (2002) show that in such a model the eddy kinetic energy below 1000 m is lower than in a σ-coordinate model, the latter being in better agreement with observations. By performing sensitivity experiments with the z model they show that the grid-scale topographic roughness is responsible for a too rapid decay of the eddy kinetic energy with depth.

Beside allowing overflows, deep passages and fracture zones often act to sharpen and focus fronts. This effect can influence the whole water column when major currents cross topographic ridges. One example is the flow of the North Atlantic current across the Mid Atlantic ridge, which seems to be distributed in three branches corresponding to three fracture zones (Bower et al., 2002). There is no parameterization of this effect in low resolution models.

4. Lateral mixing parameterizations

4.1 Prandtl number

Let us assume that lateral momentum and tracer mixing are parameterized as laplacian operators with turbulent viscosity ν and diffusivity κ. The ratio of viscosity to diffusivity is the Prandtl number, $Pr = \nu/\kappa$. For molecular viscosity and heat diffusivity in sea water, it varies from 13 (at 0°C) to 7 (at 20°C). Molecular values are irrelevant at the scale of ocean models, and the Prandlt numbers used in models parameterizations vary widely. This is not based on physics but rather the result of numerical stability constraints which seem more stringent on viscosity than diffusivity. In low resolution climate models, Prandtl numbers as high as 50 can be found.

At the scale of quasi-geostrophic eddies, the Prandtl number could be one if one accepts that QG eddies essentially mix potential vorticity. In that case, mixing of vortex stretching (with diffusivity κ) has to be the same as mixing of relative vorticity (with viscosity ν). At the submesoscale, I do not know of theories nor observations that would guide modellers in a choice of Prandtl number.

4.2 Isopycnal mixing of tracers

Let us consider lateral mixing operators in our eddy permitting models (table 2). Some models use a laplacian operator rotated to follow the isopycnal (neutral) direction, others use a horizontal biharmonic (bi-laplacian) operator. The biharmonic operator has been introduced in quasi-geostrophic models based on the properties of two-dimensional turbulence. Because it is more scale-selective, it allows a model to represent a larger part of the mesoscale spectrum at a given grid resolution, while removing variance at the grid scale at a sufficient rate to avoid grid scale noise. However, the biharmonic operator can cause spurious overshoots in tracer properties (Mariotti et al., 1994), so that it has disadvantages as well as advantages. The inconveniences are pointed out in more detail by Griffies (2004).

Examination of the basin-scale water mass properties reveals that they spread along isopycnals (not horizontally), due to advection and stirring by mesoscale eddies. Analysis of tracer release experiments (Ledwell et al., 1998) suggest that mixing is isopycnal down to scales of 100 m, so that there is no evidence to support the choice of a horizontal mixing as in PSY2 or MFS. Toole (1998) reviews the processes that may be responsible for isopycnal mixing at different scales. Shear dispersion due to near-inertial internal waves can cause an isopycnal diffusivity of ≈ 0.07 m^2.s^{-1} at scales between 100 m and 1 km. Vortical modes could be responsible for diffusivities of ≈ 2 m^2.s^{-1} at scales 1 to 30 km, and mesoscale eddies can cause diffusivities up to 1000 m^2.s^{-1} at scales larger than 300 km. The eddy resolving models using isopycnal mixing cannot be run with diffusivities as low as observed; FOAM for example uses $\kappa = 100$ m^2.s^{-1}. This large value is needed to avoid numerical accumulation of enstrophy at the model grid scale (12 km).

From the observations of Ledwell et al. (1998), it seems that isopycnal diffusivities increase roughly linearly with the length scale. This would justify the choice made by some modellers to make the diffusivity proportional to the grid scale (or the third power of the grid scale in the case of a biharmonic operator), as for example in the DYNAMO models (Willebrand et al., 2001), or in PSY2.

The above considerations would support the choice of a biharmonic operator (for its scale selectiveness), rotated along isopycnals for consistency with observations. Often though, modellers do not want to pay the computational cost of rotating the biharmonic. This explains why the two alternatives found in table 2 are a horizontal biharmonic and an isopycnal laplacian. Those two parameterizations were compared during the CLIPPER project. Two experiments were run with the ATL6 model,

one using a horizontal biharmonic coefficient (like PSY2 and MFS) with value of 5.5 10^{10}m^4.s^{-1} at the equator, and the other one using a isopycnal laplacian mixing (like FOAM) with coefficient 200 m^2.s^{-1}. Results were not conclusive. The meridional overturning was enhanced by 2 Sv with the isopycnal mixing, which was assumed to be an improvement, but the deep jets analyzed by Treguier et al. (2003) were weaker with isopycnal mixing and the Agulhas eddies seemed too stable.

To make progress with the parameterization of lateral mixing at the sub-mesoscale, we need to understand the physical processes better. Sub-mesoscales are difficult to observe, but high resolution quasi-geostrophic or two-dimensional models give us insights into their behavior. One key phenomenon in the tracer cascade to small scales is the formation of elongated filaments, which occurs preferentially at critical points around the eddies when they interact with each other. This flow structure with energetic eddy cores surrounded by filaments is found in all high resolution models. Fig. 8 shows an example in the PSY2 model without data assimilation). Recent studies help understand where and when filaments form as a function of resolved flow quantities (see for example Klein et al., 2000). Parameterizations based on such analysis in physical space (by opposition to the more usual biharmonic or hyperviscosities based only on the cascade in spectral space) look promising, like the one by Dubos and Babiano (2002). So far they have not been implemented in realistic primitive equation models.

If mixing at the submesoscale is mainly performed by the combined action of vertically sheared inertial oscillations and vertical mixing as proposed by Young et al. (1982), then a parameterization must include the effect of mesoscale eddies on the inertial oscillations. Such a parameterization is tested by Klein et al. (2003) in quasi-geostrophic models and shown to cause an assymetry between anticyclonic and cyclonic structures.

One important issue that is too often ignored in parameterizations of isopycnal mixing is the large inhomogeneity of the mesoscale eddy field, which is now very well mapped from satellite altimetry. Obviously the isopycnal diffusivity κ in non-eddy resolving model should depend on the eddy activity. One possible way to achieve that is to use a scaling based on the time scale for baroclinic instability (Treguier et al., 1997). Such spatially variable coefficients have been used to represent the dynamical effect of eddies (see next section), but their use for mixing of tracers along isopycnals is not documented. Note that sharp variations in the eddy mixing coefficient κ can increase the gradients of tracers along isopycnals, as shown in Fig. 5. It will also create an advection of tracers away from the regions of active eddies. This advection is different

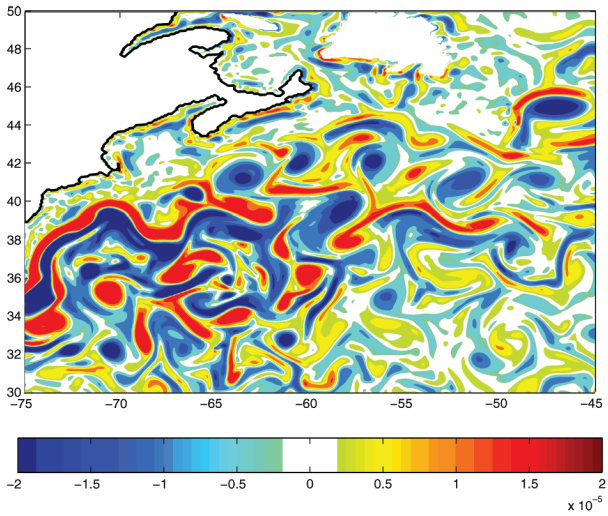

Figure 8. Relative vorticity (in s^{-1})at 17m depth in the PSY2 model without data assimilation (averaged over 5 days).

from the advective effect of the antisymmetric component of the mixing tensor (7), because contrary to V^* the velocity V_κ due to the spatial variation of κ is divergent. Both V^* and V_κ are needed to fully represent the difference between Eulerian and Lagrangian velocities (Plumb and Mahlman, 1987).

Another well known feature of diffusivity due to mesoscale eddies is its anisotropy. Because of the β-effect, mesoscale motions have longer zonal than meridional scales (Rhines, 1977). Ledwell et al. (1998) found a factor of two between the zonal and meridional diffusivity deduced from the spreading of his tracer. Despite this evidence, the use of anisotropic diffusivity in ocean models is not documented.

4.3 Lateral mixing of momentum

There are no observations similar to the tracer releases that would give us insight in the mechanisms of momentum mixing in the ocean, and theories do not help much either. Well known properties of mesoscale eddies, like their tendency to concentrate momentum in an eastward jet on a β plane (McWilliams and Chow, 1981) are very difficult to parameterize because they would involve counter-gradient fluxes.

For lack of physically motivated parameterizations, modellers generally use simple laplacian or biharmonic viscosity operators. The values of the coefficients are subject to two numerical constraints. In basin

scale models, the smallest spatial scale is often the width of the western boundary current. When it is controlled by laplacian friction it is called a Munk boundary layer. The condition that the grid scale δx be smaller than the Munk layer width results in a minimum bound for viscosity (Smith and McWilliams, 2003): $\nu > \nu_M \approx \beta \delta x^3$. On the other hand, viscosity cannot be arbitrarily large due to the stability constraint (similar to the CFL criterion for advection). This criterion is more severe in ocean models that use explicit leap-frog time stepping schemes for nonlinear advection, with the viscous terms lagged by one time step for stability. For laplacian viscosity $\nu < \delta x^2 / 8\,\delta t$. For a biharmonic operator the criterion is $\nu < \delta x^4 / 64\,\delta t$ (biharmonic coded as in the POP model) or $\nu < \delta x^4 / 128\,\delta t$ (biharmonic coded as in the OPA model).

For the laplacian operator on coarse grids, the Munk layer constraint implies very large viscosities: with δx=10 km at 45°N, $\nu_M = 16$ m^2.s^{-1}, but with δx=100 km, $\nu_M = 16000$ m^2.s^{-1}. It is impossible to represent equatorial dynamics with such a large viscosity, which is why this constraint is not always taken into account. For example, in the FOAM 1° model the viscosity is 5100 m^2.s^{-1}. In that case, some level of grid point noise usually develops near the western boundary. A second strategy is to decrease the viscosity at the equator, while increasing the meridional resolution there (parameterization of the ORCA2 model with $\nu = 2000$ m^2.s^{-1} at the equator and $\nu = 40000$ m^2.s^{-1} at mid-latitudes, Madec et al., 1998). Finally, Large et al. (2001) have proposed to make the viscosity anisotropic, noting that the equatorial current are dominantly zonal while Munk boundary currents are predominantly meridional. This solution however does not prevent the apparition of numerical noise.

For the biharmonic operation the numerical stability criterion is often more stringent than the Munk layer constraint. With δx=10 km, the latter gives $\nu > \beta \delta x^5 = 1.6\ 10^9$ m^4.s^{-1}. With a 1200 s time step, numerical stability requires that $\nu < \delta x^4 / 64\,\delta t = 1.3\ 10^{11}$ m^4.s^{-1}. A decrease of the biharmonic coefficient with the grid spacing is often needed in order to ensure stability on spatially variable grids, as in the case of PSY2 (Table 2). Based on numerical experiments with a 1/12° isopycnic model, Chassignet and Garraffo (personal communication) suggest that the Gulf stream separation is improved using together a laplacian and a bilaplacian operator. This has prompted the use of both operators in FOAM (Table 2).

Smagorinsky (1963) has proposed to make the laplacian viscosity proportional to the deformation rate times the squared grid spacing δx^2. Such a parameterization can be physically motivated in three dimensional turbulence and is used in large eddy simulations. In ocean

models it has mainly been used in MICOM (Bleck and Boudra, 1981). A study by Griffies and Hallberg (2003) suggests that using a biharmonic operator with Smagorinsky-like viscosity is better in eddy permitting simulation when the flow is non homogeneous (in the presence of western boundary currents, for instance) because it allows lower levels of viscosity in the interior.

This short review emphasizes numerical constraints as the basis for the choice of parameterizations of momentum mixing. We can hope that more physically based parameterizations will emerge in the future. Smith and McWilliams (2003) have developed a promising framework by deriving a general form for anisotropic viscosity, and an elegant functional form for the discretization following a similar work by Griffies et al. (1998) on the isoneutral diffusion.

For completeness we must mention here another approach to parameterizations. It consists in using the properties of numerical advections schemes to represent the cascade of enstrophy to small scales (this also applies to cascade of tracer variance reviewed in the previous section). With that strategy, no explicit parameterization is needed. Shchepetkin and McWiliams (1998) advocate this approach, claiming that higher Reynolds numbers can be simulated that way, compared with the combination of a classical advection scheme and hyperviscosity. Those authors also claim that it is more computationaly efficient to increase the accuracy of the advection scheme rather than increasing the spatial resolution. This is certainly true for the idealized turbulence experiments they perform, but it is probably not yet true for realistic ocean models. Subgrid scale topographic effects are the reason for this. Refining the grid offers the opportunity to better represent key straits and passages, which a higher order scheme cannot provide. This is certainly the reason why most ocean models still use second order, inexpensive advection schemes. This situation may change in the future, as higher spatial resolutions are allowed by the computational resources.

5. Dynamical effects of mesoscale eddies

5.1 Baroclinic instability

Gent and McWilliams (1990), hereafter GM, have noted that parameterizing the mixing of salinity and temperature anomalies on isopycnals by mesoscale eddies is not enough, because this leaves aside the dynamical effect of eddies on the density field. Most of the eddy energy in the ocean is believed to arise due to baroclinic instability of the mean flow. Baroclinically unstable eddies extract available potential energy from the mean flow, thus tending to flatten isopycnals. The GM parameteriza-

tion proposes that this effect is best represented using the antisymmetric part of the diffusion tensor (5), that is, by an additional advection of the density field. They propose to make this velocity proportional to the isopycnal slope (a pedagogical presentation of their parameterization is found in Gent et al., 1995). This parameterization was the first physically-based original parameterization for coarse resolution ocean model, and as such it has known a rapid success. It certainly improves the climate model solutions especially in the Antarctic circumpolar current, although too large advective velocities for the GM parameterization have negative effects there (Speer et al., 2000).

It is useful to consider the parameterizations in the quasigeostrophic limit (Treguier et al., 1997), in which case GM corresponds to a mixing of potential vorticity (more precisely, the vortex stretching contribution to potential vorticity) along isopycnals. Therefore, the coefficient used for the GM parameterization can be considered as a mixing coefficient for potential vorticity, while the coefficient used for isopycnal mixing is relevant to a passive tracer (temperature and salinity anomalies along isopycnal surfaces). Two-dimensional turbulence emphasizes the similarity between the dynamics of vorticity and passive tracers; although no similar studies exist with primitive equations in three dimensions there is no physical argument to justify widely different mixing coefficients for the two parameterizations. It is thus very surprising to find that many modellers take coefficients for their GM parameterizations that are spatially dependent on the level of baroclinic instability as proposed by Treguier et al. (1997) or Visbeck et al. (1997), thus correctly taking into account the inhomogeneity of the eddy activity in the ocean, while they keep the isopycnal mixing coefficient constant. Maybe modellers are reluctant to seek guidance from the quasi-geostrophic framework because things are indeed more complex in primitive equations: for example, the GM parameterization as usually implemented is closer to a mixing of isopycnal depth than to a mixing of potential vorticity.

A most important open question is how to represent the unresolved part of the mesoscale eddy spectrum in eddy permitting models. First, it is important to note that the unresolved spectrum varies with latitude. A typical spatial scale for baroclinic instability is the first Rossby radius R_1. Even though model grids are often of Mercator type, refined as the cosine of latitude, they still fall short of resolving R_1 in the Labrador Sea and Nordic seas where it can be a few kilometers in winter. Perhaps we should be more precise about what is meant by "resolving". A minimum requirement could be 12 grid points per wavelength (a first derivative estimated with a second order finite difference scheme still has 5% error in that case), thus $\delta x < 2\pi R_1/12 \approx R_1/2$. Chanut (2003)

finds a dramatic improvement in the representation of restratification after convection in the Labrador Sea between a 18 km grid and a 3-4 km grid ($1/3^\circ$ to $1/15^\circ$). Certainly, a full GM parameterization would be justified in the $1/3^\circ$ model, in the Labrador Sea but not elsewhere. The eddy fluxes in Chanut's high resolution case correspond to GM coefficient of up to 800 $m^2.s^{-1}$. If such a high value is used in the subtropical gyre it destroys the eddy activity in the Gulf Stream. The spatially variable form proposed by Visbeck et al. (1997) does not help in that case, because baroclinic instability growth rate is higher in the Gulf Stream than in the Labrador Sea based on the resolved flow field. Studies are under way to propose variants of the GM parameterization that would "switch on" when needed.

Considering a case where the first Rossby radius is well resolved (say, a 5 km grid where $R_1 = 40$ km), how should the dynamical effect of sub-mesoscale eddies be parameterized? Is the unresolved part of the spectrum mainly controlled by baroclinic instability? Roberts and Marshall (1998) advocate the use of a biharmonic GM parameterization, based on their wish to eliminate diapycnal mixing in the surface layers. However, this requirement may not be physically defensible, considering that eddies do perform diapycnal (horizontal) mixing across surface fronts (Treguier et al., 1997). As was the case for in the two previous parameterizations we considered (isopycnal diffusivity and lateral viscosity) we still lack observational evidence and theory to justify parameterizations of the submesoscale effects.

5.2 Other mesoscale eddy effects

Another dynamical effect of mesoscale eddies is the so-called "Neptune" effect (see for a review Alvarez and Tintoré, 1998). In the presence of bathymetry and β-effect, quasigeostrophic eddies have the tendency to generate mean flows along f/H contours. This additional mean flow must be forced by a parameterization when eddies are not represented in a model. The problem is that we do not have enough knowledge of the strength of this effect in realistic ocean circulations and neither do we know the vertical structure of the generated mean flows. Certainly, adding a parameterization forcing barotropic currents along f/H contours would help Atlantic models to improve the strength of their deep western boundary currents. However, if the models do not represent the overflow correctly, the Neptune parameterization could have the effect of generating spurious transport of water with the wrong properties.

Finally, mesoscale eddies tend to exist as coherent structures that carry water far from their generation region (well-known examples are

Meddies and Agulhas eddies). The effects of such eddies are highly non-local, and no parameterization has yet been proposed for such processes.

6. Conclusion

Parameterizations and resolution are the two fundamental characteristics of an ocean model. By choosing them, we actually pick up the "ocean" we try to model. We have reviewed different parameterizations, based on the example of the forecast models listed in tables 1 and 2. What emerges from this review is a rather unsatisfactory state of affairs. Some paramerizations are well grounded in physics (like convection) and have been evaluated by comparison with more complete models (non-hydrostatic in this case). Even then, though, we find that some features are not completely agreed upon among modellers (like the Prandtl number) and modifying them has a strong effect on the solution of low resolution models. But this is the best situation. Generally, the parameterizations do not have sound physical basis, have not been fully evaluated against laboratory experiments or more complete models, and there are strong numerical constraints limiting the choices of modellers.

The boundaries of the ocean, at the bottom and at the surface, are places where progress needs to be made. Regarding the bottom, the main problem is the representation of flow-topography interactions in z-coordinate ocean models. This is an issue of numerics rather than parameterization, which is discussed in the chapter by Griffies. The main effect of staircase topography in z-coordinate models, which is not completely alleviated by using a partial step representation, is the existence of large and noisy vertical velocities which often contaminate the upper layers (especially on the continental slopes). This is a big obstacle to the use of such models for biogeochemistry. Hybrid models like HYCOM may be better candidates for such applications, although it is not clear that numerical factors affecting the communication between the surface z layer and the interior isopcynic layers will not prove an even bigger obstacle. It is quite surprising that although σ coordinate models are extensively used for regional and coastal modelling, no larger scale σ configurations have been built for demonstration purposes, either for climate prediction or forecasting.

The representation of the surface layers in ocean models is perhaps the point where progress is the most likely in the coming years. Today, model solutions in the mixed layer critically depends on the parameterizations. This dependency may decrease as we resolve more physical processes. It is possible to do so with existing parameterizations simply by increasing the vertical resolution (to about 1 m) and using higher

frequency forcing (thus taking into account wind-forced inertial oscillations and the diurnal cycle). Improving the representation of surface layers is critical for operational forecast models because many clients need accurate surface velocities. It is also important in coupled models for climate prediction.

My personal view is that parameterization of the full mesoscale eddy spectrum is a hopeless challenge. We can certainly improve on existing parameterizations. Low-resolution climate models will still be necessary tools in the future, because (fortunately) many aspects of the long-term climate response are robust with respect to details of mesoscale eddy effects. On the other hand, growing computer power will help us to resolve a larger part of the mesocale eddy spectrum in forecast models. It is therefore very important to improve our knowledge of the sub-mesoscale dynamics and develop suitable parameterizations. In this respect, it is quite possible that progress will be easier to achieve in the ocean than in the atmosphere. Atmospheric climate models resolve a large part of the synoptic scale eddies and use crude parameterizations of the subgrid scale dynamics. This is because subgrid scale physics linked to atmospheric moisture (cloud physics, radiation) play a more important role in climate than purely dynamical subgrid scale effects. Ocean models do not have this additional level of complexity and may be a more suitable framework to develop parameterisations for the dynamics, which would fully take into account the spatial inhomogeneity of the mesoscale eddy field.

Acknowledgements

I thank Patrice Klein for useful discussions and for pointing out equation (9). Comments from Steve Griffies and GODAE school students have been very helpful.

References

Alvarez, A. and Tintoré, J. (1998). Topographic stress: importance and parameterization. In *Ocean Modeling and Parameterization*, volume 516 of *NATO Science series C*, pages 327–350. Kluwer Academic Publishers.

Arhan, M., Mercier, H., and Lujtjeharms, J. R. E. (1999). The disparate evolution of three agulhas rings in the south atlantic ocean. *J. Geophys. Res.*, 104:20987–21005.

Batchelor, G. K. (1969). Computation of the energy spectrum in homogeneous two-dimensional turbulence. *Phys. Fluids.*, 12, II:233–238.

Beckmann, A. and Döscher, R. (1997). A method for improved representation of dense water spreading over topography in geopotential-coordinate models. *J. Phys. Oceanogr.*, 27:581–591.

Blanke, B. and Delecluse, P. (1993). Variability of the tropical atlantic ocean simulated by a general circulation model with two different mixed layer physics. *J. Phys. Oceanogr.*, 23:1363–1388.

Bleck, R. and Boudra, D. B. (1981). Initial testing of a numerical circulation model using a hybrid- (quasi-isopycnic) vertical coordinate. *J. Phys. Oceanogr.*, 11:755–770.

Boer, G. J. and Denis, B. (1997). Numerical convergence of a gcm. *Climate dynamics*, 13:359–374.

Bower, A. S., Cann, B. Le, Rossby, T., Zenk, W., Gould, J., Speer, K., Richardson, P. L., Prater, M. D., and Zhang, H. M. (2002). Directly-measured mid-depth circulation in the northeastern north atlantic ocean. *Nature*, 419 (6907):603–607.

Chanut, J. (2003). *Paramétrisation de la restratification après convection profonde en mer du Labrador*. PhD thesis, université Joseph Fourier, Grenoble, France.

Chanut, J. and Molines, J. M. (2004). Implementation et validation du modèle kpp dans opa8.1. Rapport technique, LEGI, Grenoble.

Chassignet, E. and Verron, J., editors (1998). *Ocean Modeling and Parameterization*, volume 516 of *NATO Science series C*. Kluwer Academic Publishers, Cambridge.

Cox, M. D. (1984). A primitive equation, 3-dimensional model of the ocean. Technical Report 1, Geophysical Fluid Dynamics Laboratory, PO Box 308, Princeton, New Jersey, 08542.

Dengg, J., Böning, C., Ernst, U., Redler, R., and Beckmann, A. (1999). Effects of improved model representation of overflow water on the subpolar north atlantic. *WOCE Newsletter*, 37.

Dengler, M., Schott, F., Eden, C., Brandt, P., Fischer, J., and Zantopp, R. J. (2004). New mechanisms for deep water transport. *Nature*. In press.

Dubos, T. and Babiano, A. (2002). Two-dimensional cascades and mixing: a physical space approach. *J. Fluid. Mech.*, 467:81–100.

Gent, P. R. and McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation model. *J. Phys. Oceanogr.*, 20:150–155.

Gent, P. R., Willebrand, J., McDougall, T. J., and McWilliams, J. C. (1995). Parameterizing eddy-induced tracer transports in ocean circulation models. *J. Phys. Oceanogr.*, 25:463–474.

Gerdes, R. (1993). A primitive equation ocean circulation model using a general vertical coordinate transformation. *J. Geophys. Res.*, 98:14683–14701.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Woods, R. A. (2000). The simulation of sst, sea ice extents and ocean heat transports in a version of the hadley centre coupled model without flux adjustments. *Climate Dynamics*, 16:147–168.

Griffies, S. M. (2004). *Fundamentals of ocean climate models*. Princeton University Press, Princeton, U.S.A.

Griffies, S. M., Boening, C., Bryan, F.O., Chassignet, E. P, Gerdes, R., Hasumi, H., Hirst, A., Treguier, A. M, and Webb, D. (2000a). Developments in ocean climate modelling. *Ocean Modelling*, 2:123–192.

Griffies, S. M., Gnanadesikan, A., Pacanowski, R. C., Larichev, V. D., Dukowicz, J. K., and Smith, R. D. (1998). Isoneutral diffusion in a z-coordinate ocean model. *J. Phys. Oceanogr.*, 28:805–830.

Griffies, S. M. and Hallberg, R. W. (2000). Biharmonic friction with a smagorinsky-like viscosity for use in large scale eddy-permitting ocean models. *Monthly Weather Review*, 128:2935–2946.

Griffies, S. M., Pacanowski, R. C., and Hallberg, R. W. (2000b). Spurious diapycnal mixing associated with advection in a z-coordinate ocean model. *Monthly Weather Review*, 128:538–564.

Haidvogel, D. B. and Beckmann, A. (1999). *Numerical Ocean Circulation Modelling*, volume 2 of *Series on environmental science and management*. Imperial College Press, London.

Hasumi, H. and Suginohara, N. (1999). Effects of locally enhanced vertical diffusivity over rough bathymetry on the world ocean circulation. *J. Geophys. Res.*, 104:23367–23374.

Hurlburt, H. E. and Hogan, P. J. (2000). Impact of 1/8° to 1/64° resolution on gulf stream model-data comparisons in basin-scale subtropical atlantic models. *Dyn. Atmos. Oceans*, 32:283–329.

Jungclaus, J. H. and Mellor, G. (2000). A three-dimensional model study of the mediterranean outflow. *J. Mar. Sys.*, 24:41–66.

Klein, P., Hua, B. L., and Carton, X. (2003). Emergence of cyclonic structures due to the interaction between near-inertial oscillations and mesoscale eddies. *Q. J. R. Meteorol. Soc.*, 129:1–20.

Klein, P., Hua, B. L., and Lapeyre, G. (2000). Alignment of tracer gradients in two-dimensional turbulence using second order lagrangian dynamics. *Physica D*, 146:246–260.

Klinger, B. A., Marshall, J., and Send, U. (1996). Representation of convective plumes by vertical adjustment. *J. Geophys. Res.*, 101:18175–18182.

Kraus, E. B. and Turner, S. (1967). A one-dimensional model of the seasonal thermocline. ii: The general theory and its consequences. *Tellus*, 19:98–106.

Large, W. G. (1998). Modelling and parameterizing the ocean planetary boundary layer. In *Ocean Modeling and Parameterization*, volume 516 of *NATO Science series C*, pages 81–120. Kluwer Academic Publishers.

Large, W. G., Danabasoglu, G., McWilliams, J.C., Gent, P. R., and Bryan, F. O. (2001). Equatorial circulation of a global ocean climate model with anisotropic horizontal viscosity. *J. Phys. Oceanogr.*, 31:518–536.

Large, W. G., McWilliams, J. C., and Doney, S. C. (1994). Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, 32:363–403.

Laurent, L. C. St, Simmons, H. L., and Jayne, S. R. (2002). Estimating tidally driven mixing in the deep ocean. *Geophys. Res. Lets*, 29:2106.

Ledwell, J. R., Watson, A. J., and Law, C. S. (1998). Mixing of a tracer in the pycnocline. *J. Geophys. Res.*, 103:21499–21529.

Madec, G., Delecluse, P., Imbard, M., and Levy, C. (1998). Opa 8.1 general circulation model reference manual. Notes de l'ipsl, Univ. Pierre et Marie Curie, Paris.

Maltrud, M. E. and McClean, J. L. (2004). An eddy resolving global 1/10° ocean simulation. *Ocean Modelling*, 7:31–54.

Mariotti, A., Legras, B., and Dritschel, D. G. (1994). Vortex stripping and the erosion of coherent structures in two-dimensional flows. *Phys. Fluids*, 6:3934–3962.

Matteoli, O. (2003). Etude de sensibilité d'orca2 à la physique verticale des couches de surface. Rapport de stage, LODYC, Univ. Pierre et Marie Curie, Paris.

McWilliams, J. C. and Chow, J. (1981). Equilibrium turbulence i: a reference solution in a β-plane channel. *J. Phys. Oceanogr.*, 11:921–949.

Mellor, G. L. and Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. and Space Phys.*, 20:851–875.

Merryfield, W. J., Holloway, G., and Gargett, A. E. (1999). A global ocean model with double-diffusive mixing. *J. Phys. Oceanogr.*, 29:1124–1142.

Pacanowski, R. C. and Philander, S. G. H. (1981). Parameterization of vertical mixing in numerical models of the tropical ocean. *J. Phys. Oceanogr.*, 11:1442–1451.

Penduff, T., Barnier, B., Verron, J., and Kerbiriou, M. A. (2002). How topographic smoothing contributes to differentiating the eddy flows simulated by sigma- and z-level models. *J. Phys. Oceanogr.*, 32:122–137.

Phillips, O. M. (1972). Turbulence in a strongly stratified fluid - is it unstable? *Deep Sea Res.*, 19:79–81.

Plumb, R. A. and Mahlman, J. D. (1987). The zonally averaged transport characteristics of the gfdl general circulation/transport model. *J. Atmos. Sci.*, 44:298–327.

Price, J. F. and Yang, J. (1998). Marginal sea overflows for climate simulations. In *Ocean Modeling and Parameterization*, volume 516 of *NATO Science series C*, pages 155–170. Kluwer Academic Publishers.

Rhines, P. B. (1977). *The dynamics of unsteady currents*, volume 6 of *Marine Modelling, The Sea*, pages 189–318. Wiley.

Riviere, P., Treguier, A. M., and Klein, P. (2004). Effects of bottom friction on nonlinear equilibration of an oceanic baroclinic jet. *J. Phys. Oceanogr.*, 34:416–432.

Roberts, M. and Marshall, D. (1998). Do we require adiabatic dissipation schemes in eddy-resolving ocean models? *J. Phys. Oceanogr.*, 28:2050–2063.

Ruddick, B. R., McDougall, T. J., and Turner, J. S. (1989). The formation of layers in a uniformly stirred density gradient. *Deep Sea Res.*, 36:597–609.

Schmitt, R. W. (1998). Double-diffusive convection. In *Ocean Modeling and Parameterization*, volume 516 of *NATO Science series C*, pages 215–234. Kluwer Academic Publishers.

Send, U. and Käse, R. H. (1998). Parameterization of processes in deep convection regimes. In *Ocean Modeling and Parameterization*, volume 516 of *NATO Science series C*, pages 191–214. Kluwer Academic Publishers.

Shchepetkin, A. F. and McWilliams, J. C. (1998). Quasi-monotone advection schemes based on explicit locally adaptative dissipation. *Monthly Weather Review*, 126:1541–1580.

Siegel, A., Weiss, J. B., Toomre, J., McWilliams, J. C., Berloff, P. S., and Yavneh, I. (2001). Eddies and vortices in ocean basin dynamics. *Geophys. Res. Lets.*, 28:3183–3186.

Simmons, H. L., Jayne, S. R., Laurent, L. C. St., and Weaver, A. J. (2004). Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Modelling*, 6:245–263.

Smagorinsky, J. (1963). General circulation experiments with the primitive equations: I. the basic experiment. *Monthly Weather Review*, 91:99–164.

Smith, R. D., Maltrud, M. E., Bryan, F. O., and Hecht, M.W. (2000). Simulation of the north atlantic ocean at 1/10°. *J. Phys. Oceanogr.*, 30:1532–1561.

Smith, R. D. and McWilliams, J. C. (2003). Anisotropic horizontal viscosity for ocean models. *Ocean Modelling*, 5:129–156.

Speer, K., Guilyardi, E., and Madec, G. (2000). Southern ocean transformation in a coupled model with and without eddy mass fluxes. *Tellus*, 52A:554–565.

Stratford, K. and Haines, K (2000). Frictional sinking of the dense water overflow in a z-coordinate ogcm of the mediterranean sea. *Geophys. Res. Lets.*, 27:3973–3976.

Toole, J. M. (1998). Turbulent mixing in the ocean. In *Ocean Modeling and Parameterization*, volume 516 of *NATO Science series C*, pages 171–190. Kluwer Academic Publishers.

Treguier, A. M., Barnier, B., de Miranda, A. P., Molines, J. M., Grima, N., Imbard, M., Messager, C, Reynaud, T, and Michel, S (2001). An eddy permitting model of the atlantic circulation: evaluating open boundary conditions. *J. Geophys. Res.*, 106:22115–22129.

Treguier, A. M., Held, I., and Larichev, V. (1997). On the parameterization of quasi-geostrophic eddies in primitive equation ocean models. *J. Phys. Oceanogr.*, 27:567–580.

Treguier, A. M., Hogg, N. G., Maltrud, M., Speer, K., and Thierry, V. (2003). Origin of deep zonal flows in the brazil basin. *J. Phys. Oceanogr.*, 33:580–599.

Visbeck, M., Marshall, J., Haine, T., and Spall, M. (1997). Specification of eddy transfer coefficients in coarse resolution ocean circulation models. *J. Phys. Oceanogr.*, 27:381–402.

Willebrand, J., Barnier, B., Böning, C., Dieterich, C., Hermann, P., Killworth, P. D., LeProvost, C., Jia, Y., Molines, J. M., and New, A. L. (2001). Circulation characteristics in three eddy-permitting models of the north atlantic. *Prog. Oceanogr.*, 48:123–161.

Young, W. R., Rhines, P. B., and Garrett, C. J. R. (1982). Shear-flow dispersion, internal waves and horizontal mixing in the ocean. *J. Phys. Oceanogr.*, 12:515–527.

Chapter 4

ON THE USE OF HYBRID VERTICAL COORDINATES IN OCEAN CIRCULATION MODELING

Rainer Bleck
Los Alamos National Laboratory, Los Alamos, New Mexico, USA

Abstract The rationale for building hybrid-coordinate ocean circulation models is discussed in the context of various approaches presently taken to improve mathematical and physical aspects of ocean models. Design choices made in formulating the vertical grid generator, the core component of hybrid models, are laid out. A new experimental approach toward minimizing numerical errors and inconsistencies during simultaneous advection of temperature, salinity and density is presented.

Keywords: Ocean models, vertical coordinate, hybrid coordinate.

1. Introduction

Motion systems interacting in the ocean range in size from centimeters to planetary, a spread of 9 orders of magnitude. Ocean circulation modeling therefore is a textbook example of a multiscale problem. Since even our fastest computers cannot spatially resolve the earth's surface by more than a few thousand grid points in each direction, only about one-third of those 9 orders of magnitude can be explicitly resolved in state-of-the-art ocean models, meaning that the other two-thirds are relegated to the "subgrid" scale. Hence, a perpetual challenge to ocean modeling is the need to parameterize the interaction of physical processes across the threshold between resolved and unresolved scales of motion. This need will be with us until computers become $(10^6)^3 = 10^{18}$ times faster than they are today – i.e., practically forever.

The ability of today's ocean models to correctly simulate essential aspects of the global circulation has led to a proliferation of applications

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 109-126.

where the models are used outside their "design range" and therefore generate results that do not always live up to the expectations of the various user communities. Ocean models clearly need to be developed further to satisfy those communities.

There are essentially three ways in which ocean models can be improved. One can

1 increase grid resolution;

2 improve model physics;

3 improve model numerics.

Increasing grid resolution (item 1), in theory, allows a numerically obtained solution to approach that of the underlying differential equation. However, given the huge spectral range of unresolved processes in the ocean, truncation errors continue to cast their shadow even over what we call "high-resolution" models.

Efforts in model physics improvement (item 2) in most cases boil down to improvements in the parameterization of spatially unresolved processes. This is a never-ending process: parameterization schemes tend to be sensitive to where the transition between resolved and unresolved scales occurs and therefore must evolve in lockstep with refinements in mesh size.

Improved numerics (item 3) is a non-exclusive alternative to higher grid resolution; both approaches lower the truncation error in the finite difference equations that constitute the ocean model.

Model numerics can be improved in several ways. One approach is to switch to higher-order finite difference approximation ("order" here refers to the power of the spatial or temporal mesh size in truncation error expressions; the higher the power, the faster the error goes to zero with increased grid resolution). Another approach is to revive techniques developed in pre-computer days when a problem became solvable only by transforming the equations into a coordinate system that exploited some symmetry or conservation laws inherent in the underlying physics. The geophysical fluid dynamics community is doing some of this already by formulating its equations in a coordinate system whose z axis points in the direction of gravity rather than, say, the center of our galaxy. But the concept of manipulating the equations to make them easier to solve or improve the accuracy of the solutions can be extended much further.

The recent proliferation of unconventional vertical coordinates in geophysical modeling should be viewed as one particular attempt to improve ocean model numerics along the lines of item 3 above. Repeating the phrase just used, the new coordinates currently being experimented with

seek to exploit "some symmetry or conservation laws inherent in the underlying physics" with the goal of improving the accuracy of the numerical solution. Foremost among those conservation laws is the one stating that adiabatic motion follows surfaces of constant entropy or its proxy, potential density. Thus, if one uses potential density as vertical coordinate, adiabatic flow that is 3-dimensional in Cartesian space is rendered 2-dimensional in potential density space. This makes it particularly easy to satisfy adiabatic constraints while modeling lateral transport of tracers, including temperature and salinity.

The above advantage actually holds for resolved as well as a wide range of *un*resolved scales of motion. Lateral stirring processes in the ocean are known to mix properties predominantly along isentropic surfaces. Hence, a model based on potential density (which in this particular case must be locally referenced) should be able to simulate the effects of subgridscale stirring, typically parameterized as an eddy diffusion process, more accurately than a model in which lateral isentropic stirring must be projected onto the Cartesian x, y, z axes.

The principal design element of isopycnic (potential density) coordinate models, in relation to conventional Cartesian coordinate models, is that depth (alias layer thickness) and potential density trade places as dependent and independent variables. This switch does not affect the number of prognostic equations, nor does it alter the familiar mix of wave modes and processes by which information is transmitted in the ocean. This is to say that both model types solve the same physical problem. However, since the two models are based on different sets of differential equations, their numerical properties should be expected to be very different as well.

Wind-forced process models framed in isopycnic coordinates have been in use since the 1960s (e.g., Welander, 1966; Holland and Lin, 1975a,b; Bleck and Boudra, 1981). With the addition of thermohaline forcing (Bleck et al., 1992; Oberhuber, 1993; Hu, 1997; Sun and Bleck, 2001; Cheng et al., 2004), these models have become sufficiently comprehensive to be useful in studying the oceanic general circulation.

The focus in this article is on one particular extension of the isopycnic coordinate concept that addresses certain shortcomings of potential density as vertical coordinate. These shortcomings are

- coordinate surfaces intersecting the sea surface (the outcropping problem);
- lack of vertical resolution in unstratified water columns.

Note that these are two sides of the same coin. In a global model, most density surfaces required to span the top-to-bottom density range

in the subtropics are out of place in the relatively unstratified high-latitude waters and hence must exit the domain somewhere between the subtropics and the subpolar latitude bands.

The hybridized depth-isopycnic coordinate (Bleck, 2002) developed for the isopycnic model MICOM (Bleck et al., 1992) alleviates both shortcomings just mentioned. It does so by not allowing coordinate layers to outcrop, but rather forcing layers that are isopycnic in character at low latitudes to turn into fixed-depth layers at high latitudes. Stated differently, a coordinate layer associated with a chosen "target" isopycnal adheres to the depth of that isopycnal as long as the latter exists in a given water column. Near the outcrop latitude of the target isopycnal, the coordinate layer turns horizontal and becomes a fixed-depth layer. (The term fixed-depth here covers both constant-depth and bottom-following layer configurations.)

The approach just outlined creates fixed-depth coordinate layers at the same rate at which the model loses isopycnic layers between the equator and the poles. The new fixed-depth layers provide vertical resolution in unstratified high-latitude regions and on coastal shelves, allowing the hybrid coordinate model to simulate turbulent mixing and buoyant convection in a manner similar to z and σ coordinate models.

The term "hybrid vertical coordinate" has more than one meaning. Often it refers to configurations (e.g., Bleck 1978) where the model domain is divided into a stack of subdomains, each filled with coordinate surfaces defined independently of those in other subdomains. A simple example is the combination, popular in weather prediction models, of terrain-following coordinate surfaces in the lower atmosphere and isobaric surfaces in the upper atmosphere.

Our present hybrid scheme, which dates back to Bleck and Boudra (1981), does not rely on rigidly defined subdomains. Instead, it permits temporal and lateral transitions between different coordinate types – terrain-following, constant-depth, isopycnal – based on local conditions such as layer thickness and vertical density contrast. The scheme has much in common with the ALE (Arbitrary Lagrangian-Eulerian) technique of Hirt et al. (1974) but adds one important element to that scheme, namely, a mechanism for keeping coordinate layers aligned with, or for nudging them toward, their designated target isopycnals wherever possible. The original ALE scheme only concerns itself with maintaining a finite separation between adjacent coordinate surfaces. While the flexibility of coordinate placement in ALE-type schemes is disconcerting to some users because grid point location in physical space cannot be expressed in terms of a simple analytic formula, the flexibility inher-

ent in the ALE concept allows it to maximize the size of the isopycnic subdomain in the model. This is a major advantage.

2. The grid generator

At the core of ALE-type hybrid models is a utility called the vertical grid generator. Most efforts to produce a robust hybrid model are directed at refining and "shockproofing" this unique utility. A thorough discussion of the grid generator used in HYCOM is therefore in order. For the convenience of the reader, some details already presented in Bleck (2002) will be repeated here, but emphasis will be on recent improvements.

Models like HYCOM belong to the so-called *layer* model class where, as mentioned earlier in the context of isopycnic-coordinate models, the depth z of coordinate surfaces is treated as a dependent variable.[1] Having lost z as independent variable, layer models need a new independent variable capable of representing the 3rd (vertical) model dimension. This variable is traditionally called "s".

With the number of unknowns increased by one (namely, layer interface depth), the model needs one additional diagnostic equation. The logical choice is an equation linking s to the other model variables. In purely isopycnic coordinate models, s is equated with potential density. In ALE-type hybrid ocean models, s becomes a "target" potential density value assigned to each coordinate layer. The grid generator's task is to move the coordinate layer toward the depth where the target density is observed to occur. Once a layer is aligned with its target isopycnal, it becomes a material layer whose subsequent evolution is no longer governed by the grid generator until interior mixing or surface thermohaline forcing cause the density to drift away from its target value.

If the target density lies outside the density range in the water column, the layer set in motion by the grid generator will encounter the sea surface or sea floor before finding its target density, depending on whether the target density lies below or above the density range found in the column. Instead of rendering layers whose target density does not exist dynamically invisible by deflating them, the grid generator is designed to impose a minimum layer thickness. This constraint also affects layers which during their vertical migration impinge on layers already converted to fixed-depth layers.

[1] MICOM and HYCOM actually express layer depth in terms of pressure p, the weight per unit area of the water column. Replacing z by p simplifies matters if the vertical model dimension is expressed in terms of a thermodynamic variable. This also means that the *Boussinesq* approximation is not needed in these models.

In summary, grid generation in HYCOM is a two-step process. In step 1, interfaces are set in motion to restore the target density in individual layers. In step 2, this migration is checked for compatibility with an imposed minimum thickness constraint. Step 2 overrides step 1.

The details of the minimum thickness constraint are essentially the model designer's choice. For example, during HYCOM development the decision was made to impose nonzero-thickness constraints only at the surface but allow layers at the bottom to become massless. This is done for a good reason. Steeply inclined coordinate layers, like those following the bathymetry, are prone to errors in the horizontal pressure force calculation. In a layer containing no mass, such errors are dynamically inconsequential.

Complications arising from the grid generator's ad-hoc strategy for vertical grid point placement are minimal. This can be demonstrated as follows (Bleck, 1978). Recall that material vertical motion in hydrostatic models is inferred from mass continuity — specifically, from the vertically integrated horizontal mass flux divergence. The material vertical motion so diagnosed is then decomposed into motion of the coordinate surface and motion relative to the coordinate surface:

$$\left(\begin{array}{c}\text{vertical}\\ \text{motion } \underline{of}\\ s \text{ surface}\end{array}\right) + \left(\begin{array}{c}\text{vertical}\\ \text{motion}\\ \underline{through}\\ s \text{ surface}\end{array}\right) = \left(\begin{array}{c}\text{vertically integrated}\\ \text{horizontal mass flux}\\ \text{divergence}\end{array}\right) \quad (1)$$

After diagnosing the right-hand side of (1) at a given time step, the hydrostatic model needs one additional condition to distribute this quantity among the two terms on the left. In a material coordinate system, for example, the second term on the left is zero by definition; hence, the vertically integrated mass flux divergence yields the rate at which a coordinate surface moves up or down in space. The other extreme is the spatially fixed grid where, by definition, the first term on the left is zero; the vertically integrated mass flux divergence in that case yields the vertical velocity.

The point of this discussion is that compromise solutions between the two extremes just mentioned can easily be accomodated in a model on a grid-point-by-grid-point and time-step-by-time-step basis. Whatever vertical motion the grid generator prescribes for a given grid point during a given model time step is simply used in (1) in conjunction with the vertically integrated mass flux divergence to compute the appropriate generalized vertical velocity $ds/dt \equiv \dot{s}$. (By definition, $\dot{s}$ is the rate at which a fluid element moves up or down in s space. To avoid dimensional

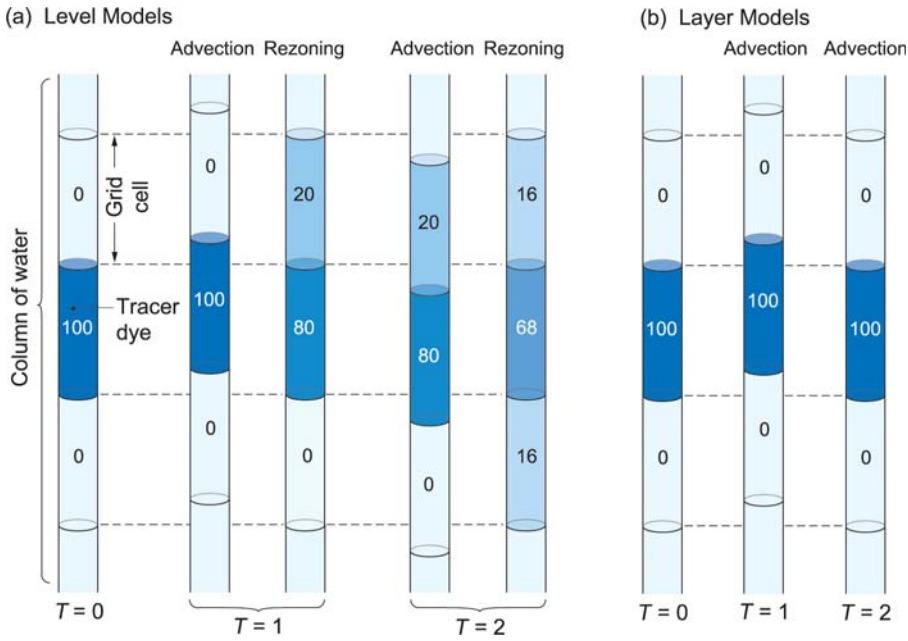

Figure 1. Schematic illustrating numerical dispersion in a water column subjected to gravity wave-induced oscillatory vertical motion. Shown is a stack of three grid cells. The initial state, T=0, is chosen to coincide with the wave trough, at which time the middle cell is assumed to be filled with a tracer of concentration 100. The approaching wave crest causes water to be advected upward by a distance chosen in this example to correspond to one-fifth of the vertical cell size (T=1). In level models (a), the clock is stopped momentarily to allow the tracer to be reapportioned ("rezoned") among the original grid cells. The next wave trough causes the water column to return to its original position (T=2). After renewed rezoning, tracer concentration in the middle cell has fallen to 68, the remainder having seeped into cells above and below. In layer models (b), the periodic rezoning steps are skipped, so tracer concentration remains unaffected by the wave motion.

ambiguities created by the lack of a physical definition of s, the quantity actually diagnosed is the interlayer mass flux $\dot{s}\partial p/\partial s$ which is always in units of pressure per time.) The latter forms the basis for vertically advecting all prognostic variables in the model grid. The simplicity of the mechanism expressed by (1) is one of the factors that make ALE-type hybrid modeling attractive.

Many ideas have been put forth on how the minimum thickness constraint in ALE ocean models should be formulated. One option is to scale the vertical spacing of nonisopycnic coordinate surfaces by the depth of the turbulent surface mixed layer, thereby ensuring that coordinate surfaces exist throughout the mixed layer for evaluating turbulent exchange processes. This concept is attractive at first sight but has shortcomings if the mixed layer depth changes rapidly, as typically happens during

transitions from surface cooling to warming. Not only do fluctuations in vertical grid spacing spawn fluctuations in the magnitude of the truncation error in the finite difference equations, but the large $\dot{s}$ values in (1) resulting from rapid coordinate movement are likely to unduly disperse water mass properties in the vertical. These effects are of particular concern if the mixed layer depth responds, as it does in nature, to the 24-hr cycle in solar radiation.

If suppression of excessive vertical migration of coordinate surfaces during the daily or annual heating-cooling cycle is deemed important, the optimal strategy is to "park" coordinate layers near the surface at those times (night or winter, respectively) when their target density does not exist. When surface warming makes the target density reappear, it will reappear at the sea surface. A coordinate layer lying in waiting near the surface can reattach itself to the target density with minimal vertical displacement. Reattachment will also take place sooner if intervening fixed-depth layers, associated with lighter target densities yet to appear, are kept as thin as possible.

False numerical dispersion of water mass properties caused by truncation errors in the finite-difference advection operators is a major concern in ocean modeling. The problem is particularly acute in the z direction where undulating vertical velocities associated with gravity wave trains can have a noticeable dispersive effect; in fact, elimination of vertical dispersion by gravity waves is often mentioned as one of the points in favor of using a material coordinate system. The dispersive effect of gravity waves is illustrated in Fig. 1.

While HYCOM's coordinate surfaces are for the most part material and thus remain unaffected by the dispersion problem just mentioned, the enforcement of minimum layer thickness constraints in the upper part of the ocean does open the door to vertical dispersion. This is of particular concern if the vertical regridding process displaces coordinate surfaces over large distances, because a large first term in (1) is likely to produce an $\dot{s}$ term of similar magnitude. Dispersion in this case likely will be larger than in a fixed-grid model whose $\dot{s}$ is given by the right-hand side of (1) and thus is bounded by dynamic constraints.

In HYCOM's original grid generator (see Bleck, 2002), the vertical "remapping" of prognostic variables following the "regridding" by the grid generator is formulated as a donor cell process: the amount of a variable X transferred from one grid cell to the next due to interface movement is computed under the assumption that property X is distributed uniformly within each grid cell. The donor cell scheme is known to be quite diffusive (as illustrated, for example, in Fig. 1), and upper-ocean vertical dispersion in HYCOM therefore has been a persistent concern.

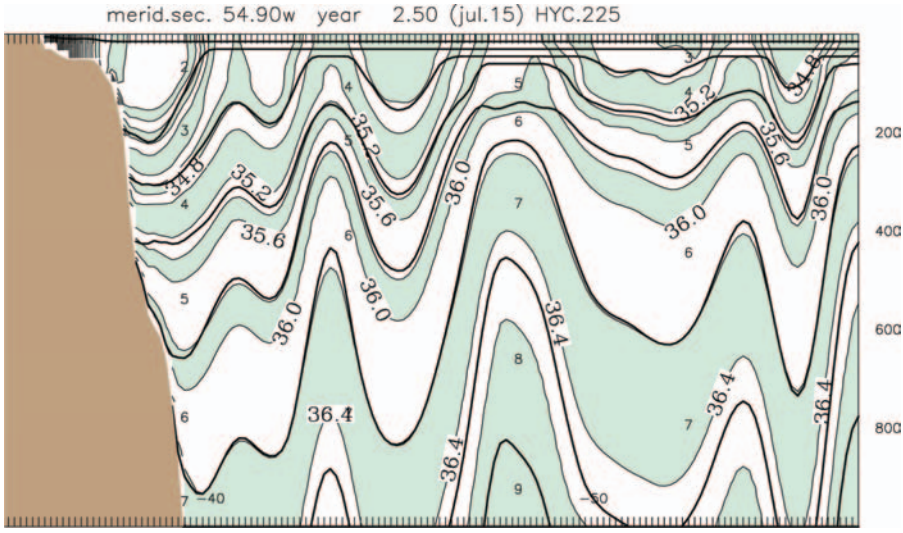

Figure 2. Sample vertical section through HYCOM solution extending south from Montevideo into the eddy-rich Brazil and Malvinas Current confluence region. Heavy lines: Layer interfaces. Shaded contours: potential density anomaly (σ_2, $\mathrm{kg\,m^{-3}}$). Tick marks along top and bottom indicate horizontal grid resolution [$0.225^\circ \times \cos(lat.)$, approx. 15 km]. Vertical scale: 1000 m. Crowded isopycnals on continental shelf (upper left corner) are due to Rio de la Plata inflow.

In the presently supported HYCOM version, the donor cell scheme has been replaced by the more accurate and less diffusive Piecewise Linear Method (PLM).

An illustration of how hybrid coordinates work in practice is given in Fig. 2.

3. Mutually consistent T/S/ρ advection

HYCOM development is by no means complete. One problem still awaiting a satisfactory solution is created by the fact that sea water potential density ρ is a function of two independent and equally influential tracers, potential temperature T and salinity S. Advecting both T and S in an isopycnic layer is at best redundant and, at worst, a source of "coordinate drift" – inconsistencies between T, S-implied potential density and the prescribed coordinate value. Since HYCOM's coordinate layers for the most part are isopycnic, HYCOM has inherited this problem from its predecessor MICOM.

Early versions of MICOM (Bleck et al., 1992) addressed this issue by invoking a "coordinate maintenance" algorithm which was also part of Oberhuber's (1993) isopycnic model. The strategy adopted in subsequent MICOM versions was to alleviate both the redundancy and the

consistency problem by advecting only one variable, S, and diagnose T from S and the coordinate value ρ. (This strategy clearly works only in isopycnic layers. MICOM's nonisopycnic slab mixed layer requires a second prognostic thermodynamic tracer aside from S.)

The pitfalls of diagnosing T from S and ρ are obvious and have prompted some MICOM users working on polar ocean circulation problems to switch from S to T as prognostic variable (e.g., Holland and Jenkins, 2001). Treating T as a diagnostic variable not subject to an explicitly enforced conservation law is also problematic in climate models used for predicting secular temperature changes in the ocean-atmosphere system. But the alternative, advecting T and diagnosing S everywhere in a global model, has its own drawbacks because of the strong correlation between ρ and T in the stratified low- to mid-latitude upper ocean which makes salinity a relatively poorly constrained diagnostic variable there.

Dispensing of a conservation equation for S may also be more detrimental to dynamic stability than dispensing of one for T because of the somewhat stronger control exerted by the atmosphere on the oceanic T field. This is to say that spurious salinity transients are harder to control in a model (in the absence of artificial restoring boundary conditions, that is) because of the lack of a natural restoring process on salinity akin to thermal relaxation. Given that salinity is more likely to act dynamically as a "loose cannon", one can argue that, globally speaking, S conservation is more important than T conservation in situations where a choice must be made between the two.

Since there is no guarantee that ρ is spatially uniform in any given HYCOM coordinate layer, HYCOM must everywhere carry two prognostic thermodynamic tracers. The strategy adopted in the production version is to treat both T and S as prognostic variables and delegate the coordinate maintenance task to the grid generator. Unfortunately, this choice is not optimal in all respects.

The real ocean has a tendency toward "density compensation", meaning that T, S fields evolve in a manner which minimizes the dynamic effects of T, S contrasts on the buoyancy field. This is to say that salinity fronts are often accompanied by compensating temperature fronts. The main problem with advecting T, S in a numerical model (any model, not just HYCOM) is that numerical shortcomings of the transport algorithm can and will destroy the spatial coherence of T, S fronts. In HYCOM, this will lead to localized ρ anomalies which the grid generator, in an attempt to restore target density, will convert into undulations in the layer thickness field. This adjustment in turn causes additional vertical dispersion.

It is for this reason that experiments continue in which T, S are replaced as prognostic variables in HYCOM by pairs of thermodynamic tracers which are more likely to maintain the coherence of T, S fronts during advection. In MICOM's isopycnic interior, this is presently achieved by advecting only one variable, S or T, and diagnosing the other one knowing ρ. The most straightforward generalization of this concept to the HYCOM case is to advect, as is done in MICOM's slab mixed layer, either the pair ρ, S or the pair ρ, T. An alternative approach, which addresses the pitfalls of diagnosing either T from S or S from T, is discussed below.

3.1 Spiciness

Because the density of near-freezing sea water is mainly a function of salinity, diagnosing T from ρ and S is an ill-posed problem in polar oceans – in the sense that small changes in S or ρ can bring about large changes in the diagnosed value of T. If advecting ρ in HYCOM is deemed important for maintaining the spatial coherence of T, S fronts, then the ideal second variable to be advected should be one whose isolines are everywhere orthogonal to isopycnals in T, S space. Such a variable exists and has become known as spiciness (Flament, 2002).

Orthogonality means that we need to construct a function χ satisfying

$$\begin{pmatrix} \partial\chi/\partial S \\ \partial\chi/\partial T \end{pmatrix} \cdot \begin{pmatrix} \partial\rho/\partial S \\ \partial\rho/\partial T \end{pmatrix} = 0.$$

Since

$$\begin{pmatrix} -\partial\rho/\partial T \\ \partial\rho/\partial S \end{pmatrix} \cdot \begin{pmatrix} \partial\rho/\partial S \\ \partial\rho/\partial T \end{pmatrix} = 0,$$

χ and ρ are connected through

$$\partial\chi/\partial S = -\partial\rho/\partial T \qquad\qquad \partial\chi/\partial T = \partial\rho/\partial S.$$

For dimensional consistency, T and S must appear in these expressions in nondimensional form. With this in mind, we can construct a perfect differential of χ,

$$d\,\chi = \left(-\frac{\partial\rho}{\partial(T/T_0)}\right) d(S/S_0) + \left(\frac{\partial\rho}{\partial(S/S_0)}\right) d(T/T_0), \tag{2}$$

which upon integration yields the sought-after spiciness function χ. A multiplicative or additive constant can be incorporated into the definition of χ at will.

The appearance of T_0, S_0 in (2) implies that orthogonality of ρ and χ is not "universal" but rather a matter of scaling. To visualize this

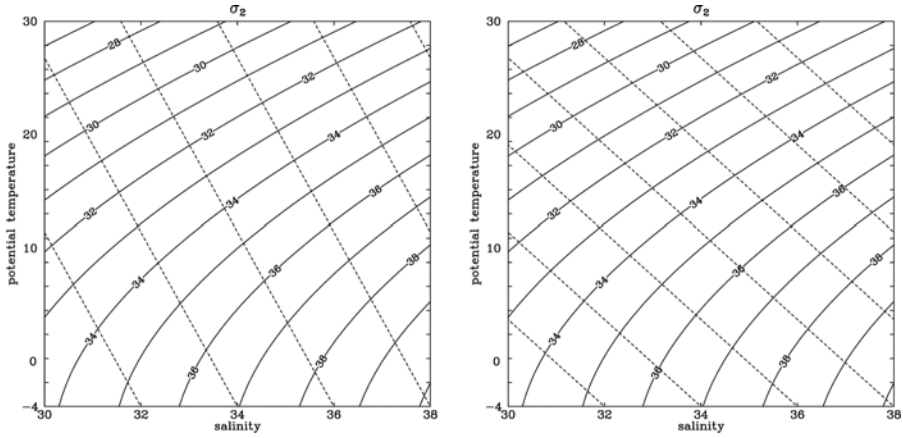

Figure 3. T, S diagrams showing ispopycnals referenced to 2000 m (solid) and two renditions of linearized spiciness χ' (dashed). Left: λ = -0.13 psu/°C; right: λ = -0.26 psu/ C.

dependence, consider two sets of orthogonal lines plotted in a T, S diagram, one set representing ρ and one representing χ. These lines lose their orthogonality as soon as the diagram is stretched in one or the other direction. The stretching operation is equivalent to changing T_0 and/or S_0.

The need in MICOM to recover T diagnostically from known values of ρ and S, and to do this in a noniterative fashion (iteratively obtained solutions tend to be unreliable in ill-posed problems), requires that the equation of state be approximated by a polynomial of at most 4^{th} degree in T. The approximation traditionally used is of 3^{rd} order in T and 1^{st} order in S (Brydon et al., 1999):

$$\rho(S,T) = c_1 + c_2T + c_3S + c_4T^2 + c_5ST + c_6T^3 + c_7ST^2. \qquad (3)$$

Assuming that T, S are already nondimensionalized, it is easy to derive from this polynomial the differential expression (2) and integrate the latter to obtain a polynomial expression for spiciness:

$$\chi(S,T) = -(c_2S + 2c_4ST + \frac{1}{2}c_5S^2 + 3c_6ST^2 + 2c_7S^2T^2) + \\ +(c_3T + \frac{1}{2}c_5T^2 + \frac{1}{3}c_7T^3). \qquad (4)$$

Recall that the goal of the present exercise is to advect buoyancy-related properties in terms of the pair ρ, χ instead of T, S. While computing ρ, χ from T, S at the beginning of each advection step is trivial, the inverse, i.e., recovering T, S from the advected ρ, χ fields by jointly

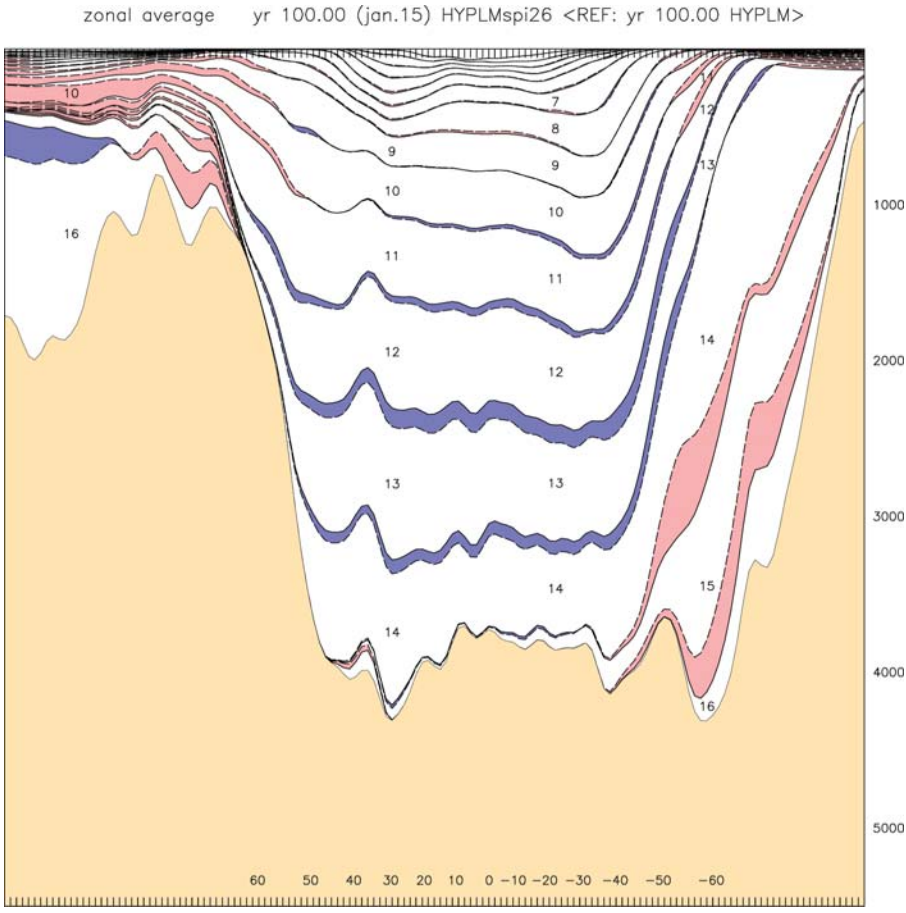

Figure 4. Vertical section through the zonally averaged density field after a 100-year, global, coarse mesh integration of HYCOM forced by monthly climatology [mesh size $2^\circ \times \cos(lat.)$]. Colors highlight differences in isopycnal layer depth resulting from using two different λ values, -0.26 psu/°C and 0. Blue/red: interfaces in λ = -0.26 psu/°C run are at shallower/greater depth, respectively, than interfaces in $\lambda = 0$ run.

solving (3) and (4) in noniterative fashion seems impossible. One way to overcome this obstacle is to use a linear approximation of (4) as the second advected variable. Let us write this variable as

$$\chi' = S + \lambda T \tag{5}$$

where λ is a free parameter which should be chosen to mimic the orthogonality of ρ and χ across the T, S range encountered in the world ocean. Solving the coupled system (3),(5) for T, S, given ρ and χ', is no more complicated than diagnosing T from (3), given ρ and S.

The dashed lines in Fig. 3 show two choices of χ' which roughly optimize $\rho - \chi'$ orthogonality at high and low temperatures. The corresponding λ values are -0.13 psu/°C and -0.26 psu/°C, respectively. (Note that χ' is expressed here in salinity units.)

The advantages of replacing ρ, S advection by ρ, χ' advection are difficult to quantify, mainly because the complexities of the wind- and thermohaline-forced ocean circulation make it hard to identify numerical solutions that are clearly impacted by errors resulting from advecting S and treating T as a diagnostic variable.

Fig. 4 illustrates the extent to which two HYCOM solutions, one based on ρ, S advection and one on ρ, χ' advection using λ = -0.26 psu/°C, diverge during a global 100-year coarse-mesh global simulation forced by monthly climatology. In this figure, the gap between isopycnal layer interfaces in the two simulations is colored red or blue depending on whether the interfaces in the χ'-based solution are at a greater or shallower depth, respectively, than the corresponding interfaces in the reference solution based on S advection ($\lambda = 0$). Fig. 4 shows that the use of spiciness leads to a very slight density increase at low-to mid-latitudes (blue coloration), while density is seen to decrease to a somewhat larger degree at high latitudes (red coloration).

Differences between a model run based on λ = -0.13 psu/°C and the S-advecting reference run are roughly half as large as those shown in Fig. 4.

It is virtually impossible to judge whether the ρ, χ'-based solution represents an improvement over the ρ, S-based one. We are able to state, however, that the switch from S to χ' advection induces much smaller changes (at least in this particular experiment) than what is typically seen in 100-year experiments when surface forcing fields or aspects of model physics are changed. This is a positive result, because it indicates that the increased robustness of the algorithm for diagnosing T from the two prognostic mass field variables is not achieved at the price of encountering new potentially harmful model sensitivities.

Perhaps the best evidence that advecting χ' instead of S leads to improved model performance is seen in a multi-century integration of a coupled ocean-atmosphere model consisting of HYCOM and the GISS atmospheric general circulation model. An early experiment based on the traditional ρ, S-advecting version of HYCOM showed incidents of anomalous ice growth occurring roughly once per century (grey curve in Fig. 5). Detailed diagnostics of these events showed that the ice growth took place in a region of the Arctic Ocean where slight errors in diagnosing T from ρ and S were compounded by a convective instability feedback which spread the error conditions over a large depth range and

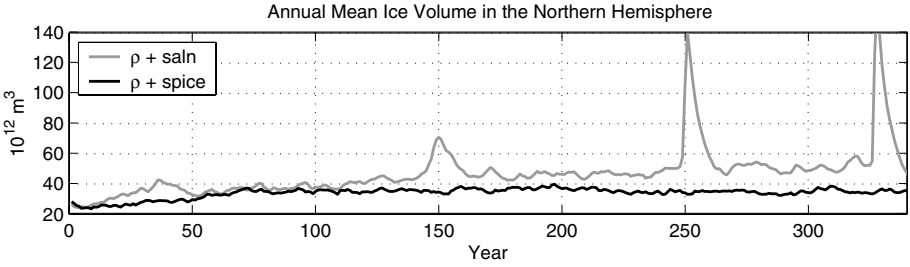

Figure 5. Time series of northern hemispheric ice mass in a coupled GISS-HYCOM simulation. Grey: results from ρ, S-advecting model; black: results from ρ, χ'-advecting model using $\lambda = -0.13\,\mathrm{psu}/^\circ\mathrm{C}$. Courtesy: Dr. Shan Sun, Goddard Institute for Space Studies.

thereby magnified and prolonged the heat loss. A subsequent run based on χ', S advection (black curve in Fig. 5) does not show this behavior. The overall lower ice mass in the second experiment is qualitatively consistent with the density loss in the high-latitude ocean seen in Fig. 4. While the ice loss happens to bring the coupled simulation closer to reality, it would be premature to conclude from this one experiment that the use of spiciness in general leads to more accurate solutions at high latitudes.

More experimentation is clearly needed before ρ, χ' advection in HYCOM can be accepted as being a better choice than T, S advection. Lack of formal conservation of quantities that are conserved in the real ocean (during advection, that is) is the biggest potential drawback of the proposed alternative. Compared to the choice of ρ, S as advected variables, however, ρ, χ' advection may well prove to be the better compromise between conflicting conservation and accuracy demands.

4. Closing remarks

The search for numerical models capable of accurately depicting the wind- and buoyancy-forced global ocean circulation has spawned a multitude of model formulations that differ, among other aspects, in the layout of the 3-dimensional computational mesh. This chapter describes some of the considerations that have led to the development of models attempting to combine the advantages of fixed-depth and isopycnic vertical coordinates.

Due to the vast range of spatial scales that cannot explicitly be incorporated into global ocean models, it is unlikely that we will ever find a model satisfying all users. However, models built around a hybrid (i.e., partially fixed-depth and partially isopycnic) vertical coordinate do provide flexibility not found in the classical, architecturally rigid x, y,

z-coordinate models in suppressing certain types of truncation errors.The newfound flexibility gives the user more choices in tayloring a model to specific physical situations, but it also means that far more experimentation is required to establish optimal model configurations than is necessary in the case of the x, y, z model class.

One example is vertical resolution. Because changes in the number or depth of coordinate surfaces in an x, y, z model often require elaborate adjustment of the model bathymetry, the grid layout in such models typically remains frozen long enough to allow thorough "tuning" of subgrid-scale physics parameterizations. In hybrid coordinate layer models, changing vertical resolution is a matter of changing a few parameters like target densities and minimum layer thicknesses. Consequently, hybrid model users are more likely to experiment with different vertical resolutions, a degree of freedom whose exploitation can have unintended consequences for subgrid-scale closure schemes.

The advantages offered by a hybrid vertical coordinate do not come without a price. Foremost among the complexities introduced by variable-depth layer models is the need to cast transport equations in flux form and to use relatively complex lateral transport operators that maintain the physical integrity of mass field tracers in situations characterized by strong (order-one) changes in layer thickness.

Another concern is the potential in hybrid-coordinate models for excessive vertical diffusion caused by the dispersive character of vertical advection schemes. If left uncontrolled, this diffusion can exceed that found in z coordinate models, for the simple reason that interlayer mass exchange can be much larger than the vertical transport rate seen in fixed-grid models. Limiting "capricious" interface movement, which according to (1) will spawn a compensating generalized vertical velocity of similar magnitude, therefore is an important step toward controlling vertical diffusion. Experiments to be reported elsewhere have shown that this diffusion is particularly noticeable in the transition zone between the isopycnic and fixed-depth coordinate subdomains, especially if the grid generator is unable to prevent abrupt vertical changes in layer thickness in the transition zone.

The areas of concern just mentioned must be weighed against the unquestionable advantages of the ALE-type vertical coordinate. Some of these advantages, listed here in no particular order, are

- substantial reduction of numerically induced diapycnal fluxes due to the presence of a sizable isopycnic subdomain;
- a larger isopycnic subdomain than can be achieved by other, more traditional "hybrid" coordinate schemes;

- smooth lateral transition between deep-ocean and coastal-shelf grid domains: no need to transfer boundary conditions between different models across the shelf break;
- retention of vertical resolution in unstratified water columns not achievable in pure isopycnic models, hence ability to incorporate turbulence closure and buoyant convection schemes developed for fixed-grid models;
- relatively simple modeling of abyssal dense overflows ("simple" compared to z models);
- the flexibility inherited from isopycnic models to accomodate massless layers on the sea floor, thereby circumventing the σ coordinate pressure gradient error.

Acknowledgements

The work presented here is being funded by the Climate Change Prediction and Ocean Science Programs of DOE's Office of Science, Climate Change Research Division. Dr. Shan Sun (NASA/GISS) contributed Fig. 5.

References

Bleck, R., 1978: On the use of hybrid vertical coordinates in numerical weather prediction models. *Mon. Wea. Rev., 106*, 1233–1244.

—, 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modelling, 4*, 55–88.

—, and D. B. Boudra, 1981: Initial testing of a numerical ocean circulation model using a hybrid (quasi-isopycnic) vertical coordinate. *J. Phys. Oceanogr., 11*, 755–770.

—, C. Rooth, D. Hu, and L. Smith, 1992: Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic. *J. Phys. Oceanogr., 22*, 1486–1505.

Brydon, D., S. Sun, and R. Bleck, 1999: A new approximation of the equation of state for sea water, suitable for numerical ocean models. *J. Geophys. Res., 104*, 1537–1540.

Cheng, W., R. Bleck, and C. Rooth, 2004: Multi-decadal thermohaline variability in an ocean-atmosphere general circulation model. *Climate Dyn., 22*, 573–590.

Flament, P., 2002. A state variable for characterizing water masses and their diffusive stability: spiciness. *Progr. Oceanog., 54*, 493-501.

Hirt, C.W., A.A. Amsden, and J.L. Cook, 1974: An arbitrary Lagrangian-Eulerian computing method for all flow speeds. *J. Comput. Phys., 14*, 227–253.

Holland, D.M., and A. Jenkins, 2001: Adaptation of an isopycnic coordinate ocean model for the study of circulation beneath ice shelves. *Mon. Wea. Rev., 129*, 1905–1927.

Holland, W.R., and L.B. Lin, 1975a: On the generation of mesoscale eddies and their contribution to the oceanic general circulation. I. A preliminary numerical experiment. *J. Phys. Oceanogr., 5*, 642–657.

—, and —, 1975b: On the generation of mesoscale eddies and their contribution to the oceanic general circulation. II. A parameter study. *J. Phys. Oceanogr., 5*, 658–669.

Hu, D., 1997: Global-scale water masses, meridional circulation, and heat transport simulated with a global isopycnal ocean model. *J. Phys. Oceanogr., 27*, 96–120.

Oberhuber, J.M., 1993: Simulation of the Atlantic circulation with a coupled sea ice-mixed layer-isopycnal general circulation model. Part I: model description. *J.Phys. Oceanogr., 23*, 808–829.

Sun, S., and R. Bleck, 2001: Atlantic thermohaline circulation and its response to increasing CO_2 in a coupled atmosphere-ocean model. *Geophy.Res.Ltrs., 28*, 4223–4226.

Welander, P., 1966: A two-layer frictional model of wind-driven motion in a rectangular oceanic basin, *Tellus, 18*, 54–62.

Chapter 5

NESTING OCEAN MODELS

Eric Blayo and Laurent Debreu
LMC-IMAG and INRIA Rhône-Alpes, Grenoble, France

Abstract: This note is focused on the problem of providing boundary conditions for regional ocean models. It is shown that usual methods generally do not address the correct problem, but more or less approaching ones. A tentative classification of these methods is proposed. Then their theoretical foundations are discussed, and recommendations are given.

Keywords: Open boundary conditions, regional models, nesting.

1. Introduction

The use of high resolution regional ocean models has become widespread in recent years, in particular due to the development of operational oceanography and coastal management systems. An important point, that has a strong influence on the quality of the results, is the way that a local model is forced at its open boundaries. Several methods, whose precise contents, theoretical justification, and practical performances are often somewhat difficult to compare precisely, are presently used in actual applications. In this context, the first aim of this note is to provide a tentative classification of these methods (section 1). Then we will discuss the one-way (section 2) and two-way (section 3) interactions, focusing on the theoretical foundations and practical use of the different approaches. Some final remarks on the available software tools and on the problem of data assimilation within nested models are given in sections 4 and 5.

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 127-146.

2. A classification of nesting problems

2.1 General framework

We are interested in representing as accurately as possible the ocean in a local domain Ω_{loc}. The circulation is supposed to be described on a time period $[0,T]$ by a model which can be written symbolically

$$L_{loc}u_{loc} = f_{loc} \quad \text{in}\Omega_{loc} \times [0,T] \tag{1}$$

with convenient initial conditions at $t = 0$. L_{loc} is a partial differential operator, u_{loc} is the state variable, and f_{loc} the model forcing. The conditions at the solid boundaries will never be mentioned in this note, since they do not interfere with our subject.

Since Ω_{loc} is not closed, a portion of its boundary does not correspond to a solid wall, and has no physical reality. This artificial interface, also called *open boundary* (OB), is denoted Γ. The local solution u_{loc} is thus in interaction with the external ocean through Γ, and the difficulty consists in adequately representing this interaction in order to get a good approximation of u_{loc} in $\Omega_{loc} \times [0,T]$.

We also assume that we have at our disposal a (probably less accurate) representation of the external ocean, either under the form of some data u_{ext} or of an external model

$$L_{ext}u_{ext} = f_{ext} \quad \text{in } \Omega_{ext} \times [0,T] \tag{2}$$

where Ω_{ext} is an external oceanic domain. Note that, in our notations, Ω_{loc} and Ω_{ext} do not overlap (Figure 1).

The best way to solve the local problem is then probably to use an inverse approach (*e.g.* Bennett, 2002), *i.e.* for example

$$\text{Find } u_{loc} \text{ that minimizes} \quad \|L_{loc}u_{loc} - f_{loc}\|^2_{\Omega_{loc}\times[0,T]} + \varepsilon \, \|u_{loc} - u_{ext}\|^2_{\Gamma\times[0,T]} \tag{3}$$

where the norms are defined conveniently and take into account some statistical knowledge on the errors on u_{ext} and on the model (1), and where ε is a weighting factor. One can also consider that the model is perfect, and minimize only $\|u_{loc} - u_{ext}\|^2_{\Gamma\times[0,T]}$, i.e. control the boundary values, under the constraint (1) (*e.g.* Taillandier *et al.*, 2004).

However solving such an inverse problem is quite difficult and expensive. That is why ocean modellers usually use direct approaches. The goal is then to find u_{loc} satisfying (1) that connects adequately to u_{ext} through Γ. The mathematical formulation of this problem is generally not expressed clearly in actual applications. Since Γ has no physical reality, the connection between u_{ext} and u_{loc} should be as smooth as possible,

i.e. generally continuous and differentiable. Therefore a correct direct formulation of the problem can be the following :

$$\left\| \begin{array}{l} \text{Find } u_{loc} \text{ that satisfies} \\ \quad \left\{ \begin{array}{ll} L_{loc}u_{loc} = f_{loc} & \text{in } \Omega_{loc} \times [0,T] \\ u_{loc} = u_{ext} \quad \text{and} \quad \dfrac{\partial u_{loc}}{\partial n} = \dfrac{\partial u_{ext}}{\partial n} & \text{on } \Gamma \times [0,T] \end{array} \right. \\ \text{under the constraint} \quad L_{ext}u_{ext} = f_{ext} \quad \text{in } \Omega_{ext} \times [0,T] \end{array} \right. \tag{4}$$

or equivalently :

$$\left\| \begin{array}{l} \text{Find } u_{loc} \text{ and } u_{ext} \text{ that satisfy} \\ \left\{ \begin{array}{l} L_{loc}u_{loc} = f_{loc} \quad \text{in } \Omega_{loc} \times [0,T] \quad \text{and} \quad L_{ext}u_{ext} = f_{ext} \\ \quad \text{in } \Omega_{ext} \times [0,T] \\ \qquad \text{with } u_{loc} = u_{ext} \quad \text{and} \quad \dfrac{\partial u_{loc}}{\partial n} = \dfrac{\partial u_{ext}}{\partial n} \quad \text{on } \Gamma \times [0,T] \end{array} \right. \end{array} \right. \tag{5}$$

where n denotes the normal direction. However, in actual applications, the external model is not always available for online interaction. Moreover it is defined generally on $\Omega_{ext} \cup \Omega_{loc}$ (*i.e.* it fully overlaps the local domain), and it would be quite expensive to modify it in order to avoid this overlapping by implementing an open boundary on Γ. Therefore most applications generally do not address thecorrect problem (5) itself, but rather more or less approaching problems.

Remark: the operators L_{ext} and L_{loc} generally differ, both in their continuous form (*e.g.* subgrid scale paramaterizations) and in their discretized form (the local numerical model often has a higher resolution than the external model). Moreover the forcings f_{ext} and f_{loc}, and the discretized bathymetries defining Ω_{ext} and Ω_{loc} can be rather different. In that case the regularity conditions in (5) cannot be satisfied, and the connection between u_{ext} and u_{loc} is unsmooth, which is of course non-physical. That is why it is recommended to define the models and forcings in order to ensure as far as possible the smoothness of the transition between the two models. This can be done for instance into a transition zone defined in the vicinity of Γ.

2.2 The different approaches

The usual approaches can be classified as follows:

The open boundary problem. This is the usual case where the local model only is used. The problem writes

$$\left\{ \begin{array}{ll} L_{loc}u_{loc} = f_{loc} & \text{in}\Omega_{loc} \times [0,T] \\ Bu_{loc} = g & \text{on } \Gamma \times [0,T] \end{array} \right. \tag{6}$$

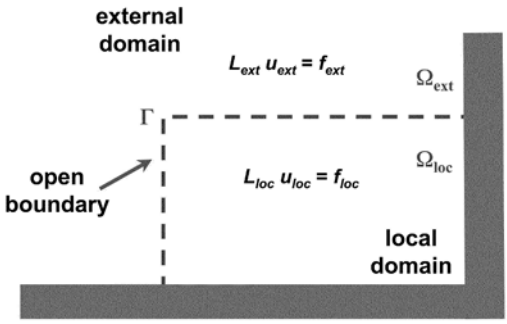

Figure 1. A schematic view of the nesting problem.

where B denotes an open boundary operator. The choice of B and g will be discussed in §3.

A particular case: one-way nesting. It is frequent that the solution u_{ext} of an external model covering an area $\Omega_{ext} \cup \Omega_{loc}$ larger than Ω_{loc} is available. Therefore this larger scale solution can be used to force the local model along Γ. The formulation of the problem which is solved in that approach is:

$$\left\| \begin{array}{l} L_{ext} u_{ext} = f_{ext} \quad \text{in } \Omega_{ext} \cup \Omega_{loc} \times [0,T] \\ \text{then} \\ \quad \left\{ \begin{array}{ll} L_{loc} u_{loc} = f_{loc} & \text{in } \Omega_{loc} \times [0,T] \\ B u_{loc} = B u_{ext} & \text{on } \Gamma \times [0,T] \end{array} \right. \end{array} \right. \tag{7}$$

This interaction between the two models can be performed on-line (the two models are run together) or off-line (the external solution is taken from an archive). In the case of an on-line interaction, u_{ext} is available at every external model timestep, while it is generally subsampled or averaged (*i.e.* of lesser quality) in the case of an off-line interaction, in order to limit the storage volume.

Note that this problem (7) is a particular case of the open boundary problem (6). It is different from the target problem (5) because u_{ext} is computed not only on Ω_{ext} but on the global domain $\Omega_{ext} \cup \Omega_{loc}$. Therefore both the external and the local equations are supposed to be relevant in Ω_{loc}. This assumption can be rather reasonable in the particular case where both models are identical except for the resolution,

and u_{ext} can be in that case a correct approximation of u_{loc}. However, as mentioned previously, L_{ext} and L_{loc} generally differ, as well as the forcing terms and the bathymetries. The quality of u_{ext} is then lesser, which will degrade the estimation of u_{loc}. Moreover, since this approach is only one-way, u_{loc} never acts on u_{ext}, and the external model cannot be improved.

Usual two-way nesting. An immediate possibility to address this shortcoming is to add a feedback from the local model onto the external one. Formulation (7) then becomes

$$\left\| \begin{array}{l} L_{ext}u_{ext} = f_{ext} \quad \text{in } \Omega_{ext} \cup \Omega_{loc} \times [0,T] \\ \text{then} \\ \left\{ \begin{array}{ll} L_{loc}u_{loc} = f_{loc} & \text{in } \Omega_{loc} \times [0,T] \\ Bu_{loc} = Bu_{ext} & \text{on } \Gamma \times [0,T] \end{array} \right. \\ \text{then} \\ u_{ext} = Hu_{loc} \quad \text{in } \Omega_{loc} \times [0,T] \end{array} \right. \tag{8}$$

where H is an update operator, mapping u_{loc} from its time and space grid onto the grid of the external model. This implies of course that the external model is fully available, and that both models are run together with on-line interaction. The update can be performed at each external model timestep, or less frequently.

In this approach, the local solution has some influence onto the external one, the goal being to get closer to the target problem (5) without having to modify the external model.

Full coupling. As mentioned previously, the correct approach should be to solve (5). However this implies first to modify the external model by defining an open boundary on Γ in order to avoid overlapping Ω_{loc}, and also to find an interaction procedure that makes u_{loc} and u_{ext} satisfy the regularity conditions on Γ. We will see in §4.2 how this can be done. Such methods are quite recent, and are not yet disseminated in the ocean and atmosphere modelling community.

2.3 A numerical example

Let us now illustrate the different preceding approaches in the very simple case of a 1-D ordinary differential equation. The problem is:

$$\left\{ \begin{array}{ll} -\nu(x)u''(x) + u(x) = \sin n\pi x & x \in]0,1[\\ u(0) = u(1) = 0 \end{array} \right. \tag{9}$$

The local domain we are interested in is $\Omega_{loc} =]a,b[$; hence $\Omega_{ext} =]0,a[\cup]b,1[$. $\nu(x)$ is displayed on Figure 2a. It is equal to ν_0 in Ω_{ext}, and

$\nu_0/\sqrt{2}$ in Ω_{loc}, except in two small transition zones of width δ, where it varies smoothly between ν_0 and $\nu_0/\sqrt{2}$. This problem has a unique solution (Brezis, 1983), denoted u_{ref}, which is plotted on Figure 2b. The elliptic nature of this problem amplifies the influence of the boundary conditions, which will help highlighting the differences between the nesting approaches described in §2.

Open boundary problem. Solve $-\nu(x)u''_{obc}(x)+u_{obc}(x) = \sin n\pi x$, $x \in]a,b[$, with OBCs at a and b. Such OBCs can be for example Dirichlet conditions $u_{obc}(a) = \alpha_0, u_{obc}(b) = \beta_0$, or Neumann conditions $u'_{obc}(a) = \alpha_1, u'_{obc}(b) = \beta_1$. If the external data are perfect ($[\alpha_0,\beta_0] = [u_{ref}(a), u_{ref}(b)]$ or $[\alpha_1,\beta_1] = [u'_{ref}(a), u'_{ref}(b)]$) then we get the true solution u_{ref}. We have plotted in Figure 2b the case of imperfect Dirichlet data $\alpha_0 = \beta_0 = 0$.

One-way / two-way nesting. Since the problem is not time dependent, both one-way and two-way approaches yield the same solution u_{nes}, defined by :

$$\begin{cases} -\nu_0\, u''_{ext}(x) + u_{ext}(x) = \sin n\pi x, \quad x \in]0,1[\\ u_{ext}(0) = u_{ext}(1) = 0 \\ -\nu(x)u''_{nes}(x) + u_{nes}(x) = \sin n\pi x, \quad x \in]a,b[\\ B_a u_{nes}(a) = B_a u_{ext}(a) \quad \text{and} \quad B_b u_{nes}(b) = B_b u_{ext}(b) \end{cases} \tag{10}$$

We have plotted in Figure 2b the cases $B_a = B_b = Id$ and $B_a = B_b = \partial/\partial n$. As can be seen clearly, these methods, which are all supposed to approximate the true problem (9), yield quite different solutions, which can differ from u_{ref} both in Ω_{loc} and Ω_{ext}. Note also that the true problem (9), reformulated as (5), requires two BCs at a and b, while the approximate formulations require only one BC.

The same type of comparison is displayed in Figure 3, but for the realistic testcase of a high resolution model of the bay of Biscay coupled with an eddy-permitting model of the North Atlantic.

3. The open boundary problem

Let us now focus on the main point, central in all approaches, namely the choice of the open boundary operators B in (6)-(7)-(8). This is a difficult problem, which has been the subject of numerous studies for more than 30 years, ranging from purely mathematical approaches to specific modelling applications. Mathematical results are often obtained for simplified equations (e.g. linearized and/or inviscid). They generally address the derivation of OBCs, and the well-posedness of the model equations using these OBCs. Note that the well-posedness of the system

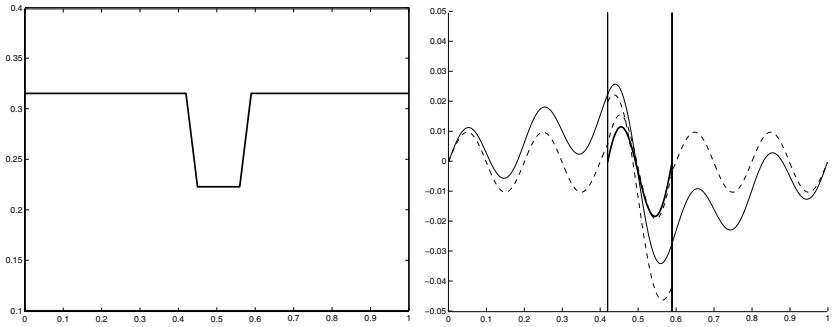

Figure 2. a) $\nu(x)$ for eq.(9); b) The different solutions (see text): u_{ref} (solid line), u_{obc} (thick solid line), u_{nes} with Dirichlet and Neumann OBCs (dashed lines). The two vertical lines correspond to $x = a$ and $x = b$.

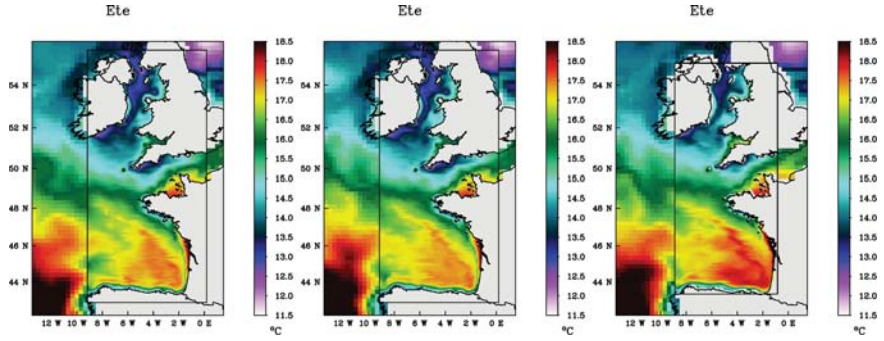

Figure 3. Averaged temperature at $z = 30m$ in spring 1998, in a $1/15°$ regional model of the bay of Biscay interacting with a $1/3°$ model of the north Atlantic. The internal rectangle corresponds to the limits of the regional model. The model of the north Atlantic is only partially shown. Three interactions procedures are compared (from S. Cailleau, 2004).

ensures the uniqueness of the solution and its stability with regard to initial datum, but does not give any information on its accuracy nor relevance with regard to the "true" solution u_{ref}. On the other hand, numerical studies can use complex realistic models, but their results seem often dependent on the test cases. We present here a brief overview of usual OBCs, and give a tentative explanation of their performance through the point of view of hyperbolic systems. The contents of this section is discussed in much more details in Blayo and Debreu (2005).

3.1 Usual OBCs

Several reviews of OBCs are available, either for ocean and atmosphere models or in a more general context. Let us mention for instance the introductory parts of the papers by Palma and Matano (1998), Marchesiello *et al.* (2001), Treguier *et al.* (2001), or the review papers by Givoli (1991), Tsynkhov (1998) or Holdstad and Lie (1999). OBCs are often classified roughly into two categories : *global* OBCs are usually accurate, but computationally expensive and difficult to implement; *local* OBCs are much cheaper and easier to implement, but also generally much less accurate and mathematically justified. We will give now briefly a list of OBCs used in the context of ocean and atmosphere modelling.

Relaxation methods. The goal of this widely used class of OBCs is to relax the model solution ϕ towards the external data ϕ^{ext} on (or in the vicinity of)Γ. The most brutal way to do this is to impose $\phi = \phi^{\text{ext}}$ on Γ, *i.e.* to use a Dirichlet (or *clamped*) boundary condition. Such a condition is often used in particular in the context of one-way nesting. However, a major drawback of this method is that the outflowing information is totally determined by these external data, and does not depend at all on the internal solution. Therefore part of the outgoing information will be reflected into the domain as soon as the external data is not perfectly consistent with the internal dynamics. One of the conclusions of a comparative study by Röed and Cooper (1987) in the context of a simple linear barotropic ocean model is that such a clamped BC should be avoided in most applications.

It is frequent in practical applications to use a more progressive method, called *flow relaxation scheme*. This approach consists in extending the computational domain Ω_{loc} by defining an additional domain Ω_s (the *sponge layer*), which interface with Ω_{loc} is Γ. In the original method proposed by Davies (1976), the model equations are numerically solved on $\Omega_{\text{loc}} \cup \Omega_s$, and the solution in Ω_s is replaced at each timestep by

$$(1 - \alpha)\, \phi + \alpha \phi^{\text{ext}} \tag{11}$$

where α is a relaxation function increasing from 0 on Γ to 1 far enough from Γ. While primarily designed for discretized equations, it can be shown easily (*e.g.* Martinsen and Engedahl, 1987) that this correction scheme can also be interpreted as adding a nudging term to the original model equations

$$\frac{\partial \phi}{\partial t} + F(\phi) = 0 \tag{12}$$

which become

$$\frac{\partial \phi}{\partial t} + F(\phi) + K(\phi - \phi^{\text{ext}}) = 0 \tag{13}$$

where K is a positive function, null on Ω_{loc} and increasing away from Γ (K depends on α and on the time-discretization scheme). Relaxation methods are often performed jointly with a sponge layer approach, which means that the model viscosity is artificially increased in Ω_s, in order to damp the local turbulent activity. Relaxation generally appears to be one of the best methods in comparative numerical studies (*e.g.* Röed and Cooper, 1987; Palma and Matano, 1998; Nycander and Döös, 2003).

Two drawbacks of these methods must however be emphasized. The first one is the increase of the computational cost induced by the additional layers Ω_s. The ratio of this additional cost to the cost of the initial model is roughly equal to $|\Omega_s|/|\Omega_{\text{loc}}|$, and can either be negligible or reach some tens of percents, depending on the configuration. The second drawback is the empirical aspect of the governing equation (13) in the sponge layer.

Finally, note also that perfectly matched layer (PML) methods, which have been proposed quite recently in the context of electromagnetism (Berenger, 1994), can be seen as an improvement of relaxation methods. This methodology consists basically in a convenient splitting of the equations with addition of relaxation terms with well-chosen coefficients. PML approach has been applied to the Euler equations (Hu, 1996, 2001) and to the shallow water equations (Darblade *et al.*, 1997; Navon *et al.*, 2004), and leads to improved results in academic test cases. It must now be validated in realistic configurations to get a better evaluation of its actual effectiveness.

Radiation methods. A very popular class of OBCs are *radiation methods*. They are based on the Sommerfeld condition :

$$\frac{\partial \phi}{\partial t} + c\,\frac{\partial \phi}{\partial n} = 0 \tag{14}$$

which corresponds to the transport of ϕ through Γ (n is the outward normal vector) with the velocity c.

Orlanski (1976) proposed a numerical implementation of this condition for complex flows, including an adaptive evaluation of c. A number of variants were then derived, using alternative computations of c, and/or taking into account the tangential derivative, and/or including an additional relaxation term (*e.g.* Camerlengo and O'Brien, 1980; Miller and Thorpe, 1981; Raymond and Kuo, 1984; Barnier *et al.*, 1998; Marchesiello *et al.*, 2001).

Such radiation methods are frequently used in ocean and atmosphere modelling. However their relevance for such complex flows is far from obvious. Their reputation is split: they have proved to give rather poor results in several comparative studies (*e.g.* Röed and Cooper, 1987; Palma and Matano, 1998; Nycander and Döös, 2003), while they seem to have some efficiency in others (*e.g.* Marchesiello *et al.*, 2001; Tréguier *et al.*, 2001). In fact the Sommerfeld condition is justified only in the context of wave equations with a constant phase velocity (Blayo and Debreu, 2005). Applying such a condition to variables which do not satisfy at all such equations results in a fundamental nonlinearity, which has been recently pointed out by Nycander and Döös (2003). Therefore this condition cannot be mathematically justified in the context of ocean and atmosphere modelling. However, its actual implementations give an important role to external data. As indicated previously, the radiation velocity c is evaluated at each timestep and at each gridpoint on the open boundary. If c is inward, the model variable is generally set to the corresponding external value: $\phi = \phi^{\text{ext}}$, or strongly relaxed towards it:

$$\frac{\partial \phi}{\partial t} = -\frac{\phi - \phi^{\text{ext}}}{\tau_{\text{in}}} \tag{15}$$

where τ_{in} is a short relaxation timescale. If c is outward, then the radiation equation is applied, but often with the addition of a relaxation term:

$$\frac{\partial \phi}{\partial t} + c\,\frac{\partial \phi}{\partial n} = -\frac{\phi - \phi^{\text{ext}}}{\tau_{\text{out}}} \tag{16}$$

where τ_{out} is a longer relaxation timescale. In their careful analysis of a simulation of the Atlantic ocean, Tréguier *et al.* (2001) have observed that c behaves in some sense like a white noise, and is directed inwards about half of the time at any location on the open boundaries. Therefore the model solution at the open boundary never departs significantly from the external data, and the radiation condition acts in fact nearly as a clamped condition. So it is probably the strong influence of the external data through the additional relaxation term in the radiation conditions that gives them most of their practical efficiency, rather than the radiation procedure.

Flather condition. Flather (1976) proposed an OBC for 2-D barotropic flows, which is often classified within the family of radiation conditions. This condition can be obtained by combining the Sommerfeld condition for the surface elevation η (with surface gravity waves phase speed)

$$\frac{\partial \eta}{\partial t} + \sqrt{gh}\,\frac{\partial \eta}{\partial n} = 0 \tag{17}$$

with a one-dimensional approximation of the continuity equation

$$\frac{\partial \eta}{\partial t} + h\,\frac{\partial v_n}{\partial n} = 0 \tag{18}$$

where g is the gravity, h is the local water depth and v_n is the normal component of the barotropic velocity. Substracting (17) to (18) and integrating through Γ, one obtains:

$$v_n - \sqrt{\frac{g}{h}}\,\eta = v_n^{\text{ext}} - \sqrt{\frac{g}{h}}\,\eta^{\text{ext}} \tag{19}$$

The Flather condition has been used in several comparative studies (*e.g.* Palma and Matano, 1998; Marchesiello *et al.*, 2001; Nycander and Döös, 2003), and it always appears to be one of the most efficient conditions.

Model adapted methods. A striking aspect of radiation and relaxation methods is that the OBCs do not depend on the model equations. On the opposite, other methods provide OBCs which are adapted to the system. However, since they are more complicated to handle, the use of such methods is quite rare and restricted to simple 1-D or 2-D models, and has never been extended to our knowledge to realistic primitive equations systems.

This is the case of **characteristic waves amplitudes methods** (sometimes called *Hedström methods*), designed for hyperbolic systems. The basic idea consists in choosing for OBCs the original set of model equations with as few approximations as possible. Since the only quantities that cannot be evaluated by the model alone are the incoming characteristics (see §3.2) the approximations must concern only these terms, and eventually the viscous terms if the model is not inviscid. This results in setting to zero (or to a value deduced from external data) the normal derivative of the incoming characteristic variables on Γ. Several papers developed this idea these last years in the context of direct numerical simulation of compressible Euler and Navier-Stokes equations, with apparently good experimental results (Poinsot and Lele, 1992; Bruneau, 2000; Bruneau and Creusé, 2001). In the context of ocean modelling, it is compared to other OBCs by Röed and Cooper (1987), Jensen (1998) and Palma and Matano (1998), and leads to rather good results.

Another important family of methods are **absorbing conditions**, which are *exact* relations satisfied by the outgoing quantities at the open boundary. In a reference paper, Engquist and Majda (1977) give a general method for obtaining such relations, using time and space Fourier transforms. However, these conditions are generally global in time and space, and cannot be used just as it is in practice. That is why they

must be approximated to give tractable local conditions. A strong interest of this approach is its sound mathematical foundation, and its practical efficiency in several domains of applications. Several papers have recently readdressed the derivation of absorbing BCs for the inviscid shallow water system, and obtain apparently quite good numerical results (Lie, 2001; McDonald, 2002, 2003; Nycander and Döös, 2003).

3.2 An hyperbolic point of view

When attempting to draw some synthesis of the numerous previous studies on OBCs, two keypoints stand out, which seem to be necessary constituents for any good OBC. The first point is that good results are obtained when taking primarily into account the hyperbolic part of the dynamics, and therefore when working on incoming characteristic variables. The second point is that this must be associated with a consistent use of some *external data*.

Incoming characteristic variables. Let us first introduce some standard definitions concerning hyperbolic systems. The general form of such a system is

$$\frac{\partial \Phi}{\partial t} + A(\Phi)\,\frac{\partial \Phi}{\partial x} = F \tag{20}$$

where $\Phi(x,t)$ is a vector of n functions, $A(\Phi)$ is a $n \times n$ matrix of functions of Φ, and F is a forcing term. For the system to be hyperbolic, A must have n real eigenvalues and n distinct eigenvectors. Let W_k the k^{th} left eigenvector of A, corresponding to the k^{th} eigenvalue λ_k: $W_k^T\, A = \lambda_k\, W_k^T$. Multipliying (20) on the left by W_k^T, one gets:

$$W_k^T\,\frac{d_k \Phi}{dt} = W_k^T\, F \qquad \text{with} \quad \frac{d_k}{dt} = \left(\frac{\partial}{\partial t} + \lambda_k\,\frac{\partial}{\partial x}\right) \tag{21}$$

The operator d_k/dt represents a total (or directional) derivative in the direction defined by $\dfrac{dx}{dt} = \lambda_k$. To the hyperbolic system (20) correspond n such families of curves, which are called *characteristic curves* of the system. If the system (20) is linear with constant coefficients, *i.e.* if A is a constant matrix, one can define the new variables $w_k(x,t) = W_k^T \Phi(x,t)$. (20) is then equivalent to the system of n uncoupled transport equations:

$$\frac{\partial w_k}{\partial t} + \lambda_k\,\frac{\partial w_k}{\partial x} = W_k^T\, F \qquad k = 1, \ldots, n \tag{22}$$

The characteristic curves in that case are the lines $x - \lambda_k t =$ constant, along which the w_k (called *characteristic variables* or *Riemann invariants*) are conserved. One can notice that, at a given boundary, these

characteristic variables will be either inflowing or outflowing, depending on the sign of λ_k.

A fundamental point is that, for a hyperbolic open boundary problem to be well-posed, one must prescribe as many boundary conditions as the number of incoming characteristics. This result is in fact quite intuitive: the solution can be decomposed into outgoing and incoming characteristics; information on the former is available within the computation domain, and no additional condition is required, while information on the latter is not available, and mustbe specified.

Consistency with external data. The second keypoint concerns the connection with external data. It appears that a reasonable choice consists in imposing the consistency locally all along the boundary. This means that the OBC is of the form

$$B\phi = B\phi^{\text{ext}} \tag{23}$$

where B is the open boundary operator. $B = Id$ corresponds to the continuity of ϕ through the boundary, and $B = \partial/\partial n$ to the continuity of the flux. Such a formulation (23) is quite natural for example if we consider that the external data ϕ^{ext} represents some steady state or far field solution ϕ^{∞}. In that case, as detailed for example by Engquist and Halpern (1988), if we want the model solution to converge to the steady state solution as $t \to \infty$, then the OBC *must* also be satisfied by ϕ^{∞}.

Used together with the point of view of characteristic variables presented previously, this condition (23) leads to recommending OBCs of the form

$$Bw = Bw^{\text{ext}} \tag{24}$$

where w is any incoming characteristic variable of the governing equations.

The extension to non-hyperbolic systems, like for example the Navier-Stokes equations, is not trivial. A logical approximation consists however in considering only the hyperbolic part of thesystem, and to use the same procedures as for the hyperbolic case.

Revisiting usual OBCs. The preceding criteria give a new light on usual OBCs. It appears indeed that:

- the Sommerfeld condition (14) corresponds to prescribing to zero the incoming characteristic of the wave equation. That is why it is legitimate for wave equations but not for other systems.

- the Flather condition (14) corresponds to specifying the value of the incoming characteristic of the shallow water system, fulfilling the criterion (24) with $B = Id$ the identity operator.

- absorbing conditions are closely linked to incoming characteristic variables, and the conditions proposed by McDonald (2002, 2003) and Nycander and Döös (2003) can be written under the form (24).

- characteristic waves amplitudes methods do also meet the preceding point of view.

- since relaxation methods are not local conditions, the criterion (24) does not apply directly. However, it is obvious from (11) that the transition from ϕ to ϕ^{ext} is smooth as soon as the additional domain Ω_s is large enough. Similarly the problem of specifying incoming characteristics and evacuating outcoming characteristics at the open boundary is treated implicitely: the values of the incoming characteristics are computed within Ω_s, using the relaxed solution, while the outgoing characteristics are not directly affected when reaching Γ but are relaxed in Ω_s towards their corresponding external values, and damped by the increased dissipation.

Details on these aspects, as well as an application of the criterion (24) to shallow-water and primitive equations systems, are discussed in Blayo and Debreu (2005).

3.3 Some practical remarks

- It is important to note that we discussed here only the continuous form of the equations. However discretized models contain spurious numerical modes, which nature is different from that of physical modes, and which have to be handled by the OBCs. Therefore, once the continuous form of the OBCs is chosen, one has to perform some specific work in order to adapt their numerical implementation to the numerical schemes of the model. This difficulty is probably also a reason for the efficiency of relaxation and radiation-relaxation methods, which tend to automatically damp these non-physical modes.

- Incoming information is entirely given by the external solution ϕ^{ext}. Therefore the quality of these data is of course an important point in the performance of a regional modelling system.

- Another important practical aspect in a regional modelling system is the initialization problem. The initial condition is generally built

by interpolation of a larger scale solution, which is not perfectly consistent with the local model. This can yield an adjustment phase which can be quite long, and which pollutes the model solution. A way to avoid (or limit) this problem is to add some relevant constraints in the computation of the initial condition, as done for instance by Auclair *et al.* (2000) using an inverse approach. This aspect is presently the subject of numerous studies.

4. Two-way interaction

4.1 Two-way nesting

As explained in §2, the usual two-way method differs from the preceding one-way method by the addition of an update procedure. This supplementary step aims at improving $u_{\rm ext}$ by modifying it locally using $u_{\rm loc}$. This retroaction from the local model onto the external model is performed every external model timestep, or less frequently. The update operator generally replaces the values of $u_{\rm ext}$ at gridpoints located in $\Omega_{\rm loc}$ by copying the corresponding values of $u_{\rm loc}$, eventually after some time and space averaging. Such an update is quite brutal, and in particular does not ensure the balance of mass and tracers fluxes through Γ. For example, $\int_\Gamma \mathbf{U}_{\rm loc}.\mathbf{n} \neq \int_\Gamma \mathbf{U}_{\rm ext}.\mathbf{n}$, where $\mathbf{U}$ denotes the velocity. That is why a flux correction step is often added, which generally modifies $u_{\rm loc}$ to distribute the flux misfit all along Γ, to get finally a local solution $u^*_{\rm loc}$ which is in flux balance with $u_{\rm ext}$.

The two-way method generally decreases the difficulties that can be encountered by the one-way method (in particular the instabilities along Γ), and seems to improve the model solution. That is why it is recommended to use it as far as possible rather than one-way interaction. However, it is clear that the solution provided by this usual two-way nesting is not solution of the original problem (5): before the flux correction step, the connection between $u_{\rm ext}$ and $u_{\rm loc}$ is not differentiable, because their fluxes are not balanced; after the flux correction step, the connection is no more continuous because $u_{\rm loc}$ has been modified into $u^*_{\rm loc}$, which in addition does not satisfy any longer the local model equations (1).

4.2 Full coupling - Schwarz methods

Obtaining a solution of the original problem (5) is much more difficult and expensive than what is done in the above usual algorithms. This is mainly due to the fact that, since the local and external model equations are different, their domains of application should not overlap. Therefore

the external model, which is generally available in a configuration fully overlapping $\Omega_{\rm loc}$, must be modified to add an open boundary. Moreover, once this is done, one has to find and implement an algorithm ensuring that the solutions $u_{\rm ext}$ and $u_{\rm loc}$ will satisfy the desired regularity conditions through Γ.

These difficulties explain that this problem has never been addressed before in ocean and atmosphere modelling. This can be done however within the mathematical framework of domain decomposition methods. These methods have been intensively studied and developed since the end of the eighties due to the advent of parallel computers. Without going into details, let us present the *global-in-time non-overlapping Schwarz algorithm*, which seems well suited for our ocean coupling problem. This iterative algorithm can be written as follows:

$$\left\{\begin{array}{rll} L_{loc}u_{loc}^{n+1} & = f_{loc} & \text{in } \Omega_{loc}\times[0,T] \\ u_{loc}^{n+1} & \text{given} & \text{at } t=0 \\ B_{loc}u_{loc}^{n+1} & = B_{loc}u_{ext}^{n} & \text{on } \Gamma\times[0,T] \end{array}\right.$$

$$\text{and} \quad \left\{\begin{array}{rll} L_{ext}u_{ext}^{n+1} & = f_{ext} & \text{in } \Omega_{ext}\times[0,T] \\ u_{ext}^{n+1} & \text{given} & \text{at } t=0 \\ B_{ext}u_{ext}^{n+1} & = B_{ext}u_{loc}^{n} & \text{on } \Gamma\times[0,T] \end{array}\right. \tag{25}$$

where the superscripts denote the number of iterations, and B_{loc} and B_{ext} are interface operators to be chosen. Note that, at each iteration, the two models can be run in parallel over the whole time window $[0,T]$. If no parallel computer is available, the interface condition for u_{ext} can be replaced for example by $B_{ext}u_{ext}^{n+1} = B_{ext}u_{loc}^{n+1}$, which prevents parallelism but increases the convergence rate of the algorithm.

This rate closely depends of the choice of B_{loc} and B_{ext}. An obvious possibility is to choose the operators Id and $\partial/\partial n$. Therefore, once the algorithm has converged, its solution will satisfy (5). However the convergence can be quite slow and, given the computational burden of ocean models, one probably cannot afford numerous iterations of such an algorithm. That is why the choice of the interface operators must be optimized. A simple but quite efficient possibility is to use Robin conditions: $B_{loc} = \partial/\partial n + r_{loc}Id$ and $B_{ext} = \partial/\partial n + r_{ext}Id$ with $r_{loc} \neq r_{ext}$. This ensures the desired regularity as previously for the converged solution, but a good choice of the coefficients r_{loc} and r_{ext} can greatly speed up the convergence. More sophisticated approaches can be used to determine good interface operators, which are closely linked to characteristic methods and absorbing conditions. Martin (2003) applied such approaches to 2-D tracer equations and to the shallow-water system. She derived very efficient operators, which ensure the convergence of the

algorithm in some very few iterations. Development of such algorithms for realistic ocean models is ongoing research work.

5. Software tools

Designing nested or coupled systems starting from existing models is quite a difficult and time-consuming practical task. However several software tools have been developed these last years, which automatically manage an important part of the job.

The **AGRIF package**[1] (Debreu *et al.*, 2004a) allows an easy integration of mesh refinement capabilities within any existing finite-difference model written in Fortran. One can therefore design one-way and two-way multiply-nested systems, with the possibility of adaptive regridding, without reprogramming the model. This package is presently implemented into several operational ocean models.

General couplers can also be used to implement nested systems, especially in the case when the local and external model codes are totally different. The user has then to prescribe the structure of the coupling algorithm and the interactions between the different objects, but at a rather high level, without having to go too much into programming details. In the context of geophysical fluids, we can cite for instance **PALM**[2] or **MpCCI**[3].

6. Data assimilation and nesting

Along with the development of nested ocean modelling systems, the problem of assimilating data within these systems is presently strongly emerging. Addressing this difficult problem is out of the scope of this note. Let us however point out a few related issues.

- The exact mathematical formulation of the data assimilation problem for one-way or two-way nested systems is far from obvious. A first attempt in this direction for the 4D-Var approach can be found in Debreu *et al.* (2004b). Concerning the stochastic approach, interesting ideas can probably be found in the theories of multiresolution stochastic models and multiscale estimation.

- Several ad-hoc procedures are already in use in numerous systems. A possibility is to perfom the assimilation only on one grid (the largest or the finest) of the system. Another way is to "hide" the

[1] `http://www-lmc.imag.fr/IDOPT/AGRIF`
[2] `http://www.cerfacs.fr/~palm`
[3] `http://www.mpcci.org`

grid interaction process and to make the assimilation globally on a multiresolution state vector (*e.g.* Barth *et al.*, 2004).

- It is possible in a variational method to manage simultaneously the coupling problem and the assimilation problem. See for instance Bounaim (1999) or Taillandier *et al.*(2004).

- In a multiresolution modelling system, one has to choose which data are assimilated on which grid. Since the model dynamics depends on the grid resolution, and since the data themselves have often been collected or processed with some spatial and temporal resolution, this choice is not obvious and has consequences on the quality of the identified solution.

References

Auclair, F., S. Casitas, and P. Marsaleix, 2000: Application of an inverse method to coastal modelling. *J. Atmos. Oceanic Technol.*, **17**, 1368–1391.

Barnier, B., P. Marchesiello, A.P. de Miranda, J.M. Molines, and M. Coulibaly, 1998: A sigma coordinate primitive equation model for studying the circulation in the South Atlantic I, Model configuration with error estimates. *Deep Sea Res.*, **45**, 543–572.

Barth, A., A. Alvera-Azcarate, J.-M. Beckers, M. Rixen, L. Vandenbulke and Z. Ben Bouallegue, 2004: Multigrid state vector for data assimilation in a two-way nested model of the Ligurian sea. *36th International Liege Colloquium on Ocean Dynamics.*

Bennett, A.F., 2002: Inverse modeling of the ocean and atmosphere. Cambridge University Press, 2002.

Berenger, J.-P., 1994: A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, **114**, 185–200.

Blayo, E., and L. Debreu, 2005: Revisiting open boundary conditions from the point of view of characteristic variables. *Ocean Modelling*, **9**, 231–252.

Bounaim, A., 1999: Méthodes de décomposition de domaine: application à la résolution de problèmes de contrôle optimal. *PhD thesis*, Université Grenoble 1.

Brezis, H., 1983: Analyse fonctionnelle. *Masson.*

Bruneau, C.-H., 2000: Boundary conditions on artificial frontiers for incompressible and compressible Navier-Stokes equations. *Math. Mod. and Num. Anal.*, **34**, 303–314.

Bruneau, C.H., and E. Creusé, 2001: Towards a transparent boundary condition for compressible Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, **36**, 807–840.

Cailleau, S., 2004: Validation de méthodes de contrainte aux frontières d'un modèle océanique: application à un modèle hauturier de l'Atlantique Nord et à un modèle régional du Golfe de Gascogne. *PhD thesis*, Université Grenoble 1.

Camerlengo, A.L., and J.J. O'Brien, 1980: Open boundary conditions in rotating fluids. *J. Comp. Phys.*, **35**, 12–35.

Darblade, G., R. Baraille, A.-Y. Le Roux, X. Carton, and D. Pinchon, 1997: Conditions limites non réfléchissantes pour un modèle de Saint-Venant bidimensionnel barotrope linéarisé. *C.R. Acad. Sci. Paris, Série 1*, **324**, 485–490.

Davies, H.C., 1976: A lateral boundary formulation for multi-level prediction models. *Quart. J.R. Meteorol. Soc.*, **102**, 405–418.

Debreu, L., C. Vouland and E. Blayo, 2004a: AGRIF: Adaptive Grid Refinement in Fortran. To appear in *Computers and Geosciences.*

Debreu, L., Y. De Visme and E. Blayo, 2004b: 4D Variational data assimilation for locally nested models. In preparation.

Engquist, B., and A. Majda, 1977: Absorbing boundary conditions for the numerical simulation of waves. *Math. Comp.*, **31**, 629–651.

Engquist, B., and L. Halpern, 1988: Far field boundary conditions for computation overlong time. *Appl. Num. Math.*, **4**, 21–45.

Flather, R.A., 1976: A tidal model of the north-west European continental shelf. *Mem. Soc. R. Sci. Liège*, **6**(10), 141–164.

Givoli, D., 1991: Non-reflecting boundary conditions. *J. Comp. Phys.*, **94**, 1–29.

Hedström, G.W., 1979: Nonreflecting boundary conditions for nonlinear hyperbolic system. *J. Comp. Phys.*, **30**, 222–237.

Holstad, A., and I. Lie, 1999: On transparent boundary conditions and nesting for ocean models. *Research report 91*, Norwegian Meteorological Institute, Oslo, Norway.

Hu, F.Q., 1996: On absorbing boundary conditions for linearized Euler equations by a perfectly matched layer. *J. Comp. Phys.*, **129**, 201–219.

Hu, F.Q., 2001: A stable perfectly matched layer for linearized Euler equations in unsplit physical variables. *J. Comp. Phys.*, **173**, 455–480.

Jensen, T., 1998: Open boundary conditions in stratified ocean models. *J. Mar. Sys.*, **16**, 297–322.

Lie, I., 2001: Well-posed transparent boundary conditions for theshallow water equations. *App. Num. Math.*, **38**, 445-474.

Marchesiello, P., J. McWilliams, and A. Shchepetkin, 2001: Open boundary conditions for long-term integration of regional oceanic models. *Ocean Modelling*, **3**, 1–20.

Martin, V., 2003: Méthodes de décomposition de domaine de type relaxation d'ondes pour des équations de l'océanographie. *PhD thesis*, Université Paris 13.

Martinsen, E.A., and H.E. Engedahl, 1987: Implementation and testing of a lateral boundary scheme as an open boundary condition in a barotropic ocean model. *Coastal Eng.*, **11**, 603–627.

McDonald, A., 2002: A step toward transparent boundary conditions for meteorological models. *Mon. Weath. Rev.*, **130**, 140–151.

McDonald, A., 2003: Transparent boundary conditions for the shallow water equations: testing in a nested environment. *Mon. Weath. Rev.*, **131**, 698–705.

Miller, M.J., and A.J. Thorpe, 1981: Radiation conditions for the lateral boundaries of limited-area numerical models. *Quart. J. R. Meteorol. Soc.*, **107**, 615–628.

Navon, I.M., B. Neta, and M.Y. Hussaini, 2004: A perfectly matched layer approach to the linearized shallow water equations models: the split equation approach. *Mon. Weather Rev.*, **132**, 1369–1378.

Nycander, J., and K. Döös, 2003: Open boundary conditions for barotropic waves. *J.Geophys. Res.*, **108**(C5), 3168–3187.

Orlanski, I., 1976: A simple boundary condition for unbounded hyperbolic flows. *J.Comp. Phys.*, **21**, 251–269.

Palma, E.D., and R.P. Matano, 1998: On the implementation of passive open boundary conditions for a general circulation model: the barotropic mode. *J. Geophys. Res.*, **103**(C1), 1319–1341.

Poinsot, T., and S.K. Lele, 1992: Boundary conditions for subsonic Navier-Stokes calculations. *J. Comp. Phys.*, **101**, 104-129.

Raymond, W.H., and H.L. Kuo, 1984: A radiation boundary condition for multi-dimensional flows. *Quart. J.R. Met. Soc.*, **110**, 535–551.

Röed, L.P., and C. Cooper, 1987: A study of various open boundary conditions for wind-forced barotropic numerical ocean models, in *Three-dimensional models of marine andestuarine dynamics*, edited by J.C.J. Nihoul and B.N. Jamart, pp. 305–335, Elsevier.

Taillandier, V., V. Echevin, L. Mortier and J.-L. Devenon, 2004: Controlling boundary conditions with a four-dimensional variational data assimilation method in a non-stratified open coastal model. *Ocean Dyn.*, **54**, 284–298.

Tréguier, A.-M., B. Barnier, A.P. de Miranda, J.-M. Molines, N. Grima, M. Imbard, G.Madec, and C. Messager, 2001: An eddy permitting model of the Atlantic circulation: evaluating openboundary conditions. *J. Geophys. Res*, **106**(C10), 22115 –22130.

Tsynkhov, S.V., 1998: Numerical solutions of problems on unbounded domains. A review. *Appl. Numer. Math.*, **27**, 456–532.

Chapter 6

SATELLITE MEASUREMENTS FOR OPERATIONAL OCEAN MODELS

Ian Robinson

Southampton Oceanography Centre, University of Southampton, U.K.

Abstract: This chapter outlines the character of ocean measurements from satellites in relation to their use by operational ocean forecasting models. Following a generic introduction to the field of satellite oceanography, it outlines the basic remote sensing methodology for measuring some key variables used by models; sea surface height from altimetry, ocean colour, sea surface temperature and ocean waves. It then presents the approach adopted by an international programme for combining sea surface temperature data from many sources, as an example of the issues involved in effectively preparing satellite data for ingestion into models. It concludes with comments on the actions needed to achieve the integration of satellite and *in situ* data with ocean models in an operational system.

Keywords: Satellite oceanography, altimetry, sea surface height, ocean colour, sea-surface temperature, ocean waves, GODAE, ocean models, GHRSST.

1. Introduction

The purpose of this chapter is to outline the basic characteristics of ocean measurements obtained from Earth-orbiting satellites. It introduces the reader to the subject within the particular context of considering how satellite ocean data can be used operationally to support model based ocean observing and forecasting systems.

After 25 years in which a number of methodologies have been developed to measure different aspects of the surface ocean, there are now many satellite-derived ocean data products available. A new generation of oceanographers takes it almost for granted that they can find global datasets of sea surface temperature and ocean colour, or detailed images of a particular ocean area, readily accessible through the Internet. The impact of the global revolution in telecommunication, capable of transporting

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 147-189.

megabytes of data around the globe in seconds, has expanded the vision of ocean scientists so that we are now contemplating the creation of ocean forecasting systems for operational applications. We envisage systems in which observational data from sensors on satellites and *in situ* platforms are fed in near-real time into numerical models which describe the state of the ocean. Just as meteorologists look to numerical models, supplied by the global meteorological observations network, to give them the most complete and reliable view of what is happening in the atmosphere, so we expect that in future the output of ocean forecasting models will greatly improve the daily knowledge of the state of the ocean needed by operational users to manage the marine environment and to save life at sea.

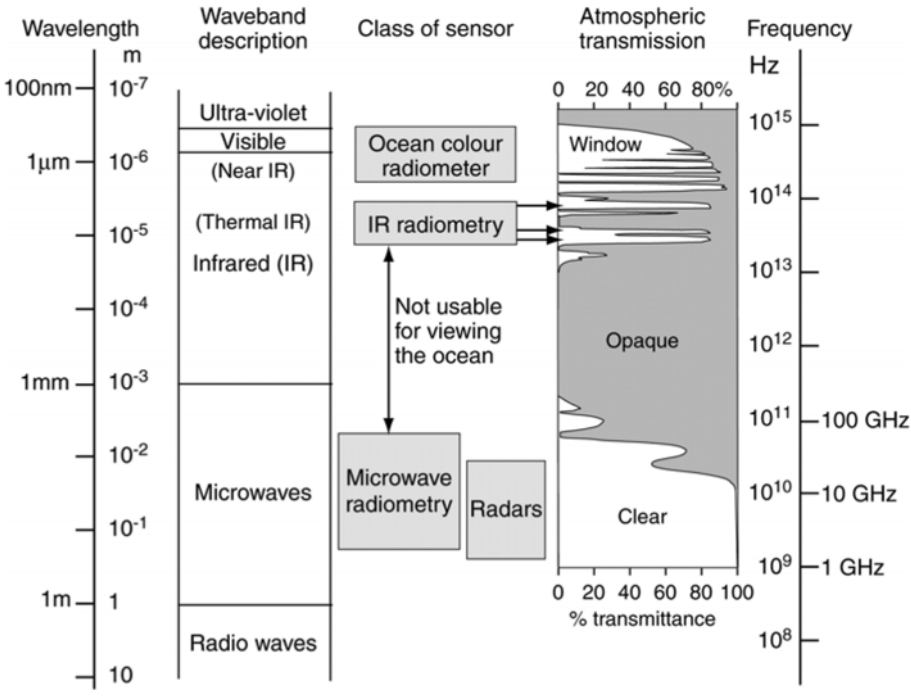

Figure 1. The electromagnetic spectrum, showing atmospheric transmission and the parts used by different remote sensing methods.

Computer models depend on observational data to ensure that they represent the true state of the ocean as closely as possible. It is therefore essential that the observational data fed into ocean forecasting systems are themselves as accurate as possible. It is also important that the limitations and inaccuracies inherent in remote sensing methods are understood and properly accounted for when such data are assimilated into models, or used to initialise, force or validate the models. If use is made of datasets broadcast on the Internet, the user should find out what processes have been

performed on them, and whether they are suitable for the purpose. For example some data products may contain "cosmetic" filling of values in locations where there would otherwise be gaps due to cloud cover or other obstructions to the remote sensing process. Ideally only true observations, and not the artefacts of data processing, should be presented to the numerical model.

This chapter is therefore written for those engaged in developing operational oceanography systems, to give them a basic background in the methods of ocean remote sensing so that they can appreciate what issues to consider as they evaluate the quality of satellite data. It is split into three main sections. The first is a generic overview of the subject. The second introduces the basic remote sensing methodology for some of the key variables used by models. These are sea surface height from altimetry, ocean colour, sea surface temperature (SST) and ocean waves. The third main section uses the example of SST to explore how measurements retrieved from several different sensor systems and supplied by different agencies can be most effectively combined to serve the needs of ocean forecasting models. It presents the methods adopted by an international programme established by the Global Ocean Data Assimilation Experiment (GODAE) for this purpose.

2. Methods of satellite oceanography: An outline

2.1 Using the electromagnetic spectrum

All satellite remote sensing sensors use electromagnetic (e.m.) radiation to view the sea. The ability of particular sensors to measure certain properties of the ocean and how well they can view through the atmosphere or penetrate clouds depends critically on which part of the e.m. spectrum they use. Figure 1 shows the section of the electromagnetic spectrum that is of relevance to remote sensing, and the four broad classes of sensors that are used. The diagram also shows how the transmittance of the atmosphere varies with e.m. wavelength, which accounts for why sensors are found only in certain wavebands. A much fuller account is given by Robinson (2004).

For much of the e.m. spectrum the atmosphere is opaque and therefore unusable for remote sensing of the ocean. However in a number of "window" regions of the spectrum most of the radiation is transmitted although it may be attenuated to some extent. These windows provide the opportunities for ocean remote sensing.

One of the windows extends from the visible part of the spectrum (between 400 nm and 700 nm, used by the human eye) into the near infrared (NIR). This is used by "ocean colour" radiometers that observe sunlight reflected from the ocean, both from the surface and from within the upper

few metres of the water column, with the potential to carry information about those contents of sea water such as chlorophyll, dissolved organic material and suspended particulates that affect the colour of sea water. Solar radiation in the near infra-red (wavelengths above 700 nm) is rapidly absorbed in water and so is not reflected out. Consequently any NIR radiation detected by an ocean-viewing radiometer is evidence of atmospheric scattering or surface reflection, and can be used to correct the visible part of the spectrum for those effects.

There are several narrow windows at wavelengths between about 3.5 μm and 13 μm that are exploited by infrared (IR) radiometers. This is the thermal IR part of the spectrum in which most of the detected radiation has been emitted by surfaces according to their temperature. In ocean remote sensing it is used for measuring sea surface temperature (SST). Like ocean colour, the presence of cloud interrupts the use of this waveband.

At much longer wavelengths, greater than a few millimetres, the atmosphere becomes almost completely transparent. This is referred to as the microwave spectral region. Between those parts of the microwave frequency spectrum allocated by international regulation to radio and TV broadcasts, telecommunications, mobile telephony and so on, a few narrow bands are reserved for remote sensing, within the broad regions indicated in Figure 1. Different bands are used for microwave radiometry and radars. Microwave radiometers are passive sensors, simply measuring the naturally ambient radiation that is emitted by the ocean, atmosphere and land surfaces. Radars are active microwave devices which emit pulses and measure the echoes from the sea surface, in order to gain information about some aspect of the surface.

There is a variety of different types of radar, which can be distinguished by the direction in which they point, the length and modulation of the emitted microwave pulse, and the way the echo from the sea surface is analysed. Radars can be classed as either viewing straight down at the nadir point below the platform, or viewing obliquely to encounter the surface at an incidence angle between 15º and 60º. The nadir sensors measure the surface height or slope and are called altimeters. Those viewing obliquely measure the surface roughness at length scales comparable to the radar wavelength. This is represented by a property of the radar interaction with the material and the geometry of the surface called σ_0, the normalised radar backscatter cross-section.

2.2 Generic processing tasks for analysing ocean remote sensing data

Figure 2 illustrates schematically what is involved in measuring properties of the ocean using a sensor that is typically hundreds or thousands of kilometres from the sea surface. An electromagnetic signal of a particular

kind leaves the sea carrying information about one of the primary observable quantities which are the *colour*, the *temperature*, the *roughness* and the *height* of the sea. This must pass through the atmosphere where it may be changed, and where noise may be added to it, before it is received by the sensor which detects particular properties of the radiation and converts each measurement into a digital signal to be coded and sent to the ground. The sensor geometry restricts each individual observation to a particular instantaneous field of view (IFOV). In order to convert the numbers received at the ground station into scientific measurements of useful precision and quantifiable accuracy, the remote sensing process represented in the left side of Figure 2 must be inverted digitally using the knowledge and information identified on the right side.

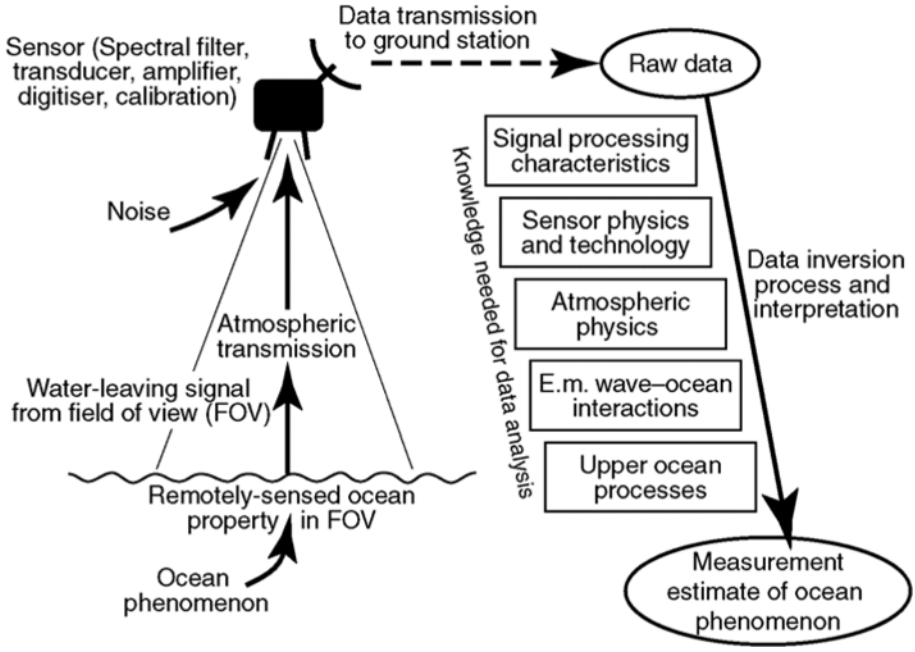

Figure 2. Schematic of information flow in ocean remote sensing.

Although there are just four observable quantities[1] and these are measured only at the very surface of the sea, apart from colour, it is surprising how much information about other properties or ocean processes can be retrieved from these four variables. Many phenomena in the upper ocean have sufficient influence on one or more of the primary measurable quantities to generate a "surface signature" in remotely sensed data and images. Some of these are obvious and predictable, such as the influence of

[1] The capacity to measure a fifth quantity from space, salinity, waits to be demonstrated by the European Space Agency's SMOS sensor.

a near-surface phytoplankton bloom on the colour of the sea. Other signatures took many scientists by surprise when they were first discovered in the satellite images. For example internal waves, a dynamical phenomenon centred tens of metres below the sea surface, can sometimes be revealed in exquisite spatial detail in the images of synthetic aperture radar (SAR), because of their surface roughness signature.

In order to extract quantitative information about an ocean phenomenon from satellite data, we need to understand the physical processes in the upper ocean that control its surface signature in one of the primary detectable variables. Several of the derived properties, such as chlorophyll concentration retrieved from colour sensors, surface wind speed from scatterometers, wave height from altimetry, wave spectra from SARs and salinity from microwave radiometry are now being used, or proposed, for ingestion into ocean models. Figure 3 summarises the different classes and types of sensors, the primary variables which they detect, and the way in which they can supply inputs to ocean models.

For many of the applications to ocean models noted above, it is possible to use ocean data products already produced by the agencies responsible for the sensors, without the user having to engage themselves in any of the processing tasks. Nonetheless, it is important for users to be aware of the calibrations, corrections, analyses and resampling that are applied to data products before they are distributed, since these processes have impacts on the quality, accuracy and timeliness of the data. Figure 4 summarises them, and also indicates what is meant by the different "levels" of data products that may be available. Robinson (2004) provides a detailed explanation of what is involved in each of these processes.

2.3 The sampling constraints imposed by satellite orbits

The use of Earth orbiting satellites as platforms for ocean-viewing sensors offers a number of unique advantages such as the opportunity to achieve wide synoptic coverage at fine spatial detail, and repeated regular sampling to produce time series several years long. However, these benefits are won at the cost of being tied to the unavoidable constraints imposed by the physical laws of satellite orbital dynamics.

There are just two basic types of orbit useful for ocean remote sensing, geostationary and near-polar. The geostationary orbit, at a height of about 36000 km, has a period of one sidereal day (~23.93 hr). Placed over the Equator, the satellite flies West to East at the same rate as the Earth's rotation, so it always remains fixed relative to the ground, allowing it to sample at any frequency. Being fixed it can view only that part of the world within its horizon, which is a circle of about 7000 km radius centred on the Equator at the longitude of the satellite. Its great height also makes it difficult for sensors to achieve fine spatial resolution.

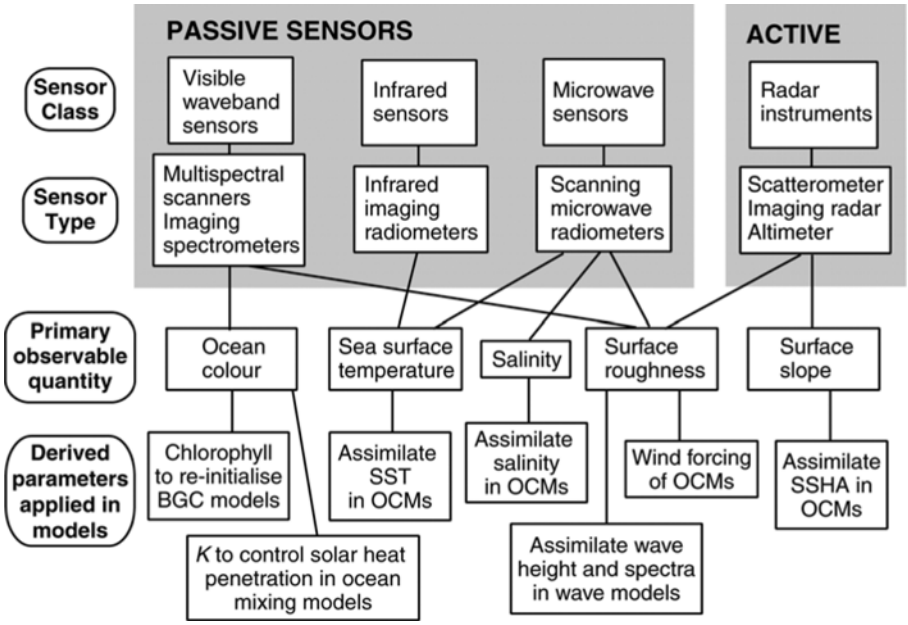

Figure 3. Summary of the different classes and types of ocean sensors carried on satellites, indicating the primary quantity which each sensor type measures and ways in which the derived parameters are used in numerical models including ocean circulation models (OCM) and biogeochemical models (BGC).

In a near-polar orbit the satellite flies at a much lower altitude, typically between about 700 km and 1350 km, for which the orbital period is about 100 min. It thus completes between 14 and 15 orbits a day, during which the Earth rotates once, so the satellite marks out a ground track crossing about 14 times northeast to southwest (descending tracks) and the same number of southeast to northwest ascending tracks. The tracks are distributed evenly around the globe, with successive orbits following a track about 24º of longitude to the east of the previous orbit. A wide-swath sensor that can scan across about 2800 km will thus view every part of the Earth twice a day, once from an ascending and once from a descending orbit. An even wider swath permits more samples per day as swaths from successive orbits overlap at the Equator, while at higher latitudes overlapping occurs for much narrower swaths. However, the global coverage is won at the price of a much reduced sampling frequency compared to the geostationary orbit.

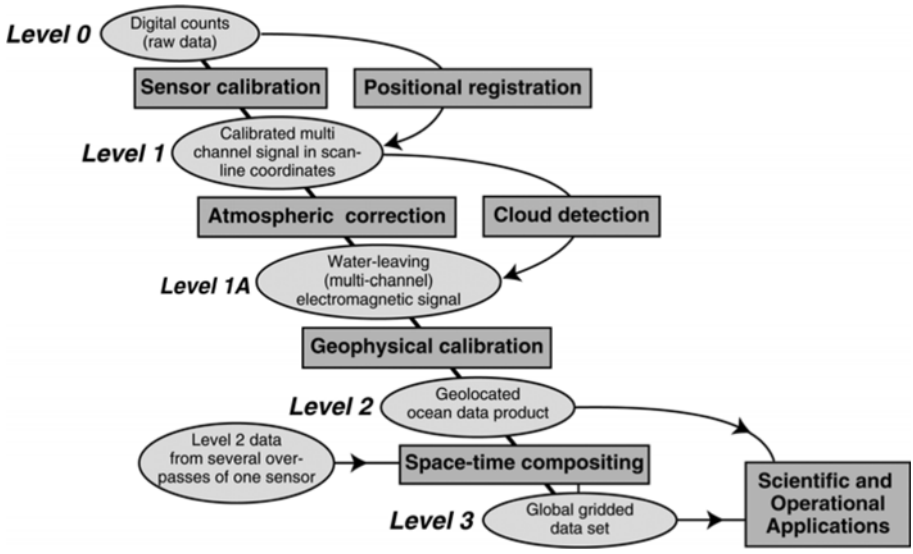

Figure 4. Outline of data processing tasks to convert raw satellite data into ocean products suitable for operational applications, showing the different "levels" of processed data which are produced at each stage.

For much narrower swaths (normally associated with fine resolution imaging sensors) or for non-scanning instruments such as the altimeter that sample only along the ground track, the time between successive views of the same location depends on the precise way in which the orbit repeats itself. If the orbit repeat period is just a few days then the sensor revisit interval will be the same, but in this case a narrow swath sensor will miss many parts of the Earth surface altogether. Global coverage by a sensor whose swath is about 200 km would take about 15 days to accomplish. A non-scanning sensor builds up a sampling pattern that progressively fills the gaps left by previous orbits until one orbit repeat cycle is completed when the tracks repeat. For scanning and non-scanning sensor alike, there is evidently a well defined trade-off between spatial and temporal sampling capability, which is discussed in more detail by Robinson (2004). It is important to appreciate these fundamental constraints when designing an ocean observing system for operational purposes. For example, the only way to ensure that even a wide swath sensor can sample every six hours is to fly sensors on two satellites. Ideally a combination of spatial and temporal resolution should be selected in order that important phenomena can be adequately sampled. If mesoscale eddies are to be monitored then the spacing between orbit tracks should not be wider than their variability length scale, nor should the repeat cycle be longer than the characteristic lifetime of an eddy. Otherwise some eddies may be missed altogether.

Most satellites in a low, near-polar orbit are sun-synchronous. By choosing an inclination that is slightly greater than 90º (i.e. their path does

not quite reach the poles) the orbit plane is constrained to precess at a rate of once per year relative to the stars. This locks the overpasses to the position of the sun and means that every orbit always crosses the Equator at the same local solar time. For most ocean observing sensors this is very convenient, since it ensures that the longitudinal position of the sun does not change from one sample to the next, even though the solar latitude inevitably changes with the annual cycle. However, for altimetry the sun-synchronous orbit is to be avoided since it aliases the solar semidiurnal tidal constituent.

2.4 Strengths and weaknesses of ocean remote sensing

The global, spatially detailed and regularly repeated views of the oceans that have been obtained from satellites for more than a decade have made them an important part of the design of operational ocean monitoring systems. It is therefore worth summarising the benefits that satellite ocean data bring as well as noting their limitations.

The importance of satellite data to oceanography can be highlighted by the way they have opened up the study of global ocean phenomena. We can now ask questions about large scale processes which could not properly be addressed scientifically until remote sensing methods allowed us to make observations of ocean scale phenomena which test and stretch the theoretical models. A good example of this is the study of oceanic Rossby waves (Challenor *et al*., 2004). To some extent 21st Century Oceanography has become dependent on satellite observations. All branches of ocean science now expect to use satellite image data and interest in the subject is no longer limited only to specialist "satellite oceanographers". Another powerful impact has come from the immediacy of satellite data. Observations from all around the world are now being made available within hours, minutes in some cases, of their acquisition by the sensor, and this has reinforced their importance for use in operational ocean monitoring and forecasting

At the same time we must not overlook the fundamental limitations of satellite ocean remote sensing methods. They can observe only some of the ocean's properties and variables. They measure the ocean only at or near the surface although it can be argued that, of all the parts of the ocean, the surface is the most critical place to be able to measure. Most critically, ocean measurements may be corrupted by the atmosphere and some methods cannot see through clouds at all. Moreover, measurements cannot be made "to order" but only when the satellite is in the right place. Finally it must not be overlooked that all measurements from satellites require calibration and / or validation using *in situ* data. While it might be carelessly thought that satellites can remove the need for measurements at sea, the reverse is in fact the case. The full benefit of the wider and higher perspective achieved from satellite data will only be realised when combined with an integrated array of *in situ* sensors interfacing with operational ocean models.

3. Some satellite measurements used for operational models

3.1 Sea surface height anomaly from altimeters

3.1.1 The principles of altimetry over the ocean

A satellite altimeter is a nadir-viewing radar which emits regular pulses and records the travel time, the magnitude and the shape of each return signal after reflection from the earth's surface. The travel time is the essential altimetric measurement, leading to a determination of the ocean surface topography at length scales longer than about 100 km. Ocean surface topography contains information about ocean dynamical and geophysical phenomena. If the travel time can be measured to a precision of 6×10^{-11} s then, knowing the speed of light, the distance can be calculated to a resolution of 1 cm. Corrections have to be made to allow for the changed speed of light through the ionosphere and the atmosphere, and for delays associated with reflection from a rough sea surface (Chelton *et al.*, 2001). It is generally agreed that for these corrections to approach the target accuracy of 1 cm a dual frequency altimeter must be used (to determine the ionospheric refraction), and a three channel microwave radiometer is needed to sound the water vapour in the atmosphere.

The altimeter is not an imaging sensor. Viewing only the nadir point below the satellite, it simply records measurements of distance between the satellite and the sea surface along the ground track. As discussed in 2.2, the spatial and temporal sampling characteristics therefore depend entirely on the exact orbit repeat cycle of the satellite. This was chosen to be about 10 days for the TOPEX/Poseidon (T/P) and Jason altimeters which fly on platforms dedicated to the altimetric mission, although for other altimeters it has ranged between 3 days, 17 days and 35 days. The longer the revisit interval the finer the spatial sampling grid.. Typically, ocean topography data are interpolated onto a geographical grid and composited over the period of an exact repeat cycle, to produce "images" which are comparable with global SST or ocean chlorophyll composite images although produced in a completely different way.

By itself, knowing the distance, R_{alt} between the ocean surface and a satellite is of limited value. Figure 5 shows what else needs to be defined or measured for this to yield an oceanographically useful property. First of all, when the height of the satellite, H_{sat}, is known relative to a reference level, then the height, h, of the sea above the reference level can be determined. The reference level is a regular ellipsoid-shaped surface defined within a frame of reference fixed in the rotating earth. It is chosen to match approximately the shape of the earth at sea level, and provides a convenient datum from which to measure all other heights.

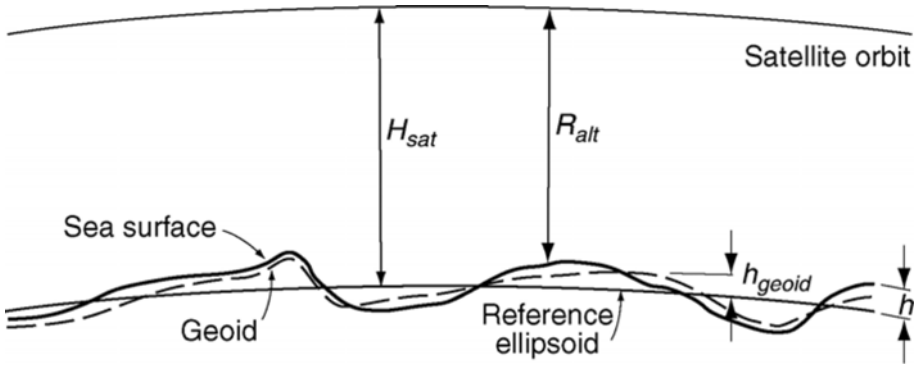

Figure 5. The relationship between different distance quantities used in altimetry.

Several physical factors contribute to *h,* which is called the ocean surface topography. The first is the distribution of gravity over the earth, as represented by the geoid, at height h_{geoid} above the reference ellipsoid in Figure 5. The geoid is the equipotential surface, at mean sea level, of the effective gravitational field of the earth which incorporates earth-rotation forces and the gravitation of the solid earth, the ocean itself and the atmosphere. By definition it is normal to the local effective gravity force, and if the ocean were everywhere in stationary equilibrium relative to the earth, its surface would define the geoid.

Another factor which contributes to h is h_{tide}, the instantaneous tidal displacement of the sea surface relative to its tidally averaged mean position, including the contribution of the Earth tide. A third is the local response, h_{atm}, of the ocean to the atmospheric pressure distribution over the ocean, approximated by the inverse barometer effect in which an increased pressure of 1 mbar lowers sea level by 1 cm. The remaining factor is the displacement of the sea surface associated with the motion of the sea, called the ocean dynamic topography h_{dyn}. Thus:

$$h = h_{dyn} + h_{geoid} + h_{tide} + h_{atm} \quad (1)$$

The dynamic topography is the property which is of most relevance for ocean modelling since it contains information about the ocean circulation. Rearranging (1) and substituting h = $H_{sat} - R_{alt}$ yields:

$$h_{dyn} = H_{sat} - R_{alt} - h_{geoid} - h_{tide} - h_{atm} \quad (2)$$

The accuracy and precision of the estimated ocean dynamic height depends not only on the altimetric measurement itself but also on the other four terms in (2). For dedicated altimetry missions flying at a height of

about 1340 km where atmospheric drag is minimal, the height of the satellite in orbit, H_{sat}, can now be predicted to a precision of 2 cm (Tapley *et al*., 1994) using a combination of laser and microwave tracking devices and an orbit model using precise gravity fields. The tidal contribution has been evaluated along the repeat orbit track by tidal analysis of the altimeter record spanning several years (Le Provost, 2001). Because the tidal frequencies are very precisely known the response to each constituent can be evaluated to an accuracy better than 2 cm in the open ocean, even though the sampling interval of about 10 days is longer than most of the tidal periods. This is only possible when the precise period of the repeat cycle is chosen to avoid any serious aliasing with one of the major tidal constituents. For this reason a sun-synchronous orbit, which aliases the S_2 (solar semidiurnal) tidal signal, should not be used. Over shelf seas where tides are very high and can vary rapidly over short distances it is not so easy to remove the tides and so the estimate of dynamic height is less accurate. The atmospheric pressure correction is based on the output of atmospheric circulation models.

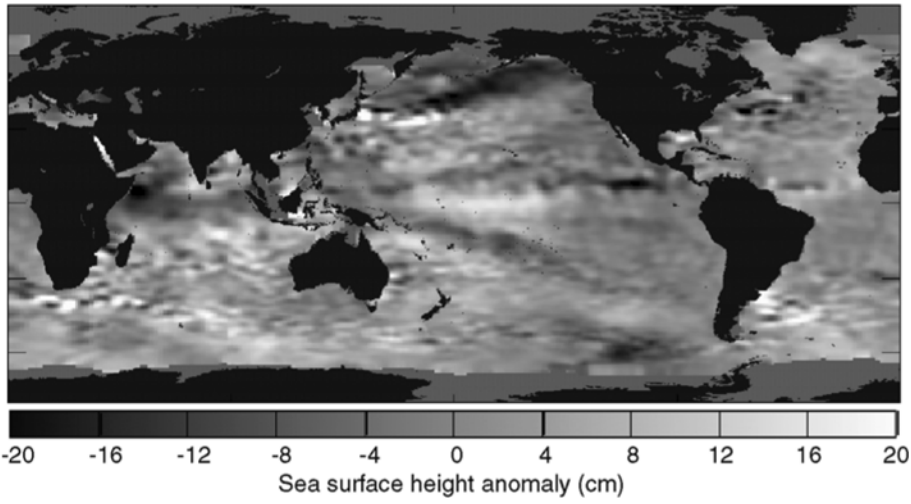

Figure 6. The spatially averaged SSHA field from TOPEX/Poseidon for 31 Dec, 2001. (Image generated with data obtained from JPL at podaac.jpl.nasa.gov/poet).

3.1.2 Evaluating sea surface height anomaly

At present, the geoid is not known independently and so oceanographers must be content with measuring the combined $h_{dyn} + h_{geoid}$. Of these, the typical magnitude of the spatial variability of h_{geoid} is measured in tens of metres, about ten times greater than that of h_{dyn}, which is why until recently the time-mean ocean topography from altimeters provided geophysicists with the best measure of the geoid. However, h_{geoid} does not vary with time, at least not sufficiently to be detected by an altimeter over tens of years,

whereas the time-variable part of h_{dyn} is comparable in magnitude with the mean component, of order metres over a few months. Therefore the time variable part of h_{dyn}, called the sea-surface height anomaly, SSHA, can be separated from the measured $h_{dyn} + h_{geoid}$ by simply subtracting the time mean over many orbit cycles. To enable a time-mean to be produced, the orbit track must be precisely repeated to within a kilometre and the data must be accumulated from several years of a ten-day cycle. For this reason it is essential to fly a new altimeter in precisely the same orbit as its predecessor so that the mean surface topography of the earlier mission can be used straight away. Then SSHA can be calculated from the first orbit cycle of the new altimeter, without having to wait another few years to build up a new mean topography for a different orbit track.

It is important to remember that the SSHA, which is widely used and assimilated into ocean models, does not contain any information about the dynamic height of the ocean associated with the mean circulation. Thus in Figure 6, which is an example of the SSHA from T/P observed during a ten-day period in December 2001, the dominant features are mesoscale eddies. The dynamic topography signatures of the strong ocean currents are not seen at all, apart from the fact that the eddy-like activity is strongest where the major currents tend to meander.

Altimeter	Agency	Dates	Height	Orbit	Accuracy
TOPEX/ Poseidon	NASA/ CNES	1992-present	1336 km	9.92 day repeat non-sun-synchronous	2-3 cm
Poseidon-2 on Jason-1	NASA/ CNES	2001-present	1336 km	9.92 day repeat non-sun-synchronous	~2 cm
Radar altimeter (RA) on ERS-1	ESA	1991-2000	780 km	3 & 35 day repeat sun-synchronous	~5-6 cm
RA on ERS-2	ESA	1995-2003	780 km	35 day repeat sun-synchronous	~5-6 cm
RA2 on Envisat	ESA	2002-present	800 km	35 day repeat sun-synchronous	3 cm
Geosat	U.S. Navy	1986-89	800 km	17.05 day repeat sun-synchronous	10 cm reanalysis
Gesosat Follow-on	U.S. Navy	2000-present	880 km	17.05 day repeat sun-synchronous	~10 cm

Table 1. Recent and current series of satellite altimeters.

There are presently three families of altimeters in operation, as listed in Table 1 with the details of their altitude, orbit repeat and the approximate accuracy (root mean square) of an averaged SSHA product. The T/P–Jason family is a joint French/U.S.A. dedicated altimetry mission in a high non-sun-synchronous orbit. In contrast the Geosat and ERS series are on lower sun-synchronous platforms for which the orbit prediction accuracy would be, on their own, much poorer. However, because these satellites cross over

each other's orbits tracks it is possible, over an extended time span, to significantly improve their orbit definitions by cross-referencing to the better known T/P or Jason orbits. The accuracy quoted for the SSHA applies only after this procedure has been performed, and would otherwise be much worse for the ERS and Geosat families. The specification of errors for an altimeter must be handled with care because the error magnitude relates very much to the time and space scale over which it is being averaged. The lower error attached to larger-scale / longer-period averaging must be offset against the lesser utility of the averaged SSHA field, especially in the context of operational oceanography.

3.1.3 Variable currents from sea surface height anomaly

To determine an estimate of the time-variable part of ocean surface currents the geostrophic equations are used:

$$fv = g\frac{\partial h_{SSHA}}{\partial x}$$
$$fu = -g\frac{\partial h_{SSHA}}{\partial y} \qquad (3)$$

where (u,v) are the East and North components of the geostrophic velocity, f is the Coriolis parameter, g is the acceleration due to gravity and x and y are distances in the East and North direction respectively.

From a single overpass, only the component of current in a direction across the altimeter track can be determined, but where ascending and descending tracks cross each other the full vector velocity can be estimated. Because Eq. (3) assumes geostrophic balance, if there is any ageostrophic surface displacement it will lead to errors in (u,v). However, ageostrophic currents should not persist for longer than half a pendulum day $(1/f)$ before adjusting to geostrophy. Thus the spatially and temporally averaged SSHA maps produced from all the tracks acquired during a single repeat cycle (10, 17 or 35 days depending on the altimeter) should represent a good approximation to a geostrophic surface that can be inverted to produce the surface geostrophic currents.

This raises an interesting question when SSHA is to be assimilated into an ocean circulation model. Most models which use altimeter data presently assimilate the global or regional field from a whole orbit cycle, and perhaps combine the data from the different altimeter families shown in Table 1. But this means some of the data may be many days old by the time they can be ingested into the model and therefore of less utility in improving the accuracy of operational nowcasts or forecasts. Alternatively models can in principle assimilate the SSHA along track records within a few hours of

acquisition. While the data errors will be greater, the input will be timely. If the model variables are properly matched to the altimeter data it may even be possible to use the ageostrophic information contained in the non-averaged along track record.

Close to the Equator the SSHA cannot be interpreted directly in terms of surface currents since here f is very small and the geostrophic Eq. (3) cannot be applied.

In the relatively near future it is hoped that the lack of knowledge about the Geoid can be remedied. What is needed is a means of measuring h_{geoid} without using altimetry, and this is provided by the measurement of the gravity field above the Earth from satellites. Both the presently operating Gravity Recovery And Climate Experiment (GRACE) and the Gravity and Ocean Circulation Explorer (GOCE) mission which is due for launch by 2007 measure elements of the gravity field from which it is possible to recreate the sea-level Geoid. At the required accuracy of about 1 cm the GRACE can achieve this only at a length scale longer than at least 1000 km, but it is expected that the GOCE can do so once-for-all down to a length scale of about 100 km. This will allow the steady state ocean currents to be derived from archived altimetric data and greatly improve the capacity to utilise altimetric data in near-real time.

3.1.4 The impact of altimeters on oceanography

Despite the present limitations of altimetry to measuring just the time-variable part of surface currents in the open ocean only, this achievement by itself has made a tremendous difference to Oceanography in the 21st Century. It has opened up the whole approach to operational ocean forecasting based on numerical modelling, since without the capacity of altimetry to monitor the mesoscale turbulence of the ocean at time scales of days to weeks there would be little hope of ensuring that ocean models remain consistent with the real ocean. The ability to measure changes in the absolute height of the sea to an accuracy of 2-3 cm (a measurement uncertainty of just 2 parts in 10^8) is technologically quite breathtaking, yet instrument engineers and data analysts do not believe the limits to further improvement have been reached. The measurement of the geoid independently of altimetry will liberate more information from the altimeter record within a few years from now. Meanwhile, there remains the challenge to develop methods of assimilating SSHA into ocean models which fully recognizes the character of the altimeter data and maximises the utilisation of the available information.

3.2 Ocean colour

The light measured by an ocean colour sensor pointing towards the sea comes originally from the sun. Photons of light on their path from the sun to the sensor have encounters with the medium, such as reflection at the sea surface and scattering in the atmosphere or ocean, while some photons are absorbed and never reach the sensor. To the extent that the outcome of these encounters are spectrally sensitive, the resulting colour (spectral distribution) of the light reaching the sensor contains information about some aspects of the sea and the atmosphere. Figure 7 summarises the factors which affect the colour. Direct solar reflection, sun glitter, tends to dominate all other signals when it is present and so it is avoided as far as possible by the choice of orbit geometry and overpass time, and in some cases by deliberately tilting the sensor away from the specular reflection of the sun over that part of the orbit where it would otherwise be a problem. The rougher the surface the wider becomes the sea area affected by sun glitter but the magnitude of the reflected radiance in any particular direction is reduced.

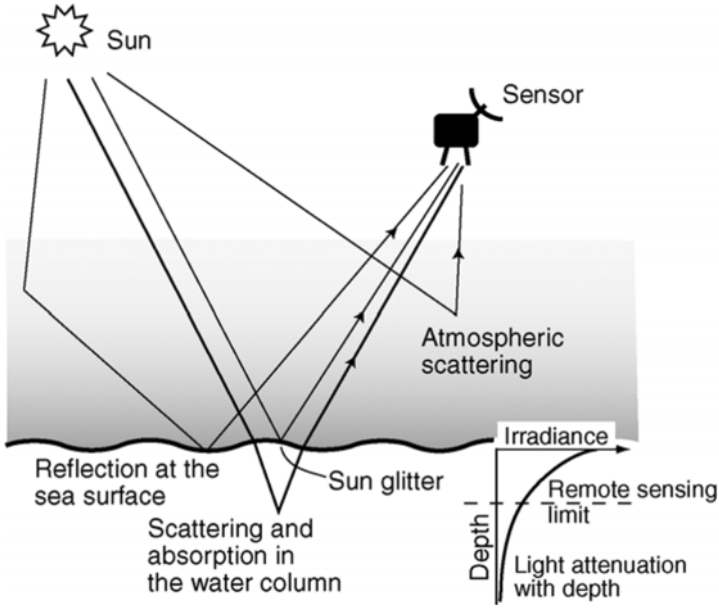

Figure 7. Interaction of sunlight with the atmosphere and ocean

3.2.1 Atmospheric correction

For the satellite oceanographer, the "signal" consists of the light reflected from below the sea surface since its colour can be interpreted in relation to the water content. All the other interactions cause additions or alterations to

the signal, producing noise which requires correction. The greatest contribution comes from light scattered by the atmosphere into the field of view of the sensor, which may make up over 90% of the measured radiance, including skylight reflected by the sea surface into the sensor. The atmospheric correction procedure must account for this in order to estimate the water leaving radiance for each of the spectral bands recorded by the sensor.

Scattering by the air gas molecules themselves can be directly calculated for each pixel in the field of view of an imaging sensor, but scattering by larger particles of aerosols such as water vapour or dust particles cannot be calculated because their distribution in the atmosphere is unknown and impossible to predict. Instead this part of the atmospheric correction uses the radiance measured in two spectral bands from the near infra-red part of the spectrum. Because the sea absorbs almost all incident solar near-infrared radiation, any measured at the top of the atmosphere must have been scattered by the atmosphere. This is then used to estimate how much aerosol scattering has occurred in the visible channels where the water leaving radiance is not zero, and so the correction is accomplished.

3.2.2 Estimating water content from its colour

When the atmospheric correction has been successfully applied to satellite ocean colour data, the result is an estimate of the water-leaving radiance in each spectral channel in the visible waveband, normalised to reduce dependence on the sun's elevation and the viewing incidence angle. Effectively the normalised water leaving radiance should represent what a sensor would measure if looking straight down from an orbit that takes it just above the sea surface at the bottom of the atmosphere. This is what our eyes would detect as the colour and brightness of the sea, ignoring any light reflected from the surface. The primary challenge of ocean colour remote sensing is to derive quantitative estimates of the type and concentration of those materials in the water which affect its apparent colour.

Photons of visible wavelength e.m energy from the sun that enter the sea will eventually interact with molecules of something in the sea. The outcome will be either that the photon is scattered, in which case it may change its direction with a chance of leaving the sea and contributing to what the sensor sees, or it will be absorbed. The probability of scattering or absorption depends on the wavelength of the light and the material which it encounters. The molecules within sea water tend to preferentially scatter shorter wavelengths of light (the blue part of the spectrum) and preferentially absorb longer wavelengths (the red end). This is why pure sea water with little other content appears blue.

The pigment chlorophyll-*a* which is found in phytoplankton has a strong and fairly broad absorption peak centred at 440 nm in the blue, but not in the

green. Therefore, as the chlorophyll concentration increases, more blue light is absorbed while the green light continues to be scattered and so from above the sea water looks greener. This is the basis for many of the quantitative estimates of sea water content derived from satellite ocean colour data. The typical form of an algorithm to estimate the concentration of chlorophyll (C) or phytoplankton biomass is:

$$C = A(R_{550}/R_{490})^B \tag{4}$$

where A and B are empirically derived coefficients and R_λ is the remote-sensing reflectance (radiance coming out of the sea towards the sensor, normalised by ingoing irradiance) over a spectral waveband of the sensor centred at wavelength λ. When using the wavelengths indicated in Eq. (4) this is described as the green / blue ratio. In the open sea it is possible to estimate C to an accuracy of about 30% by this means. Most algorithms presently in use are somewhat more complex than Eq. (4) but still closely related to it. If the sample data from which the coefficients A and B etc. are derived is representative of many different open sea situations then such algorithms can be applied widely in many locations.

Other substances which interact with the light and so change the apparent colour of the sea are suspended particulate material (SPM) that has a fairly neutral effect on colour except in the case of highly coloured suspended sediments, and coloured dissolved organic material (CDOM, sometimes called "yellow substance") which absorbs strongly towards the blue end of the spectrum. Both of these affect the light along with the chlorophyll "greening" effect when there is a phytoplankton population. However because the chlorophyll, CDOM and SPM all co-vary within a phytoplankton population the green-blue ratio effect dominates the colour and each of these materials can be quantified by an algorithm such as Eq. (4), as long as phytoplankton are the only major substance other than the sea water itself which is affecting the colour. Such conditions are described as being Case 1 waters, and it is here that the ocean colour algorithms work fairly well to retrieve estimates of C from satellite data.

However, if there is SPM or CDOM present from a source other than the local phytoplankton population, for example from river run-off or resuspended bottom sediments, then we can no longer expect any simple relationship between the concentrations of these and C. In this situation the green-blue ratio algorithms do not perform very well, if at all, and it becomes much harder to retrieve useful quantities from ocean colour data using universal algorithms. These situations are described as Case 2 conditions. Unfortunately it is not easy to distinguish between Case 1 and Case 2 waters from the satellite data alone. This can result in very degraded accuracy with errors of 100% if the chlorophyll algorithms are applied in Case 2 waters. It is prudent to classify all shallow sea areas as Case 2,

particularly where there is riverine and coastal discharge or strong tidal currents stirring up bottom sediments, unless *in situ* observations confirm that Case 1 conditions apply.

Another useful measurement that can be derived from the ocean colour is the optical diffuse attenuation coefficient, K, usually defined at a particular wavelength such as 490 nm (i.e. K_{490}). This is also inversely correlated with the blue-green ratio because the less the attenuation coefficient, the deeper the light penetrates before it is scattered back out, the more of the longer wavelengths are absorbed, and the bluer the water appears. The algorithms for K are similar in form to (4) and are less sensitive to whether Case 1 or Case 2 conditions are found.

3.2.3 Ocean colour sensors and products

Although visible wavelength radiometers were among the very first Earth observing sensors flown in the 1970s the development of ocean colour sensors is less mature than that of other methods of satellite oceanography. After the Coastal Zone Color Scanner (CZCS) proved the concept of measuring chlorophyll from space in 1978-1986, there was a long pause until the Ocean Colour and Thermal Sensor (OCTS) was launched in 1996, followed by the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) in 1997, which has provided the first reliable, long-term fully operational delivery of ocean colour data products. Since then two Moderate resolution imaging spectrometers (MODIS), the Medium resolution imaging spectrometer (MERIS) and the Global Imager (GLI) have been launched (see Table 2). All fly in low (~800 km) sun-synchronous polar orbits, providing a resolution at nadir of about 1.1 km and almost complete Earth coverage in 2 days. Other colour sensors have also been flown by individual countries offering less comprehensive coverage and poorer data availability than those listed.

Sensor	Agency	Dates	No. of visible channels	No. of near-IR channels
CZCS	NASA	1978-86	4	-
OCTS	NASDA	1996-97	6	2
SeaWiFS	NASA	1997-present	6	2
MODIS/Terra	NASA	2000-2004	7	2
MERIS	ESA	2002-present	8	3
MODIS/Aqua	NASA	2002-present	7	2
GLI	NASDA	2002-2003	12	3

Table 2. Details of the major satellite ocean colour sensors.

All of the sensors listed in Table 2 are supported by a calibration and validation programme and their data are worked up by the responsible agency into derived oceanographic products at level 2 and in some cases

level 3. In all cases some measure of *C* is produced and an estimate of K is derived globally. These are generally reliable products and C approaches the target accuracy of 30% in open sea Case 1 waters. However great care must be taken when using the products over coastal and shelf sea (possibly Case 2) waters where the potentially large errors could give misleading information. Some of the agencies attempt to provide a number of other products such as SPM and CDOM but these are yet to be proven.

3.2.4 Using satellite ocean colour data in ocean models

The use of ocean colour derived data products in ocean models is in its infancy. The long and successful deployment of SeaWiFS has given some confidence that ocean colour sensors are capable of supplying data for operational applications. However the disappointing loss of two excellent sensors (OCTS and GLI) through spacecraft failure, and the difficulties with calibrating the MODIS/Terra products, have slowed down any moves in this direction. It is to be hoped that, when MODIS/Aqua and MERIS are fully proven and delivering data products routinely within a few hours of acquisition, they will establish an even better operational supply of data than SeaWiFS which is approaching the end of its operational life.

There are two main ways in which ocean colour data are likely to be used operationally. The first is to use measurements of *C* to improve the modelling of phytoplankton biomass in numerical ocean models which contain a biogeochemical, phytoplankton or carbon cycle component. The uncertainties associated with modelling biological populations are such that improvements can be gained by assimilating or otherwise ingesting measurements of *C* even when their accuracy is no better than 30%. So far the most promising approach has been to use the satellite observations to identify when and where a phytoplankton bloom emerges, since it is particularly difficult for a model to trigger the initiation of a bloom. Satellite data are used to update or re-initialise a model with the newly emerged bloom conditions. The problem of cloud cover and the uncertain accuracy make the data less useful to the model for the stage when the modelled bloom has peaked and then gradually decays.

The other main use of colour data is in providing *K* for physical models which need to know how far solar radiation penetrates into the sea. Thus models of mixed layer development and the occurrence of a diurnal thermocline, which are sensitive to *K*, can benefit from being regularly updated with information about how *K* is distributed and varies in space and time.

Another factor which still needs more investigation is the relationship between the near-surface measurements of *C* which satellites provide and the distribution of *C* with depth. Because phytoplankton need light to thrive, then it is reasonable to suppose that most will be detected by ocean colour

sensors. However, as indicated in Figure 7 a visible waveband radiometer will "see" only down to the level where the irradiance is about 1/3 of its surface value. The satellite measurements are unlikely to record accurately, if at all, any phytoplankton below this level, such as those contributing to the deep chlorophyll maximum found at the base of a mixed layer after nutrients have been used up from the mixed layer. Similarly if a second, low-light, species develops below the main bloom the satellite will not be able to detect them.

There is undoubtedly a large amount of valuable information for ocean models to be found in the ocean colour data products from satellites. However considerably more research is needed to learn how best to inject that information into the models. A particularly enticing prize is to combine the satellite measurements with ocean carbon cycle models to be able to estimate with some confidence the rates of primary production occurring in the sea. Allied to this is the potential for improving our knowledge of how pCO_2 (representing the amount of CO_2 dissolved in the surface water) is distributed, leading to better estimates of air-sea fluxes of CO_2. Finally we should not overlook the much simpler application of using ocean colour as a tracer of mesoscale eddies. There is a need to develop techniques to assimilate this information so that eddy-resolving ocean circulation models are guided to present eddies in the right place at the right time.

3.3 Sea surface temperature

3.3.1 Diverse methods for measuring sea surface temperature

Sea surface temperature (SST) can be measured in a variety of ways, using sensors on both satellites and *in situ* platforms (Robinson & Donlon, 2003). Sampling from *in situ* platforms can generally be performed at high frequency whereas most satellite methods are severely restricted by orbit constraints to long sampling intervals of several hours or more. On the other hand remote sensors are capable of wide synoptic spatial coverage at fine spatial detail down to 1 km resolution when unobstructed by clouds, while all the *in situ* methods sample very sparsely, and may miss some regions altogether. Table 3 lists the different classes of satellite-based methods and the typical absolute accuracy of measurements which they can achieve. Relative accuracy (that is the smallest temperature difference that can be detected confidently within a given image from a single overpass) may be somewhat better than the absolute value quoted.

Instrument	Spatial coverage and nadir resolution	Time sampling	Accuracy
Polar orbiting IR radiometer (e.g. AVHRR)	Global; 1.1 km,	12 hr; cloud-limited	0.3 - 0.5 K
Polar orbiting dual view IR radiometer (e.g. AATSR)	Global; 1 km	Twice in 2-4 days, cloud-limited	0.1 - 0.3 K
Polar-orbiting microwave radiometer (e.g. AMSR-E)	Global; 25 - 50 km	12 hr - 2 days	0.3 - 0.5 K
Geostationary orbit IR sensor (e.g. SEVIRI on Meteosat S.G.)	50ºS – 50ºN; 2-5 km	30 min, cloud-limited	0.3 - 0.5 K

Table 3. Classes and characteristics of satellite temperature sensors.

Space methods for measuring SST are differentiated both by the part of the electromagnetic spectrum used and by the orbit of the platform from which the Earth is viewed. Sensors placed on geostationary satellites such as Meteosat and GOES are capable of regular and frequent (15-30 min) sampling throughout every 24 hr period but are limited in spatial coverage by the horizon at 36,000 km altitude. Because they are so far above the Earth a very fine angular resolution is required to achieve useful spatial resolution at the sea surface. This presently rules out the use of microwave radiometers and so all SST sensors in geostationary orbit use the infrared. This makes them vulnerable to cloud cover, but they are able to take advantage of any clear skies which may develop at any time of the day or night.

The sensor type used most for global SST monitoring is the infrared scanner on polar orbiting satellites. The NOAA Advanced Very High Resolution Radiometer (AVHRR) series has been routinely flown since 1978, with normally two satellites operational at any time in a sun synchronous orbit providing morning and afternoon overpasses plus two night-time overpasses (Kidwell, 1991). Since 1991 a series of along-track scanning radiometers (the ATSR class) has been flown on ESA polar platforms. Using the same infra-red wavebands as the AVHRR, these sensors have a unique design allowing them to observe the same part of the sea surface twice, once looking almost straight down and the other viewing obliquely. This dual view capability significantly improves the atmospheric correction.

The third approach is to use microwave radiometers in polar orbit, operating at 6 or 10 GHz. Although not so sensitive or easy to calibrate as infrared instruments, and having much coarser spatial resolution, microwave radiometers have the advantage over infrared of being able to view through clouds and are insensitive to the presence of atmospheric aerosol.

3.3.2 Satellite infrared sensors

An infrared sensor records the radiance detected at the top of the atmosphere in specific wavebands, λ_n. The individual measurements in each channel, *n*, can be expressed as an equivalent black body brightness temperature, T_{bn}, that is the temperature required for a black body with 100% emissivity to emit the measured radiance. At a particular wavelength, black body emission is defined by the Planck equation:

$$L(\lambda,T)=\frac{C_1}{\pi\lambda^5\left[\exp\left(C_2/\lambda T\right)-1\right]}. \tag{5}$$

where L is the spectral radiance, per unit bandwidth centred at λ, leaving unit surface area of the black body, per unit solid angle (W m-2 m-1 str-1), λ is the wavelength (m), T is the temperature (K) of the black body, C1 = 3.74 × 10-16 W m2, and C2 = 1.44 × 10-2 m K. This must be integrated with respect to wavelength over the measured waveband and convoluted with the spectral sensitivity of the sensor in order to represent the radiance intercepted by a particular spectral channel.

To obtain T_{bn} from the digital signal S_n recorded by the sensor for waveband *n* requires direct calibration of the sensor using two on-board blackbody targets of known temperatures which straddle the range of ocean surface temperatures being observed. This is the method adopted by the ATSR class of sensor, whereas the AVHRR uses the simpler but less accurate alternative of a single on-board black body with a view of cold space serving as an alternative to the second black body.

Ideally we wish to measure the radiance leaving the water surface, which is determined by the skin temperature of the sea, T_s, and by the emissivity of seawater. In the thermal infrared this is greater than 0.98, but a small contribution to the satellite detected radiance comes from the reflected sky radiance, for which allowance must be made. Because of absorption by greenhouse gases T_{bn} is cooler than T_s by an amount which varies in time and place, mainly with the amount of atmospheric water vapour. It is the task of the atmospheric correction procedure to estimate T_s given top of atmosphere measurements of T_{bn}.

A well-established method of atmospheric correction is to make use of the differential attenuation in different wavebands. When viewing the same ground cell, different wavebands (*i*, *j* etc.) of the sensor would record the same temperature ($T_{bi} = T_{bj}$) if there were no atmospheric attenuation. The difference between the top of atmosphere brightness temperatures T_{bi} and T_{bj} is related to the amount of absorbing gases in the atmospheric path, so that algorithms of the form

$$T_s = aT_{bi}+b(T_{bi} - T_{bj}) + c \quad (6)$$

where a, b and c are coefficients to be determined, provide a good basis for atmospheric correction of the AVHRR (McClain et al., 1985). During the day the "split-window" algorithm uses wavebands at 10.3-11.3 μm and 11.5-12.5 μm, while at night the 3.5-3.9 μm channel can also be used. This 3.7 μm channel is corrupted by reflected solar radiation in the daytime. A number of non-linear variants of this basic form have also been developed (Barton, 1995). The algorithm is also supposed to accommodate the non-blackness of the sea.

Common to each of these approaches for AVHRR is the requirement for the coefficients to be determined by a best fit between the satellite predictions and coincident observations of SST from a number of drifting buoys. The match between the buoys and satellites has a variance of more than 0.5 K, applicable only to the regions populated by the buoys (Podesta *et al*., 1995). The same algorithms are assumed to apply to parts of the ocean where there are no buoys, although the validity of this assumption needs to be quantified. Regional algorithms matched to local data may achieve greater accuracy.

Although the instantaneous distribution of water vapour and aerosols in the atmosphere are not known, the radiation transfer physics of the atmosphere is well understood and can be modelled with some confidence in fine spectral detail. It is therefore possible to simulate T_b for a given combination of T_s, atmospheric profile and viewing angle for the spectral characteristics and viewing geometry of each channel of a particular sensor. This offers an alternative strategy for atmospheric correction in which an artificial dataset of matching T_s and T_{bi}, T_{bj}, etc. is created using a wide variety of typical atmospheric water vapour and temperature profiles. The coefficients for an equation of form similar to (6) are generated by a regression fit to the artificial dataset. The resulting algorithm should be applicable to all atmospheric circumstances similar to those included in the modelled dataset, leading to an estimate of the skin SST. It is independent of coincident *in situ* measurements, although they are needed for validation.

This was the approach adopted for the along track scanning radiometer (ATSR) flown on the ERS polar orbiting satellites (Edwards *et al*., 1990). The ATSR scans conically to observe a forward view at about 60° incidence angle and a near-nadir view about two minutes later. Using the same three spectral channels as AVHRR for each of the views, it thus acquires six measures of brightness temperature. The different path lengths for forward and nadir views make for a more robust algorithm (Zavody *et al*., 1995), less reliant on the spectral dependence of the atmospheric attenuation. The single-view approach was rendered inoperative when large volumes of volcanic dust were temporarily injected into the stratosphere by the eruption of Mt. Pinatubo in 1991 (Reynolds, 1993). Although the first ATSR

algorithms were also affected, a reworking of the semi-physical model by including the stratospheric aerosols in the radiation model led to algorithms which cope well with the volcanic problem (Merchant *et al*. 1999). This approach should also be robust in situations such as those where dust from the Sahara is lifted into the troposphere over the Atlantic.

The atmospheric correction algorithms produce maps of SST at fine resolution (about 1.1 km) for each overpass. However, the atmospheric correction methods cannot retrieve SST when cloud wholly or partly obstructs the field of view. Therefore at this stage cloud must be detected using a variety of tests (e.g., Saunders and Kriebel, 1988), so that only cloud-free pixels are retained for oceanographic applications, such as assimilation into models. The most difficult cloud contamination to identify is that by sub-pixel size clouds, thin cirrus or sea fog where only small deviations of temperature occur. Failure to detect cloud leads to underestimation of the SST and can produce cool biases of order 0.5 K. Thus confidence in the cloud detection procedure is just as important as atmospheric correction for achieving accurate SST. Where uncertainty remains in cloud detection, this should be flagged in the error estimate fields attached to SST products. Cloud detection is generally more successful during daytime, when visible and near-IR image data can be used, than at night.

The SSTs measured in individual overpasses are incorporated into global composite datasets by averaging all individual pixel contributions to each larger cell over a period of a few days. The larger cells are defined by longitude and latitude on a grid with spacing typically 1/2 or 1/6 degree (about 50km or 16 km at the equator). The multi-channel sea surface temperature (MCSST) (Walton *et al.*, 1998) was the standard global composite product derived from AVHRR, until superseded by the Pathfinder SST (Vasquez *et al.*, 1998). This is a re-processing of the archived pixel-level AVHRR data with algorithms incorporating the best knowledge of sensor calibration drift and making full use of the available drifting buoy dataset (Kilpatrick *et al*., 2001). It also makes more use of night-time data than previous analyses, and aims for long term consistency. Reynolds and Smith (1994) developed an OI-SST archive that is an optimal interpolation of both *in situ* and satellite data, and should therefore provide more climatological continuity with pre-satellite SST records before 1980. The global composite product from ATSR is the ASST (Murray, 1995) which is being re-processed using the more robust atmospheric algorithms (Merchant *et al.*, 1999).

3.3.3 Microwave radiometers on satellites

Microwave radiometers detect the brightness temperature of microwave radiation which, like the infrared, depends on the temperature of the emitting

surface. Their great benefit is that their view is not impeded by cloud and very little attenuation occurs in the atmosphere, although water present as large liquid drops in precipitation does attenuate the signal. However, the emissivity, ε, of the sea surface in the microwave part of the spectrum is less than 0.5. ε also depends on factors such as the temperature, the salinity and the viewing incidence angle. This in turn means the brightness temperature is also a function of the mean square slope and hence of the sea surface roughness and wind speed. While this complicates the retrieval of SST from microwave radiometry compared with infra-red methods, the corollary is that microwave sensors can be used to measure the surface roughness, rainfall or even salinity as well as SST.

It is possible to distinguish between the different contributions to the brightness temperature of SST, surface roughness and salinity, as well as to identify atmospheric contamination by liquid water, because each factor differentially affects different microwave frequencies. For example SST strongly affects wavebands between 6 and 11 GHz whereas the effects of salinity are found only at frequencies below about 3 GHz. Surface roughness effects influence frequencies at 10 GHz and above, and are also polarisation specific. Thus a multi-frequency and multi-polarisation radiometer can, in principle, be used to measure SST, surface wind and precipitation (see chapter 8 of Robinson (2004)). Each of these has potential for use in ocean models, and coincident measurement of SST and winds has potentially useful applications in the estimation of air-sea fluxes.

Despite a long series of microwave sensors flown for atmospheric remote sensing, serious consideration of microwave measurements of SST from space started only when a microwave radiometer having a 10.7 GHz channel was flown on the Japanese-US Tropical Rainfall Mapping Mission. Called the TRMM microwave imager (TMI) it has a spatial resolution of 0.5° (about 50 km) and because it over-samples it is capable of mapping mesoscale eddies quite effectively using a grid scale of 25 km. It lacks the preferred SST waveband of 6.6 GHz, but its 10.7 GHz channel is sensitive to SST in tropical water temperatures (Donlon *et al*., 2001), and its usefulness for measuring the thermal signatures of tropical instability waves has already been demonstrated by Chelton *et al*. (2000). It covers only latitudes lower than 40º.

In 2002 the Japanese Advanced Microwave Scanning Radiometer (AMSR-E) was launched into a near-polar orbit on the NASA Aqua satellite. This sensor includes a channel at 6.6 GHz, which is effective over the full range of sea temperatures, and has opened the way for routine, high quality, global mapping of SST by microwave radiometry. AMSR-E is now providing global cloud free SST to an accuracy of ~0.3 K derived from over-sampled 76 km resolution data. The composite daily, weekly and monthly SST products are supplied on a ¼° grid (Wentz and Meissner, 2000).

Microwave radiometers cannot be used within about 100 km of the coast because of the side-lobe contamination of microwave sources on land leaking into the antenna reception. This, with their low spatial resolution, severely limits their usefulness in coastal and shelf seas.

3.3.4 The character of the ocean surface thermal structure

The ocean modeller requiring measurements of SST, for assimilation or to validate the temperatures in the top layer of an ocean GCM, may conclude that their task is greatly simplified by the wide choice of different types of observations of SST now available from both *in situ* and satellite platforms. However, there is a pitfall for the unwary user of SST data, arising from the detailed character of the thermal structure in the top few metres of the ocean. Two distinct factors create near-surface vertical temperature gradients. Firstly on sunny calm days a diurnal thermocline tends to develop above which a top layer is found, a metre or so thick and up to about 1 K warmer than below (although exceptionally it can be several K warmer). At night the warm layer collapses. Secondly (and independently of the first effect) the top skin layer of the sea, a fraction of a millimetre thick, tends to be a few tenths of a Kelvin cooler than the water immediately below. Both these effects, and especially the first, may be horizontally variable, leading to spatial patchiness of SST.

Neither of these processes is normally represented in the physics of ocean models for which the topmost layer typically corresponds to the upper mixed layer assumed to be uniform above the seasonal thermocline. The different types of measurement of SST also sample at different levels of the near-surface thermal structure. In other words the definition of "SST" is different for the thermometer on a buoy's hull, for a sensor in a ship's cooling water intake, for an infrared radiometer, for a microwave radiometer, and for an ocean model. These differences are important when accuracies of a few tenths of a Kelvin are required. They may also vary considerably during the day so that a single daily measurement used may be aliased depending on the time in the diurnal cycle at which it is sampled. It is therefore necessary to harmonise SST data from different sources before they are introduced to an ocean model. This is one of the issues discussed in section 4 of this paper.

It is certainly worth taking the trouble to resolve these issues because SST observations can provide a very useful constraint on models. Surface ocean dynamical features often have thermal signatures. Major ocean currents are normally associated with thermal fronts. Ocean eddies are often visible in satellite SST images. Thus the assimilation of SST should in principle help to constrain the modelled evolution of mesoscale variability. In the case of coupled ocean-atmosphere models the interface temperature gains even more importance for constraining the model. In this case

particular care must be taken in defining which type of SST is needed, since the atmosphere is in contact with the skin temperature rather than the upper mixed layer temperature normally represented in the ocean model.

3.4 Ocean waves

The nowcasting and forecasting of ocean waves is an operational task which benefits many different users of the sea and is essential for the safety of mariners and for cost-effective navigation. Ocean wave forecasting models depend on good wind forecasts from numerical weather prediction models. Their performance can also be improved if good observations of wave data can be assimilated in a timely way. In situ measurements from wave buoys offer a means of testing and validating model predictions, but they are isolated and too few to make much impact if assimilated into a model. This requires the wide area coverage offered by satellites. Here we consider two ways in which ocean wave data potentially useful for models are measured from space.

3.4.1 Significant wave height measured by altimeters

When an altimeter measures the time for an emitted pulse to return, it tracks in detail the shape of the leading edge of the echo, from which it is possible to make a very good estimate of the significant wave height, $H_{1/3}$, within the pulse-limited footprint illuminated by the altimeter. For a perfectly flat calm surface the return echo has a very sharp edge. If there are large waves, several metres in height from trough to crest, then the return signal starts to rise earlier, as the first echoes are received from the crests, but takes longer to reach its maximum, when the first echoes are received from the wave troughs. The rising edge of the echo is modelled by a function in terms of the root mean square ocean wave height, so that by matching the observed shape to the model function it is easy to gain an estimate of $H_{1/3}$. This method has delivered robustly accurate measurements of $H_{1/3}$ for more than twenty years from different altimeters (Cotton & Carter, 1994) and comparison with buoys shows root mean square differences of only 0.3 m (Gower, 1996), which is the limit of the buoy accuracy.

It therefore provides an excellent source of information to be used in wave models. At present there are at least two altimeters in operation (see chapter 11 of Robinson (2004)), each crossing the Earth with 14-15 orbits per day. However, the altimeter footprint measures waves only along a 10 km wide swath along the satellite ground track. At the Equator the tracks of successive orbits are nearly 3000 km apart, and so in one day the available samples are quite sparse and it is possible to miss altogether local regions of high waves associated with a recent storm. Even with several altimeters in

complementary orbits it would be difficult to provide a reliable operational wave monitoring service based on observations alone. However, when the satellite data are assimilated into ocean wave forecasting models, their global coverage, their accuracy and their detailed resolution along the track ensure that they make a measurable improvement to the skill and reliability of wave forecasts. The major shortcoming of this type of wave measurement is that it contains no information about the wave direction.

3.4.2 Directional wave spectra from synthetic aperture radars

When a serious discrepancy is found between a wave model forecast and the measured wave height, it is desirable to know the directional properties of the waves so that incorrect wave energy sources in the model can be located and corrected. Directional wave spectra can be estimated using synthetic aperture radar (SAR). SARs view the sea surface obliquely and produce image maps of the backscattered microwave energy, at a spatial resolution of about 25m. The signal processing required to achieve this resolution is computationally intensive and can generate interesting artefacts, especially when the scattering surface is in motion (see Chapters 9 and 10 of Robinson (2004) for an introduction to SAR ocean imaging). The radar backscatter signal itself needs to be interpreted carefully. It represents a measure of the roughness of the sea surface at length scales similar to the radar wavelength, since a type of Bragg scattering mechanism is responsible for the radar echo from obliquely incident microwave pulses.

Swell waves appear on SAR images as approximately linear patches of bright and dark corresponding to regions where the backscatter is greater or less than the mean. There are three different mechanisms by which the swell waves modulate the short waves which determine the backscatter. A two-dimensional spectrum of the image amplitude is related to, but not the same as, the directional wave spectrum. The modulation transfer function which relates the ocean wave spectrum to the image spectrum depends on the swell wave frequency, their amplitude and their direction relative to the radar azimuth. The problem of accurately and reliably inverting image spectra to retrieve wave spectra has been a challenging field of research for two decades, although recent techniques which use the complex form of the SAR image (including the phase as well as the amplitude of the backscatter) look most promising (see Chapron *et al*. (2001) for a review of the subject). It is most important that users of SAR-derived wave spectra should understand that SAR can provide very little information about the shorter waves with period less than abut 10 s, and this frequency cut-off is worse for waves propagating in a direction parallel to the satellite track.

The SARs of the European Space Agency can operate in a wave-mode in which small (5 × 10 km) imagettes are acquired every 200 km along track. Thus wave spectra can be sampled from across the global ocean every day.

As yet there is no operational ingestion of such data into wave models, but the improved spectra have raised new interest in doing so.

4. Preparing satellite SST for assimilation into models

4.1 Introduction

If SST data from space are to be used operationally as part of ocean observing systems and for creating a reliable, stable climate time series, there is a need to harmonise and inter-calibrate the SST products already being produced by several different agencies. Specifically for use in ocean forecasting models there is also a need to precondition the data to make them more immediately usable for assimilation into numerical models. This is true of most types of satellite data required by models although the processing tasks for preconditioning vary according to the parameter of interest. The blending and preparation of SST data differs from what has to be done for altimetric measurements of sea surface height anomaly or for parameters such as chlorophyll concentration derived from satellite ocean colour data. The rest of this chapter describes a new international initiative to perform the intermediate processing tasks needed to generate the best coherent products from complementary SST data delivered by several agencies. It serves as a specific illustration of the organisation and initiative that is desirable in order to enhance the usability in model assimilation of any observational parameter derived from several independent measurement programmes, particularly from satellites.

In 2001 the Global Ocean Data Assimilation Experiment (GODAE) prompted the creation of a working group which has developed into the GODAE high resolution sea surface temperature pilot project (GHRSST-PP). In the rest of this chapter, section 4.2 identifies the particular issues relevant to SST data and how these have been rationalised by the GHRSST-PP and formalised in the GHRSST Data Processing Model. Section 4.3 describes how a European project called Medspiration is about to start producing products conforming to the GHRSST-PP specifications. Section 4.4 concludes with a discussion of what can be learned more generally from the example of GHRSST-PP and Medspiration.

4.2 The challenge of assimilating SST data from many sources

4.2.1 Sampling and resolution capability of existing sensors

Section 3.3 noted that satellite sensors for measuring SST may be differentiated into four classes. The derivation of SST from the top-of-

atmosphere brightness temperatures recorded by the sensors identified in Table 3 is performed by a number of different agencies around the world, leading to a variety of SST data products. For example Table 4 lists the SST products available for European seas and the Atlantic Ocean. These are produced in near–real time, most are publicly available and can be served for use by operational models. Other SST sensors such as the infrared channels on MODIS have not been included because at present they are not processed within an operational timeframe.

Each product can be considered to be independent of the others. Even those derived from the same satellite source by different agencies are by no means identical because each agency has its own protocols regarding matters such as cloud detection, atmospheric correction algorithms, rules for compositing, confidence flags and error statistics. However there is at present little, if any, independent validation of most of the products, although ESA does have a formal AATSR product validation process.

The type of SST (see section 4.2.3) also differs according to the producer. Note for example that although AATSR and AVHRR measure radiation emitted from the sea-surface skin, the SST products from AVHRR are classified as either subskin or bulk, and from AATSR as skin, because of the different ways each producer calibrates the atmospheric correction.

The wide choice and apparent redundancy offered by the different sensors in Table 3 and SST data products in Table 4 prompts the question of which is the best to use for assimilation into ocean forecasting models. Because the measurement of global SST from space using polar orbiting infra-red sensors is a well established mature observational system, having acquired useful data for more than 20 years, it might seem reasonable to assume that it is ready to provide data for assimilation into ocean models.

However, stringent sampling requirements and a higher degree of accuracy are now demanded for applications in both climate monitoring and operational oceanography (Robinson and Cromwell, 2003). On closer inspection it seems increasingly difficult to meet these requirements using any one of the SST data products currently produced by several different agencies. No matter what improvements are made to sensor technology or atmospheric correction algorithms, the problem of cloud cover imposes unavoidable limits on the use of infra-red sensors, while microwave sensors which can penetrate the cloud are not capable of the required spatial resolution.

The most promising way to obtain the best SST data for input to models is by combining data from the different sensor types of Table 3 so that each product from Table 4 complements the others (Robinson and Donlon, 2003). Data from the AATSR provides the best absolute accuracy through that sensor's dual view, but coverage suffers from the narrow swath inherent in the viewing geometry and so it cannot achieve a revisit interval appropriate to operational applications at all latitudes. In contrast this is achieved by the

AVHRR with its less certain radiometric accuracy, although its wide swath and more frequent revisits can still not view the sea when cloud is present. This problem can be partly overcome in Equatorial regions by the use of geostationary sensors, which are available at all times of the day and night to see the sea whenever there are gaps in the cloud. Ultimately, however, the absence of IR observations caused by persistent cloud can be overcome only by the use of microwave sensors. Although their spatial resolution is very much inferior to that of infrared sensors, the recent improvement in their radiometric performance should enable them to contribute unique information to operational systems under persistently cloudy conditions.

Sensor name	Product name and resolution	Spatial coverage in specified time	Type of SST	Data provider
AATSR	ATS_NR_2P 1km	Global, pseudo 3 day repeat	Skin	ESA
AATSR	ATS-Meteo 10 arc min	Global, pseudo 3 day repeat	Skin	ESA
AVHRR NOAA16	GAC (1) 9km	Global, daily repeat	At 1m (2)	NAVOCEANO
AVHRR NOAA16	LAC (3) 2 km	Regional, at least daily repeat, day and night	At 1m (2)	NAVOCEANO
AVHRR NOAA16	NAR 2 km	Regional, at least daily repeat, day and night	Subskin	EUMESAT O&SI SAF
AVHRR NOAA17	GAC (1) 9km	Global, daily repeat	At 1m (2)	NAVOCEANO
AVHRR NOAA17	LAC (3) 2 km	Regional, at least daily repeat, day and night	At 1m (2)	NAVOCEANO
AVHRR NOAA17	NAR 2 km	Regional, at least daily repeat, day and night	Subskin	EUMESAT O&SI SAF
MSG SEVIRI	SEVIRI 0.1º lat./lon.	Regional, 3 hour repeat	Subskin	EUMETSAT O&SI SAF
GOES-East	GOES 4 km	Regional, 3 hour repeat	At 1m (2)	EUMETSAT O&SI SAF
GOES-East	GOES 6 km	Regional, 1/2 hour repeat	At 1m (2)	NAVOCEANO, PO.DAAC
TMI	0.25º lat./lon. grid	40ºN - 40ºS, daily	Subskin	REMSS
AMSR-E	0.25º lat./lon. grid	Global, daily repeat	Subskin	REMSS

1. GAC means the SST product is produced from the global area coverage, a reduced dataset stored on board and delivered once per orbit.
2. This expresses the definition of the product given by the data provider.
3. LAC means the SST product is produced from the local area coverage, which is the full resolution data stream downloaded on acquisition to a ground station within line of sight of the satellite. LAC is not available globally, but there is good provision of receiving stations to cover European coastal and shelf waters and much of the North Atlantic Ocean.

Table 4. Summary of satellite SST data products available over European seas.

4.2.2 Combining data from different sources

It is one thing to decide in concept that the data from different types of SST sensors should be combined in order to benefit from their complementary sampling characteristics; it is another to implement a system in which this can be done without introducing more errors that result in a poorer assimilation performance than using a single SST data product from a single sensor. If each of the sensors measured precisely the same value of SST with the same accuracy then each data value could be used with equal weight, although the coarser spatial resolution of the microwave data would still have to be taken into account.

In practice, there are at least three factors in addition to their different spatial resolutions which cause the different data products to deliver different values of SST. These are:

a) different accuracy, with errors in calibration and noise leading to a different bias and standard deviation when compared against an *in situ* "standard" of SST;
b) the fact that different methods of measuring SST result in different values even when there are no measurement errors, because they sample different parts of the near surface thermal microstructure; and
c) there are circumstances where the measured SST varies considerably throughout the day because of a strong diurnal variability signal.

Factor (a) is unavoidable when different sensor measurement techniques are being used. It is best treated by obtaining a confident knowledge of the errors (mean bias and standard deviation) associated with each measurement type. However, it is important when estimating the errors that factors (b) and (c) are also taken into account and do not create additional bias or variance. Factors (b) and (c) both relate to the physical behaviour of the upper layer of the ocean and are discussed in the next paragraph.

A particular problem that faces the use of SST measurements is that of the difference between precisely which part of the sea surface is being measured, as mentioned already in Section 3.3.4. Figure 8 identifies the difference between the skin SST which is measured by an infrared radiometer and the sub-skin SST a short distance below the surface (of order tenths of a millimetre). They are separated by the thermal skin layer where heat transport is restricted to molecular conductivity because of the suppression of turbulence close to the surface. The sub-skin is typically a few tenths of a degree warmer than the actual skin. Microwave radiometers measure the temperature at approximately the same depth as the sub-skin. Thus when microwave and infrared SST measurements are compared the thermal skin layer difference must be modelled and allowed for.

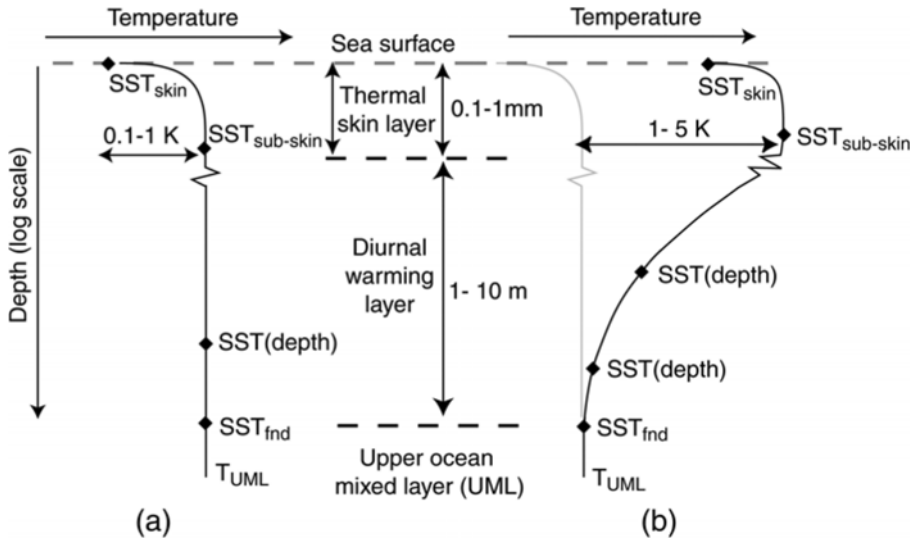

Figure 8. Schematic diagram showing characteristic temperature profiles at the sea surface for (a) night time conditions or daytime with moderate to strong winds and (b) daytime calm to light wind conditions and direct solar heating.

Both infra red and microwave measurements are different from the temperature sampled by an *in situ* sensor, normally used for calibration purposes. A thermometer in contact with surface sea water is typically mounted on the hull of a buoy or ship and located a distance of order 1 m below the surface. In circumstances of daytime cloud-free conditions and low wind the sun tends to heat up the water near the surface, inducing a thermal gradient in the upper metre or so. Consequently an *in situ* measurement of SST may be cooler than the satellite measurement. This diurnal thermocline develops to a maximum in the early afternoon and then reverts to a mixed layer of uniform temperature during the night. Thus the difference between the satellite measurement and an *in situ* measurement used to calibrate it can vary by up to several K through the day, depending on the surface solar irradiance and the wind history during the day.

None of the SST data products listed in Table 4 take these factors into account in their processing. Consequently to use the data products as they are presented could lead to additional errors in the data being fed to the model, especially when they are combined, since the different response of each type of data to these factors introduces a spurious variability. This is the underlying problem which led to the development of the GHRSST-PP project.

4.2.3 The GHRSST data processing model

The GHRSST-PP project has approached the problems identified in the previous section by adopting a particular model of the thermal structure of the near surface of the ocean and with it defining clear terminology that is intended to avoid confusion and misunderstanding in the use of the loose term "SST" which can have several meanings. A distinction is made between skin and sub-skin SST as outlined in 4.2.2, and a new term, the "foundation SST" or SST_{fnd} is used to describe the temperature on which the diurnal warming (if any) is built each day. It is most clearly specified as the temperature of the well mixed layer found just below the skin layer at dawn, when any diurnal thermocline structure from the previous day has collapsed. At this time of day it is equivalent to the subskin SST. The SST_{fnd} is defined on a daily basis. It also corresponds closely to the temperature of the upper layer of the ocean most commonly represented in numerical ocean circulation models that do not attempt to represent diurnal warming. The use of this name is preferred instead of the term "bulk" SST. The latter is very imprecise because it tends to be used also for any *in situ* measurement made by a thermometer in contact with the water (as distinct from a radiometric measurement which observes the skin SST) which may or may not be influenced by the diurnal thermocline, depending on the depth of the thermometer and the character of a particular day's diurnal variability.

A key aspect of the GHRSST processing model is that satellite measurements are recognised to be either skin or subskin temperatures, while it is assumed that the temperature normally required by models is the SST_{fnd}. The relationships between each of the different SST quantities are parameterised in terms of the controlling air-sea interaction quantities, particularly the wind speed and the solar irradiance, and these are expected to vary during the day. Note that the difference between skin and subskin (the thermal skin deviation), and between subskin and foundation (the diurnal warming) are different and controlled by largely independent processes. It is envisaged that when SST values are needed to high accuracy for ingestion into a numerical ocean model, adjustments will be applied to convert the satellite measurement of skin or subskin SST into the appropriate value of SST_{fnd}. However, there are circumstances where a model may need to use the skin temperature, for example when estimating air-sea fluxes, and in this case the GHRSST-PP processing model specifies how conversion can be made from SST_{fnd} to SST_{skin}.

The approach outlined in the previous two paragraphs provides a framework in which problems (b) and (c) from 4.2.2 are handled. It is also intended that errors caused by (b) and (c) should be isolated from the more general biases and errors of (a). The latter need to be provided along with the basic temperature data to allow the model assimilation system to weigh the importance to be attached to a particular SST ingest. Therefore

GHRSST-PP has specified a core SST product which consists of the same SST data as received from the original processing agency (e.g. one of the data products listed in Table 4) with the addition of:

i. a confidence value associated with the probability of the data being corrupted by such things as cloud cover (IR only), high winds (m/w only), diurnal warming and proximity to land;
ii. single sensor error statistics (SSES) defining the bias and standard deviation applicable to that particular sensor depending on the confidence flag; and
iii. sufficient information to be able to calculate the parameterized values of the thermal skin deviation and diurnal warming at the precise time of the satellite sample (allowing conversion to SST_{fnd}).

Note that for this basic product there is no intention of resampling or regridding the data specified in Table 5, which are typically Level 2 image datasets (presented in the sensor co-ordinates). Hence it is referred to as the GHRSST-PP L2P product (level 2 pre-processed). Neither is there any merging of data from different sources. It is assumed that the model assimilation scheme can handle the native grids of the inputs, and any differences in bias between them, better than any pre-processing or blending which risks introducing unnecessary additional errors. However, the addition of (iii) allows the assimilation system to make the conversion to whatever definition of SST is appropriate for the model, while the provision of (i) and (ii) allows the influence of data assimilated into a model to be weighted according to its quality. Note that the SSES in (ii) should be calculated on the basis of matching against independent validation measurements of temperature, after making the appropriate correction to adjust between skin, subskin or foundation SST so that comparison is made between like quantities. Thus the bias and standard deviation should no longer contain an element caused by inappropriately comparing unlike quantities (although they may contain a contribution from errors in the method used to perform the conversions).

There are four main product types proposed by the GHRSST-PP processing model:

- The L2P product described in the previous two paragraphs, intended as the primary source of data for assimilation into numerical ocean forecasting models, and therefore required in near-real time.
- A match-up data base (MDB) that pairs spatially and temporally coincident values of SST independently measured *in situ* and by satellite.
- Diagnostic data sets (DDS) will be produced for a number of small regions chosen to represent different ocean and atmosphere conditions around the world.
- An analysed SST product at level 4 that is a blend of all the available SST data, each converted to SST_{fnd} and then used in an optimal

interpolation (OI), or other analysis technique, to create the best daily fields of SST_{fnd}, without any cloud gaps.

The processing models required to generate these products are fully defined in the GHRSST Data-processing System document, the current version of which is GDS v1.5, available online from http://www.ghrsst-pp.org. The model is an active document, still being refined in the light of working experience.

4.3 Putting the GHRSST processing model into action

GHRSST-PP has evolved rapidly to meet the perceived need for a new class of merged SST products. After four Workshops the Science Team produced an implementation plan that was approved by GODAE, its parent body. GHRSST-PP is now established through the creation at the UK Met. Office of an International Project Office sponsored by the European Space Agency (ESA). While this provides leadership and co-ordination, the development comes from work at a regional level, through the GHRSST-PP regional data assembly centres (RDAC)s. One of these already existed through a Japanese SST programme for the Asian Pacific area. The function of a European RDAC is now being fulfilled by a project within the ESA's Data Utilisation Envelope (DUE) Programme called "Medspiration". The Multi-sensor Improved SST Project (MISST), has been established by the National Ocean Partnership Programme in USA to support the function of an American RDAC. When fully implemented GHRSST-PP should make a considerable impact on the quality of SST data available for operational oceanography worldwide.

As the European RDAC for GHRSST-PP the Medspiration project, which commenced in January 2004, contributes to the development of an operational oceanography system within the European Union's Global Monitoring for Environment and Security (GMES) programme. The core user need identified for Medspiration was for SST products at high resolution (~10 km) over the Atlantic Ocean and adjoining Seas, which combine the best of all the SST data products already available from individual sensors. The desired temporal resolution is 6 hrs. The required accuracy of SST data is to be better than 0.4 K, with a goal of 0.1 K. Since Medspiration will use the SST sources in Table 4, which are already supplied by various agencies, it will always depend on the quality of the data being ingested.

However, it will for the first time provide European operational users not only with a near-real time single point access to data from all the different SST sources but also with robust error and confidence statistics attached to the SST products. This capacity to assign quality values to all the SST data sources being used, and to any future SST products that may be generated, is a fundamental requirement of GHRSST-PP. It is needed in order to

facilitate the assimilation of SST products into operational ocean models and forecasting systems.

Medspiration will produce four types of data product. These are elaborated as follows:

Level 2 pre-processed SST data: ESA's Medspiration project will generate L2P data, as defined in Section 4.2.3, within a few hours of the input L2 SST products being served by their respective agencies. The area coverage is the whole Atlantic Ocean and adjacent seas. In the User Requirements Document (URD) produced by GHRSST-PP the EurDAC area is defined as encompassing the area 70ºS to 90ºN (to the ice limit) and from 100ºW to 45ºE (to include the adjacent seas but excluding any part of the Pacific Ocean).

Level 4 gridded, interpolated SST data product: An L4 product will be produced for the Mediterranean Sea, consisting at each grid cell of a single daily value of SSTfnd produced by the optimal interpolation process, plus an estimate of the expected diurnal stratification and skin effects every two hours. Error estimates will also be provided for every temperature estimate. The ultra-high resolution (UHR) grid for this product is 2 × 2 km. Delivery is due by noon on the day following that to which it refers.

The software for generating the L4 product is designed with sufficient flexibility to allow the parameters in the OI scheme to be changed. Similarly the form of the model defining the differences between skin, subskin and foundation SST can be adjusted in future following operational experience and analysis of the resulting L4 products. The main purpose of the L4 product in GHRSST is to meet the demand from general users of SST for a product which best combines all of the available primary sources of SST from different sensors.

A match-up data base (MDB) will be produced, which assembles matched pairs of all available in situ measurements of SST that are coincident with satellite measurements from the ingested L2 SST data within the Medspiration domain. It is most convenient to create the MDB in near-real time when the L2P data have just been produced. Medspiration will make use of the French Coriolis Data Centre used to assemble in situ data although, the difficulty of timely delivery of some in situ SST samples is likely to delay the MDB creation by up to a week after real time.

The MDB will be analysed regularly to allow GHRSST to update frequently (say once per week or every few weeks) the sensor specific error statistics (SSES) attached to the L2P product, needed for assimilation and also in the OI procedure. It is also important that a broad spatial coverage is achieved by the MDB so that the SSES are properly spatially representative and can also be stratified by confidence value.

A diagnostic data set (DDS) of SST observations will be compiled for the Medspiration L2P product area by extracting 1/100º × 1/100º gridded sub scenes for a set of predefined small areas (typically 2º × 2º in extent) from

every L2P and L4 product produced by the Medspiration processor. Each product will be resampled (nearest neighbour) onto a common grid to facilitate analysis. Where possible the DDS sites have been chosen to coincide with known locations of regular in situ measurements (e.g. PIRATA buoys, Ferry-Box routes).

Within Medspiration the DDS will allow close comparison to be made between the different input streams of SST and with the L4 output. One objective is to provide a prompt alert should a particular sensor start to deviate seriously from the others. Another is to facilitate the research needed to test and improve the data merging methods, diurnal variability parameterisations and OI configuration used both in Medspiration and more widely in GHRSST-PP.

4.4 Lessons to be learned from the GHRSST-PP and Medspiration approach

Although still in the development stage and not scheduled to commence a demonstration of operational capability before early 2005, GHRSST-PP and its European contributor Medspiration have been singled out for this case study because they provide a pathfinding example of what can be achieved when data providers and the remote sensing scientists working with them turn their attention to the needs of users, and in this case the requirements of the ocean modelling community for SST data to be assimilated into ocean forecasting models. It is worth noting that, although GHRSST-PP has been driven largely by the science team of satellite oceanographers and data providers, it was spawned in the first place by the modellers in GODAE and has at all stages interacted strongly with its intended users in the ocean modelling community, who have attended the science team meetings and endorsed the implementation plan and the data processing model. This was recognised by ESA when they decided to initiate the Medspiration Project within their DUE Programme. The Medspiration project is thus not so much a science development project as primarily a software contract to develop a system to meet the needs identified in the Medspiration User Requirements Document.

Stepping back from the details which are specific to SST measurements, there are more general lessons to be learned about the preparation of satellite data in order to make it more useful to the modelling community, as follows.

1. Attempts should be made to widen the spatial and temporal coverage of remote sensing by using data from different sources, especially when they offer complementarity of sampling. Ideally use should be made of the products being produced by agencies already, without changing them apart from adding error statistics and confidence flags. This implies an active partnership with the data supplying agencies.

2. The different characteristics of data from different sources needs to be recognised. The goal should be to harmonise them, accommodating the differences in a suitable physical model that parameterises them in terms of measurable and routinely available quantities.
3. Merging of data from different sources is not necessarily the best approach. Dialogue with the model assimilation specialists is essential before decisions are made about this.
4. If data are to be used for assimilation it is essential to provide error statistics and quality flags in near real time attached to the primary data. The error statistics must be based on independent measurements of the quantities actually represented in the data.
5. The concept of creating a match-up database appears to be a very useful approach to achieving validation by regularly updated error statistics for each sensor and data source. To be useful it needs to be populated in near-real time by matching satellite and *in situ* data pairs. However, care must be taken not to utilise data that is already being used as a means of calibration or fine tuning of processing algorithms by the agency responsible for the primary processing.
6. Diagnostic data sets, which assemble all the data available from different sources resampled to a common grid, provide a valuable resource for evaluation of the products. They will provide the ground on which to base research for further improvements.

Collaboration is essential between all the players concerned. Handled properly, this approach should be welcomed by the data producers as a means for making their products more useful, and also because rapid feedback from the MDB and DDS provides an external quality control. The potential users of SST in the modelling community find a group of scientists within GHRSST-PP ready to work with them. Within this partnership the modellers should be able to use the data successfully without themselves needing to become experts in all aspects of the remote sensing methodology. Finally the scientists are able to contribute more effectively to the application of their work, which is not only professionally stimulating but ultimately enhances the profile of this field of research and improves the prospects for funding of further work.

5. Conclusion

This chapter has offered a brief introduction to ocean remote sensing and especially those methods which make measurements of parameters such as SSH, SST and colour that can usefully contribute to operational monitoring and forecasting of the ocean. There is considerable potential to enhance the usefulness of satellite ocean data by assimilating them in near-real time into

ocean models. Given recent advances in both numerical ocean modelling and satellite observing systems, it is expected that fully operational ocean forecasting systems will be in place within a few years. In future it is likely that oceanographers will turn to operational models, supported by a global ocean observing network of satellite and *in situ* sensors, to obtain the best knowledge of the ocean state at any time or place, in much the same way as meteorologists rely today on the analysis delivered by atmospheric numerical weather prediction models.

Further research and development are needed before this can be achieved. There is scope to refine the remote sensing methods described above, not only to improve the accuracy of measurements but also to specify the errors more confidently, which is important if they are to be assimilated into ocean models. Ways should be found to harmonise the data provided by different ocean colour sensors, comparable to what the GHRSST project is doing for SST data products. A secure future for operational ocean forecasting systems also depends on planning now for continuity of appropriate sensors in space and needs commitment by funding agencies to the long term support of ocean monitoring satellites and buoy/drifter programmes which this implies. Such a commitment must be justified by sensitivity studies that clearly demonstrate the impact which the assimilation of particular ocean observations makes to the forecasting / nowcasting skill of ocean models. We can no longer expect new ocean monitoring satellites to be provided automatically as part of national or international space technology programmes. Alternative funding routes must be established so that ocean monitoring satellites can in future be commissioned by their users. It is therefore vital to spell out the benefits of large scale ocean observing and forecasting systems (e.g. Johannessen *et al.* 2003) so that society at large and especially those who use the sea will appreciate their potential benefits.

References

Barton, I.J. (1995) Satellite-derived sea surface temperatures: Current status. *J. Geophys. Res.,* **100**: 8777-8790.

Challenor P. G., Cipollini, P., Cromwell, D.D., Hill, K.L., Quartly, G.D., and Robinson, I.S. (2004) Global characteristics of Rossby wave propagation from multiple satellite datasets", *International Journal of Remote Sensing*, **25**(7-8): 1297-1302.

Chapron, B., Johnsen, H. and Garello, R. (2001) Wave and wind retrieval from SAR images of the ocean. *Ann. Télécommun.*, **56**(11-12): 682-699.

Chelton D.B., Wentz, F.J., Gentemann, C.L., De Szoeke, R.A. and Schlax, M.G. (2000) Satellite microwave SST observations of transequatorial tropical instability waves. *Geophys. Res. Lett.*, **27**: 1239-1242

Chelton, D. B., Ries, J. C., Haines, B. J., Fu, L.-L. and Callahan, P. S. (2001) Satellite Altimetry. In: Fu, L.-L. and Cazenave, A. (eds) *Satellite Altimetry and Earth Sciences*, pp. 1-131. Academic Press, San Diego.

Cotton, P. D. and Carter, D. J. T. (1994) Cross calibration of TOPEX ERS-1 and Geosat wave heights. *J. Geophys. Res.*, **99**: 25025-25033.

Donlon, C. J., Gentemann, C. and Wentz, F. (2001) Measuring surface temperature with microwave sensors , *Backscatter*, **12**(2): 37-39.

Edwards, T., Browning, R., Delderfield, J., Lee, D. J., Lidiard, K. A., Milborrow, R. S., McPherson, P. H., Peskett, S. C., Toplis, G. M., Taylor, H. S., Mason, I., Mason, G., Smith, A. and Stringer S. (1990) The Along Track Scanning Radiometer measurement of sea-surface temperature from ERS-1. *J. British Interplanetary Soc.* **43**: 160-180.

Gower, J. F. R. (1996) Intercalibration of wave and winds data from TOPEX/Poseidon and moored buoys off the west coast of Canada. *J. Geophys. Res.*, **101:**, 3817-3829.

Johannessen, J.A., Ian Robinson, Kostas Nittis (2003) MERSEA strand 1, In: H. Dahlin, N.C. Flemming, K. Nittis, S.E. Petersson (eds.) *Building the European Capacity in Operational Oceanography, Proceedings of the Third International Conference on EuroGOOS, 3-6 December 2002, Athens, Greece*, Elsevier Oceanography Series, No. 69, 714 pp., ISBN 0-444-51550-X.

Kidwell, K. (1991) *NOAA Polar Orbiter User's Guide*. NCDC/NESDIS, National Climatic Data Center, Washington, D.C.

Le Provost, C. (2001) Ocean Tides,. In: Fu, L.-L. and Cazenave, A. (eds) *Satellite Altimetry and Earth Sciences*, pp. 267-303. Academic Press, San Diego.

Kilpatrick, K.A., Podesta, G. P. and Evans, R. H. (2001) Overview of the NOAA/NASA Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res*., **106**: 9179-9198.

McClain E. P., Pichel, W. G. and Walton, C. C. (1985) Comparative performance of AVHRR based multichannel sea surface temperatures, *J. Geophys.. Res*., **90**: 11587-11601.

Merchant C.J., Harris, A.R., Murray, M.J. and Zavody, A.M. (1999) Toward the elimination of bias in satellite retrievals of sea surface temperature 1. Theory, modeling and interalgorithm comparison, *J. Geophys. Res*., **104**(C10**)**: 23565-23578.

Murray, M.J. (1995) *Sea surface temperatures from ATSR (August 1991 to July 1995*), (CD-ROM), Rutherford Appleton Lab., Chilton, England.

Podesta, G.P., Shenoi, S., Brown, J.W., and Evans, R.H. (1995) *AVHRR Pathfinder Oceans Matchup Database 1985-1993 (Version 18)*, draft technical report of the University of Miami Rosenstiel School of Marine and Atmospheric Science, June 8, 33pp.

Reynolds, R.W. (1993) Impact of Mt. Pinatubo aerosols on satellite-derived sea surface temperatures, *J. .Climate*, **6**: 768-774.

Reynolds, R. W. and Smith, T. S. (1994) Improved global sea surface temperature analyses, *J. Climate*, **7**: 929-948.

Robinson, I. S. (2004) *Measuring the Oceans from Space: The principles and methods of satellite oceanography*, Springer-Praxis, Chichester and Heidelberg, 670 pp.

Robinson, I.S. and Cromwell, D.D. (2003) *Observational Requirements for inputs to European Ocean Forecasting System Models*, Report D 3.1 for the E.U. Framework 5 Project MERSEA Strand-1. Southampton Oceanography Centre, Southampton, U.K., 57 pp.

Robinson, I. S. and Donlon, C. J. (2003) Global measurement of sea surface temperature from Space: some new perspectives. *Global Atm. Ocean System*, **9**(1-2): 19-37.

Tapley, B. D. and fourteen co-authors (1994) Precision orbit determination for TOPEX/Poseidon. *J. Geophys. Res.*, **99**: 24383-24404.

Vasquez, J., Perry, K. and Kilpatrick, K. (1998) *NOAA/NASA AVHRR Oceans Pathfinder Sea Surface Temperature Data Set; User's Reference Manual Version 4.0*. JPL Publication D-14070.

Walton, C. C., Pichel, W.G., Sapper, J. F. and May, D. A. (1998) The Development and Operational Application of Nonlinear Algorithms for the Measurement of Sea Surface

Temperatures with the NOAA Polar-Orbiting Environmental Satellites,. *J.Geophys. Res*., **103**: (C12) 27999-28012.

Wentz, F. J. and Meissner, T. (2000) *AMSR Ocean Algorithm (Version 2)*, Algorithm Theoretical Basis Document (ATBD), Remote Sensing Systems, Santa Rosa, CA, Revised edition, November 2, 2000.

Zavody, A.M., Mutlow, C.T. and Llewellyn-Jones, D.T. (1995) A radiative transfer model for SST retrieval for the ATSR, *J. Geophys. Res*., **100**: 937-952.

Chapter 7

IN-SITU OBSERVATIONS: PLATFORMS AND TECHNIQUES

Uwe Send

Scripps Institution of Oceanography, San Diego, California, USA

Abstract: In-situ observations and satellite remote sensing together need to be viewed as an integrated system to provide observational data required by operational ocean modelling and forecasting. This chapter outlines the methods and platforms available to operational oceanography and how they complement remote sensing and each other.

Keywords: In-situ data, platforms, instruments, observing methods.

1. Observing needs of operational oceanography

The goal of operational oceanography is to provide routine ocean forecasts on timescales of days to seasonal, to detect and predict short-term changes in the ocean (turbulence and "ocean weather") all the way to regime shifts and climatic changes, including the associated impacts on coupled ocean-atmosphere and coupled biogeochemical systems.

Such routine modelling and forecasting requires sustained observations for initializations and validation/ground truthing, for keeping the models on the correct trajectory, and in the development phase also for model testing and calibration. This chapter addresses the options and considerations for choosing platforms and techniques to provide these data. The focus will be on complementarity between the different in-situ methods, and between in-situ and satellite observations. Not covered here is the important issue (for operational systems) of sustained routine observing system *operation*, of quality control and data dissemination. This is addressed in the companion chapter by S. Pouliquen (this volume).

Some differences need to be recognized between the data requirements of global, regional, and coastal applications, in terms of resolution, accuracy, and variables needed. Global systems do not require accurate representation of individual small-scale features but need to have the correct "bulk"

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 191-206.

properties like heat content, stratification, transports (mass and heat), mixing, and air-sea fluxes. They need to have accurate enough forcing and fluxes to run without bias for weeks and months. The biogeochemical components of global models are still uncertain enough to make data requirements less stringent – the most useful contribution of data is a good representation of the overall biogeochemical regime conditions on the large scale for a few basic variables. Coastal models usually are run only for a few days but focus on small scales, and often try to represent detailed ecosystem species, both of which require more detailed observations. These models are strongly affected by the small-scale advective processes and less sensitive to small errors in the forcing. However, mixing, which changes stratification but also affects e.g. nutrients, is equally critical as in global models. Regional applications are somewhere in the middle. They typically address conditions and changes in sub-basins (e.g. Nordic Seas) or marginal seas (e.g. the Mediterranean). Mesoscale features generally still need to be resolved and correctly represented.

1.1 Variables

The description of the physical state of the ocean requires the density and temperature (T)/salinity (S) fields (not independent of course), as well as the absolute currents. Closely coupled to this is the physical forcing at the surface (e.g. wind, radiation, heat), which is covered in the chapter by W. Large (this volume). For biogeochemical models, the basic state is described by the variables, such as nutrients, phytoplankton, zooplankton and detritus in the simplest cases. These state variables, which take values everywhere in the model domain, need to be distinguished from the quantities which are forecast – these are sometimes derived quantities or not predicted everywhere (e.g. only at the surface).

The prime variables *forecast* in current operational systems are temperature and currents, with a focus on surface fields. The rationale for this emphasis is a combination of models being primarily physical, remote sensing delivering these variables at the surface, and many applications *needing* this information at the surface. For predicting ocean circulation, however, the associated *interior* density field needs to be known. Many applications also require the vertical stratification of temperature (i.e. heat content), and density stratification (for pollutant dispersal/mixing or fisheries). Some defense applications also seek the interior sound speed distribution which is calculated from subsurface T and S. Thus a minimum data requirement for physical models is the full density field and absolute currents at some level, unless data are only used for validation (in this case, selected locations or layers may be sufficient). Conceptually also integral properties should be important for constraining and initializing models, like a water-mass or basin heat content, or transports of mass and heat in major

current systems or passages. This however is not exploited in current assimilation and forecasting approaches.

Increasingly, operational systems also need to and do address ecosystem dynamics and biogeochemical cycles in the ocean. This is especially true for the regional and coastal applications. In such cases, a larger range of variables is required from an observing system. At the minimum level this includes the prime variables like oxygen, nutrients, chlorophyll or phytoplankton biomass, and zooplankton biomass (e.g. Fasham et al, 1993). In more small-scale applications, it may require the knowledge of individual species of plankton or fish, or of certain chemicals (specific nutrients, trace elements). For very specific purposes, long lists of variables can be drawn up, which however is not useful here since the goal of this presentation is the nature of typical observing systems and not the exhaustive coverage of singular cases.

1.2 Coverage and time-space sampling

The ideal observing system, both for research and for operational applications, covers the three space (x,y,z) and time (t) dimensions "completely". The word 'covering' usually denotes the extent/reach in the four dimensions. All current sampling techniques are discrete, however, in these dimensions, and 'completely' therefore must also be interpreted as having sufficient resolution to reveal the smallest scales of the variabilities of interest.

Thus, rigorously, for each application, a new system would need to be designed which can deliver the needed observations with the accuracy and the sampling specific to the needs. In general, this is not feasible. Also the envisioned 4-D sampling is not possible with current technology. There are, however, various techniques which provide different sections through this 4-D space with useful resolution in at least some of the dimensions, see Figure 1. Satellites have excellent x-y-t coverage, and sufficient x-y-t resolution for many applications. However, the sampling is provided only at or near the surface, and is restricted to few variables. The ARGO float network provides good sampling in the vertical (profiles), with global coverage and hopefully long (sustained) coverage in time, but has sparse horizontal and temporal resolution despite the large number of platforms (e.g. not eddy-resolving, unable to resolve the timescale of short events, etc). Also the number/types of variables observable with the ARGO system will remain very limited. Fixed (moored) instruments can deliver excellent (probably complete) sampling of the time domain and may have good coverage of the z-dimensions, but can only be installed in a few number of x-y locations. Therefore, a coordinated and deliberate use of several observing techniques often is required to provide the information needed.

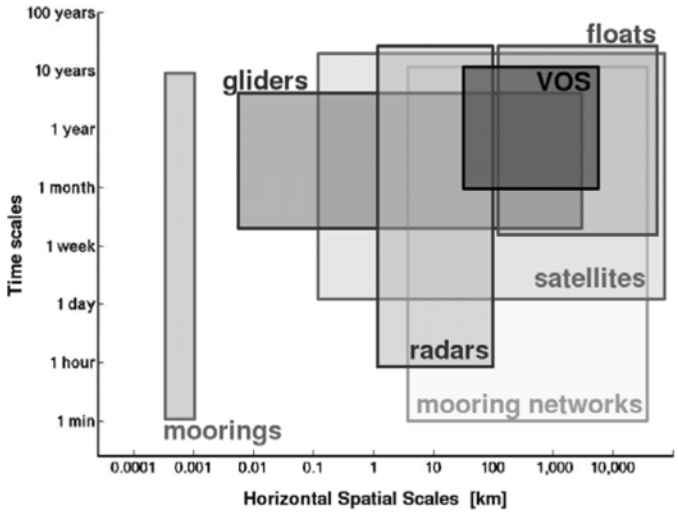

Figure 1. Spatial and temporal resolution (lower side of boxes) and coverage (upper side) of the observing methods discussed in this chapter (after T.Dickey). This representation does not include the vertical dimension, thus giving an imcomplete impression.

In all cases, aliasing in time and space should be a real concern. This results from not resolving the smallest scales of variability, and thus creating the false impression of larger/longer scale variability. A good example was provided by S.Rintoul (pers. comm.) shown in Figure 2. Also scanning satellite sensors which revisit a certain location only every 10-30 days have aliasing problems (tides, diurnal signals, etc).

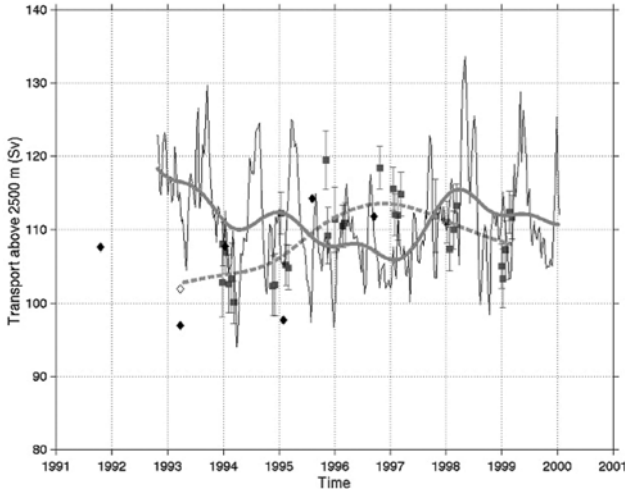

Figure 2. Interannual variability of transport south of Tasmania inferred from sporadic XBT transects (black symbols) and filtered (dashed line). The same from frequent estimates from altimetry (solid thin line) and filtered (heavy solid). The long-term trends are opposite due to aliasing. From S.Rintoul.

2. In-situ techniques and platforms

As also highlighted in the chapter by I. Robinson in this volume, a very tightly integrated approach is needed between remote sensing and in-situ observations, in order to provide the data necessary for the modelling procedures in operational ocenography and ocean forecasting. None of the two methodologies alone can acquire the ocean data required with sufficient accuracy, 4-D coverage and 4-D resolution. This section is organized on a platform basis for clarity, but the guiding principle in all cases will be the way in which in-situ data are indispensable for providing information impossible to obtain with remote sensing, or for complementing, validating and calibrating remote sensing data.

At the end of this section, Table 1 provides an overview of the platforms with their costs, strengths and weaknesses. This can help in guiding through the following sections, comparing the different methods. More importantly, however, the table is meant to emphasize the complementarity between all these elements and technologies.

2.1 Profiling floats

Description:

These are platforms that passively follow the horizontal flow in the ocean interior and periodically rise to the surface for satellite positioning and to collect profile data on the way up. Originally designed to give the current field (e.g. a deep reference flow for geoid estimation), they are presently used more for the T/S profiles they provide. The original rationale was to have a platform that is so cheap and long-lived (while requiring no other infrastructure) that it can be used in large numbers anywhere around the globe. This is still the philosophy, now implemented in the ARGO program, so most of the floats have only standard sensors. Heavy or power-hungry sensors cannot be incorporated. ARGO plans to deploy and sustain such floats on a 3°x3° grid globally, which needs a total of over 3000 instruments (see also the chapter by S. Pouliquen in this volume). The standard drift depth is 1000m in ARGO, but for profiling the floats dive down to 2000m before ascending to the surface. The cycle period is 10 days, and with a targetted capacity of 180 cycles, the floats are designed to last for 4-5 years.

Application:

Profiling floats in the ARGO approach are intended to complement satellite altimetry in two ways. From the latter, anomalies of sea surface height (SSH) can be derived which consists of the steric (dynamic height) contribution of T and S (H_{dyn}) and a reference level pressure (P_{ref}) related to a barotropic flow component. Symbolically (strictly these are different quantities which cannot be added), the contributions are

$$SSH = P_{ref} + H_{dyn} = SSH' + \langle SSH \rangle \quad (1)$$

where <…> is the mean and SSH' are the fluctuations observed by altimetry. Altimetry has good spatial and temporal coverage but cannot determine/ differentiate the

- steric and non-steric components
- mean SSH field (relative to the geoid)
- T and S contributions (spiciness)
- interior structure (vertical distribution) of H_{dyn}

The float profiles of T and S provide the H_{dyn} component globally (i.e. the steric component), as well as the spiciness and the interior structure. The trajectory data provide the absolute flow at a reference level and thus an estimate of the mean P_{ref} field. As a residual in (1) then the mean SSH field can be determined and thus the geoid.

Strengths and weaknesses:

The strength of these platforms is the broad (basin-scale or global) spatial coverage achievable, as in the ARGO program, and the vertical information provided. While at first sight they tend to spread randomly with time, there are regions (divergences, passages) that are impossible to sample. Some sampling biases can exist, like convergences towards regions with larger velocities (giving too high mean flows), Stokes drift in oscillating flows with

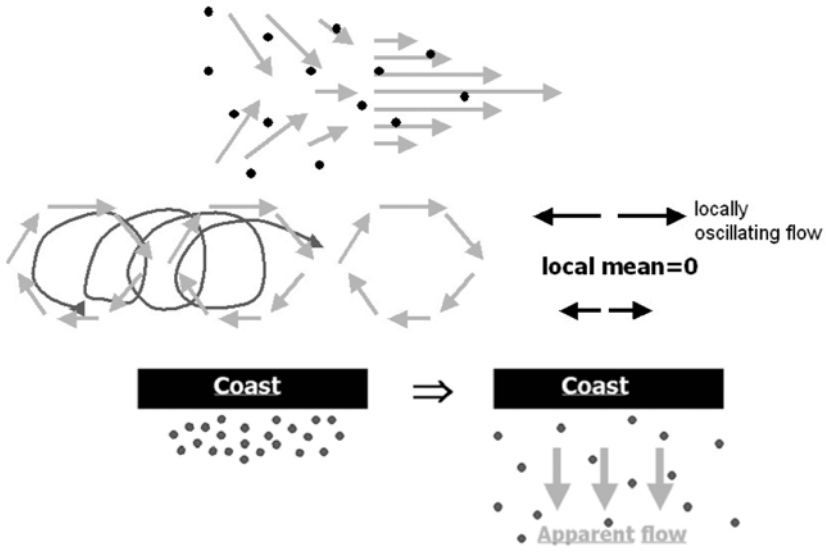

Figure 3. Three types of biases that can occur with lagrangian platforms (floats). Top: convergences accumulating floats in regions of larger flow. Middle: Stokes drift in oscillating flows with spatial gradients. Bottom: Diffusion bias due to spreading in a preferred direction.

spatial gradients, and diffusion bias if high float concentrations spread preferentially in one direction (e.g. near a coast), see Figure 3. The spatial and temporal resolution, as implemented in ARGO, is coarse. Floats are expendable so can not be post-calibrated, thus a sensor drift is difficult to detect or correct.

Further readings: Davis (1991), Davis et al (2001), ARGO website.

2.2 Surface drifters

Description:

Surface drifters are cheap and light-weight platforms that passively follow the horizontal flow at the surface via a drogue/sail at usually 15m depth. The drogue is connected to a small surface float which carries the satellite transmitter and other electronics. All of them measure SST and many also air pressure.

Application:

While profiling float data have the strongest synergy with satellite *altimetry* measurements, sea surface *temperature* (SST) observations from space are best complemented by surface drifters. The chapter by I. Robinson in this volume explains in detail the difficulty in defining and observing the different types of SST. Surface drifters are an important resource for collecting such data.

A ten year long data set now also seems to allow an estimate of the mean geostrophic surface circulation, after subtracting the Ekman component, which can also serve to determine absolute SSH and thus the geoid. There is a global operational drifter program under way, which maintains on the order of 1000 drifters in the ocean.

Strengths and weaknesses:

The strengths and weaknesses of drifters are similar to the ones of floats (section 2.1), but are restricted to the surface. In addition they normally do not measure salinity, yet.

Further readings: Niiler et al (1995), Niiler et al (2003), Global drifter center website.

2.3 Ship sections

Where better horizontal resolution is desired than achievable with floats or drifters, especially on regular transects across ocean basins or across boundary currents, ship sections are currently the best way to obtain this.

2.3.1 Research Vessels

Description:

Research vessels are still the backbone for much of ocean research. Ships are usually needed to deploy heavy or big instrumentation (geophysical equipment, moorings, nets) or to collect samples for chemical and biological analyses. These are very important applications, but the vessels are not suited for routine, frequently repeated, operational observations – they are very expensive and e.g. the one-time survey of the oceans with hydrographic sections under the WOCE program took 10 years! An exception is the use of research vessels as "ships of opportunity", where they are used to collect underway samples as they go to remote and ill-sampled areas for other work. The instrumentation used then are XBTs (expendable, free-falling temperature probes), thermosalinographs analyzing water pumped into the ship from near the surface, and Acoustic Doppler Current Profilers (ADCP) built into the hull of the vessels.

Application:

When using research vessels for operational underway measurements, continuous surface temperature/salinity measurements are usually obtained with the thermosalinograph. These data have value in adding to the in-situ reference data base for surface temperatures, but it is important for operational use that the data are transmitted to shore at least daily. An increasing number of vessels also deploys XBT probes now routinely when steaming on long sections or in transit. This provides more valuable temperature *profiles*. Depending on need, funding, and watch schedule of the crew, the probes are launched 2-4 times per day, and exceptionally every few hours. For ships steaming at 10 knots, this gives a horizontal spacing of 50-200km.

Strengths and weaknesses:

Research vessels are extremely expensive, and most cruises have a length of 2-4 weeks. Thus it is not realistic to use such ships purely for operational routine measurements. When a ship is on the way for another purpose, however, measurements that *do not require stops or extra labour* can be collected nearly for free.

The advantage of these vessels is that they often go to remote areas where no other in-situ observations are available. A drawback is the lack of regularity, since most research ships do not routinely go along the same sections (an exception are Antarctic supply ships).

Further readings: Routine research vessel website.

2.3.2 Volunteer observing ships

Description:

Volunteer observing ships (VOS) are merchant vessels which are willing to collect underway sampling (or deploy instruments) for free as they transit along their most economical paths (which are not always the same, and they do not stop or slow down for measurements). Thus they are useful for performing repeat section measurements on trans-basin paths, with a repeat interval of usually 2-4 weeks. Most measurements are only at the surface of the ocean (sampling water from the engine intake), but this is done for an increasing number of variables. For example complete CO_2 analyzers are now installed on various ships. Depth profiles are limited to temperature, for which the expendable XBT probes are used (normal profile depth is 800m), employing automatic launching system. XCTDs to measure also conductivity are used more rarely because of their high cost.

Application:

The so-called low-density sampling of the GOOS XBT program is carried out on 70 VOS on 26 transects in the three ocean basins at a density of usually 4 profiles per day (note that commercial vessel are usually twice as fast as research ships). These data are used operationally for weather/climate forecasting (e.g. by NCEP) and for climate research. Further VOS sampling is organized by the Ship of Opportunity Program (SOOP). These include high-resolution XBT lines, which sample every 50km in the open ocean and every 10-30km in boundary currents, some ADCP measurements, and pumped surface observations of temperature, salinity, pCO_2, and chlorophyll.

Strengths and weaknesses:

This is a cost-effective methods for collecting data with high resolution along repeated trans-oceanic tracks. For the surface layer, a wide range of variables can be measured now. However, the initial installation of equipment may be difficult, and then there is no guarantee that a ship operator will not change ship routes or destinations. Spatial coverage is limited to commercial ship routes, and subsurface sampling (profiles) is restricted to temperature and usually not deeper than 800m.

Further readings: Smith at al (2001), Upper ocean thermal center website, SOOP website.

2.4 Moorings and fixed platforms

Description:

In many applications, temporal resolution much higher than with satellites, floats, and repeat ship sections is needed, as well as measurement of a wider range of variables. This requires timeseries observations in fixed locations, and for operational purposes a sustained mode of sampling is a prerequisite. This leads to the useage of moored sensors or bottom-mounted systems. The more generic modern expression is "ocean observatories". Sampling is possible, depending on sensors, from minutes to years, and from the surface to the ocean bottom. There are "subsurface" and "surface" moorings, depending on where the top buoy is located.

Moorings can carry heavy sensors and thus observe, in case an autonomous instrument exists, nearly everything. Apart from physical sensors for T, S, currents, there now are optical sensors (for radiation measurements, chlorophyll fluorescence, oxygen), optical plankton counting and video instruments, chemical sensors (analyzers with wet reagants, or samplers), acoustic instruments for zooplankton backscatter or long-range tomography transmissions, and more.

Mooring networks are a special case and provide high time resolution at a set of fixed locations covering an ocean region. For dense networks like the tropical TAO/TRITON array in the Pacific, spatial gradients are sought, while more widely spaced systems sometimes only intend to contrast differences between areas or to occupy different parts of an ocean region.

Application:

Since moorings can only be installed and maintained in a discrete number of selected locations, the rationale normally is to use them in locations with critical ocean processes or in places that are expected to be representative of larger areas of an ocean basin. Examples are water mass formation regions, where the location of the deep mixing process is well known, and where a single mooring with sensors for water mass properties (T, S, etc) and possibly vertical currents (ADCP sensor) is sufficient. Similarly, flows and transports through important straits and passages, like Denmark Strait, the Indonesian Passages, or the Strait of Gibraltar, could be monitored by fixed observatories. Observing the uptake of CO_2 on the global scale is also a crucial type of information, which can be provided by a network of moorings with CO_2 sensors in the major regions of uptake or release by the ocean. The concept of ecological ocean provinces (Longhurst 1995) helps to identify locations which may be representative of larger areas in terms of chlorophyll and nutrient concentrations/distributions, mixed-layer depth, and other aspects. Maintaining observatories in each of these global provinces might enable the detection of variability or regime shifts in the different ecosystems.

In obtaining in-situ chlorophyll data, moorings will become an important complement to satellite chlorophyll estimates, which are very difficult to determine and have an accuracy of 30% in the best of cases. In addition,

moorings are already able to provide other biogeochemical variables, like nutrients and O_2 which are critical for biogeochemical models (see chapter by A. Oschlies in this volume) but are not available from remote sensing.

Strengths and weaknesses:

Moorings are expensive to build and maintain, need a lot of technical effort, and require regular visits by research vessels. Therefore, only a limited number of distinct locations can be monitored by moorings. They have no x-y coverage or resolution, thus normally should be complemented with other techniques. On the other hand, they are ideal for sampling in the time domain, covering many multidisciplinary variables, and measuring in difficult fixed locations (straits, boundary currents). Moored instruments can be re-calibrated so may serve as in-situ reference stations both for satellite data and other types of sensors like floats and drifters.

Further readings: Tupper et al (2000), Dickey et al (2001), Dickey (2003), OceanSITES website.

2.5 Gliders and AUVs

Description:

A new class of platforms are autonomous gliding or self-propelled vehicles. These navigate under water and can be programmed (or "steered") to sample along specific mission tracks. AUVs have propellers, usually not a very long range or endurance (order of days) and need support ships. Gliders on the other hand propel themselves by buoyancy changes and wings, thus they undulate up/down through the ocean. They are still in the prototype stage. Current versions have a limited speed of 20-25cm/s, a depth range of 1000m, and endurance of 6-12 months. Like floats they are very restricted in terms of additional payload mass and energy consumption, but usually carry more sensors than floats. Therefore they have the potential to provide biogeochemical data like fluorescence (for chlorophyll) and other optical measurements in a spatial mode and thus to greatly complement timeseries data from moorings.

Application:

Gliders can be used for repeat transects in remote areas or to complement VOS lines, either on orthogonal tracks or by providing additional variables. Every 2 weeks a glider could cover the equivalent of a 300 km XBT section (though not synoptic, i.e. not a snapshot, but for assimilation into models this makes little difference). Useage under the ice is also imaginable. Apart from running along repeat sections, holding position like a "virtual mooring" is also possible, and even entering a float mode may be feasible soon.

Strengths and weaknesses:

A main limitation of gliders is the maximum current they can stem due to their low speed. While the range is already a few 1000km, this still limits the ability to reach a mission area from shore, to carry out survey work, and return to a base. A very strong point is the flexible useage, both in terms of sampling and sensors, of being able to choose tracks that are defined by science not by merchant ships, of mission type, and the ability to steer the glider from shore.

Further readings: Davis et al (2002), Seaglider website, Spray website.

2.6 Integrating techniques

Description:

A few methods exist which inherently provide spatially integrated information, rather than data at the location of the platform. One is acoustic tomography, which samples the ocean horizontally with sound over large distances. This is usually done between pairs of moorings, which are fixed, but the information extracted from the traveltime of the sound (temperature, current along the path) represents an average over the entire section between the moorings. This technique has been used successfully in a number of experiments.

The other approach exploits the geostrophic relation and the principle that the average (or integrated) geostrophic current can be determined alone from the pressure distribution (as a function of depth) at the endpoints of a section. This is usually calculated from density profiles, traditionally collected with shipboard CTD's, but can now be done with self-recording sensors on a mooring. One thus obtains timeseries of mass transport, integrated over the section between the moorings again. As in the traditional geostrophic method, there is still a reference level problem, since the pressure field determined from the density measurements is relative to a pressure level whose depth and inclination is not known. For this, high-precision bottom pressure measurements are now possible to within a few millimeters of equivalent sea surface elevation, which at least give the *fluctuations* of the pressure gradient at a reference level.

Application:

Tomography is not much used in the "imaging" sense anymore, i.e. trying to extract horizontal mapping resolution from the integrals along the transmission paths. Instead, it is most useful where heat content or currents along a section are of interest (water mass formation regions, straits). Over long ranges it is also sometimes called "thermometry" and can then provide basin-scale temperature changes.

The geostrophic transport monitoring is suitable for timeseries of transports over entire sections. These maybe confined currents (passages), wide boundary currents, or meridional flows across entire ocean basins. In a German CLIVAR application (MOVE project), this was carried out successfully over a 1000km long section.

Strengths and weaknesses:

Acoustic tomography is an expensive technique requiring highly specialized teams and equipment. The niche in providing large-scale integrals has become smaller with the advent of ARGO, but the strengths remain full depth coverage and occupation of specific sections of interest. Both tomography and geostrophic integral techniques require specific geometry and bathymetry, thus cannot be used anywhere. However, they are remote sensing approaches, providing integral information about ocean regions without the need to deploy instruments everywhere.

Further readings: Kanzow et al (2005), Dushaw et al (2001).

2.7 Coastal radars

Description:

Radar installations with typically 50-150km range are able to sense the surface currents in the vicinity of coasts, by analyzing the doppler shift from surface waves which Bragg-scatter the radar signal. Each piece of ocean surface to be sensed needs to be covered by two separate radars. The variables that can be extracted are the very near-surface current vectors, and as a second-order quantity, the wave height. The spatial resolution is 2-3km, and the time resolution typically 1 hour. Shorter-range systems also exist.

Application:

For operational applications like ship routing, prediction of pollutant transport, harmful algal blooms, etc, this is is a method of increasing interest. To date, most installations are only in select locations of specific interest. However, some countries are starting to set up radar networks along entire coastlines. These would contribute to monitoring systems of coastal or near-shore ocean processes.

Strengths and weaknesses:

An advantage of radars is that they are entirely land-based and have useful spatial and temporal resolution. However, they coverage is limited to near-coast, they require elevated terrain for the installations, and can only sense currents at the surface.

Further readings: Essen et al (2000), EuroROSE website.

3. Conclusions

An overview of the most widely used or most promising in-situ observing techniques for operational applications has been provided. The main intention was to emphasize the differences in terms of sampling and capabilities, in order to give appreciation of the complementarity of the approaches. No one method can usually fulfil the observational needs of any operational (or science) application. Table 1 is meant to summarize the main characteristics of the platforms discussed, and to help in guiding to the most appropriate choice of observing means for specific observational requirements. More importantly, however, it is meant to emphasize the complementarity between all these elements and technologies which exist.

It is clear that many observing techniques had to be omitted here. Acoustically tracked floats, electromagnetic methods to sense currents, or inverted echosounders are some of them. They are not, however, used operationally and there seems to be no plan at present to include them in operational systems.

To highlight the sampling and complementarity in space and time, Figure 1 at the beginning of this chapter summarized the spatial and temporal resolution and coverage of the observing methods discussed. While the figure does not do justice to various methods by omitting the depth dimension (where satellites would just provide a single horizontal layer/slice), it is helpful to think in terms of this horizontal and temporal sampling. There seems to be a gap on scales of 10 m - 1 km and on short timescales, but most processes of interest to operational oceanography can be observed with suitable combinations of existing methods.

One aspect that was not addressed above is that of data delivery. Operational systems require in-situ data with minimal delay, usually within one day. Some approaches inherently have a built-in data telemetry capability, like drifters, floats, and gliders. Coastal radars and vessels just need to be equipped with the required transmission systems, which is no problem in principle. For moored or bottom-mounted instruments, and even more so under the ice, data telemetry is not easy. Either seafloor cables need to be available, or surface buoys are required, or telemetry packages that occasionally come to the surface need to be attached to moorings. All these exist or are under development.

The challenge in collecting data for operational applications is to combine the available methods in the most efficient way, in order to provide the observing system – together with remote sensing – that really samples all four dimensions and the variables of interest such that models can make maximum use of them.

Platform/cost	Strengths	Weaknesses
Research vessels ($25,000/day)	- taking samples - deployment of heavy equipment - reach remote areas (VOS-like)	- sparse sampling (operational) - too expensive for operational obs (but needed for servicing of operational installations)
VOS (free)	- high resolution along repeat tracks - many variables (for surface measurements)	- not always where wanted - tracks may change, they don‘t stop - subsurface only for T (800m)
Surface drifters ($3,000)	- global coverage - rapid sampling in time - low-cost, robust technology	- sparse spatial sampling - only surface observations - limited variables (T, air-p, S)
Floats ($15,000)	- global coverage - vertical profiling to mid depth - large numbers since “cheap”	- coarse x,y,t resolution - limited weight/power (sensors) - avoid/quickly leave passages, divergences, places of interest
Moorings ($250,000)	- high time resolution, surface to bottom - many variables possible - difficult locations possible - re-calibrations, referencing	- no x,y resolution - expensive, including the need for ships - large technical effort
Gliders ($70,000)	- good sampling along tracks - free choice of track, can be steered/controlled - small sensor suite feasible	- very slow (20-25cm/s) - limited depth range and variables
Integrals	- integrate over long distances - good time resolution	- expensive - limited variables and places possible
Coastal radars	- good x,y,t resolution - land based	- limited coverage - only currents, waves (surface)

Table 1. Typical costs and tradeoffs of the observing methods discussed in this chapter.

4. Study and discussion questions

The following example applications are meant to motivate discussion and critical evaluation of the observing methods, including remote sensing, for achieving the operational goals. Therefore consider the following needs:

1) Monitoring of water mass formation in specific regions
2) Detection of coastal eddies and their impact on the ecosystem
3) Observation of the outflow through the Strait of Gibraltar
4) Collection of routine observations under the ice in the Arctic and Southern Ocean.

Even though at first sight some of these may appear obvious, it is helpful to consider each approach in the earlier table and diagram, and also remote sensing, and discuss why a certain method may not be suitable or less so than another. In many cases, the requirement of real-time data transmission poses a particular challenge. The problems do not have a unique answer, and the solutions will evolve with the implementation of new technologies.

References (including WWW sites)

Davis, R.E., 1991: Observing the general circulation with floats. *Deep Sea Res.*, 38(Suppl. 1), 531-571.

Davis, R.E., J.T. Sherman and J. Dufour, 2001: Profiling ALACEs and other advances in autonomous subsurface floats. *J. Atmos. Ocean. Tech*., 18, 982-993.

Davis, R.E., C.E. Eriksen, and C.P. Jones, 2002: Autonomous buoyancy-driven underwater gliders. *The Technology and Applications of Autonomous Underwater Vehicles.* G. Griffiths, Ed., Taylor and Francis, London, 324 pp.

Dickey, T., 2003: Emerging ocean observations for interdisciplinary data assimilation systems, J. Mar. Syst., 40-41, 5-48.

Dickey, T., S. Zedler, D. Frye, H. Jannasch, D. Manov, D. Sigurdson, J.D. McNeil, L. Dobeck, X. Yu, T. Gilboy, C. Bravo, S.C. Doney, D.A. Siegel, and N. Nelson, 2001: Physical and biogeochemical variability from hours to years at the Bermuda Testbed Mooring site: June 1994 - March 1998, *Deep Sea Res. II*, 48, 2105-2131.

Dushaw, B. D., G. Bold, C.-S. Chui, J. Colosi, B. Cornuelle, Y. Desaubies, M. Dzieciuch, A. Forbes, F. Gaillard, J. Gould, B. Howe, M. Lawrence, J. Lynch, D. Menemenlis, J. Mercer, P. Mikhaelvsky, W. Munk, I. Nakano, F. Schott, U. Send, R. Spindel, T. Terre, P. Worcester, and C. Wunsch, 2001: Observing the ocean in the 2000's: A strategy for the role of acoustic tomography in ocean climate observation. In: *Koblinsky, C.J., and N.R. Smith, N.R. (Eds),* "Observing the Oceans in the 21st Century". Melbourne. GODAE Project Office, Bureau of Meteorology, 391-418.

Essen, H.-H., K.-W. Gurgel, and T. Schlick, 2000: On the accuracy of current measurements by means of HF radar. *IEEE Journal of Oceanic Engineering*, 25, 4, 472-480.

Fasham, M.J.R., J.L. Sarmiento, R.D. Slater, H.W. Ducklow, and R. Williams, 1993: Ecosystem behaviour at Bermuda Station "S" and OWS "India": a GCM model and observational analysis. *Global Biogeochemical Cycles*, 7, 379-415.

Kanzow,T., U. Send, W. Zenk, A. Chave, and M. Rhein, 2005: Monitoring the integrated deep meridional flow in the tropical North Atlantic: Long-term performance of a geostrophic array. *Deep Sea Res.,* in press.

Longhurst A.R., 1995: Seasonal cycles of pelagic production and consumption. *Prog. Oceanogr.* 36, 77-167.

Niiler, P., A.Sybrandy, K.Bi, P.Poulain, and D.Bitterman, 1995: Measurements of the water-following capability of holey sock and TRISTAR drifters, *Deep Sea Res.*, 42, 1837-1858.

Niiler, P.P., N. Maximenko, and J.C. McWilliams, 2003: Dynamically balanced absolute sea level of the global ocean derived from near-surface velocity observations. *Geoph. Res. Letters*, 30 (22), 2164.

Smith, N.R., , D.E. Harrison, R. Bailey, O. Alvez, T. Delcroix, K. Hanawa, B.Keeley, G. Meyers, B. Molinari, and D.Roemmich, 2001: The role of XBT sampling in the ocean thermal network. In: *Koblinsky, C.J., and N.R. Smith (Eds),* "Observing the Oceans in the 21st Century". Melbourne. GODAE Project Office, Bureau of Meteorology, 248-258.

Tupper, G., I. Waddington, and G.Ferrari, 2000: Surface and subsurface moorings at Woods Hole. In: Proceedings Inmartech 2000, Royal Netherlands Institute for Sea Research (NIOZ), Texel.

Argo website: www.argo.net

EuroROSE website: http://ifmaxp1.ifm.uni-hamburg.de/EuroROSE/index.html

Global drifter center website: http://www.aoml.noaa.gov/phod/dac/gdc.html

OceanSITES webite: http://www.oceansites.org/OceanSITES/index.html

Routine research vessels: www.coriolis.eu.org/english/research-vessels/research-vessel.htm

SOOP website: http://www.brest.ird.fr/soopip/

Seaglider website: http://www.apl.washington.edu/projects/seaglider/summary.html

Spray website: http://spray.ucsd.edu/

Upper ocean thermal center website: http://www.aoml.noaa.gov/goos/uot/

Chapter 8

IN-SITU OBSERVATIONS: OPERATIONAL SYSTEMS AND DATA MANAGEMENT

Sylvie Pouliquen

IFREMER, Centre de Brest, Plouzané, France

Abstract: This paper presents, through existing examples, the main characteristics of operational in-situ observing systems and the data management issues to be addressed for operational oceanography needs. It provides the main characteristics of an operational in situ observing system in comparison with a research one in term of sustainability, coverage, timeliness, implementation issues and international coordination. It highlights the main features that have to be put in place for operational system data management and differences between different architectures that are nowadays operated.

Keywords: In-situ, observing systems, data management, quality control, data formats, ARGO, GOSUD, OceanSites.

1. Introduction

Scientists, fishermen, navigators… have been observing the oceans since the middle of the 19th century for their own needs (to enhance safety, to improve transit time, to understand some phenomena, etc). But this has often been done in an unorganized way, shared only among small communities, measured over limited areas and periods of time: a lot of data have thus been lost or are too incomplete to be used by the community nowadays.

Because it has been demonstrated that ocean and atmosphere behaviour are clearly linked together, it is mandatory to observe and understand the oceans the same way it has been done for the atmosphere since the 20th century. That is why individuals, research groups, and nations have started to work together to overcome the technical and logistical challenges associated with carrying out joint routine measurements of the global-ocean.

While satellites are providing a global view of the surface of the ocean, it is important to set up in-situ systems to monitor their interior (e.g. Send this volume). Basically, the following are needed:

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 207-227.

- Autonomous instruments (moorings, drifters, profiling floats, gliders, etc) to monitor on long period of times
- Regular ship measurements to monitor long repeat sections,
- In order to have all these data available for operational models: a well-designed and robust observing system, good communication to shore to deliver data rapidly,
- Real time operational data centres,
- Suitable data protocols to distribute data to operational centres in a timely way,
- International cooperation to achieve a global coverage, set up an adequate system and maintain it in the long term.

2. Essential features of operational oceanography systems

The goal of operational oceanography is to provide routine ocean nowcasts and forecasts and analysis on timescales of days to seasons, from global to regional and coastal regions. To address the operational oceanography needs, in-situ observing systems must comply with the following requirements.

2.1 Coverage

The observing systems to be put in place are different depending on the area and the phenomena to be sampled. We usually sort observing system into 3 categories:

- Global: System designed to provide data all over the ocean (e.g ARGO for general circulation). Such a system can only be built at the international level and is complementary to observations made from space. It is built to resolve climate scale phenomena with sufficient resolution and accuracy and provides systematic upper ocean observations of a limited number of parameters (temperature, salinity, …) on a time scale from 10 days to 1 month. International collaboration is the key factor for the success of such a network because none of the countries is able to cover the globe alone, but each country has to set up elements compatible and guaranteed on the long term for their contribution to the system.

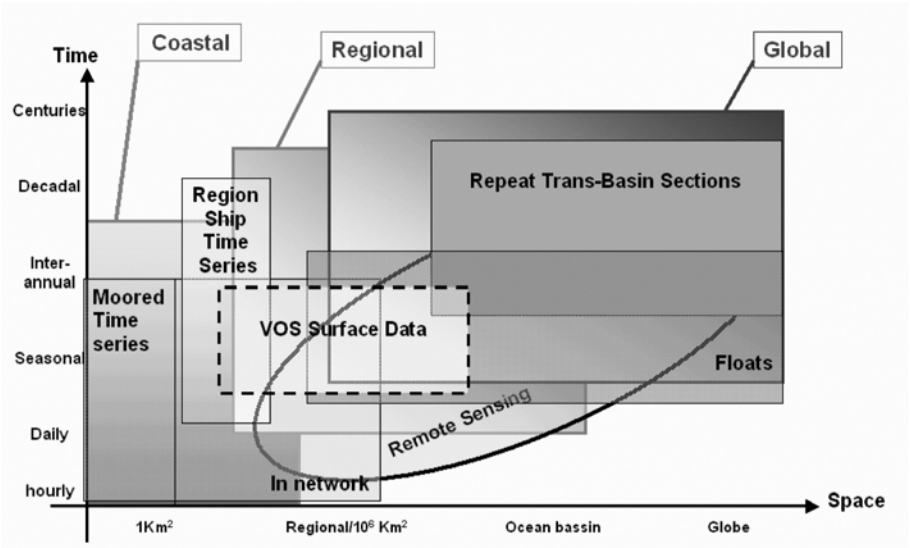

Figure 1. Platforms to use according to phenomena to sample and the kind of network to be set up (Global, Regional, Coastal).

- Regional: System designed to provide data in a specific area to monitor specific phenomena (e.g., TAO/TRITON/PIRATA Array for El Nino detection, Artic buoy network for Ice monitoring, etc). Generally it is set up in collaboration with few countries (less than 10) and the number of parameters is more important (between 10 and 20), including ocean (both physical and bio-chemical) and meteorological measurements. The time sampling is often higher rate: from hours to days.
- Coastal: These observing systems are usually set up at the national level to answer very specific questions such as coastal monitoring of the water quality or wind/wave/tide monitoring in harbour areas, etc. There is very poor collaboration among countries and these data are often used exclusively by the coastal models that have led to setting up the system. The technical issues to be solved are much more complicated such as bio-fouling (micro-organisms growing on the sensors and perturbating the measurements), interference by fishermen or other ships in area. A lot of technological work is under development in this field especially to set up cabled systems linked to shore with very high-speed networks.

2.2 Timeliness

What does real time mean for operational oceanography? The main criterion is to define the delay between measurements and assimilation beyond which the measurement adds nothing to the performance of the model. There is no unique answer: this depends on the type of models, the variables that are assimilated, the forecast product and the application for which it is produced. For instance, assimilated information of deep ocean temperature and salinity will persist within an ocean global circulation model for weeks or months and so a delay of several days in supplying data can be acceptable. On the other hand, oceans mixed layers vary on more rapid timescales in response to the diurnal heating and to storms. The impact of such data will probably not persist more than 3-5 days after assimilation, so measurements are needed within a day. As a compromise, real time for operational oceanography generally means availability within 1 or 2 days from acquisition, to allow data centres to better qualify the data even if it takes a bit more time. For climate applications larger delays are acceptable, but the length of the observation period is critical.

2.3 Agreed procedures and standards

Operational models use a wide variety of data from a diverse sources including buoys, drifters, ARGO floats, regional ships of opportunity, coastal observatories and even isolated local measurements made either by nations or scientists, as long as the data are easily available and quality-controlled in a timely way. Sea observations are very expensive and no country is able to sustain alone, the network needed by operational oceanography at the global level. Moreover it is important to design a system able to serve different communities: e.g. research, climate and operational communities. Therefore, an international coordination is needed.

In 1950, the meteorological community has set up an organisation WMO (World Meteorological Organisation) to organize this partnership for the meteorological needs. Under the auspice of Unesco, IOC(Intergovernmental Oceanographic Commission) has played an essential role in defining measurement standards and formats. The JCOMM (Joint WMO/IOC commission for operational oceanography and marine meteorology) has been set up to strengthen the role of WMO and IOC in the field of ocean and marine meteorology. It is involved in the main observing systems used nowadays by operational models:

- Surface data: DBCP (Drifters), VOS (Voluntary Observing ships),

- Sub-surface: ARGO (Profiling floats), TAO, GTSPP (Global Temperature and Salinity Pilot Project)
- Sea-Level: GLOSS (Sea Level)

Being able to collect and share the acquired data and distribute them to the user community requires significant work of normalisation/coordination on data collection and format (from metadata to profile and timeserie datasets), on quality control procedures, as well as on networking organisation to make these data circulate efficiently. Several concurrent attempts of normalization for metadata description (ISO 19115, GXML, XML-Marine, etc.) or data format and access systems are underway, both at national and international levels, but there is still no convergence towards a unique general agreement.

3. Implementation issues

When an observing system, often set up and maintained by scientific teams, moves to operational status there are some requirements that need to be fulfilled.

3.1 Sustainability

First, an operational system is sustained in different ways. This regards funding of course, as they are often expensive networks: new funding mechanisms have to be set up coming from sources other than R&D. Not all countries are organized in such a way that a transition to operational is easy: for example it is the case between NSF or NASA and NOAA in the USA, between ESA and Eumetsat for earth observations in Europe. Systems must be sustained also in terms of the operation: this goes from deployment planning, at-sea servicing (this requires ship and engineering teams to perform these activities), to data processing that has to move from R&D laboratories to operational data centres who are committed to do such tasks in the long-term. It is not always easy to find the institutions that are, in each contributing country, mandated or capable or willing to perform these tasks.

3.2 System maintenance

Second, work to maintain and operate such a network has to be coordinated at the international level with a clearly identified Project Office. This Project Office interacts with the contributing countries to update the implementation plans and secure the fundings. It coordinates the national

activities with an internationally agreed framework. It interacts with other international bodies to integrate this system in a wider perspective.

3.3 Data management

Finally data processing and distribution must be designed properly to be able to deliver the data in time for operational use. First, data have to be publicly available in real-time for forecasting activities, and within a few months for re-analysis purposes. This is a revolution in a scientific community where scientists have kept data private for years until they publish and sometimes forever. This is an important data policy element to be solved by the funding agencies at national and international levels. Second is the organisation of the data flow among the different contributors in order to have an efficient data management network able to answer the operational needs listed above. For a long time, data management aspects have been neglected in projects and a too small funding was devoted to this activity both for in-situ and satellite data processing. With the arrival of operational ocean systems, the question has started to be crucial and examples like WOCE have shown that it was very energy demanding to get integrated quality-controlled data sets when it is not organized from the beginning. It is now clear that operational observing systems have to be processed by professional data centres that are sustained in the long term, that distribution has to be tailored to fulfil operational user needs

All the above has lead to the fact that attached to an observing system, there must be an effective management structure to address the implementation, coordination, data management, advertising and funding issues.

4. Prime examples of observing systems

4.1 ARGO project

To monitor and understand ocean circulation and water mass characterization on a global scale, systematic observation of temperature and salinity are essential. In the 90's, during the WOCE program, a new instrument was developed: the autonomous profiling float. Now the technology has become mature enough to start implementing an ambitious program that would deploy a large number of these instruments to cover the global ocean: the ARGO program was born. It aims to deploy and maintain an array of 3000 autonomous floats, one per 3°x3° box, measuring temperature and salinity from 2000m to the surface every 10 days for 3 to 5

years. Assimilated in models together with Sea Surface Height Anomalies from altimetry, they have become an essential network for operational models. This program started from an initiative of a group of scientists who were convinced of the importance of such a network. It was and is still partially funded on research money but a lot of work is done at the inter-governmental level to find funds to sustain such a network. A float costs about 15.000$, so the setting up of the array will cost about 50.000.000$ and about 10.000.000$ is needed each year to maintain it (700 new floats each year to replace the dead ones). These numbers do not include any additional cost to deploy these floats as the deployments are often done through free opportunities.

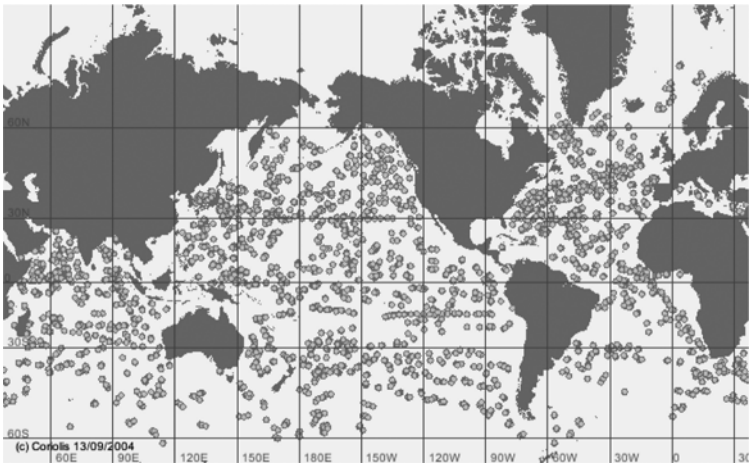

Figure 2. ARGO coverage in mid-September 2004: 1366 active platforms

In 2004 about 40 % of the network is deployed with a good coverage of the northern hemisphere and more work is to be done in the southern ocean. The keys to the success include:

- First an efficient coordination at the implementation level: deployment plans are consolidated by ocean basin to achieve uniform coverage of the float array. Good collaboration has also been set up to facilitate these deployment activities among the countries involved.
- A collaboration at the scientific and technological levels to improve the quality of the instruments, to detect deficiencies in time to avoid to deploy platforms when technological problems have been detected, and collaborative scientific work to develop delayed mode quality control methods to be used by the ARGO community.

- Finally an efficient data management system able to distribute the ARGO data in real-time within 24h from acquisition both on GTS (Global Telecommunication System used by the meteorological agencies) for meteorological community and in FTP for other operational users. This system is based on collaborative work between national data centres that process the data from the float deployed by their country, or partner countries, and a centralized data distribution through two Global Data Centres (one in CORIOLIS/France and one in FNMOC/USA). The architecture of this data management network will be presented at §5.1.1. Since 2004, the Data Management team is putting into operation the delayed mode procedures developed by the Science team.

The real challenges are now to secure funds on an operational and sustained budget to maintain this observing system. It is also to improve the technology to increase the lifetime of the platforms as well as their ability to survive in dangerous area such as partially ice covered regions.

4.2 SOO/VOS and GOSUD: Surface data

Merchant vessels are doing long ocean transects on regular basis and are good platforms to implement repetitive measurements. On the other hand, research vessels frequently traverse the oceans on routes where few other in-situ ocean observations are available. As such, they represent a natural and cost-effective mechanism to deliver routine oceanographic data along their way. These data include temperature and salinity at the surface as well as currents. Sea surface salinity (SSS) is an important parameter that is not yet measured from space.

In 2000, the GOSUD project was set up, under the IOC umbrella, as an end-to-end system for data collected from ships along their cross-ocean tracks. The goal of GOSUD is to develop and implement a data system for ocean surface data, to acquire and manage these data and to provide a mechanism to integrate with other types of data collected in the world oceans. It is complementary of the SOOP/VOS projects that under JCOMM umbrella that organize the data collection from Voluntary Observing Vessels or Ships of Opportunity. The project seeks to organize underway-surface data that are currently collected and to work with data collectors to improve data collection. These data, complementary to ARGO and OceanSites (see figure 3), will be one of the major ground truths for the calibration of the Salinity satellites SMOS and Aquarius.

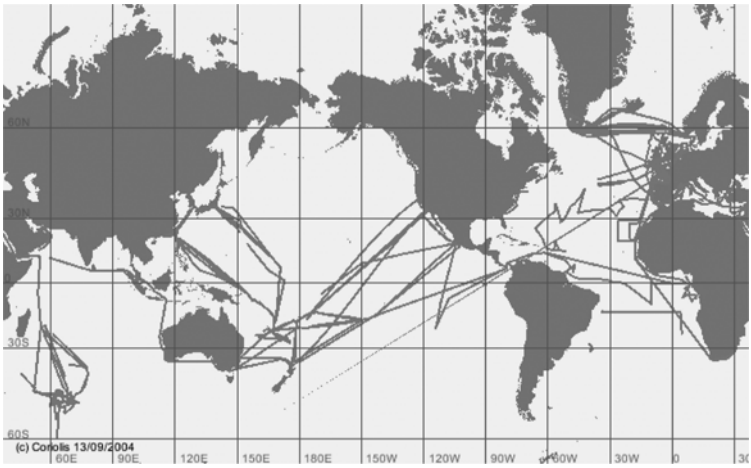

Figure 3. SSS data available for one year at CORIOLIS GDAC from 14 different vessels

In contrast to ARGO, GOSUD is not dealing with the implementation issues, which are handled by SOOP/VOS projects or national initiative (such as CORIOLIS for French research vessels). Moreover, it is building upon existing data centres that have to harmonize their quality control processing and coordinate the data distribution to ease the access to these data. The strategy used for GOSUD data distribution is similar to ARGO with distributed national data centres and two global data centres that act as "one stop shopping" points for users. 90% of the available data are distributed by Global Data Centers, as shown in figure 3.

4.3 OceanSITES

Another and complementary way to sample the space and time variability in a routine and sustained mode is to collect timeseries information at fixed locations in the ocean. The measured parameters are physical (temperature, salinity, current, etc), biochemical (oxygen, nutrients, fluorescence, carbon dioxid, etc) and atmospheric (air temperature and humidity, wind speed, etc).

In order to complement the good spatial coverage provided by ARGO and GOSUD, an international pilot project is under way for a global network of timeseries observatories, called OceanSITES. It plans to coordinate and implement a system of multidisciplinary deep-ocean sites, where sustained and publicly available data will be collected in a timeseries mode. A goal is to telemeter data in real-time (where feasible), and to interface and form

synergies with the developing US OOI initiative. An important factor in turning these measurements into an operational system will be the harmonization, integration, and dissemination of the data collected. This effort is under way.

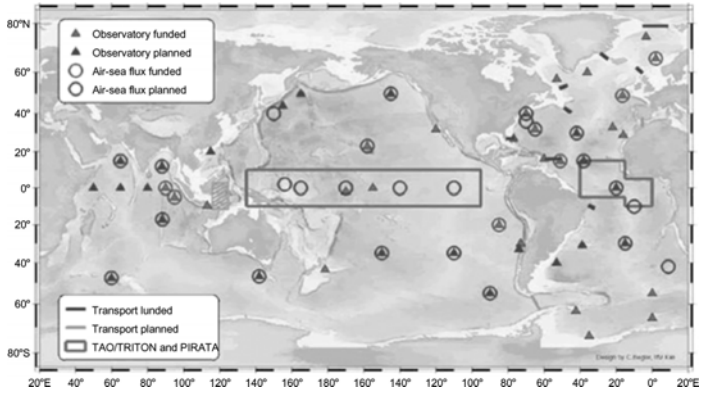

Figure 4. OceanSITES map

The current and planned state of the system is shown in figure 4. New sites have come online in the past 5 years, and an increased awareness of the importance of timeseries data has been created since OceanObs99. Two data centres are under development and a draft timeseries data format has been formulated.

At present only the surface met and TAO/TRITON/PIRATA data are used operationally, but the hope is that other data will be used for ocean forecasting once they are routinely available, once models are able to better assimilate point or integral timeseries data, and once more biogeochemical models are run operationally.

4.4 Comparison of these three systems

If we look at the criteria cited at the beginning of this paper, it is interesting to see how these three networks comply with these requirements.

	ARGO	GOSUD	OceanSITES
Sustainability	Only small part of it is sustained	Part on operational funds, part on R&D	R&D funding

Coverage	Global Network homogeneous coverage	Global to regional network. Good trans-basin coverage	Global network but very sparse coverage
Timeliness	Operational within 1 day for 85% of the data Used by operational models within Godae	From 1-2 days for vessels that have realtime transmission to month for the others. Data used for validation purposes at the moment	Operational for TAO/TRITON/PIRATA array that is used in operational models Mostly R&D for the rest of the sites because data not easily available.
International coordination	Well organised both for implementation and data management	Implementation organized at national level for implementation (France, USA, etc.). Starting to be well organized on data management level	International organisation is trying to be organized but it's hard to achieve.

5. Data management

At present, there is no consensus on data management and communication strategy for effectively integrating the wide variety of complex marine environmental measurements and observations across disciplines, institutions, and temporal and spatial scales. Data are obtained by diverse means (ships, drifters, floats, moorings, seafloor observatories, etc.), they come in very different forms, from a single variable measured in a single point to multivariate, four dimensional collections of data, that can represent from a few bytes a day to gigabytes

Even if an in-situ observing system were to make great measurements in a sustained way, if the data are not available easily to the operational users, they will not be used because they will not meet the operational modellers basic requirements: a data system for operational oceanography must provide quality controlled data, in a timely way, on a regular basis, according to procedures that are clearly documented and evolve upon common agreed decisions between user and provider.

There are three main characteristics for a data management system:

1. Its architecture
2. The quality control procedures
3. Data format and metadata attached to the data

5.1 System architectures

A data management system is designed according to the type of data handled (images/profiles/timeseries/kilobytes versus gigabytes, etc), the users access needs (individual measurements, geographical assess, integrated datasets, etc), the level of integration needed, etc.

In the past decade, with the improvement of the computer technology, the internet revolution, the increase of network speed and capacity, data management systems have been progressively moving from centralized to distributed systems. Two main architectures are nowadays commonly used:

- Distributed processing and centralized distribution: data are processed in different places and are than copied in a single place for distribution to users.
- Distributed processing and distribution: data are processed in different paces and stay where they are. To ease user access a virtual WWW portal is implemented that use networking techniques to find the data that fit the user needs.

Each system has its advantages and drawbacks, depending on the type of datasets to distribute and the contributors to the network. These different architectures will now be quickly described through examples operating at present.

5.1.1 ARGO data system: Distributed processing and centralized distribution

Within the ARGO data system, the float data processing is distributed among the contributing national data centres. They feed two global data centres (GDACs) automatically with the latest version of their float profiles, trajectories and metadata. Both GDACs are updated simultaneously to ensure consistency between the two datasets. They synchronise their holdings each night [in case a DAC (Data Assembly Centre) has updated one GDAC and not the other one].

Individual agencies (some acting on behalf of several countries) assemble the data collected from the communications system and carry out the initial processing of the data. Each file is under the responsibility of a single DAC (i.e. the data provider) who guarantees the quality and integrity of the data.

Data exchanges between DACs and GDAC are performed using a common data format. The main objective is for the users to access a unique data source (in this case, we have two servers for reliability/redundancy). A central website provides an extensive set of tools to query, retrieve, plot and compare the profiling float data dynamically. They also provide an FTP access for easy automatic data retrieval by users.

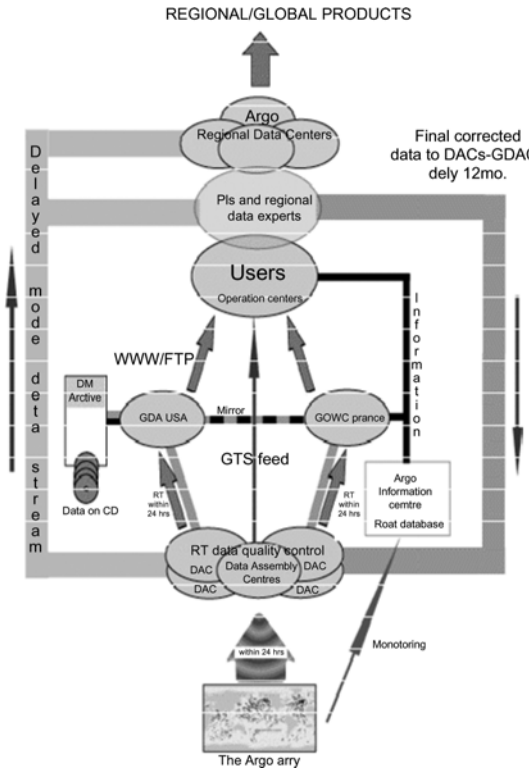

Figure 5. ARGO data management flowchart. On the right, the real time data stream; on the lef the delayed mode data stream.

In the ARGO data flow, there are two loops. One, in real time, on the right on figure 5: DAC qualify data in real time (see §5.2.1) in a semi-automated way and feed two GDACs at the same time that good measurements are put on GTS (Global Telecommunication System used by all meteorological offices). The second loop is in delayed mode, on the left on the figure 5, within one year, data are scientifically quality-controlled (see §5.2.2) and eventually corrected by scientists before being sent again to GDACs by the DACs in charge of these floats processing.

The advantages of such a system are:

- One stop shopping for the users where they get the best available data for ARGO in an unique format
- Data discovery and sub-setting tools are easy to implement as all the data are in the same place
- A robust system, as the probability that both GDACs fail is very small
- Easy to guaranty a quality of service in data delivery because GDAC have the control of all the elements in-house

The disadvantages are:

- Data are moved around the network and must rely on the "professionalism" of the DACs involved in the system to be sure that GDACs have the best profiles available.
- Additional work at DAC level to convert their data from their home format to the ARGO format. This may be hard to do for small entities.
- Data format used for data exchange cannot evolve easily as it requires coordination among all actors before implementation. Since users, especially operational ones, do not like format changes it is not such a big problem.
- If only one main server is set up than the system is fragile. Setting up a mirroring system can over pass this problem with additional synchronisation mechanisms.

5.1.2 Ocean US: Distributed data processing and distribution

The USA are developing an Integrated Ocean Observing System (IOOS) ranging from global to regional to coastal. The purpose is to integrate existing and planned observing systems that address both research and operational needs. Considering the diversity of actors and of parameters involved, this system must be a cooperative integration of independent systems that will continue their missions independently while participating in an integrated data system.

It is clear that in such a system the data processing is distributed and the data stay on physically distributed repositories, some containing huge amounts of data. The user connecting to the Ocean.US website will be able to query for data without knowing where they physically reside.

The key elements of such a system are the metadata management, the data discovery system and the data transport protocols.

- **Metadata management**: Metadata describes the data. Certain classes of metadata (variable names, units, coordinates, etc.) are mandatory to any utilization of the data, and must be tightly bound to data transport as an integral part of the delivery protocols. Other types of information, such as descriptions of measurement and analysis techniques, help to place the data in context and are essential to the overall understanding and usefulness. To be able to share data among a network it is mandatory to have a common vocabulary. Some international groups are working together to build such norms: FGDC and ISO19115 are the most common for geospatial data. As a lot of system pre-exist to Ocean US, it is mandatory to develop translation mechanism to build metadata catalogues that will be used by the Data discovery system.

- **Data discovery**: it is the way to locate data of interest for the user. This search is done by scanning the metadata catalogue. Depending on the possibilities that the system wants to offer the user, the metadata data stored in the catalogues can be more or less precise. This use of metadata is comparable to the indexing of catalogue records within a library to help users to locate books of interest. The common data discovery systems typically allow selecting the available data for a set of parameters, on a geographical area, within a period of time. In future, "data mining" techniques will offer search on semantic criteria ("I want a cloud free AVHRR image of SST over this area in March 2004 together with SST from drifters acquired in same area at same time").
- **Data transport protocols**: these are protocol between a user or a system who wants data and a data repository that stores the data. It is in this field that significant improvements have been made with Internet revolution and the increase in network speed. These protocols are mainly based on available technologies ("web services", cgi, scripts, etc) based on current transfert protocols (HTTP, FTP, etc). Each data provider needs to serve its data and metadata through a common access interface, which can be achieved with existing softwares such as OpeNDap (alias DODS) Live Access software, etc. Direct access to these interfaces may require also specific software or libraries for the user. Although the datasets are distributed through various nodes within the system, setting up a centralized query system that will redirect the user requests to the relevant node can hide this to users.

The advantages of such system are:

- Optimisation of the resources (network, CPU, Memory, etc) among the contributors,
- Data stay where they are generated preventing non compatible duplicates among the network
- Built on internationally agreed standards that guaranty its efficiency in the long term and its adaptability because it will benefit from international shared developments.

The disadvantages are:

- The system is not easy to set up because it needs a lot of international coordination, especially for metadata.
- Even more work for small contributors because it requires important computer expertise
- It can be unreliable if some data providers cannot guaranty data service on the long term. To be reliable such a system must rely on sustained data centres.

5.2 Quality control procedures

These procedures have to be adapted to the allowed delay of the delivery. In real-time, most of these QC are made automatically and only outliers are rejected. In delayed mode, more scientific expertise is applied to the data and error estimation can be provided with the data.

Data quality control is a fundamental component of any ocean data assimilation system because accepting erroneous data can cause incorrect forecast, but rejecting extreme data can also lead to erroneous forecast by missing important events or anomalous features.

The challenge of quality control is to check the input data against a pre-established "ground truth". But who really knows this truth when we know that the ocean varies in time and space, but also that no instrument gives an exact value of any parameter but only an estimation of the "truth" within some error bars.

For operational oceanography, other problems must be solved. First, the forecast requires quality-controlled data within one day. This means that only automated or semi-automated quality control procedures can be applied. Second, most of the data are processed by different actors, but used all together by operational models: this implies a clear documentation of the quality control procedures, an homogenisation of the quality flags, a reliability of different actors in applying these rules. Third, for re-analysis

purpose, the models need better QC'd data for which methods employing scientific expertise are used to correct the data (drift and offset) and to provide error estimates of the corrections. ARGO quality control procedures will be discussed to highlight the different aspects.

5.2.1 Real-time quality control procedures for ARGO

Because of the requirement for delivering data to users within 24 hours of the float reaching the surface, the quality control procedures on the real-time data are limited and automatic. 16 automatic tests divided in 4 categories:

- Gross error tests: date, position, float speed at drift, temperature, Salinity
- Profile coherence: decrease of the pressure, spike detection, excess gradient between two points, density inversion, constant value or overflow for T or S
- Coherence between profiles: jump or big drift in temperature or salinity between two cycles (see figure 7)
- Grey List: For the float in this list, all profiles must be checked by an operator because their behaviour is "strange"

5.2.2 Delayed mode quality procedure for ARGO

The free-moving nature of profiling floats means that most float measurements are without accompanying *in situ* "ground truth" values for absolute calibration (such as those afforded by shipboard CTD measurements). In general pressure sensors are regarded as good even if time drift may be possible; no agreed method exist yet for ARGO but the impact of pressure drift is not negligible: 5 dbar will result in a salinity drift of 0.003psu. Temperature sensors perform pretty well and similar method could be applied to detect temperature drifts.

ARGO salinity delayed-mode procedures rely on statistical methods for detecting artificial trends in float salinity measurements. However, since the ocean has inherent spatial and temporal variability, ARGO delayed-mode quality control is accurate only to within the associated statistical uncertainties.

Using 2-stage objective mapping methods, salinity data mapped from a historical database of existing profiles can be compared to float measurements. Careful analysis of the spatial and temporal scales of the mapping gives realistic confidence levels for the mapped values. A weighted average in the vertical (giving more weight to stable water masses) results in a single salinity offset for each float profile, as compared with the mapped

data. Looking at the trend of these residuals allows detection of a sensor offset or a drift and quantification within error bars.

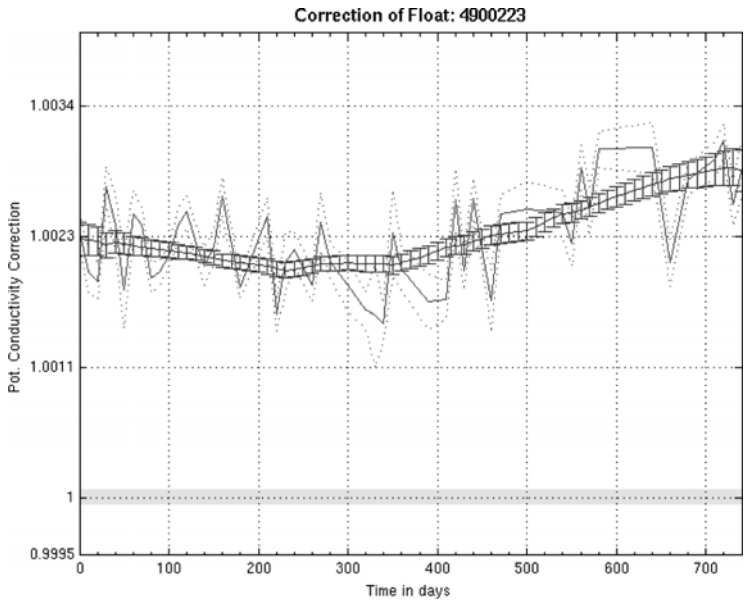

Figure 8. A float with an initial offset and which started to drift after one year at sea. The black line corresponds to individual cycle calculated corrections, The bars correspond to the proposed correction calculated by linear fit on a 6 month sliding window. When the proposed correction stays within the grey box limits (+/- 0.01PSU), no correction is applied.

Another statistical method also used to estimate sensor drift consist of calculating weekly analysis with all the available QC'd profiles coming from CTD, moorings, floats and monitoring the error residual for each float over time both in temperature and salinity by averaging these residuals on a number of levels. Such a method combines three methods: reference to climatology and history of the float as before but also collocation with other neighbouring floats. As it is unlikely that all floats in same area drift the same way, this method should help to QC float data even in areas where the climatology is poor.

5.3 Data formats

Data must be preserved in such a manner that they will still be useful in the future when the Pi that acquired the data may have moved somewhere else. They must also be distributed in a way that a user can easily merge it

with other datasets relevant for his application. They must help to find the data among the network (data catalogues). That is the purpose of defining correctly distribution data format as well as the metadata (data on the data) that need to be preserved for future processing.

Data format have always been a nightmare both for users and data managers and they are both dreaming of the "Esperanto" of data format. Computer technology has improved a lot in the past decade and we are slowly moving from ASCII format (easy to use by human eyes but not for softwares), to binary format (easy for software but not shareable among platforms (Windows, Unix, etc), and self-descriptive, multiplatform formats (Netcdf, Hdf, etc) that allow more flexibility in sharing data among a network and are read by all softwares that are commonly used by scientists.

Depending on who is using the oceanographic data, the information stored in a dataset can be more or less precise. When a scientist is using data that he has acquired himself on a cruise, he has a lot of additional information (often in his head) and he is mainly interested by the measurements themselves. When he starts to share with other persons from his laboratory he has to tell them how he took the measurements, from which platform, what the sea-state was that day, what are the corrections he applied on the raw data, etc in order for his colleagues to use the data properly and understand differences with other datasets. When these data are made available to a larger community the number of necessary additional information, to be stored with the data themselves, increase, especially when climatological or long-run re-analysis are some of the targeted applications. This is why nowadays a lot of metadata are attached to any data shared among a community.

One important point for metadata is to identify a common vocabulary to record most of these information. This is pretty easy to achieve for a specific community such as ARGO, but it starts to be a bit more difficult when we want to address multidisciplinary datasets such as mooring data. To help community in this area some metadata standards are emerging for the marine community with Marine XML under ICES/IOC umbrella and ISO19115 norm.

Another important point in data format is to keep, together with the data, the history of the processing and corrections that have been applied to it. This is the purpose to the history-records that track what happened and allow going back to data centres to ask for a previous version if a user wants to perform his own processing from an earlier stage.

6. Conclusion

This paper has shown that the expectations regarding in-situ observing systems are very high and that they are not easy to set-up: in-situ observations are very expensive, diverse and made by laboratories all around the world. Some pre-operational systems, such as the TAO/TRITON/PIRATA array or ARGO float program, are managing to comply with some of the operational oceanography requirements which are sustainability in time, adequate coverage timeliness of data delivery, coordination both at implementation and data management level.

This paper has also addressed some issues related to data management such as the different data distribution architecture, the necessity of common agreed quality control procedures both in real time and delayed mode and the importance of data and metadata standardisation if we want to be able to share efficiently these data among the network.

References (including WWW sites)

Böhme L., 2003: Quality Control of Profiling Float Data in the Subpolar North Atlantic. Diploma thesis, Christian-Albrechts-Universität Kiel.

Briscoe, M., et al., 2001: Round Table 3: Access to oceanographic data, GODAE, Observing the ocean in the 21st century, p 419.

IGOS, 2004: A Strategy to Realise a Coordinated System of Integrated Carbon Cycle Observations.

Le Traon, P.Y., et al., 2001: Operational Oceanography and Prediction: a GODAE perspective, GODAE, Observing the ocean in the 21st century, p 529.

Ocean US, 2004: Data Management and Communication Plan for Research and Operational Integrated Ocean Observing Systems, http://www.dmac.oceans.us.

Robinson, I., et al, 2003: Observational requirements for inputs to European Ocean Forecasting System Models, Mersea-Strand1 EU project.

Robinson, I., et al, 2004: Measurement Challenges for European Ocean Observing System, Mersea-Strand1 EU project.

Roemmich, D., et al., 2001: The Global Array of Profiling Floats, GODAE, Observing the ocean in the 21st century, p 248.

Roemmich, D., S. Riser, R. Davis, and Y. Desaubies, 2004: Autonomous profiling floats: Workhorse for broad scale ocean observations. *Marine Technology Society Journal,* **38**, 31-39.

Send, U., et al., 2001: Oceanographic Time Series Observatories, GODAE, Observing the ocean in the 21st century, p 376.

Smith, N.R., and C.J Koblinsky, 2001: The Ocean Observing System for the 21st Century: a Concensus Statement, GODAE, Observing the ocean in the 21st century, p 1.

Wong. A.P.S., G.C. Johnson and W.B. Owens, 2003: Delayed-Mode Calibration of Autonomous CTD Profiling Float salinity Data by Theta-S Climatology. *J. Atmos. Oceanic Technol.*, **20**, 308-318.

ARGO WWW site: www.argo.net

CORIOLIS WWW site: www.coriolis.eu.org

Gosud WWW site: http://www.ifremer.fr/sismer/program/gosud/

OceanSITES www site: http://www.oceansites.org/OceanSITES/index.html

GODAE WWW site: http://www.bom.gov.au/bmrc/ocean/GODAE/

JCOMM WWW site: http://www.wmo.ch/web/aom/marprog/index.htm

JCOMMOPS WWW site: http://w4.jcommops.org/cgi-bin/WebObjects/JCOMMOPS

Chapter 9

SURFACE FLUXES FOR PRACTITIONERS OF GLOBAL OCEAN DATA ASSIMILATION

William B. Large
National Center for Atmospheric Research, Boulder, Colorado, USA

Abstract The ability of available ocean surface fluxes to meet the demands of the Global Ocean Data Assimilation Experiment (GODAE) for global, near real time, fields of known uncertainty is examined. Surface flux problems that are discussed in detail include the lack of direct surface measurements to serve as a standard, the difference between fluxes measured at height above the sea and the desired surface fluxes, the complications posed by the need for ocean-ice fluxes, and the large number of global fields required to describe the fluxes. The formulation of the air-sea, ocean-ice and air-ice fluxes of momentum, heat and freshwater, in terms of these fields is detailed from the measurements (including satellite based flux estimates) to the parameterizations. Air-ice fluxes are included to cover the possibility of coupling a sea-ice model within the data assimilation system. The position that there is no one set of flux products that represents the best possible choice for GODAE in all regions and all components is adopted. An alternative merger of a variety of different datasets is described along with objective corrections based on regional and/or short term observations, and ocean model behavior. A flux climatology based on these datasets and observed sea surface temperature is presented as the mean and variability from the seasonal to inter-annual, that GODA flux products should strive to reproduce. The necessary condition of near zero net global heat and freshwater climatological fluxes is demonstrated.

Keywords: Ocean flux measurement, parameterization, fields, climatology.

1. Introduction

Global Ocean Data Assimilation Experiment (GODAE) places rigorous demands on the surface flux forcing, not all of which can be entirely met. Most obvious, is the need for global coverage, including the poorly

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 229-270.

sampled Southern Ocean and ice covered polar seas. Another challenge is proper specification of uncertainties. Also, the requirement for "near real time" fluxes precludes the use of some data, and makes this application very different from its cousin; Ocean Reanalysis. In both cases, flux variability should be properly represented across all the important ocean time scales; from the 12-hour polar inertial period, to the decadal, but GODAE may place a higher premium on time scales comparable to an assimilation cycle, such as the diurnal solar and three to seven day synoptic cycles.

Of primary importance are the surface fluxes of heat, Q, freshwater, F, and momentum, $\vec{\tau}$, with components τ_λ in the zonal and τ_ϕ in the meridional directions. Nearly all aspects of this rather complicated subject are comprehensively covered in the final report of the WCRP/SCOR Working Group on Air-Sea Fluxes (WGASF, 2000). With this report as a solid base, it is possible here to focus on what practitioners of GODAE should understand about the ocean surface fluxes they may utilize.

Perhaps the most well known constraint on surface fluxes is that the global long term heat and freshwater fluxes into the ocean should both be near zero. Observations of long-term changes in ocean heat content suggest a heat flux of at most a few W/m^2. For example, Levitus et al. (2000) find that the temperature of the world's oceans increased from the 1950s to the 1990s at a rate equivalent to a surface heat flux imbalance of only $0.3W/m^2$. Similarly, the global ocean salinity record doesn't support a significantly non-zero global long-term freshwater flux. However, fluxes on shorter time and space scales are not known nearly so well. This issue and other surface flux problems are outlined in Section 2. Sections 3 and 4 deal with the measurement and parameterization of the turbulent flux components. Satellite techniques have been most successful at estimating the radiative heating, precipitation and wind stress, so only these products are discussed in Section 5.

Section 6 is concerned with what the various surface flux fields might really look like. Any specified surface forcing will be incremented by the data assimilation cycle. The resulting fluxes should at least satisfy global constraints and should be compared to existing air-sea flux climatologies to see if there are unacceptably large discrepancies in seasonal cycles, variability on inter-annual and longer time scales, and the mean. A number of such climatologies are compared by Beranger et al. (1999); including the SOC (Southhampton Oceanography Centre; Josey et al. 1998), COADS (University of Wisconsin-Milwaukee version of the Cooperative Ocean Atmosphere Data Set; Da Silva et al., 1994), NCEP (NCEP/NCAR reanalysis fluxes; Kalnay et al., 1996) and ERA-15 (15 year ECMWF reanalysis fluxes, Gibson et al., 1997). Others can be

found in WGASF (2000) and fluxes from the recent 40 year ECMWF reanalysis (ERA-40) are beginning to appear. Different fields from various datasets have been corrected for known biases and merged (Large and Yeager, 2004) to produce the fluxes presented in Section 6. Although sea-ice concentration has been well observed from satellites since the late 1970s, other necessary ice data (e.g. thickness and surface temperature) are not available to compute companion air-ice and ice-ocean flux climatologies.

2. Ocean surface flux problems

The fundamental problem of ocean surface forcing is that it can not be directly observed. In the case of the momentum flux between the atmosphere and a uniform land surface, it is possible to construct a "drag plate" (Bradley, 1968) and actually observe the wind stress over a small area, at high frequency. This standard can then be transferred to less direct, more easily implemented measurement techniques such as eddy correlation (Haugen et al., 1971), which themselves can be used as a transfer standard to even more indirect and easily implemented methods, such as inertial dissipation (Large and Pond, 1981). The problem for GODAE is that there are no prospects for an ocean surface drag plate. It is conceivable that detailed measurements of heat and moisture changes over a small patch of ground could yield direct estimates of the heat and water fluxes that could serve as a basic truth standard. Unfortunately, a similar approach in the ocean appears hopeless, largely because of transports across the non-surface boundaries of the control volume, and also because of the vertical penetration of solar radiation and the presence of bubbles that are carried to depth after wave breaking. At the end of the transfer of standard chains are the bulk aerodynamic formulae (Roll, 1966), whose inputs are relatively easily measured mean quantities such as atmospheric wind speed, U, potential temperature, θ, and specific humidity, q, as well as sea surface temperature (SST). But, without a direct standard there always remains an unquantified uncertainty in estimates of ocean surface fluxes.

Since the nature of the sea surface dictates that flux measurements be made at some distance, d of order 10 m above the sea surface, there is another basic problem; namely, this flux is not equal to the surface flux. To estimate the fractional error in wind stress, $\delta\tau = |\tau - \tau(d)|/\tau$, consider the momentum equation for a mean near surface wind, U, aligned with the x-coordinate axis. The dominant terms in horizontally homogeneous flow are (Busch, 1977)

$$\rho \ \partial_t U \ = \ \partial_z \tau(z) \ - \ \partial_x P_o \ , \tag{1}$$

where ρ and P_o are atmospheric density and pressure, respectively, and $\tau(z)$ is the downstream stress as a function of height, z. Except near the equator, the geostrophic winds aloft, U_g, are found empirically to be about 30% greater than $U(d)$, and rotated by about 16° (Deacon, 1973):

$$\rho \ f \ U_g \ = \ \partial_n P_o \ = \ \partial_x P_o / \ sin(16^\circ) \ = \ \rho \ f \ 1.3 \ U(d) \ , \qquad (2)$$

where f $\approx 10^{-4} s^{-1}$, is the Coriolis parameter and n is a horizontal coordinate perpendicular the direction of U_g. In steady flow, $\partial_t U = 0$, substitution of (2) into (1) gives

$$\begin{aligned} \delta\tau \ &= \ 1.3 \ d \ f \ sin(16^\circ) \ \frac{\rho \ U(d)}{\tau} \\ &\approx \ 0.04 s^{-1} \ d/U(d) \ , \end{aligned} \qquad (3)$$

where measurements over the sea have been used to approximate $\rho \, U(d)/\tau$ with $1000/U(d)$. Thus, ship measurements of stress at say d = 15 meters should be systematically biased low by 15% at a wind speed of $U(d) = 4m/s$. This problem becomes more complicated in unsteady winds, with (1) showing the error due to flux divergence increasing during falling winds and decreasing on the rising wind. Despite the approximations, the above exercise does illustrate that there is no such region as a "constant flux layer". Indeed, Lumley and Panofsky (1964) introduce the term only to describe the region where measurement error is expected to be greater than the loss of flux with height. Similar consideration of the heat and moisture equations show that heat and moisture fluxes are not, in general, constant with height either.

A third problem is that truly global ocean data assimilation is greatly complicated by the presence of sea-ice at polar latitudes. Not only do ice-ocean fluxes become involved, but the freezing of sea-water and brine rejection need to be accounted for. An attractive, but complicated approach is to include a Sea-Ice Model (SIM) in the assimilation system, so that ice-ocean fluxes are explicitly computed and exchanged as part of the coupling. However, it then becomes necessary to force the SIM with air-ice fluxes of heat, freshwater and momentum. This option places very high demands on flux accuracies, because of positive feedbacks associated with sea-ice. A simpler procedure is just to specify reasonable ice-ocean fluxes even though they are neither routinely observed, nor well known.

The final Ocean Surface Flux problem to be considered is the proliferation of fields that are required to capture the physics and to take full advantage of the available observations. Clearly, the simplest GODAE scheme would follow the Stammer et al. (2002) Ocean Reanalysis example and utilize global fields of the four fluxes; Q, F, τ_λ and τ_ϕ from

Numerical Weather Prediction (NWP), but important physics is lost. In general, a fraction, f_i, of the ocean surface may be covered by sea-ice, leaving a fraction, $f_o = 1 - f_i$, exposed to the atmosphere above. The ocean surface fluxes are, therefore, given by

$$\begin{aligned} Q &= f_i\, Q_{io} + f_o\, Q_{as} \\ F &= f_i\, F_{io} + f_o\, F_{as} + R \\ \vec{\tau} &= f_i\, \vec{\tau}_{io} + f_o\, \vec{\tau}_{as}\,, \end{aligned} \tag{4}$$

where the subscripts "as" and "io" denote air-sea and ice-ocean fluxes, respectively. In (4), R is the continental runoff. The number of required fields has already expanded from 4 to 10, and there are further increases as the air-sea, ice-ocean and air-ice fluxes are explored individually below.

2.1 Air-sea fluxes

The air-sea heat flux has radiative (shortwave and longwave), turbulent (sensible and latent) and particle components, which are all defined here as positive when they act to heat the ocean. Similarly, water fluxes (precipitation, evaporation and runoff) are positive when they add water to the ocean. The wind stress is a turbulent flux aligned to the vector difference between $\vec{U}(d)$ and the ocean surface current, $\vec{U}_0$. A major complication for GODAE is that air-sea heat and freshwater fluxes need to be broken down into estimates of their components, namely;

$$\begin{aligned} Q_{as} &= Q_S + Q_L + Q_E + Q_H + Q_P \\ F_{as} &= P + E, \end{aligned} \tag{5}$$

because the penetration of solar radiation, Q_S into the upper ocean is an important process governing the evolution of SST (Denman and Miyake, 1973) and evaporation, E, from the ocean surface is accompanied by a latent heat flux, $Q_E = -\Lambda E$, where $\Lambda = 2.5 \times 10^6 j/kg$ is the latent heat of vaporization. Delivery of near real time estimates of all these components to a GODAE system is a daunting challenge. Instead, at least some degraded products will need to be used, but then the difficulty will be to specify the appropriate uncertainty. The growth of air-sea flux fields from 4 to as many as 14 is shown in Table 1.

Solar radiation includes wavelengths between 0.3 and 3μ and is always positive. It passes through the atmosphere where it is attenuated, mainly by clouds, before reaching the surface as solar insolation, Q_I, and the surface albedo (α) is the fraction that is reflected back to the atmosphere. About 40% of this insolation is diffuse, with an albedo $\alpha_{df} = 0.06$ (Payne,

Fluxes	Air-sea	Components	Fields	Bulk	
Q					Net Surface Heat Flux
	Q_{as}				Air-Sea Heat Flux
		Q_S			Net Solar Radiation
			Q_I		Solar Insolation
			α		Solar Albedo
		Q_L			Net Longwave Radiation
			Q_A		Downwelling Longwave
			SST_s		Skin SST
		Q_E^*			Latent Heat Flux
		Q_H^*			Sensible Heat FLux
F					Net Surface Freshwater Flux
	F_{as}				Air-Sea Freshwater Flux
		P			Total Precipitation
			P_R		Rainfall
			P_S		Snowfall
		E^*			Evaporation
		R	R		Continental Runoff
$\vec{\tau}$					Surface Wind Stress
	$\vec{\tau}_{as}{}^*$				Air-Sea Wind Stress
				$\vec{U}$	Wind Vector
				θ	Potential Air Temperature
				q	Air Specific Humidity
				P_o	Atmospheric Pressure
				SST_b	Bulk SST
				$S\vec{S}U$	Sea Surface Current

Table 1. Proliferation of air-sea flux fields from the four in the AIR-SEA column. Component fluxes denoted with an asterisk are parameterized by Bulk Aerodynamic Formulae in terms of the 7 parameters, including 2 wind components, in the BULK column. Other flux components are given by the 7 specified fields in the FIELD column.

1972). The albedo for the remaining direct solar radiation, α_{dr} varies with the solar zenith angle. The solar energy transferred to the ocean is then given by

$$Q_S \;=\; Q_I \;[\; 0.6\;(1-\alpha_{dr}) \;+\; 0.4\;(1-\alpha_{df})] \;\;=\;\alpha\; Q_I. \qquad (6)$$

At the temperature of the ocean surface and of the atmosphere above, blackbody radiation occurs at longer wavelengths (up to 50μ) and comprises the net longwave radiation,

$$Q_L \;=\; Q_A \;-\; \epsilon\;\sigma\;\;(SST)^4 \;\;. \qquad (7)$$

In (7), the downwelling (positive) radiation from the atmosphere, Q_A, increases with cloud cover. The upwelling radiation from the ocean is given by $(-\epsilon\sigma(SST)^4\;)$, where $\sigma = 5.67 \times 10^{-8} W/m^2/K^4$ is the Stefan-Boltzmann constant and taking the surface emissivity ϵ as 1, accounts for reflected Q_A (Lind and Katsaros 1986). This high emissivity and the usually warmer SST compared to the radiating atmosphere and clouds, makes Q_L negative.

Since SST is a product of GODAE, the radiative flux problem can be thus reduced to one of specifying Q_I, α and Q_A (Table 1). Fortunately, these fields have been derived from satellite observations over the ocean (Section 5), so that there is no need to use empirical formulae, such as those developed by Smith and Dobson (1984) and examined by Fung et al. (1984). Nonetheless, these expressions do quantify to first order the decrease in Q_I and the increase in Q_A due to clouds. The compensating effect on the daily surface heat flux can be nearly complete in some situations. Therefore, it is important to use consistent data sets for both radiation components, so that errors due to clouds are minimized.

The sea surface is usually warmer than the overlying air, leading to an upward (negative) molecular diffusion of heat across the surface. At the same time evaporation from the surface is assumed to keep the surface air saturated, and hence usually more moist that the air above. These gradients are maintained by the vertical turbulent transports of sensible heat flux and moisture (latent heat), which in steady state must match the surface sensible heat flux, Q_H, and E (Q_E), respectively. Except in the rare circumstance of very warm moist air over a cold sea, evaporation takes water from the ocean and cools.

The major particle fluxes are precipitation, P, over the ocean due to rain, P_R and snow, P_S:

$$P \;=\; P_R \;+\; P_S. \qquad (8)$$

These fluxes are positive definite and account for about 90% of the evaporation (WGASF, 2000). Snowfall has an associated negative particle

heat flux, because the ocean must lose heat to melt it according to

$$Q_P \;=\; -\Lambda_f \; P_S, \tag{9}$$

where $\Lambda_f = 3.337 \times 10^5 j/kg$ is the latent heat of fusion. This flux can also include the temperature difference between the precipitation and SST. The in situ measurement of marine precipitation is very difficult (Knox, 1991), and further complicated by very serious sampling issues arising from its intermittency in both time and space. Also, problematic is the partitioning into rain and snow, because some data sets just give the total.

2.2 Ice-ocean fluxes

A comprehensive review of ice-ocean coupling is provided by Schmidt et al. (2004). Over much of the ice pack, especially in the Southern Ocean, the force balance of sea-ice is primarily between wind stress and ocean drag, which is referred to as free drift. For GODAE purposes it may be sufficient to assume that the air-ice stress, $\vec{\tau}_{ai}$ passes unchanged to the ocean:

$$\vec{\tau}_{io} \;=\; \vec{\tau}_{ai} \tag{10}$$

Otherwise internal ice rheology becomes involved in very computationally expensive calculations of internal ice stresses that are often small.

In general the ocean gains heat and salt where ice is formed and is cooled and diluted where the ice melts. The heat and salt fluxes are given by :

$$\begin{aligned} Q_{io} &= Q_M \;+\; Q_F \;+\; Q_B \;+\; Q_{PS} \\ F_{io} &= F_M \;+\; F_F \;+\; F_B. \end{aligned} \tag{11}$$

The fluxes associated with the melting of both sea-ice and accumulated snow are a cooling melt flux $Q_M < 0$ and positive freshwater flux $F_M > 0$. Solar radiation is able to penetrate thin ice to become an ocean surface flux, Q_{PS}. The usually cold atmosphere above the ice produces basal ice and a freshwater flux, $F_B < 0$, from the ocean. The corresponding ocean surface heat flux, Q_B, is usually near zero, because the basal cooling is balanced by the latent heat of fusion of the basal ice. Direct air-sea cooling of the ocean allows the possibility for ocean temperatures to fall below freezing. The physical process that relieves this unphysical condition is the formation of frazil sea-ice. A heat flux $Q_F > 0$ effectively restores the ocean temperature to the freezing point. Most of the salt is left behind and is represented by a negative freshwater flux, $F_F < 0$. Frazil ice formation can be a dominant process along the

coast of Antarctica where offshore winds can keep significant portions of the ocean surface ice-free.

2.3 Air-ice fluxes

Including a SIM within the GODAE system eliminates the need to specify ice-ocean fluxes, though observed sea-ice fraction should probably be assimilated, but requires data sets for computing all the air-ice fluxes. In general a sea-ice model has a number of ice categories. The air-ice heat and water fluxes for a particular ice category, n, are given by

$$\begin{aligned} Q^n_{ai} &= Q^n_S + Q^n_L + Q^n_E + Q^n_H \\ F^n_{ai} &= P^n + E^n \end{aligned} \tag{12}$$

Ice models differentiate between snow and rain. The former becomes P^n in (12), which accumulates in the snow layer atop ice category, n, with surface temperature, T^n_S, while the rain is often passed directly to the ocean below, as a contribution to F_{io} in (4).

Given a shortwave insolation data set, ice models typically split this radiation into four components; visible (wavelengths from 0.3 to 0.7μ) direct, visible diffuse, near-infrared (0.7 to 3μ) direct and near-infrared diffuse. An ice model then must provide the respective albedos; α^n_{Vdr}, α^n_{Vdf}, α^n_{Ndr} and α^n_{Ndf}, for each ice category, n. In order to use Q_I data the respective fractions are needed, with $.29, .24, .31, .16$ an acceptable partitioning:

$$Q^n_S = Q_I\,[.29(1-\alpha^n_{Vdr})+.31(1-\alpha^n_{Vdf})+.24(1-\alpha^n_{Ndr})+.16(1-\alpha^n_{Ndf})]. \tag{13}$$

It is simpler to utilize downwelling long-wave radiation, because there is no need to do any partitioning. The net long-wave radiation becomes:

$$Q^n_L = \epsilon_i\, Q_A - \epsilon_i \sigma (T^n_S)^4, \tag{14}$$

where $\epsilon_i = 0.95$, is the ice/snow emissivity. This emissivity is also used in the first term on the right-hand-side of (14) as 1 minus a longwave albedo.

2.4 Continental runoff

Very complex estuarine process govern runoff into the ocean, and may be beyond the scope of GODAE. However, it is important to capture the concentration near coastlines, especially at the mouths of large rivers. There should also be a representation of seasonality, including the storage of snowfall. For example, extratropical runoff can be made to peak

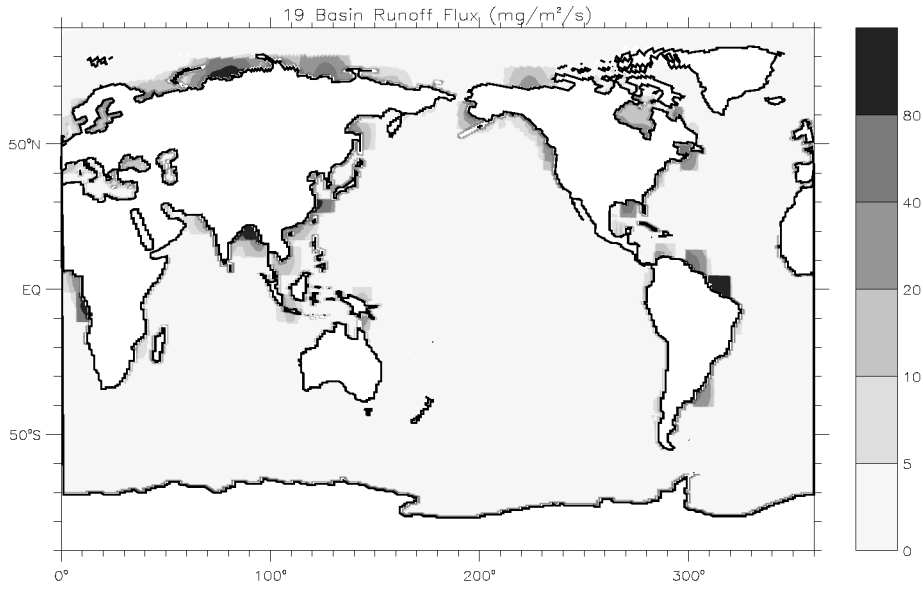

Figure 1. Distribution of the freshwater flux due to climatological runoff from continents. The flux unit of $1 mg/m^2/s$ ($0.0864 mm/day \approx 31 mm/year$) is used, because it makes about the same contribution to the density flux as 1 W/m^2 of heat flux. Note the non-linear scale.

in late spring, and the Amazon in February (UNESCO, 1985). There are runoff data for approximately the largest 1000 of the world's rivers (e.g. Perry et al. 1996), which accounts for as much as 95% of the runoff from both South America and Europe into the Atlantic Ocean, but less than 5% from Australia and Antarctica. The remainder enters the ocean as ground water seepage, or as ungauged surface runoff, including icebergs. Therefore, it is necessary to have estimates of the net excess of precipitation over evaporation from each continent (e.g. Baumgartner and Reichel, 1975) and to distribute this excess as runoff into the bordering ocean basins using river routing schemes and flow estimates (e.g. Fekete, et al., 1999).

A practical treatment of runoff from 19 drainage basins (Large and Yeager, 2004), as an ocean surface flux is shown in Fig. 1. The ungauged runoff is distributed evenly along the coast of each basin. The gauged river runoff is spread over ocean grid-cells near the measurement site, so as to give an ocean surface salinity signature that is similar to that seen observations such as the Levitus et al. (1998) World Ocean Atlas 98 (WOA98). This procedure mimics the flow of fresh estuarine

water over the salty coastal sea-water and avoids excessively low salinities at OGCM grid-points at river mouths. Rivers such as the Amazon, Ganges/Brahmaputra, Zaire, Yenisei/Ob and Yangtze, contribute to the surface density flux in an major way. They each give local freshwater fluxes in excess of $100 mg/m^2/s$ ($1 mg/m^2/s \approx 31 mm/year$), which produces a density flux approximately equivalent to that of a $100 W/m^2$ heat flux. Averaged over the entire ocean area of $3.523 \times 10^8 km^2$ the equivalent freshwater flux is $3.57 mg/m^2/s$, or about 11 cm/year.

3. Measuring turbulent air-sea fluxes

The portion of a geophysical fluid that is directly influenced by the presence of a boundary is referred to as a planetary boundary layer (PBL) and the two most notable examples are the atmospheric (ABL) and oceanic (OBL). Air-sea turbulent fluxes are actually measured in the atmospheric surface layer which begins above the direct influence of the surface roughness elements and ends at about 10% of the ABL height. The semi-empirical Monin-Obukhov similarity theory is the basis of our understanding of the physics of this turbulent layer. The theory argues that in the surface layer the only important turbulence parameters are the height, z, and the air-sea fluxes. The fundamental turbulent parameters that can be formed from the fluxes are the friction velocity, u^*, the scales of the turbulent fluctuations of scalars, such as potential temperature, θ^*, and specific humidity, q^*, and the Monin-Obukhov length scale, L:

$$\begin{aligned} u^{*2} &= |\vec{\tau}|/\rho \\ u^*\theta^* &= Q_H/(\rho\ c_p) \\ u^*q^* &= Q_L/(\rho\ \Lambda) \\ L &= u^{*3}/(\kappa\ B_o), \end{aligned} \tag{15}$$

where $\kappa = 0.4$ is von Karman's constant and c_p is the specific heat of the air. The surface buoyancy flux is given by the latent and sensible heat fluxes and can be expressed as

$$B_o = g\ u^* \left[\frac{\theta^*}{\theta_v} + \frac{q^*}{(q + 0.608^{-1})} \right], \tag{16}$$

where g is gravitational acceleration, $\theta_v = \theta\ (1. + .608q)$ is the virtual potential temperature and the factor 0.608 is the ratio of the molecular weights of dry air and of water vapor minus 1.

Dimensional analysis is very powerful in the surface layer because the number of parameters is small and comparable to the number of physical dimensions. It predicts that layer structure when appropriately scaled

(non-dimensionalized) by z, u^*, θ^*, q^* and L should be functions of the non-dimensional group $\zeta = z/L$. With all the scales common to all surface layer flows, the structure of any layer and its turbulence, according to the theory, must always be "similar", with any dependencies on ζ universal. The most important consequences are the familiar logarithmic vertical profiles of mean wind (Tennekes, 1973), temperature and humidity. Scaling the mean property gradients gives :

$$\frac{\kappa\ z}{u^*}\ \partial_z U\ = \phi_m(\zeta)\ ;\quad \frac{\kappa\ z}{\theta^*}\ \partial_z \theta\ = \phi_s(\zeta)\ ;\quad \frac{\kappa\ z}{q^*}\ \partial_z q\ = \phi_s(\zeta)\ , \tag{17}$$

where κ sets the non dimensional profiles for momentum, $\phi_m(\zeta)$, and for scalars, $\phi_s(\zeta)$ equal to 1 at neutral stability, $\zeta = 0$. These functions of the stability parameter, ζ, have been determined empirically, and acceptable fits to the data for present purposes are (Högström, 1988) :

$$\begin{aligned} \phi_m = \phi_s &= 1\ + 5\ \zeta \qquad \zeta > 0 \\ \phi_s\ = \phi_m^2 &= (1 - 16\zeta)^{-\frac{1}{2}}\ -1 < \zeta < 0\ . \end{aligned} \tag{18}$$

The uncertainty in these functions are a major problem with inferring fluxes (u^*, θ^* and q^*) from measurements of mean property gradients. Other serious issues with "the profile method" are the need for very accurate instrumentation, the small signals (gradients) except near the surface, and the danger of measuring below the surface layer.

Integration of these gradient relations (Paulson, 1970) gives the mean property profiles in the surface layer, but not below,

$$\begin{aligned} U(z) &= SSU + \tfrac{u^*}{\kappa}\ [ln(z/z_o) - \psi_m(\zeta)] \\ \theta(z) &= SST + \tfrac{\theta^*}{\kappa}\ [ln(z/z_\theta) - \psi_s(\zeta)] \\ q(z) &= SSQ + \tfrac{q^*}{\kappa}\ [ln(z/z_q) - \psi_s(\zeta)]. \end{aligned} \tag{19}$$

The constants of integration, z_o, z_θ and z_q, are the roughness lengths, that fully describe the surface as seen by the flow in the surface layer. The integrated non-dimensional profiles for $Y = (1 - 16\ \zeta)^{1/4}$ are:

$$\begin{aligned} \psi_m &= \psi_s = -5\ \zeta && \zeta > 0 \\ \psi_m(\zeta) &= 2ln(\tfrac{1+Y}{2}) + ln(\tfrac{1+Y^2}{2}) - 2\ tan^{-1}(Y) + \tfrac{\pi}{2} && -1 < \zeta < 0 \\ \psi_s(\zeta) &= 2\ ln(\tfrac{1+Y^2}{2}) && -1 < \zeta < 0. \end{aligned} \tag{20}$$

3.1 Eddy covariance

Vertical transport through a PBL is accomplished by three dimensional turbulent eddies. At a fixed point their frequency is related to the

downstream radian wavenumber, k, by Taylor's frozen turbulence hypothesis, $k = 2\pi f/U$. The time average of a vertical flux of property X at a point is the integral over frequency, f, of the cospectrum, $\Phi_{wx}(f)$, of the property fluctuations, x, and the vertical velocity, w. Similarly, the spatial average at a point in time is the integral of the cospectrum over k. In practice, the eddies occupy a finite range of wavenumber space, say k_1 to k_2, and hence frequency space, say f_1 to f_2, so the fluxes become

$$u^*\ x^* = \int_{f_1}^{f_2} \Phi_{wx}(f)df\ = \int_{k_1}^{k_2} \Phi_{wx}(k)dk\ = \int_{f_1}^{f_2}\int_{k_1}^{k_2} \Phi_{wx}(f,k)dk\ df. \tag{21}$$

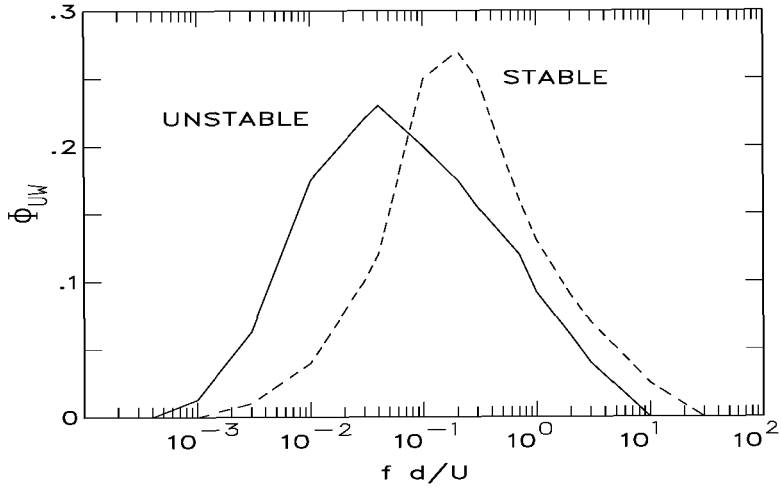

Figure 2. Observed momentum flux cospectra at 13m height in the ABL ensemble averaged over 88 stable and 108 unstable realizations. The ordinate is normalized and is variance preserved by multiplying by f/u^{*2}. The non-dimensional abscissa is natural frequency, fd/U.

Eddy covariance measurements of the sensible heat flux ($X = \rho\, c_p\, \theta$), latent heat ($X = \rho\, \Lambda\, q$) and momentum ($X = \rho\, U$) are not as straightforward as they might appear. In addition to having to work at a height, d, it is often necessary to compromise between sampling long enough to compute a statistically representative flux while not sampling through a change in flow regime. To illustrate, an ensemble average of cospectra, Φ_{uw}, observed at about 13m height over the ocean are shown in Fig. 2 for both stable ($\zeta > 0$) and unstable ($\zeta < 0$) conditions. As the boundary is approached the flux transporting eddies become confined to "fit" within the height, d, and so become smaller in wavelength (higher wavenumber and frequency). This effect is removed in the ensemble by scaling the frequency by a factor, d/U. The Fig. 2 ensembles

indicate that the limits of integration should be about $f_1 = 4 \times 10^{-4} U/d$ and $f_2 = 30U/d$. An ensemble of heat flux cospectra is similar, with the same stability dependence (*e.g.,* Large and Pond, 1982). Thus at a given height the same five decades of eddy sizes transport both the heat and the momentum vertically. However, these limits may not apply to an individual realization, where the low frequency cospectrum can be large and of opposite sign than the integral. There is no standard treatment of such cases, which adds an element of uncertainty that is difficult to quantify without the benefit of a direct flux measurement. Practical problems with eddy covariance measurements include the accuracy, response (set by f_2) and orientation of the sensors, especially on ships and buoys at sea. It is critical to know the alignment with the vertical so that the measured vertical velocity is not seriously contaminated by the much larger horizontal velocity. This problem is particularly acute in the case of the momentum flux, because erroneous U in the w measurement correlates perfectly with u. The distortion of boundary layer flow by large platforms such as ships raises similar issues as orientation.

3.2 Inertial dissipation

The inertial dissipation method is particularly well suited to moving platforms at sea, because the vertical velocity is not involved, and a considerable degree of flow distortion can be tolerated. It is relatively indirect, but can be regarded as an acceptable standard because of extensive comparison with eddy correlation measurements from stable platforms where flow distortion was minimal (e.g. Large and Pond, 1981; Yelland and Taylor, 1996). Measurements of the various terms in the turbulent kinetic energy equation find that to a very good approximation the dissipation, ϵ, equals the mechanical production minus the buoyant suppression, B_o :

$$\epsilon \;=\; u^{*2}\;\partial_z U \;-\; B_o \;\;=\;\; \tfrac{u^{*3}}{\kappa\; z}\;\phi(\zeta) \;-\; \tfrac{u^{*3}}{\kappa\; L} \tag{22}$$

$$u^{*3} \;=\; \kappa\; z\; \epsilon \;\; (\phi_m(\zeta) - \zeta)^{-1} \tag{23}$$

where (15) and (17) have been used for substitution. Similarly, dissipation equals production of scalar variance for X=$[\theta, q]$ gives

$$N_x = \; u^*\; x^*\; \partial_z X \;\;=\;\; \frac{u^*\; x^{*2}}{\kappa\; z}\;\phi_s(\zeta) \tag{24}$$

$$x^{*2} \;=\; \kappa\; z\; N_x \;\; [u^*\;\phi_s(\zeta)]^{-1}\;. \tag{25}$$

Thus, the surface flux measurement is transformed into measuring ϵ, and the dissipation rates of scalar fluctuations, N_θ and N_q. Direct dissipation measurements are difficult, because they involve centimeter scales.

Fortunately, they can be inferred from spectra of fluctuations at frequencies in the Kolmogoroff (-5/3) range. Invoking Taylor's hypothesis the frequency spectra in this range become:

$$\Phi_u(f) = K' \, \epsilon^{\frac{2}{3}} \, (\tfrac{2\pi}{U})^{-\frac{2}{3}} \;\; f^{-\frac{5}{3}}$$
$$\Phi_x(f) = \beta_x \; N_x \; \epsilon^{-\frac{1}{3}} \; (\tfrac{2\pi}{U})^{-\frac{2}{3}} \;\; f^{-\frac{5}{3}} \; . \qquad (26)$$

The empirical Kolmogoroff constants are $K' = 0.55$, and $\beta_\theta = \beta_q = 0.80$.

The Kolmogoroff range begins at about $fd/U = 0.2$, so the sensors must have a relatively fast sampling capability, of order 10Hz. Multiple measurements across frequency allows the existence of the $-5/3$ range to be checked, and contaminated data, by flow interference for example, to be discarded. The method can be regarded as a physically based parameterization of the fluxes in terms of high frequency turbulent fluctuations, with eddy correlation comparisons essential for verifying the assumptions and empirical constants, and for transferring the measurement standard.

4. Bulk aerodynamic formulae

How well can the turbulent surface fluxes be estimated from time and/or space average (bulk) measures (Table 1) from the turbulent surface layer? Since such data are likely to be a major input of GODAE systems, this question needs to be addressed in detail. Bulk transfer coefficients for momentum, sensible heat and moisture transfer, are defined, respectively, as

$$C_D \; = \; \frac{u^{*2}}{(\Delta U)^2} \quad ; \quad C_H \; = \; \frac{u^* \; \theta^*}{\Delta U \;\; \Delta\theta} \quad ; \quad C_E \; = \; \frac{u^* \; q^*}{\Delta U \;\; \Delta q} \quad , \qquad (27)$$

where $\Delta U = |\vec{U}(z) - \vec{U_o}|$, $\Delta\theta = \theta(z) - SST$ and $\Delta q = q(z) - SSQ$ are the air - sea differences. Substitution of the profile equations (19) gives

$$C_D \; = \frac{\kappa^2}{[ln(\frac{z}{z_o}) - \psi_m]^2} \; ; \quad C_H \; = \frac{\kappa\sqrt{C_D}}{[ln(\frac{z}{z_\theta}) - \psi_s]} \; ; \; C_E \; = \frac{\kappa\sqrt{C_D}}{[ln(\frac{z}{z_q}) - \psi_s]} \; , \qquad (28)$$

which demonstrates the dependencies on height, stability and the roughness lengths. In an ideal world of plentiful, reliable measurements the coefficient estimates would be binned according to height and stability and the roughness dependencies determined for each bin. Unfortunately, even the above indirect flux estimates are too difficult, expensive and rare. Therefore, most coefficient determinations are shifted to 10m

height and neutral stability, where the three coefficients become;

$$C_{DN} \;=\; \frac{\kappa^2}{[ln(\frac{10m}{z_o})]^2} \;\; ; \; C_{HN} \;=\; \frac{\kappa\sqrt{C_{DN}}}{ln(\frac{10m}{z_\theta})} \;\; ; \; C_{EN} \;=\; \frac{\kappa\sqrt{C_{DN}}}{ln(\frac{10m}{z_q})} \, . \quad (29)$$

The transfer coefficients and their 10m, neutral equivalents can be related by eliminating z_o, z_θ and z_q from (28) and (29) :

$$\begin{aligned} C_{DN} &= C_D \; (1 + \tfrac{\sqrt{C_D}}{\kappa} \; [ln(\tfrac{10m}{z}) + \psi_m(\zeta)])^{-2} \\ C_{HN} &= C_H \; \sqrt{\tfrac{C_{DN}}{C_D}} (1 + \tfrac{C_H}{\kappa\sqrt{C_D}} \; [ln(\tfrac{10m}{z}) + \psi_s(\zeta)])^{-1} \\ C_{EN} &= C_E \; \sqrt{\tfrac{C_{DN}}{C_D}} (1 + \tfrac{C_E}{\kappa\sqrt{C_D}} \; [ln(\tfrac{10m}{z}) + \psi_s(\zeta)])^{-1} \; . \end{aligned} \quad (30)$$

The roughness length dependencies of these coefficients have been explored using many data sets, but rarely with combined data. This search has not been conducted in a single standard way, so often the procedure, rather than the data, is responsible for differences in results. The better approaches begin by defining the equivalent 10m, neutral wind, U_N, temperature, θ_N, and humidity, q_N, relative to the sea surface, in terms of the turbulent flux scales and 10m, neutral transfer coefficients :

$$U_N^2 \;=\; \frac{u^{*2}}{C_{DN}} \; ; \; \theta_N \;=\; \frac{u^* \; \theta^*}{C_{HN} \; U_N} \; ; \; q_N \;=\; \frac{u^* \; q^*}{C_{EN} \; U_N} \quad (31)$$

$$U_N^2 \;=\; \frac{C_D}{C_{DN}} (\Delta U)^2 \; ; \; \theta_N \;=\; \frac{C_H}{C_{HN}} \; \frac{\Delta U}{U_N} \; \Delta\theta \; ; \; q_N \;=\; \frac{C_E}{C_{EN}} \; \frac{\Delta U}{U_N} \; \Delta q \; . \quad (32)$$

There is no consensus on how to proceed from this point. To illustrate, consider the following drag coefficient formulation. Perform a multiple regression analysis of u^{*2} on U_N, U_N^2, U_N^3, ... , to find coefficients of the polynomial

$$u^{*2} \;=\; a_0 \;+\; a_1 \; U_N \;+\; a_2 \; U_N^2 \;+\; a_3 \; U_N^3 + ... \quad (33)$$

After combining data from multiple sources to span a range of wind speeds from less than $1m/s$ to more than $25m/s$, Vera (unpublished manuscript, 1986) found that coefficients of the fourth power and higher were not statistically significant. Consistent with the principle of no net stress over space and/or time of zero wind speed neither was a_0, leaving $a_1 = 0.00270m/s$, $a_2 = 0.000142$ and $a_3 = 0.0000764s/m$ as the only nonzero coefficients. Division by U_N^2 yields

$$C_{DN} \;=\; a_1/U_N \;+\; a_2 \;+\; a_3 \; U_N \; , \quad (34)$$

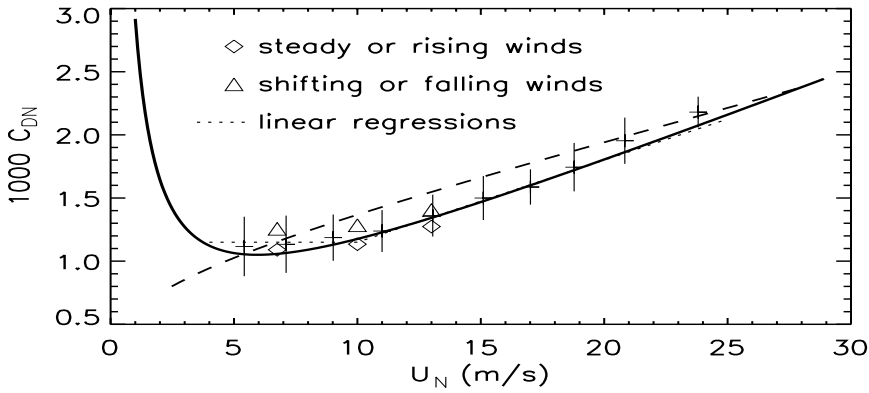

Figure 3. Neutral 10m drag coefficient as a function of equivalent neutral 10m wind speed; as observed (pluses extending ±1 standard deviation about the mean over wind speed bands), the observed means during steady or rising winds (diamonds) and for shifting and falling winds (triangles), and as formulated from piecewise linear regression of all these data (dotted lines), from Eq. (34) (solid curve) and from Eq. (36) (dashed curve).

which is plotted in Fig. 3 (thick solid curve) in a form that can be compared to the alternatives shown and discussed by WGASF (2000).

At low winds the surface stress is supported by molecular viscous stress, independent of the roughness elements. In such aerodynamically smooth flow, the emergence of kinematic viscosity, ν, as a parameter leads to the non-dimensional group :

$$z_o \; u* / \nu \;=\; \alpha_s \quad , \tag{35}$$

where $\alpha_s \approx 0.11$ is an empirical constant. From (29) C_{DN} becomes inversely proportional $(z_o \, U_N)^2$, and grows without bound, as this factor approaches zero at very low wind speeds, consistent with (34) and Fig. 3.

A more common practice has been to linearly regress C_{DN} on U_N, but data from higher winds (e.g. $U_N > 12 m/s$) give a steeper slope than lower winds. Therefore, linear regressions are meaningful only over narrow wind speed ranges, and could be used to form a piecewise linear formulation (Fig. 3, dotted; Trenberth et al., 1989). Another approach is to fit data to proposed functional forms of the roughness length. An early form (Charnock, 1955) assumes that all important features of the ocean surface wave field are captured by gravitational acceleration, which leads to the non-dimensional group :

$$z_o \; g \; u^{*-2} \;=\; \alpha_c \quad , \tag{36}$$

where α_c is constant. Garratt (1977) fits a variety of data and suggests $\alpha_c = 0.0144$ (Fig. 3 dashed curve), but Stewart (1974) notes that for

winds below $10m/s$ such a representation predicts a much more rapid increase in C_{DN} than is supported by data, including those used by Garratt (1977) and Fig. 3. More recently, Smith (1988) suggested adding the two z_o expressions (35) and (36), although there is no evidence that addition is the physically correct operation:

$$z_o \;=\; \alpha_s \;\nu\; u^{*-1} + \alpha_c \;g\; u^{*2}. \tag{37}$$

In contrast, manipulating (34) leads to an exponential dependence:

$$z_o \;=\; 10m \; e^{u^*/(\kappa \; U_N)} \;=\; 10m \; e^{\sqrt{C_{DN}}/\kappa} \; . \tag{38}$$

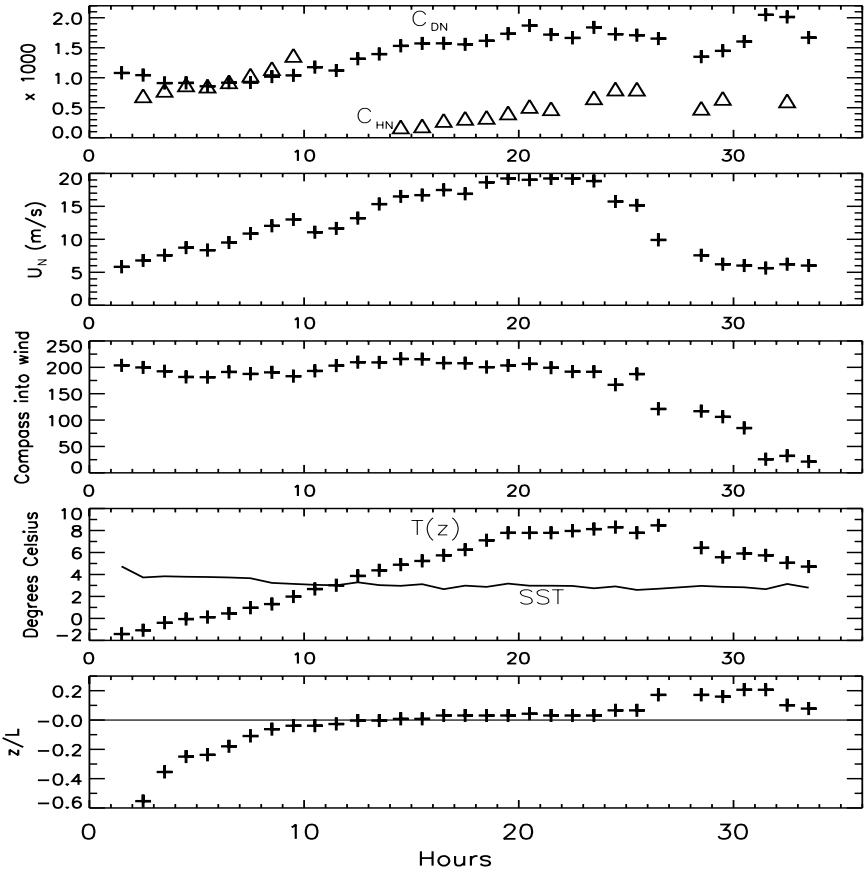

Figure 4. Time series of transfer coefficients C_{DN} (top panel pluses) and C_{HN} (top panel triangles) and corresponding wind speed, wind direction, air temperature, SST and stability (z/L).

All large data sets of measured C_{DN} display considerable scatter at any given wind speed, but the standard deviations of Fig. 3 show that

Fetch (km)	Data hours	Mean	Standard deviation	Minimum	Maximum
10 - 20	200	1.14	.18	.75	2.03
20 - 100	54	1.10	.22	.73	1.87
100 - 200	85	1.13	.24	.64	1.76
unlimited	291	1.14	.21	.62	1.75
all	590	1.13	.21	.62	2.03

Table 2. Fetch dependency of the neutral 10m drag coefficient C_{DN} for winds between 4 and 10m/s.

it tends to be greater at wind speeds below about $10m/s$. However, an error analysis suggests only about ± 20% of this can be attributed to the measurements. The remainder appears to be directly related to varying wind/wave conditions, which underlines the inadequacy of (36). This position is supported by the time series of drag coefficient measurements shown in Fig. 4. During the period of approximately steady winds between hours 15 and 24, the drag coefficient measurements display little variability, suggesting that relatively invariant coefficients may be obtained for equilibrated wind/wave conditions. However, comparing the coefficients less than 0.001 (hours 3 to 8), with those greater than 0.002 (hours 31 to 33) for similar moderate wind speeds, suggests that the latter result from the rapid decrease in wind speed and reversal of wind direction. However, when observations over many storms and frontal passages are grouped according to the wind behavior the effect is measurable, but not large. Figure 3 shows only about a 10% increase in observed C_{DN} when winds have rapidly decreased and/or changed direction over periods of rising and/or steady winds.

Quantification of other wave effects has proved elusive. Table 2 shows that between 4 and 10m/s, C_{DN} does not depend strongly on fetch, nor, by implication, on wave parameters that do, such as amplitude and wavelength. However, the ratio of these two parameters is the wave steepness which is independent of fetch, and a C_{DN} dependency is consistent with observed higher values in shallow water (steep) wave regimes.

It is interesting to note that historically, formulations of heat and moisture coefficients have more closely followed (33), which is rarely used to formulate the drag coefficient. Specifically, measured fluxes $u^*\ \theta^*$ and $u^*\ q^*$ have been regressed on $U_N\ \theta_N$ and $U_N\ q_N$, respectively. In the case of moisture, the offset is not significantly non-zero, so the slope gives C_{EN} directly from (31). However, in the heat flux case there is a significant positive offset, and furthermore, the slope is found to be steeper in unstable conditions ($u^*\ \theta^* > 0$), than in stable. Thus,

it is necessary to treat stable and unstable heat fluxes separately. The positive offset is consistent with an unbounded transfer coefficient (slope) as wind speed approaches zero, but the flux, as in the case of (8), should diminish. This behavior can also be achieved by combining (29) and (31), then using fluxes to compute the roughness lengths from

$$\frac{\kappa}{ln(10m/z_\theta)} = \frac{1}{\sqrt{C_{DN}}} \frac{u^* \ \theta^*}{\theta_N \ U_N} \ ; \tag{39}$$

$$\frac{\kappa}{ln(10m/z_q)} = \frac{1}{\sqrt{C_{DN}}} \frac{u^* \ q^*}{q_N \ U_N} \ . \tag{40}$$

Empirically, (39) is found to average about 0.0180 in stable conditions and 0.0327 in unstable, while a typical value of (40) is about 0.0346. There is relatively little scatter in these values, because of the observed variability in measured C_{HN} and C_{EN} accounted for in the drag coefficient on the right hand sides of (39) and (40). Once determined they directly give the formations of C_{HNu} (unstable), C_{HNs} (stable) and C_{HE} shown in Fig. 5.

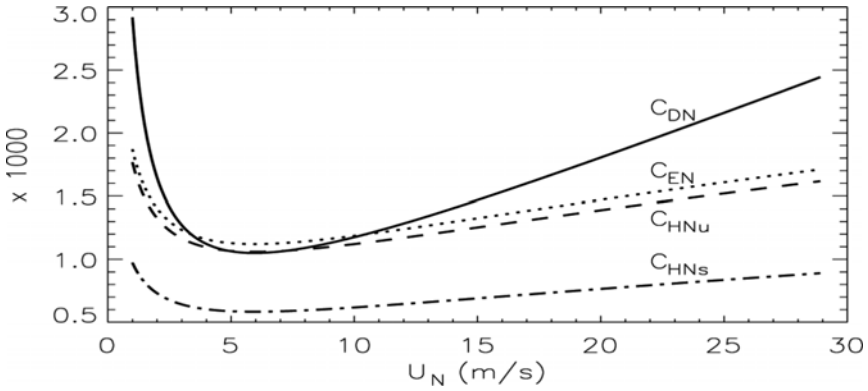

Figure 5. Neutral 10m transfer coefficients as a function of equivalent neutral 10m wind speed; C_{DN} from Eq. (34) (solid curve), C_{EN} from Eqs. (40) and (34) (dotted curve), and unstable C_{HNu} (dashed) and stable C_{HNs} (dot-dashed) from Eqs. (39) and (34).

A dramatic illustration of the effect of stability on C_{HN} is the decrease by more than a factor of two in Fig. 4 (triangles) between hours 10 and 14. During this time the increasing air temperature surpassed the SST, causing the stability parameter to change sign. The indications are that the change is abrupt at $\zeta = 0$, with high values persisting in the earlier near neutral, but still unstable conditions. The effect is smaller in the mean, but still considerable, with the 1.8 the ratio of 0.0327 to 0.0180.

Otherwise, the slow changes in C_{HN} generally follow the evolution of C_{DN}, as expected from (29) and Fig. 5.

4.1 Bulk flux estimates

Estimating the turbulent fluxes from average atmospheric and sea surface properties is essentially the inverse of determining transfer coefficients, and accordingly the bulk formulae are the inverse of (27):

$$\vec{\tau} \;=\; \rho\, C_D\, |\Delta\vec{U}|\, \Delta\vec{U}, \tag{41}$$

$$Q_H \;=\; \rho\, c_p\, C_H\, (\theta(z_\theta) - SST)\, |\Delta\vec{U}| \tag{42}$$

$$E \;=\; \rho\, C_E\, (q(z_q) - q_{sat}(SST))\, |\Delta\vec{U}| \tag{43}$$

$$Q_E \;=\; \Lambda\, E \tag{44}$$

where $\Delta\vec{U} = \vec{U}(z_u) - \vec{U}_0$. The mean wind, $\vec{U}(z_u)$, potential air temperature, $\theta(z_\theta)$, and specific humidity, $q(z_q)$, may be from different heights, z_u, z_θ and z_q, respectively. The surface current is often, but not always, negligible relative to the wind. A constant density, $\rho \approx\; 1.22 kg/m^3$ may be sufficient, but a more accurate expression is

$$\rho \;=\; \frac{P_o}{R_{gas}\; \theta(z_\theta)\; (1. + .608\; q(z_q))}, \tag{45}$$

where $R_{gas} = 287.04 J/kg/K$ is the gas constant for dry air. The order 1% effects of temperature on Λ and of humidity on c_p are usually ignored.

Although it is clear that a skin temperature should be used to compute the longwave radiation (7), a definitive study has not been conducted to determine whether a skin or bulk SST is more appropriate for (42). At present the determining factor is to use the same temperature as was used to compute C_H, which is usually a bulk temperature, because skin temperature is a very difficult measurement. This becomes an issue when satellite estimates of skin SST are available. In (43), the air at the sea surface is assumed to be saturated, with a specific humidity given well enough by

$$q_{sat}(SST) =\; 0.98\; \rho^{-1}\;\; 640380 kg/m^3\; e^{(-5107.4K/SST)}, \tag{46}$$

where the factor 0.98 applies only over sea-water. The saturation assumption has not been verified, but since it is used in formulating C_E, it should be applied in the inverse (43). The exact relation between evaporation and surface latent heat flux (44), depends on the fact that the latent heat required to evaporate comes from the liquid (the ocean). In the case of evaporation from spray this still holds unless droplets entirely

evaporate or are suspended and there is effectively a latent cooling of the lower atmosphere. The significance of this effect remains outstanding because of the lack of a direct measurement standard.

Three practical ways computing the fluxes are: A) shift the wind, temperature and humidity to 10m and neutral stability so that neutral 10m coefficients can be used directly, B) shift the coefficients to the height, and stability of the atmospheric state variables. and C) shift the temperature and humidity to the height of the wind, z_u, then shift the coefficients to this height and to the atmospheric stability. The details of (C) follow, but if the atmospheric state variables are given at the same height, it just becomes equivalent to a particularly efficient version of (B). This efficiency can also be achieved by shifting the temperature and humidity to the wind height off-line. The calculations are most efficient when the wind height, z_u, equals the 10m reference height of the transfer coefficients.

The iterative procedure for (C) is as follows:
1) Assume $\theta(z_u) = \theta(z_\theta)$ and $q(z_u) = q(z_q)$ and compute the virtual potential temperature, $\theta_v = \theta(z_u)\ (1. + .608 q(z_u))$. Then make a first guess of neutral stability and $U_N = |\Delta\vec{U}|$ to give the transfer coefficients (Fig. 5). The initial turbulent scales are then computed as:

$$\begin{aligned} u^* &= \sqrt{\rho_a^{-1}\ |\vec{\tau}|} = \sqrt{C_D}\ |\Delta\vec{U}| \\ \theta^* &= \frac{Q_H}{\rho_a\ c_p\ u^*} = \frac{C_H}{\sqrt{C_D}}\ [\theta(z_\theta) - SST] \\ q^* &= \frac{E}{\rho_a u^*} = \frac{C_E}{\sqrt{C_D}}\ [q(z_u) - q_{sat}(SST)] \end{aligned} \tag{47}$$

2) Begin the iteration loop with estimates of the stability parameters $\zeta_u = z_u/L$, $\zeta_q = z_q/L$ and $\zeta_\theta = z_\theta/L$, where L is the Monin-Obukhov length (22):

$$\zeta(z) = \frac{\kappa\ g\ z}{u^{*2}} \left[\frac{t^*}{\theta_v} + \frac{q^*}{(q(z_q) + 0.608^{-1})} \right] \tag{48}$$

For each of the these, use (20) to find the integrals of the dimensionless flux profiles of momentum, $\psi_m(\zeta)$, and of heat and moisture, $\psi_s(\zeta)$.
3) Shift the wind speed to $10m$ and neutral stability, and the temperature and humidity to the wind height :

$$U_N(10m) = |\Delta\vec{U}|\ \ (1 + \frac{\sqrt{C_D}}{\kappa}[ln(z_u/10m) - \psi_m(\zeta_u)]\ )^{-1} \tag{49}$$

$$\theta(z_u) = \theta(z_\theta) - \frac{\theta^*}{\kappa}\ [ln(\frac{z_\theta}{z_u}) + \psi_h(\zeta_u) - \psi_h(\zeta_\theta)] \tag{50}$$

$$q(z_u) \;=\; q(z_q) \;-\; \frac{q^*}{\kappa}\,[ln(\frac{z_q}{z_u}) + \psi_h(\zeta_u) - \psi_h(\zeta_q)] \quad . \tag{51}$$

4) Update the neutral 10m transfer coefficients (Fig. 5), then shift them to the measurement height, z_u, and stability, ζ_u, using:

$$\begin{aligned} C_D(z_u,\zeta) &= C_{DN}(1 \;+\; \tfrac{\sqrt{C_{DN}}}{\kappa}\,[ln(z_u/10m) \;-\; \psi_m(\zeta_u)])^{-2} \\ C_H(z_u,\zeta) &= C_{HN}(1 \;+\; \tfrac{C_{HN}}{\kappa\,\sqrt{C_{DH}}}\,[ln(z_u/10m) \;-\; \psi_h(\zeta_u)])^{-2} \\ C_E(z_u,\zeta) &= C_{EN}(1 \;+\; \tfrac{C_{EN}}{\kappa\,\sqrt{C_{DN}}}\,[ln(z_u/10m) \;-\; \psi_h(\zeta_u)])^{-2} \end{aligned} \tag{52}$$

5) Using these transfer coefficients, $\theta(z_u)$ from (50) and $q(z_u)$ from (51), recompute the virtual potential temperature, θ_v, and update the turbulent scales in (47).
6) Start the next iteration loop at step (2) above. Over the ocean stability is usually near neutral, and 2 iterations are all that is necessary. In very stable conditions up to five iterations may be needed. After the last iteration the fluxes are computed from the last set of turbulence scales according to (15).

Even with perfect atmospheric inputs and SST, the flux estimates are only as representative of true conditions as the bulk transfer coefficients. Figure 3 shows that after the wind speed and direction shifts the momentum (heat and moisture) flux would be underestimated by nearly a factor of 2 ($\sqrt{2}$), and overestimated early in the time series. When averaged over the 3 - 7 day synoptic period, the uncertainty should diminish as wave effects start to average out. But, because of the seasonal cycle of storms, annual averaging is probably required before the uncertainty in bulk fluxes is minimized.

5. Satellite flux estimates

Satellite retrieval products are attractive for GODAE, because of the frequent, global sampling, and the possibility of near real time delivery; continued availability notwithstanding. Over the ocean satellites have been most successful at producing wind, radiation and precipitation fields. However, these are necessarily very indirect measures, and in situ measurements, are essential for providing the calibration standard, at least at a few select locations. Therefore, the accuracy of these observations will be only briefly included below, with WGASF (2000) a more complete reference. There are schemes for estimating monthly mean latent heat flux (evaporation), but this is likely too infrequent for GODAE.

5.1 Wind stress

Satellite wind products have been derived from both passive microwave radiometers and active radars (altimeters and scatterometers), but the principal source of wind direction data is scatterometers. In situ buoy measurements of U_N have been empirically related to scatterometer backscatter, σ_o, such that the accuracy of scatterometer wind speed, U_S, can be rather good; -.3m/s bias and 1.3 rms difference (Freilich and Dunbar, 1999). Taking U_S as an estimate of U_N gives the stress directly from (31) as

$$u^{*2} = C_{DN}(U_S) \;\; U_S^2 \;\; . \tag{53}$$

But is this the best calculation? From section 4 it can be argued that any drag coefficient formulation is a representation of "average" wave conditions. There are recent formulations which purport to account for at least some wave effects, given local wave conditions, which we denote as C_{DN}(waves). Figure 4 suggests that at the same wind speed, rougher seas produce greater stress that a C_{DN}(waves) formulation should be able to capture, at least in large measure. However, the observations of Li et al. (1987) suggest that the rougher seas themselves produce more backscatter and a high satellite wind , so that in order not to account for the rough surface twice, C_{DN} as in (34) should be used in (53), as given.

5.2 Radiation

The in situ measurement of radiation both incident on and reflected from the ocean surface is essentially a direct measurement. The solar and longwave components are measured by a pyranometer and a pyrgeometer, respectively. The two instruments are similar, consisting of a blackened thermopile covered by a hemispherical dome whose transmissivity matches the desired measurement. In ideal conditions pyranometer accuracy can be excellent (2 %), but variations in the temperature of the dome can be problematic. At sea conditions are far from ideal, with platform motion a serious problem unless mitigated by use of a gimbal. Other sources of error are calibration error, contamination of the dome, so all together measurement accuracy is unlikely to be no better than 5%. Longwave radiation is naturally emitted near typical pyrgeometer temperatures, so corrections need to be applied for dome temperature and emissivity to transmissivity ratio, and for the temperature of the instrument housing. Although the thermopile is shielded by a shortwave filter, the solar radiation can be so large that even a small amount of leakage can be serious. Again measurement errors of at least 5% are expected.

Satellites derive surface radiation from measurements of incoming and outgoing radiation at the top of the atmosphere. These are generally combined with information on the atmospheric column in a radiative transfer model (RTM), with the treatment of clouds of critical importance. For example, the recent ISCCP-FD dataset (Zhang et al., 2004) contains flux profiles, every 3 hours for 18 years at 280 km spatial resolution. By improving the treatment of cloud vertical structure, revising the aerosol chemistry and water vapor profiles, and accounting for diurnal skin and air temperatures, the overall uncertainty in surface fluxes is claimed to be reduced to $10 - 15 W/m^2$. The standard of comparison is the Baseline Surface Radiation Network over land (Gilgen et al., 1995), and ship radiometers as part of the SEAFLUX project (Curry et al., 2004).

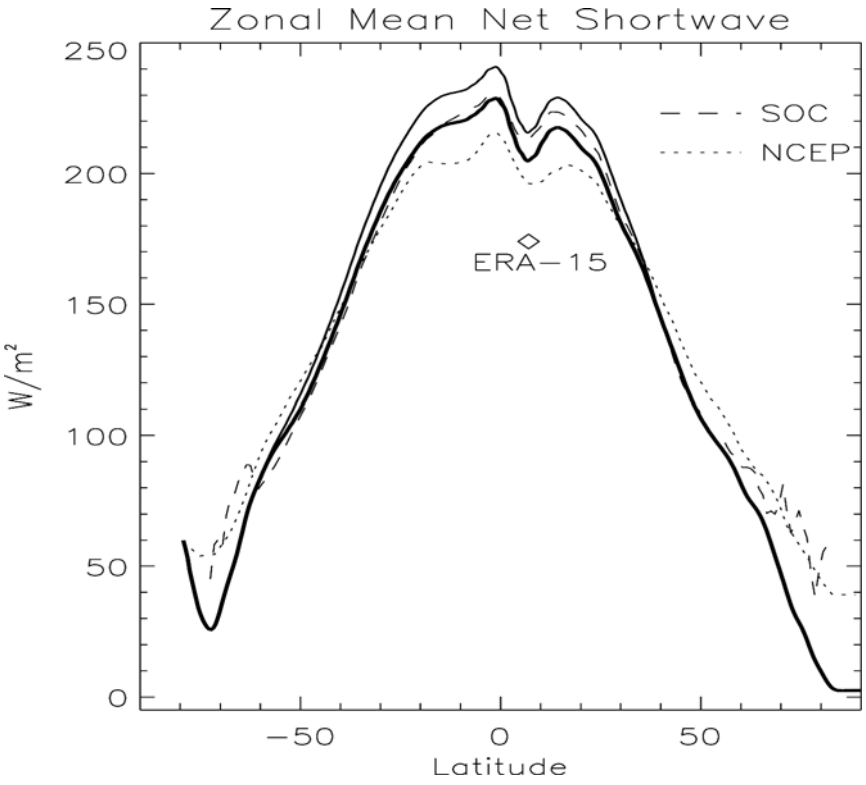

Figure 6. Zonally averaged climatological net solar radiation into the ocean from the ISCCP-FD (Zhang et al., 2004) both uncorrected (thin solid) and corrected (thick solid, see Section 6.1), compared to SOC (dashed) and NCEP reanalysis (dotted). Only one value at 5° N is shown for ERA-15.

Both Beranger et al. (1999) and WGASF (2000) compare zonal average net shortwave radiation flux from a variety of sources. Two of these (NCEP and SOC) are shown in Fig. 6 along with the ISCCP-FD product (thin solid). In the tropics, the latter is about $10 - 20 W/m^2$ greater than the Langley Research Center (Darnell et al., 1996) satellite product (not shown). There is much better agreement at higher latitudes. Although the two dataset agree to within their uncertainties, differences are significant in the tropics, where the two observational datasets, SOC (Fig. 6) and COADS (not shown) both fall about midway between the two satellite products.

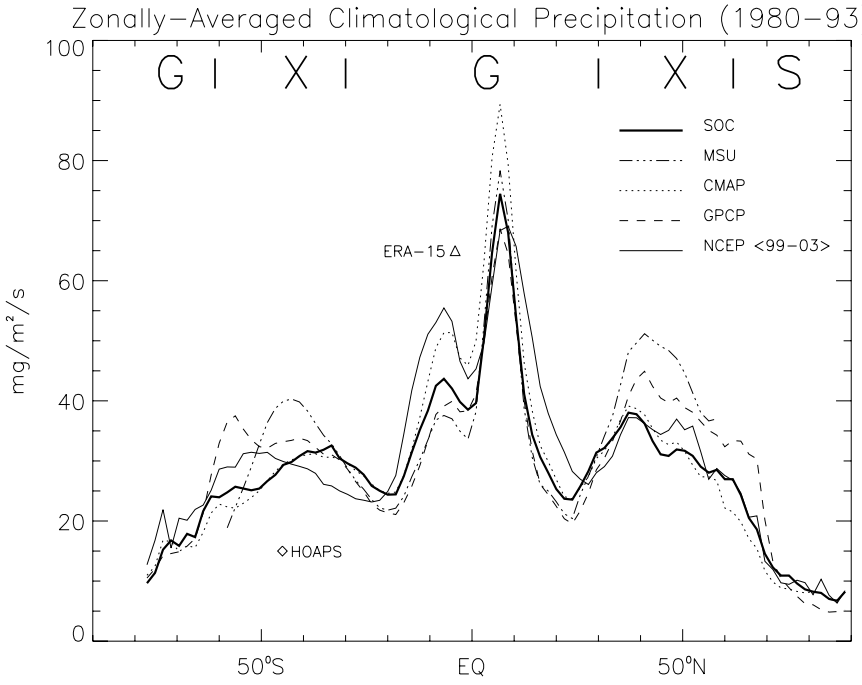

Figure 7. Comparison of zonally averaged precipitation into the ocean from a variety of climatologies.

5.3 Precipitation

The precipitation sampling problem is much less severe in satellite estimates, but the lack of accurate in situ observations makes the development and validation of algorithms difficult. Therefore, satellite precipitation data sets suitable for GODAE must be regarded as highly uncertain. Freshwater fluxes change sea surface salinity, but because these fluxes do not depend on the salinity there is no direct negative

feedback to retard the accumulation of flux error, as there is in the case of the heat flux dependence on SST. Therefore, large systematic errors are particularly problematic. The comparison of zonal averages from different precipitation data sets in Fig. 7 illustrates the problem, even though large regional differences are lost, because of compensation along latitude circles. The best agreement is found around 30° latitude in both hemispheres. Ideally, one data set would be applicable globally, but it appears that each precipitation product is demonstrably deficient in one or more of the following five zones; the Antarctic (poleward of $61°S$), the Southern ($65°S$ to $30°S$), the Equatorial ($30°S$ to $30°N$), the Northern ($30°N$ to $62°N$), and the Arctic (poleward of $62°N$).

Comparing the satellite products in the tropics, the CMAP values of Xie and Arkin (1996) are about 30% greater than GPCP (Global Precipitation Climatology Project) (Huffman et al., 1997), and more than an ocean climate model can deal with. The MSU (Microwave Sounding Unit; Spencer, 1993) falls between. In contrast, the subtropical ocean simulations are much better when forced with CMAP than either GPCP, or MSU. Polar latitudes are not sampled by some satellites, including the MSU, so CMAP becomes reliant on NWP model output, complete with spectral ringing signals in the precipitation. Therefore, the only viable Antarctic product is GPCP. It is also available in the Arctic, but doesn't compare all that well to the observed compilation of Serreze and Hurst (2000). Therefore, as denoted GXGXS in Fig. 7, a possible, but far from ideal, compromise, is to use GPCP(G), CMAP(X), GPCP(G), CMAP(X) and Serreze/Hurst(S), respectively, in the five zones noted above. Another satellite based product is HOAPS (Hamburg Ocean Atmosphere Precipitation System), but it appears to be an outlier, giving only 15 $\mathrm{mg/m^2}/s$ at 44° S.

6. A merged flux climatology

According to WGASF (2000), "there is presently no one flux climatology which does not exhibit significant errors in one region or another in each of the various flux components." In particular, there are serious issues with the flux products from NWP reanalysis, which are also expected in the operational fluxes that would be very convenient for GODAE. Most problematic are fields like radiation and precipitation which strongly depend on the cloud field. For example, tropical radiation from NCEP (Fig. 6) is significantly lower than all satellite products, with ERA-15 producing even less solar heating. Also, NCEP precipitation (Fig. 7) around $5°S$, is much greater than other datasets, with the better agreement with CMAP due to CMAP's blending of station

	CORRECTIONS			
	NONE	WIND	HUMIDITY	ALL
$f_o\ Q_S$	173	173	173	165
$f_o\ Q_L$	-52	-52	-52	-52
$f_o\ Q_E$	-76	-84	-89	-97
$f_o\ Q_H$	-13	-14	-13	-14
$f_o(Q_{as} - Q_P)$	31	23	18	1
$f_o\ E$	-30.5	-33.4	-35.6	-39.0
$f_o\ P$	30.4	30.4	30.4	35.4
R	3.5	3.5	3.5	3.5
$f_o\ F_{as} + R$	3.4	0.5	-1.7	-0.1

Table 3. Global mean air-sea fluxes 1984–2000 and continental runoff with no correction (NONE), only the wind speed correction (WIND), only the relative humidity correction (HUMIDITY) and with all the corrections of Section 6.1 (ALL). Heat fluxes are in W/m^2, and freshwater fluxes are in $mg/m^2/s$ ($0.0864 mm/day \approx 31 mm/year$). If the true ocean area of $3.523 \times 10^8 km^2$ is used to divide the sum of all runoff in Fig. 1 the Runoff becomes, R = 3.574 $mg/m^2/s$, or $1.26 Sv$.

and satellite rainfall with NCEP. Again, ERA-15 is even further afield. A major improvement in these NWP fields would seem to require the proper assimilation of observed cloud into NWP analyses. In principle, the accuracy of the radiative fluxes should then approach that of the off-line RTE fluxes.

WGASF (2000) recommends merging a number of different datasets in order to assemble a complete collection of surface flux data that is superior to all the individual datasets. One such example has been developed by Large and Yeager (2004) and is presented below. Their choices were made on the basis of global coverage, frequency duration and the behavior of ocean and sea-ice models. Other justifiable choices could have been made and most are discussed at length by WGASF (2000). The global NCEP/NCAR reanalysis gives the atmospheric state, the ISCCP-FP product provides the radiation fields, the precipitation is the blend of multiple products denoted (GXGXS) in Fig. 7, the continental runoff is shown in Fig. 1, the sea-ice concentration comes from the National Snow and Ice Data Center, historical SST is a reconstruction (Rayner et al. 2003) that has been made compatible with sea-ice concentration.

6.1 Empirical corrections

The most important aspect of flux dataset development is the use of other limited, but higher quality data sets to determine objective corrections to the flux data sets. The corrected/adjusted forcing is used

in conjunction with observed SST to produce an observationally based air-sea flux climatology over 43 years. A necessary achievement of the exercise is to lower the global air-sea heat flux over 17 years (1984–2000) from $31 W/m^2$ heating to a more reasonable $1 W/m^2$ (Table 3). A freshwater imbalance of $3.4 mg/s/m^2$ is over compensated by increased evaporation, so an overall increase in precipitation is used to give a nearly balance ($-0.1 mg/s/m^2$) global mean budget. The impact on these global budgets of the following corrections is summarized in Table 3.

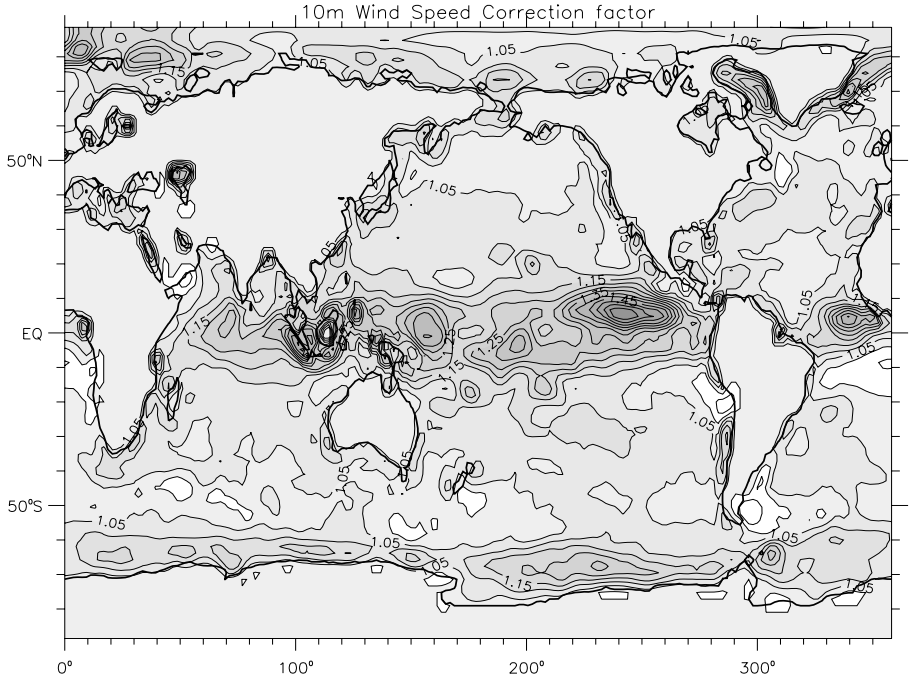

Figure 8. Ratio of two-year mean QSCAT and NCEP wind speeds. Contour interval is 0.05. Regions of weaker NCEP winds (ratio greater than 1.0) are shaded.

A recent comparison of NCEP reanalysis fields with research vessel observations shows evidence of a consistent low bias in NCEP wind speeds at all latitudes (Smith et al. 2001). The advent of satellite wind speed products makes a more global assessment possible. Physically, these speeds should best correspond to $\Delta\vec{U}$ and are a more straightforward scatterometer measurement than the wind direction. We utilize the QSCAT (QuikScat scatterometer) wind fields over a two-year period (2000-2001). These are constructed 6-hourly on a half degree grid following

Chin et al. (1998). Figure 8 shows the ratio of the two year mean QSCAT speed to that from NCEP. A low bias is evident in the NCEP winds, with this ratio greater than 1 everywhere except a few isolated (unshaded) regions off the coasts of South America and Africa. The QSCAT winds are between 5 and 10% higher than NCEP winds over most of the mid-latitude ocean, and the ratio is considerably higher in select regions near the equator and the poles. To correct the NCEP wind speed bias, both reanalysis vector wind components are multiplied by the spatially-dependent factor plotted in Fig. 8. An important effect is to make the evaporation and turbulent heat fluxes more negative, thereby improving the global heat balance by 8 W/m^2, and the freshwater balance by 2.9 mg/m^2/s (Table 3).

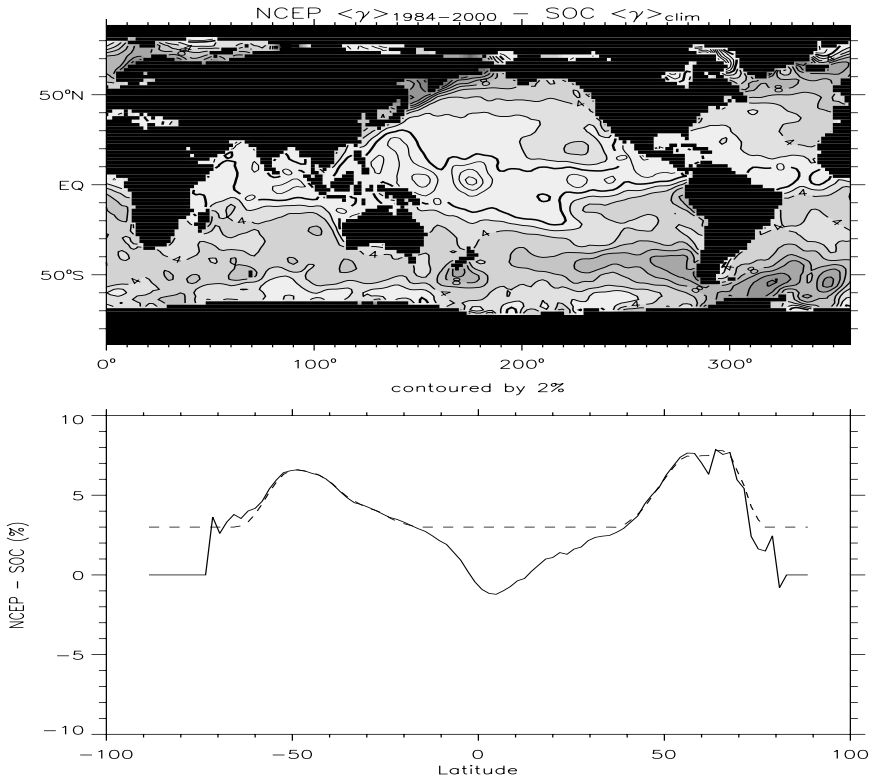

Figure 9. Comparison of NCEP/NCAR Reanalysis relative humidity with Southampton Oceanography Center (SOC) climatology as a mean difference (top panel) and as a zonally-averaged mean difference (bottom panel). Relative humidity values and differences are in %. The dashed line in the bottom panel shows the modified, smoothed version of the zonal-average difference used for correcting NCEP relative humidity as a function of latitude.

The near surface humidity in the NCEP reanalysis is too high (Fig. 9), with the relative humidity, γ, seldom less than 80%. Large et al. (1997) were able to match the Kent et al. (1993) observed annual cycle of latent heat flux in the North Atlantic only after reducing the NCEP specific humidity by a factor of 0.93. In the western tropical Pacific, a comparison of NCEP relative humidity to TAO mooring data in the equatorial Pacific shows a year-round positive bias of $\approx 2-3\%$ (Wang and McPhaden 2001). Plausible reasons for this bias are that ship based ocean humidity measurements that are assimilated into the reanalysis are more likely too wet than too dry (Kent et al. 1993), and that the NCEP model evaporates too much (Smith et al. 2001) and transports too little vertically out of the boundary layer. The excess evaporation is accomplished despite the low wind speeds and moist near surface by an excessively large C_E (Smith et al. 2001). The TAO data suggests reducing NCEP relative humidity by 3% in the tropics and this is used as a minimum correction at all latitudes. Figure 9 shows that NCEP relative humidity exceeds the SOC climatology almost everywhere outside the tropical band, with local differences exceeding 14%. The SOC humidities should be most reliable in the northern Hemisphere, because of the greater number of ship reports and the extensive work of Kent et al. (1993) in the North Atlantic. Therefore, in the north a smoothed version of the zonally averaged difference is used as a correction wherever it exceeds the minimum 3%. In the absence of any further information, and because of the rough symmetry in Fig. 9 about the equator, a similar correction is applied in the south, but with much less confidence. It is a maximum of about 6% at $50°S$. The net result is the NCEP bias $\delta\gamma$ shown as a function of latitude by the dashed line in Fig. 9. Overall, lowering the relative humidity increases the latent heat flux loss by $13W/m^2$ (Table 3).

Over most of the globe NCEP surface air temperatures are not corrected. However, comparison of NCEP temperatures with weather station and drifting buoy data from Antarctica reveals that a persistent very cold bias exists in the reanalysis product at extreme southern latitudes, especially in the winter. In the Arctic, the POLES (Polar Exchange at the Sea Surface) project has combined buoy and land station temperature data within an optimum interpolation scheme (Rigor et al., 2000). Over the Arctic cap north of $70°N$, the annual average POLES and NCEP air temperatures differ randomly by only about $\pm 0.5°C$, but NCEP air-temperatures are corrected monthly by the mean monthly climatological difference between POLES and NCEP. The 12 monthly corrections (January through December) are 0.49, 0.06, -0.73, -0.89, -0.77, -1.02, -1.99, -0.91, 1.72, 2.30, 1.81 and $1.06°C$.

Shortwave radiation is greatest at latitudes where there is a lot of ocean. Therefore, small tropical and sub-tropical errors in this field can significantly impact the global heat balance. Comparisons of Q_S such as Fig. 6, and the need to achieve a better global heat balance suggest reducing Q_I from ISCCP-FP by 5% for all latitudes from $50°S$ to $30°N$. For smoothness, the correction is linearly diminished between $50°S$ and $60°S$ and between $30°N$ and $40°N$. The resulting net shortwave, Q_S, is shown by the thick solid curve in Fig. 7. The global mean is reduced by 8 W/m^2 (Table 3).

With the above wind and humidity corrections and uncorrected GXGXS precipitation, the global ocean freshwater budget (4) is about -5 mg/m$^2/s$. To better balance the global oceanic water budget, both a gain (1.1417) and bias (0.7 mg/m$^2/s$) are applied to the GXGXS precipitation fields. The global mean freshwater flux then becomes -0.1 mg/m$^2/s$ (Table 3). This correction is designed to shift the GPCP curve of Fig. 8 into agreement with SOC and MSU at the 7° N peak in precipitation. By chance, the improved SOC agreement holds equatorward of about 35°. However, farther poleward the corrected precipitation becomes higher than SOC, but is still less than GPCP in both hemispheres, as favored by ocean model salinity results.

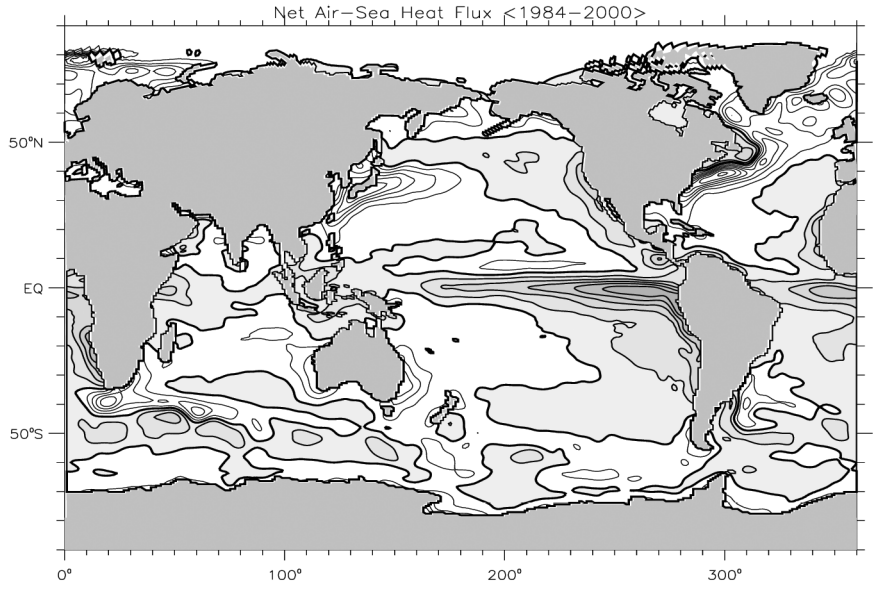

Figure 10. Climatological (1984-2000) mean total air-sea heat flux. The contour interval is 25 W/m^2, the zero contour is thicker and shaded regions indicate positive net flux into the ocean. A 5x5-point boxcar smoother has been applied twice after a uniform subtraction of the global mean imbalance of $\approx$ 1 W/m^2.

6.2 The mean flux fields and balances

The ocean surface fluxes forcing any GODAE system should not only produce mean net fields comparable to climatologies, but the component balances should also be similar. Significant departures may be indicators of errors in the data and/or the model. The climatological mean air-sea heat flux ($f_o\ Q_{as}$) computed from 1984 through 2000 is shown in Fig. 10. All the expected features are evident; strong heating along the equator with a maximum in the eastern Pacific in excess of $125 W/m^2$, a band of predominant heating along $50°S$, heating along the eastern boundaries of the subtropical gyres of the Pacific and Atlantic, strong cooling in the Nordic seas between Greenland and Europe and between Greenland and Labrador, and strong cooling over the western boundary currents and their extensions, including the Agulhas retroflection. The uncertainties in all fluxes climatologies means that there are always similarities and differences in detail. However, one large scale difference between Fig. 10 and the SOC climatology is the much greater SOC heating in the tropics from Africa east to the dateline. The area is great enough to factor significantly in the SOC global heat imbalance of $30 W/m^2$. It includes the TOGA COARE region of the western Pacific where SOC shows $60 W/m^2$ heating compared to less than $25 W/m^2$. Ocean budget (e.g. Gent, 1991) and observational estimates place the net heat flux at between 10 and $20 W/m^2$.

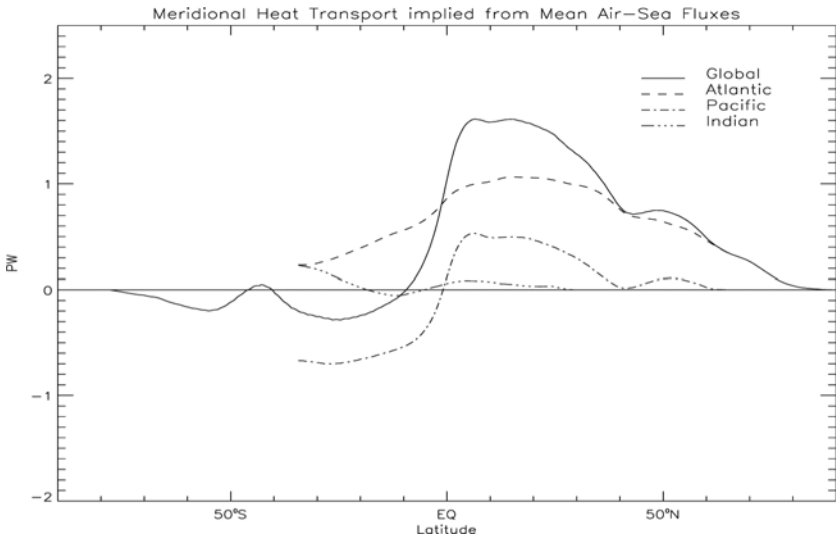

Figure 11. Implied meridional heat transport from the Fig. 10 heat flux.

Figure 11 shows the northward heat transport (global and by basin) implied by the air-sea heat fluxes of Fig. 10. In such a plot, the zonal average heat flux gives the slope of the curves (positive for heating, negative for cooling), while the value at a particular latitude represents the integrated heat flux from all ocean areas farther north, or equivalently in a balanced system, minus the integrated heat flux from all ocean areas farther south. The missing ocean heat loss via ice-ocean fluxes are believed to be only few W/m^2, but these would steepen the negative slopes at high latitudes and give larger poleward maxima in both hemispheres. Figure 11 suggests that the heating around $50°S$ is approximately equal to all the heat loss farther poleward, so that very little heat is transported across $40°S$. Similarly, there is little transport across $10°S$, because of nearly equal, but opposite transports in the Pacific and Atlantic. Note that the sum of Indian plus Pacific transport gives the correct Indo-Pacific implied transport, but partitioning between the two basins assumes no heat transport via the Indonesian Throughflow, which is not correct.

Figure 12 shows the mean (1984–2000) global distribution of air-sea freshwater flux (1984–2000), plus the climatological continental runoff from Fig. 1. The obvious freshwater source regions (unshaded in Fig. 12) are the intertropical convergence zones (ITCZs), the mid-latitude storm tracks of both hemispheres, and the mouths of large rivers. There is net water loss (evaporation) from the Arabian Sea and the subtropical gyres of each ocean basin. These features are common to all freshwater flux climatologies, although the magnitudes of the precipitation and evaporation differ substantially, with no way of determining reality. The zonal averages of the freshwater flux and its components are shown in Large and Yeager (2004). The runoff is most significant where Siberian rivers discharge between 60 and $70°N$.

Integration of fluxes in Fig. 12 gives the implied northward transport of freshwater shown in Fig. 13. Again partitioning between the Indian and Pacific incorrectly assumes no transport through the Indonesian Throughflow. The total transport can be compared to several other estimates and direct observations given be (Wijffels, 2001). In general there is better agreement with the direct observations in the northern Hemisphere than in the southern.

Climatologies (1984–2000) of the zonal and meridional wind stress components are shown in Fig. 14. The pattern and magnitudes of both components are remarkably similar to the SOC climatology, which has generally larger stress magnitudes than the COADS (Beranger, 1999). Since the SOC and COADS are based on essentially the same ship observations this difference is likely due to use of a different drag coefficient

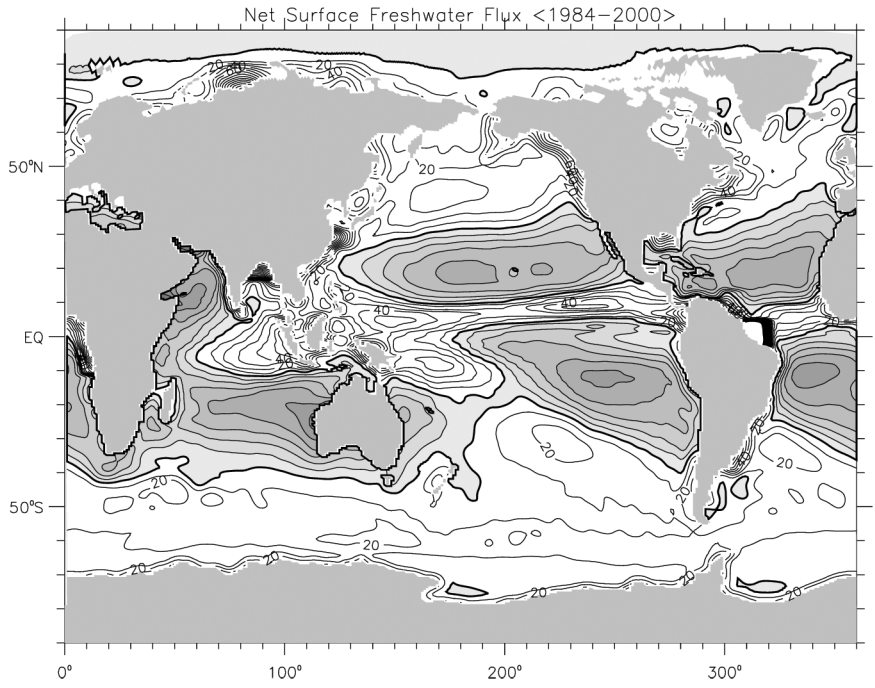

Figure 12. Climatological (1984-2000) mean total air-sea freshwater flux. Contour interval is 10 $mg/s/m^2$, up to ± 100. Unshaded regions indicate positive net water flux into the ocean. A boxcar smoother has been applied twice.

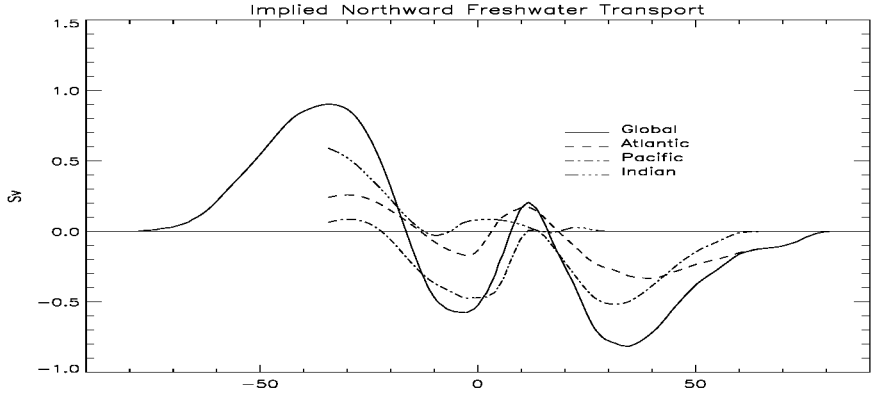

Figure 13. Implied meridional freshwater transport from the Fig. 12 freshwater flux.

formulation. It also appears that NCEP reanalysis employs a larger drag coefficient than Eq. (35), because a comparison (not shown) of NCEP stress and Fig. 14 does not show the differences implied by the wind speed correction of Fig. 8.

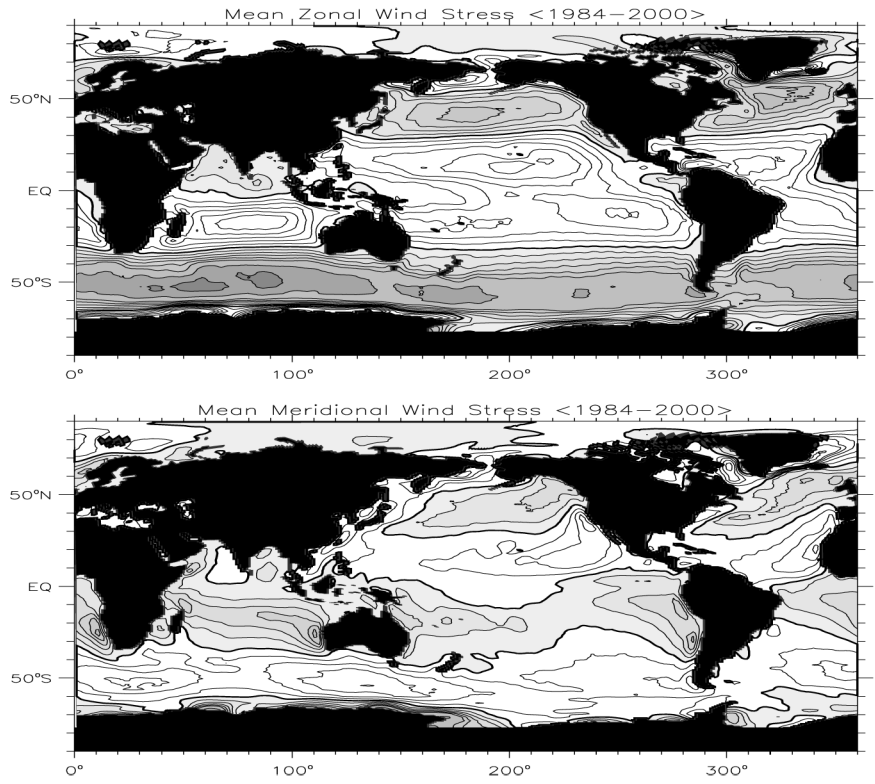

Figure 14. Climatological (1984-2000) mean zonal wind stress (top) and mean meridional wind stress (bottom) in N/m^2. Contour intervals are $\pm$(.02,.04,.06,.08,.1,.15,.2,.3) with a thick contour at 0. Positive (eastward and northward) stress values are shaded.

6.3 Interannual variability

Much of the observed interannual to decadal climate variability is displayed in Fig. 15, as time series of several large scale indices from 1950 - 2000. The North Atlantic Oscillation (NAO) and North Pacific (NP) indices are winter averages of December through March and November through March, respectively, but are plotted at mid-year. The El Nino - Southern Oscillation (ENSO) signals are the SST Anomaly from the Nino 3.4 region and the Southern Oscillation Index. These values and the Antarctic Oscillation (AAO) are 5 month running means of monthly values. All the indices display significant interannual and longer variability, that should also be prominent in incremented GODAE fluxes.

Large and Yeager (2004) computed annual means of all the fluxes for the 43 years from 1958 through 2000. On these time scales, different air-

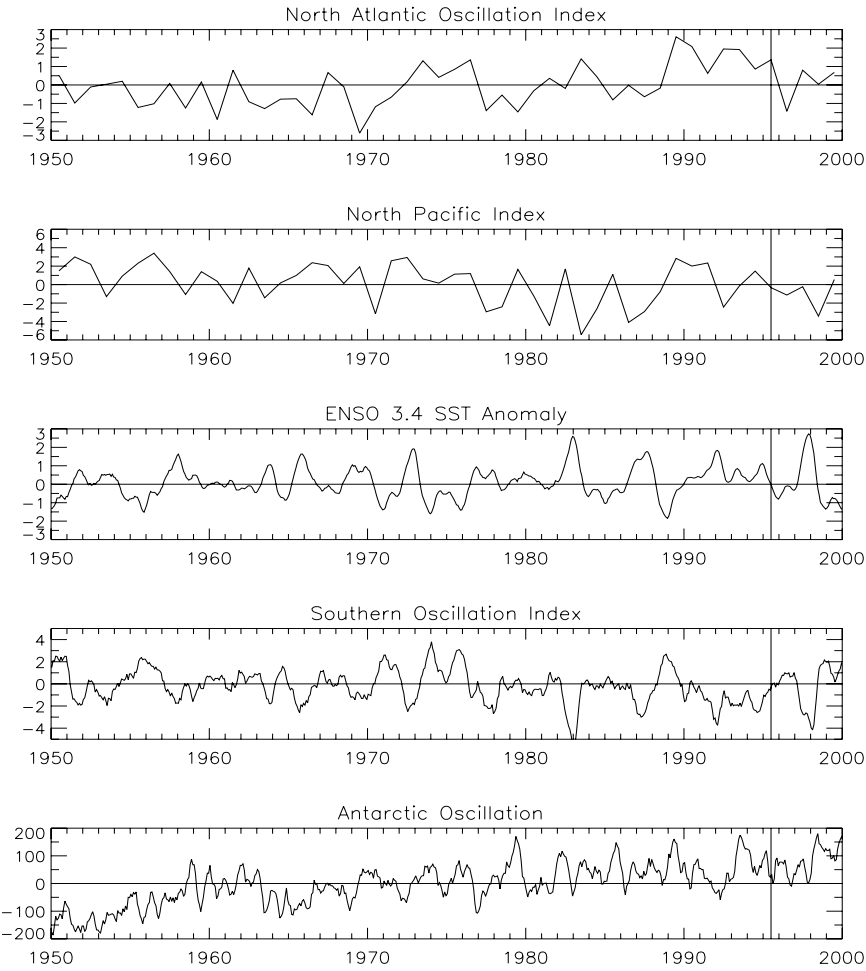

Figure 15. Interannual variability (1950-2000) of large scale climate indices.

sea interaction processes may be at work in the different ocean basins, so after removal of the climatological means, they constructed zonally averaged flux anomalies for the Indian, Pacific and Atlantic Oceans. An important caveat is that only the atmospheric state and SST datasets vary throughout the 43 years. Other dataset have been extended back in time using climatological annual cycles of monthly means from the period of time for which they are available. There is no variability in runoff.

The full variability of zonally averaged annual air-sea heat flux anomalies after 1983 is not noticeably different than the earlier partial variability suggesting that the radiation does not contribute much to interannual

variability even though it dominates the seasonal cycle. This notion is confirmed by the annual anomalies of Q_S. The maximum anomaly is only $-10W/m^2$. It is evident in 1992 between the equator and $20°S$ in the Pacific and Atlantic Oceans, and to a lesser extent in the Indian. This signal appears to be mostly a response to the Pinatubo volcanic eruption of late 1991. The interannual variability of the air-sea freshwater flux is dominated by fluctuations in tropical precipitation. Therefore, the interannual variability of freshwater forcing prior to the satellite era remains obscure, because the dominant signal is missing.

The interannual AAO variability is most clearly reflected in the zonal wind stress anomalies south of $40°S$ in all three basins, but especially the Indian ocean. In particular, weaker eastward stress in the 1960s corresponds to negative AAO, while the large positive AAO peaks in 1979, 1985, 1989, 1993 and 1998 are all associated with positive wind stress anomalies. Meridional anomalies are weaker, but positive northward wind stress anomalies tend to be associated with negative AAO (e.g. the 1960s). Interestingly, the changing winds are not consistently reflected in the heat flux anomalies.

El Nino/Southern Oscillation (ENSO) variability is reflected in the dominant tropical precipitation signals, with annual mean anomalies from years following warm events being strongly positive in the Pacific, weakly positive in the Indian, and negative in the Atlantic. The heat fluxes have a corresponding signal only in the Pacific where they respond directly to warmer SST by becoming more upward (negative). Thus, they retard ENSO warming and cooling, as they do in a forced ocean hindcast (e.g. Doney et al., 2003).

Zonal wind stress anomalies, and to a lesser extent meridional anomalies in the North Atlantic poleward of about $40°N$ reflect the North Atlantic Oscillation (NAO), with positive stress when the NAO index is high. Corresponding negative heat flux anomalies (more latent and sensible cooling when the westerly winds are stronger) are evident, but less pronounced during the period of generally high NAO index during the 1990s.

Neither the heat flux, nor wind stress anomalies in the North Pacific appear to be directly related to the North Pacific Index (NPI). Low values of the index correspond to both positive (e.g. 1970) and negative (e.g. 1981 and 1983) zonal stress variations. The heat flux anomalies tend to follow the wind and are, therefore, of the opposite sign. There is no persistent anomaly in either the heat flux, or zonal stress through the period of high NPI from 1989 through 1991.

7. Discussion and conclusion

The uncertainties in air-sea fluxes are difficult to quantify, with the lack of direct measurements a major problem, but they decrease as the time and space scales expand. Best known are the global, long term mean heat and freshwater fluxes, which ocean inventories show are near zero. At the other extreme are turbulent fluxes whose sign at a point can be uncertain on time scales less than about 10 minutes. Perhaps most relevant to GODAE are hourly fluxes on a spatial scale of 10 km or more. Some of these might be tractable from satellite measurements, but whenever bulk aerodynamic formulae are involved there is at least a factor of 2 uncertainty due to transfer coefficient variability on these scales.

The temptation for GODAE to simply utilize global fields of Q, F and wind stress should be resisted, because too much good physics is lost. Notable examples include the penetration of solar radiation into the upper ocean, the relationship between evaporation and the latent heat flux, sea-ice coverage and the strong dependence of fluxes on SST. However, the price is high because of the large number of forcing fields necessary to prescribe the flux components.

Unfortunately, there is no single flux data suitable for GODAE. Instead, various data sets should be merged within an assimilation data stream. However, a crucial procedure within the stream should be the correction of most fields for known biases, as revealed by comparisons with more limited, but more accurate and better understood observations. These adjustments should result in global heat and freshwater fluxes that are nearly in balance, and these budgets should not be overly upset by any increments made to the fluxes during the assimilation cycle. In addition the incremented fluxes ought to display the observed features of global variability, especially those portrayed by large scale climate indices.

The forcing of a truly global ocean data assimilation system that incorporates all the details of the known physics would be extremely complicated. Therefore, simplifications should be sought, but implemented only if well understood and demonstrated to be tolerable from the viewpoint of assimilation products. A prime example is high latitude forcing where representation of much of the physics requires a coupled sea-ice model. Another is forcing near coastlines, where some satellite products are contaminated by the land and continental runoff is complicated. Also, some fluxes may not need to be accounted for; such as the temperature of precipitation and runoff, the latent heat of fusion of glacier and ice-sheet runoff, and the melting of snow.

References

Baumgartner, A. and E. Reichel, 1975: The World Water Balance. Elsevier, New York, 180pp.

Beranger, K., K. Viau, B. Barnier, E. Garnier, J.M. Molines, and L. Siefridt, 1999: An atlas of climatic estimates of air-sea fluxes. MEOM, Laboratoire des Ecoulements Geophysiques et Industriels, Institute de Mechanique de Grenoble, 19pp + figures.

Bradley, E.F. A shearing stress meter for micro-meteorological studies. *Quart. J. Roy. Met. Soc.*, **94**, 380–387.

Busch, N.E., 1977: Fluxes in the surface boundary layer over the sea. Modelling and Prediction of the Upper Layers of the Oceans, E.B. Krauss (ed.), Pergamon Press, Oxford, 72–91.

Charnock, H., 1955: Wind stress on a water surface. *Quart. J. Roy. Met. Soc.*, **81**, 639–640.

Chin, T.M. R.F. Milliff and W.G. Large, 1998: Basin-scale high-wavenumber sea surface wind fields from multiresolution analysis of scatterometer data. *J. Atmos. Oceanic Technol.*, **15**, 741–763.

Curry, JA., A. Bentamy, M.A. Bourassa, D. Bourras, E.F. Bradley, M. Brunke, S. Castro, S.H. Chou, C.A. Clayson, W.J. Emery, L. Eymard, C.W. Fairall, M. Kubota, B. Lin, W. Perrie, R.A. Reeder, I.A. Renfrew, W.B. Rossow, J. Schulz, S.R. Smith, P.J. Webster, G.A. Wick and X. Zeng, 2002: Seaflux. *Bull. Amer. Meteor. Soc.*, **85**, 409–424.

Darnell, W.L., W.F. Staylor, N.A. Richey, S.K. Gupta and A.C. Wilber, 1996: Surface radiation budget: A longterm global dataset of shortwave and longwave fluxes. *Eos, Trans. Amer, Geophys. Union.* February.

da Silva, A., C. Young and S. Levitus, 1994: Atlas of surface marine data 1994. NOAA Atlas NESDIS 6. (6 volumes), U.S. Dept. of Commerce, NODC, User services branch, NOAA/NESDIS/E/OC21.

Deacon, E.L., 1973: Geostrophic drag coefficients. *Boundary-Layer Meteorol.*, **5**, 321–340.

Doney, S.C., S.G. Yeager, G. Danabasoglu, W.G. Large and J.C. McWilliams, 2003: Modeling global oceanic interannual variability (1958–1997): Simulation design and model-data evaluation. NCAR Tech. Note, TN-452+STR, 48pp.

Fekete, B.M., C.J. Vorosmarty and W. Grabs, 1999: An improved spatially distributed runoff data set based on observed river discharge and simulated water balance. Complex Systems Research Center, U. New Hampshire.

Freilich, M.H. and R.S. Dunbar, 1999: On the accuracy of the NSCAT 1 vector winds: Comparisons with National Data Buoy Center buoys. *J. Geophys. Res.*, **104**, 11231–11246.

Fung, I.Y., D.E. Harrison and A.A. Lacis, 1984: On the variability of the net longwave radiation at the ocean surface. *Rev. of Geophys.*, **22**, 177–193.

Garratt, J.R., 1977: Review of drag coefficients over the oceans and continents. *Mon. Weather Rev.*, **105**, 915–929.

Gent, P., 1991: The heat budget of the TOGA-COARE domain in an ocean model. *J. Geophys. Res.*, **96**, 3323–3330.

Gibson, J.K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura and E. Serrano, 1997: ECMWF re-analysis project, 1. ERA description. ECMWF, Project Report Series.

Gilgen, H.C., C.H. Whitlock, F. Koch, G. Mueller, A. Ohmura, D. Steiger and R. Wheeler, 1995: Technical plan for Baseline Surface Radiation Network (BSRM) Data Management. WCRP, WMO TD-No. 443, WMO, Geneva.

Haugen, D.A, J.C. Kaimal and E.F. Bradley, 1971: An experimental study of Reynolds stress and heat flux in the atmospheric surface layer. *Quart. J. Roy. Met. Soc.*, **97**, 168-180.

Högström, U., 1988: Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation. *Boundary-Layer Meteorol.*, **42**, 55–78.

Huffman, G.R., R.F. Adler, P. Arkin, A. Chang, R. Ferraro, R. Gruber, J. Janowiak, A. McNab, B. Rudolf and U. Schneider, 1997: The global precipitation climatology project (GPCP) combined precipitation data set. *Bull. Amer. Meteor. Soc.*, **78**, 5-20.

Josey, S., C.E. Kent and P.K. Taylor, 1998: The Southampton Oceanography Centre (SOC) ocean-atmosphere heat, momentum and freshwater flux atlas. Report No.6, Southampton Oceanography Centre, 30pp.

Kalnay, E., M. Kanamitsu,R. Kistler, W. Collins, D. Deaven, L. Gandin , M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu , M. Chelliah, W. Ebisuzaki , W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski , A. Leetmaa, R. Reynolds and R. Jenne, 1996: The NCEP/NCAR 40-Year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

Kent, E.C., P.K. Taylor, B.S. Truscott and J.S. Hopkins, 1993: The accuracy of voluntary observing ships' meteorological observations: Results of the VSOP-NA. *J. Atmos. Oceanic Technol.*, **10**, 591–608.

Knox, J.L., 1991: An assessment of the 27-year record of measured precipitation at Ocean Weather Station "P" in the northeast Pacific Ocean. *Clim. Bull. Canadian. Met. Ocean. Soc.*, **25**, 65–80.

Large, W.G. and S. Pond, 1981: Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr.*, **11**, 324–336.

Large, W.G. and S. Pond, 1982: Sensible and latent heat flux measurements over the ocean. *J. Phys. Oceanogr.*, **12**, 464–482.

Large, W.G., G. Danabasoglu, S.C. Doney and J.C. McWilliams, 1997: Sensitivity to surface forcing and boundary layer mixing in a global ocean model: Annual-mean climatology. *J. Phys. Oceanogr.*, **27**, 2418–2447.

Large, W.G. and S.G. Yeager, 2004: Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies. NCAR Technical Note, NCAR/TN-460+STR, 105pp.

Levitus, S., T. Boyer, M. Conkright, D. Johnson, T. O'Brien, J.Antonov, C. Stephens and R. Gelfeld. World ocean database 1998. NOAA, NESDIS Atlas.

Levitus, S., J. I. Antonov, T. P. Boyer and C. Stephens, 2000: Warming of the World Ocean. *Science*, **287**, 2225–2229.

Li, F., W.G. Large, W. Shaw, E. Walsh, and K. Davidson).1989: Ocean radar backscatter relationship with near surface winds: A case study during FASINEX. *J. Phys. Oceanogr.*, **19**, 342-353.

Lind, R. and K. Katsaros, 1986: Radiation measurements and model results from R/V Oceanographer during STREX 1980. *J. Phys. Oceanogr.*, **91**, 13308–13314.

Lumley, J.L. and H.A. Panofsky, 1964: The Structure of Atmospheric Turbulence. Interscience, New York, 239pp.

Paulson, C.A., 1970: Representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteorol.*, **9**, 857–861.

Payne, R.E., 1972: Albedo of the sea surface. *J. Atmos. Sci.*, **29**, 959–970.

Perry, G., P.B. Duffy and N.L. Miller, 1996: An extended data set of river discharges for validation of general circulation models. *J. Geophys. Res.*, **101**, 21339–21349.

N.A. Rayner, D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander and D.P. Powell, 2003, Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407.

Rigor, I., R. Colony and S. Martin, 2000: Variations in surface air temperature observations in the Arctic 1979–97. *J. of Climate*, **13**, 896–914.

Roll, H.V., 1965: Physics of the Marine Atmosphere. Academic Press, 426pp.

Schmidt, G.A., C.M. Bitz, U. Mikolajewicz and L.B. Tremblay, 2004: Ice-ocean boundary conditions for coupled models, *Ocean Modelling*, **7**, 59–74.

Serreze, M.C., and C.M. Hurst, 2000: Representation of mean Arctic precipitation from NCEP-NCAR and ERA reanalyses. *J. of Climate*, **13**, 182–201.

Smith, S. and F.W. Dobson, 1984: The heat budget at Ocean Weather Station Bravo. *Atmos. Oceans*, **22**, 1-22.

Smith, S.R., D.M. Legler and K.V. Verzone, 2001: Quantifying uncertainties in NCEP reanalyses using high-quality research vessel observations *J. of Climate*, **14**, 4062–4072.

Spencer, R.W., 1993: Global oceanic precipitation from the MSU during 1979–91 and comparisons to other climatologies. *J. of Climate*, **6**, 1301–1326.

Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C.N. Hill and J. Marshall, 2002: The global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.*, **107**, 3118.

Stewart, R.W., 1974: The air sea momentum exchange. *Boundary-Layer Meteorol.*, **6**, 151–167.

Tennekes, H., 1973 : The logarithmic wind profile. *J. Atmos. Sci.*, **30**, 234–238.

Trenberth, K.E., W.G. Large and J.G. Olson, 1989: The effective drag coefficient for evaluating wind stress over the ocean. *J. of Climate***12**, 1508–1516.

UNESCO, 1985: Discharge of selected rivers of the world, vol. I, II, III (parts I, II, III, IV), UNESCO, Paris.

Wang, W. and M.J. McPhaden, 2001: What is the mean seasonal cycle of surface heat flux in the equatorial Pacific? *J. Geophys. Res.*, **106**, 837–857.

WGASF, 2000: Intercomparison and validation of ocean-atmosphere energy flux fields. Final Report of the Joint WCRP/SCOR Working Group on Air-Sea Fluxes. P.K. Taylor (Ed.), 358pp.

Wijffels, S.E., 2001: Ocean transport of freshwater. In "Ocean Circulation and Climate", Eds. G. Siedler, J. Church and J. Gould, International Geophysics Series, **77**, Academic Press, 475–488.

Xie, P. and P.A. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. of Climate*, **9**, 840–858.

Yelland, M.J. and P.K. Taylor, 1996: Wind stress measurements from the open ocean. *J. Phys. Oceanogr.*, **26**, 541–558.

Zhang, Y.C., W.B. Rossow, A.A. Lacis, V. Oinas and M.I. Mishchenko, 2004: Calculation of radiative flux profiles from the surface to top-of-atm osphere based on ISCCP and other global datasets: Refinements of the radiative t ransfer model and the input data. *J. Geophys. Res.*, **109**, In Press.

Chapter 10

OCEAN DATA ASSIMILATION USING SEQUENTIAL METHODS BASED ON THE KALMAN FILTER

From theory to practical implementations

Pierre Brasseur
CNRS/LEGI, Grenoble, France

Abstract: The main purpose of this chapter is to review the fundamentals of the Kalman Filter for ocean data assimilation and to expose the basic ingredients of practical assimilation algorithms developed for applied ocean research and operational forecasting, focusing mainly on high-resolution applications. Important implementation issues such as the reduction in dimensionality of the estimation problem, the simplification of the schemes based on static error covariances, the formulation of low-rank filters, the problem of consistency verification, and the concepts of adaptivity and incremental analysis updating will be addressed using scientific and operational examples. Finally, the discussion will conclude with a number of key questions related to the assimilation challenges of the next decade.

Keywords: Data assimilation, mesoscale ocean circulation, reduced-order Kalman filters, statistical estimation, ocean forecasting.

1. Introduction

Operational ocean prediction systems are being developed with a variety of objectives in mind, such as ocean current hindcasting and short-range forecasting, estimation of the thermodynamic state of the ocean for seasonal and climate predictions, and production of retrospective analyses of the changing ocean through the merging of models and data. In addition to simply combining model estimates with observations, data assimilation provides a means to systematically compare theoretical

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 271-316.

models with reality, leading to potential improvements in modelling and observing systems. It is therefore likely that applied ocean research will benefit strongly from operational progress, and vice versa. The question concerning the dominant energetic activity of the mesoscale ocean, its non-deterministic nature and the interactions with the large-scale circulation make the challenge unique, requiring sophisticated numerical models and assimilation methods that make the best use of sparse observations. To produce reliable forecasts, the models must be initialized with conditions that represent as accurately as possible the actual state of the ocean at eddy-resolving resolution. Due to the chaotic properties of ocean dynamics, the forecast range cannot be extended beyond the limit of predictability of the system and the model has to be re-initialized intermittently by correcting the forecast with the most recent observations. Fortunately, the arrival of satellite observations, in particular satellite altimetry, has provided the observational basis needed to respond appropriately to the "high-resolution challenge". In order to extract the best possible information from the new data, it is necessary to assess how reliable the model forecast and the observations are. Therefore, error estimates on the measurements and the model prediction are inherent in the assimilation process.

Data assimilation is traditionally formulated as a least-squares estimation problem. Among the various methodological approaches, the theory of optimal statistical estimation, and more specifically the Kalman filtering approach, is well suited to provide a solution to the Best Linear Unbiased Estimation. Since *Kalman* in 1960, sequential filtering methods have been thoroughly explored and applied to state estimation. An extended version of the Kalman Filter (KF) has been derived for non-linear models, known as the Extended Kalman Filter (EKF) [*Jazwinski*, 1970; *Gelb*, 1974]. In spite of a fairly simple theoretical framework, the question of its applicability in assimilating observations into high-resolution, non-linear numerical models of the ocean circulation is far from trivial. As stated by *Courtier* [1997], the scientific difficulty associated with data assimilation is in finding algorithms which simplify the search for an affordable solution in terms of computer resources, while preserving some of the essential characteristics. A hierarchy of approximations to the Kalman filter has been defined to make the methodology suitable for solving large-dimension problems. These developments represent a substantial part of the research effort devoted to oceanic and atmospheric data assimilation over the past 10 years.

Compared with variational approaches such as the 4D-VAR, statistical algorithms require less initial investment in terms of coding and are naturally designed to incorporate gradual developments. This is one of

several reasons why most assimilation methods used today in operational forecasting systems are inspired by the statistical approach. In the context of GODAE (Global Ocean Data Assimilation Experiment), one objective will be to test and compare the bunch of algorithms implemented in operational systems, in order to better understand the importance of the various possible approximations made in each of these.

The objectives of this chapter are to review the fundamentals of sequential data assimilation for ocean state estimation and to expose the basic ingredients of practical assimilation algorithms developed for applied ocean research and operational systems, focusing mainly on high-resolution applications. Section 2 is dedicated to the fundamentals of applied estimation methods leading to the KF equations. Numerical representations of the mathematical objects introduced by the theory will then be illustrated using oceanographic examples in Section 3. Section 4 will provide a brief description of traditional simplifications of the KF. In Section 5, we discuss various approaches to reduce the size of the estimation problem and, in Section 6, we derive the framework of low-rank Kalman filters. The important question of the verification of consistency will be addressed in Section 7, where the concept of adaptivity is also mentioned. Finally, a number of advanced implementation issues such as the transition to incremental/smoothing algorithms will be discussed in Section 8, before concluding the chapter.

2. Kalman filtering: Fundamentals

2.1 Problem definition

In this section, we introduce the basic assimilation problem in the state space using the conventional notations proposed by *Ide et al.* [1997]. The goal here is not to present a rigorous and comprehensive derivation of the Kalman filter, which can be found elsewhere in dedicated text books (e.g., *Gelb* [1974]), but rather to introduce a simplified framework that still contains the essential characteristics needed to illustrate the more advanced concepts and implementation issues discussed in the following sections.

To start, let us assume that some *a priori* knowledge about the state of the ocean is available at time t_i, represented by vector $\mathbf{x}_i^a$. A physical model is also available to describe the transition of the state vector from time t_i to time t_{i+1}, which is represented by a numerical (matrix) operator $\mathbf{M}(t_i, t_{i+1})$. A linear model will be considered at this stage. The dimension of the state space is noted as n. The state vector $\mathbf{x}$ contains the minimum set of independent variables needed to perfectly characterize the state of the system at any time. The model can be used

to simulate the transition of the state vector up to time t_{i+1},

$$\mathbf{x}_{i+1}^f = \mathbf{M}(t_i, t_{i+1})\mathbf{x}_i^a \tag{1}$$

where $\mathbf{x}_{i+1}^f$ is the vector describing the "forecast" state of the system. At instant t_{i+1}, another piece of useful information is available about the state, collected in the observation vector $\mathbf{y}_{i+1}$ of dimension p. From the two independent pieces of information $\mathbf{x}_{i+1}^f$ and $\mathbf{y}_{i+1}$, how can the true state of the system $\mathbf{x}_{i+1}^t$ be best estimated at time t_{i+1}? To answer this question, we need to know more about the precision of the different pieces of information.

2.2 Uncertainties and PDFs

The precision of the forecast $\mathbf{x}_{i+1}^f$ can be quantified in terms of errors on the initial guess and numerical model errors. The difference between the initial guess $\mathbf{x}_i^a$ and the true state at time t_i is the error vector noted $\epsilon_i^a = \mathbf{x}_i^a - \mathbf{x}_i^t$. Of course, its value is unknown but we can make a number of assumptions about its statistical properties: we will assume that the estimation $\mathbf{x}_i^a$ is unbiased ($\overline{\epsilon_i^a} = 0$, where the overbar represents the expected value), and its error ϵ_i^a is distributed as a gaussian, multivariate random variable. The corresponding probability density function (pdf) is

$$\epsilon_i^a \rightarrow N(0, \mathbf{P}_i^a) \sim \exp\left[-\frac{1}{2}\epsilon_i^{a^T}\mathbf{P}_i^{a^{-1}}\epsilon_i^a\right] \tag{2}$$

where $\mathbf{P}_i^a = \overline{\epsilon_i^a \epsilon_i^{a^T}}$ is the $n \times n$ error covariance matrix associated with $\mathbf{x}_i^a$ and ^{T} denotes the transpose. Error covariances are formally obtained by multiplying an error vector by its transpose and averaging over many realizations, leading to symmetric and positive definite matrices. *Morrison* [1988] contains excellent background information on multivariate statistical methods. Similarly, the model operator $\mathbf{M}(t_i, t_{i+1})$ is imperfect and the simulation error is noted as follows:

$$\eta = \mathbf{M}(t_i, t_{i+1})\mathbf{x}_i^t - \mathbf{x}_{i+1}^t \tag{3}$$

Again, the individual realization of this error is unknown (otherwise a perfect model operator could be run) but its statistical distribution is assumed to be gaussian and centered ($\overline{\eta} = \mathbf{0}$):

$$\eta \rightarrow N(0, \mathbf{Q}) \sim \exp\left[-\frac{1}{2}\eta^T\mathbf{Q}^{-1}\eta\right] \tag{4}$$

where $\mathbf{Q} = \overline{\eta\eta^T}$ is the $n \times n$ model error covariance matrix. We assume in addition that ϵ_i^a and η are uncorrelated: $\overline{\epsilon_i^a \eta^T} = 0$. In general, these

statistical assumptions are quite crude approximations of the actual error distributions (in particular, a bias in the model is very common), but they are very convenient in deriving a baseline of optimal estimation. A schematic description of the error diagram in the state space is illustrated in figure 1.

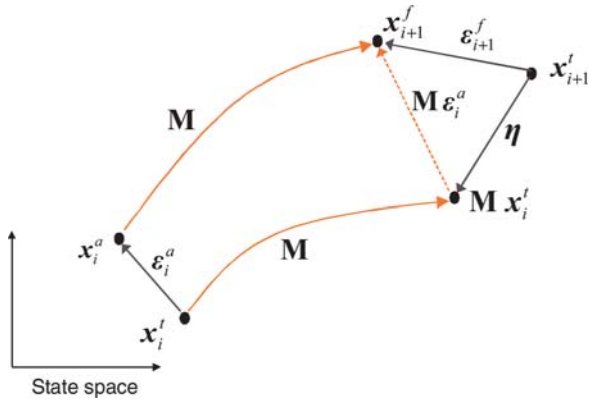

Figure 1. Vector diagram of analysis and forecast errors in the state space.

With the definitions introduced above, the forecast error ϵ^f_{i+1} can be broken down as :

$$\epsilon^f_{i+1} = \mathbf{x}^f_{i+1} - \mathbf{x}^t_{i+1} = \mathbf{M}\mathbf{x}^a_i - (\mathbf{M}\mathbf{x}^t_i - \eta) = \mathbf{M}\epsilon^a_i + \eta \tag{5}$$

and the statistical properties of the forecast error can be determined easily if the model is linear. Indeed, Eqs. (5), (2) and (4) implies that the forecast state is unbiased ($\overline{\epsilon^f_{i+1}} = \mathbf{M}\overline{\epsilon^a_i} + \overline{\eta} = 0$) and normally distributed, i.e.

$$\epsilon^f_{i+1} \to N(0, \mathbf{P}^f_{i+1}) \sim \exp\left[-\frac{1}{2}\epsilon^{f^T}_{i+1}\mathbf{P}^{f^{-1}}_{i+1}\epsilon^f_{i+1}\right] \tag{6}$$

with the forecast error covariance matrix given by:

$$\mathbf{P}^f_{i+1} = \overline{\epsilon^f_{i+1}\epsilon^{f^T}_{i+1}} = \mathbf{M}\overline{\epsilon^a_i\epsilon^{a^T}_i}\mathbf{M}^T + \overline{\eta\eta^T} = \mathbf{M}\mathbf{P}^a_i\mathbf{M}^T + \mathbf{Q} \tag{7}$$

This equation is the first fundamental equation of the KF which can be interpreted as follows: the error on the initial state is transformed during the forecast step by the model dynamics (the error being amplified by unstable modes, while it is damped by stable modes) and by the model

imperfections which increase the forecast error covariance. Understanding the actual benefits and the practical limitations of error covariance propagation through model dynamics has been a major research issue over recent years. Although simple in its algebraic form, this equation contains several major difficulties that prevent an explicit computation in the context of realistic models, as discussed in the following sections.

In order to optimally combine the forecast with new data, the precision of the observations arising at time t_{i+1} must be quantified. The vector of observations is related to the true state as follows:

$$\mathbf{y}_{i+1} = \mathbf{H}\mathbf{x}^t_{i+1} + \epsilon^o_{i+1} \tag{8}$$

where $\mathbf{H}$ is the observation operator which computes the equivalent of the observations from the model state. The observational errors ϵ^o_{i+1} are assumed to be centered ($\overline{\epsilon^o_{i+1}} = 0$), uncorrelated with the forecast error and having a covariance matrix $\mathbf{R} = \overline{\epsilon^o_{i+1}\epsilon^{oT}_{i+1}}$. Observation errors measure the misfit between the data and the equivalent of the observations in the true state, i.e. $\mathbf{H}\mathbf{x}^t_{i+1}$. They include not only the errors of the observational system but also the errors associated with the operator $\mathbf{H}$ arising, for example, from the numerical interpolation of the data. Again, a gaussian pdf can be assumed for the statistical distribution of the observation errors

$$\epsilon^o_{i+1} \rightarrow N(0, \mathbf{R}) \sim \exp\left[-\frac{1}{2}\epsilon^{oT}_{i+1}\mathbf{R}^{-1}\epsilon^o_{i+1}\right] \tag{9}$$

to make the rest of the development easier and derive the KF equations.

2.3 Optimal analysis

The pdf given by (6) determines the *a priori* statistical distribution of the true state $P(\mathbf{x}^t_{i+1})$, while (9) provides the probability of getting measurements $\mathbf{y}_{i+1}$ given the true state, i.e. $P(\mathbf{y}_{i+1} \mid \mathbf{x}^t_{i+1})$. It is then straightforward to deduce the *a posteriori* probability of the truth, given the observations, by using the Bayes formula:

$$P(\mathbf{x}^t_{i+1} \mid \mathbf{y}_{i+1}) = \frac{P(\mathbf{y}_{i+1} \mid \mathbf{x}^t_{i+1}) \cdot P(\mathbf{x}^t_{i+1})}{P(\mathbf{y}_{i+1})} \tag{10}$$

The state maximizing the posterior probability distribution is the maximum likelihood solution of this inverse problem. A comprehensive formulation of data assimilation and inverse methods using the bayesian approach has been proposed by *van Leeuwen and Evensen* [1996]. In Eq. (10), the denominator is just a scaling factor (the integral of the

numerator for all possible states) which can be ignored in determining the maximum of the posterior pdf. The gaussian distributions (6) and (9) imply that

$$P(\mathbf{y}_{i+1} \mid \mathbf{x}^t_{i+1}) \,.\, P(\mathbf{x}^t_{i+1}) \sim \exp\left[-\frac{1}{2}\left\{\epsilon^{oT}_{i+1}\mathbf{R}^{-1}\epsilon^{o}_{i+1} + \epsilon^{f\,T}_{i+1}\mathbf{P}^{f^{-1}}_{i+1}\epsilon^{f}_{i+1}\right\}\right] \tag{11}$$

and the optimal estimation of $\mathbf{x}^t_{i+1}$ is the state vector maximizing (11) or, equivalently, minimizing

$$J = \left\{\epsilon^{oT}_{i+1}\mathbf{R}^{-1}\epsilon^{o}_{i+1} + \epsilon^{f\,T}_{i+1}\mathbf{P}^{f^{-1}}_{i+1}\epsilon^{f}_{i+1}\right\} \tag{12}$$

As a result of error definitions, the optimal combination of the forecast and observed information corresponds to the minimum of the cost function

$$J(\mathbf{x}) = (\mathbf{y}_{i+1} - \mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}_{i+1} - \mathbf{H}\mathbf{x}) + (\mathbf{x}^f_{i+1} - \mathbf{x})^T\mathbf{P}^{f^{-1}}_{i+1}(\mathbf{x}^f_{i+1} - \mathbf{x}). \tag{13}$$

This quadratic form contains two terms measuring the misfit with the data and the misfit with the forecast, weighted by the inverse of their respective error covariances. Using the calculus of variations, an implicit equation for the optimal state noted $\mathbf{x}^a_{i+1}$ is obtained:

$$\delta J(\mathbf{x}) = 0 \Rightarrow \mathbf{x}^a_{i+1} = \mathbf{x}^f_{i+1} + \mathbf{P}^f_{i+1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}_{i+1} - \mathbf{H}\mathbf{x}^a_{i+1}) \tag{14}$$

which can be solved explicitly after some algebra, yielding

$$\mathbf{x}^a_{i+1} = \mathbf{x}^f_{i+1} + \mathbf{P}^f_{i+1}\mathbf{H}^T(\mathbf{H}\mathbf{P}^f_{i+1}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}_{i+1} - \mathbf{H}\mathbf{x}^f_{i+1}) \tag{15}$$

The optimal state is obtained by correcting the forecast with a weighted measure of the misfit between the observations and the prior estimate (i.e. the innovation vector $\mathbf{d}_{i+1} = \mathbf{y}_{i+1} - \mathbf{H}\mathbf{x}^f_{i+1}$). The analysis is thus simply the result of the combination of two Gaussian probability density functions. The weight matrix of dimensions $n \times p$

$$\mathbf{K}_{i+1} = \mathbf{P}^f_{i+1}\mathbf{H}^T(\mathbf{H}\mathbf{P}^f_{i+1}\mathbf{H}^T + \mathbf{R})^{-1} \tag{16}$$

is the so-called Kalman gain, which involves the forecast and observation error covariance matrices. It can be interpreted as a ratio between the error variance of the forecast and the total error variance (the sum of the forecast and the observation error variance) projected in the observation space: the larger the forecast errors, the larger the correction to the forecast. In the limit of perfect observations ($\mathbf{R} \sim \mathbf{0}$) of the whole state vector ($\mathbf{H} \sim \mathbf{I}$), the Kalman gain matrix converges towards the identity

and the optimal estimate becomes a perfect fit of the observations. By contrast, if the forecast is extremely accurate ($\mathbf{P}^f \sim \mathbf{0}$) compared with the observations, the correction is negligible. Interesting similarities between this equation and scalar formulations of least squares problems can be found in *Kalnay* [2003]. Equation (15) is the second fundamental equation of the Kalman filter.

2.4 The sequential assimilation cycle

The optimal state estimate (15) at time t_{i+1} can be used as the initial conditions for a new forecast up to time t_{i+2} when new observations become available, and the process can be repeated recursively. To sum up, the algorithm of an assimilation cycle contains two main steps: the forecast step for transitioning the model state and the associated error covariance between time t_i and time t_{i+1}, and the analysis step for correcting the forecast using the data available at time t_{i+1}. We reproduce here the complete set of the KF equations extended to non-linear models M and observation operators H.

Starting from initial conditions $\mathbf{x}_i^a$ and $\mathbf{P}_i^a$, the forecast step equations are:

$$\mathbf{x}_{i+1}^f = M(t_i, t_{i+1})\left\{\mathbf{x}_i^a\right\} \tag{17}$$

and

$$\mathbf{P}_{i+1}^f = \mathbf{M}\mathbf{P}_i^a\mathbf{M}^T + \mathbf{Q} \tag{18}$$

where $\mathbf{M}$ is the tangent linear operator derived from $M(t_i, t_{i+1})$. Thus, a linearization of the model about the non-linear evolution between t_i and t_{i+1} is performed to propagate the error covariance.

The forecast step is followed by an analysis step in which $\mathbf{y}_{i+1}$ is used to correct $\mathbf{x}_{i+1}^f$:

$$\mathbf{x}_{i+1}^a = \mathbf{x}_{i+1}^f + \mathbf{K}_{i+1}\left(\mathbf{y}_{i+1} - H\left\{\mathbf{x}_{i+1}^f\right\}\right) \tag{19}$$

using the Kalman gain

$$\mathbf{K}_{i+1} = \mathbf{P}_{i+1}^f\mathbf{H}^T[\mathbf{H}\mathbf{P}_{i+1}^f\mathbf{H}^T + \mathbf{R}]^{-1} \tag{20}$$

where $\mathbf{H}$ is the gradient of H computed about $\mathbf{x}_{i+1}^f$. It can be demonstrated that the matrix $\mathbf{K}_{i+1}$ corresponds to the minimization of the trace of the analysis error covariance on $\mathbf{x}_{i+1}^a$ [*Miller* 1989], given by

$$\mathbf{P}_{i+1}^a = \mathbf{P}_{i+1}^f - \mathbf{P}_{i+1}^f\mathbf{H}^T[\mathbf{H}\mathbf{P}_{i+1}^f\mathbf{H}^T + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P}_{i+1}^f = [\mathbf{I} - \mathbf{K}_{i+1}\mathbf{H}]\mathbf{P}_{i+1}^f \tag{21}$$

This allows us to write the gain also as

$$\mathbf{K}_{i+1} = \mathbf{P}^a_{i+1}\mathbf{H}^T\mathbf{R}^{-1}, \tag{22}$$

provided that $\mathbf{R}$ is invertible. Equation (21) shows that the uncertainty in the forecast is reduced during the analysis according to the amount of additional information assimilated in the system.

A sequential assimilation run is then conducted by repeating this forecast/analysis cycle in sequence. Since only data from the past influence the best estimate at a given time, the assimilation procedure belongs to a class of filtering methods. These contrast with smoothing methods (e.g. *Fukumori* [2001]) in which data from both the past and the future are used to estimate the optimal state of the system at a given time. The analysis error covariance reflects the competition in the Kalman filter between this accumulation of past information and the error growth due to instability mechanisms and model imperfections. Figure 2 conceptually illustrates the filtering process in sequential data assimilation.

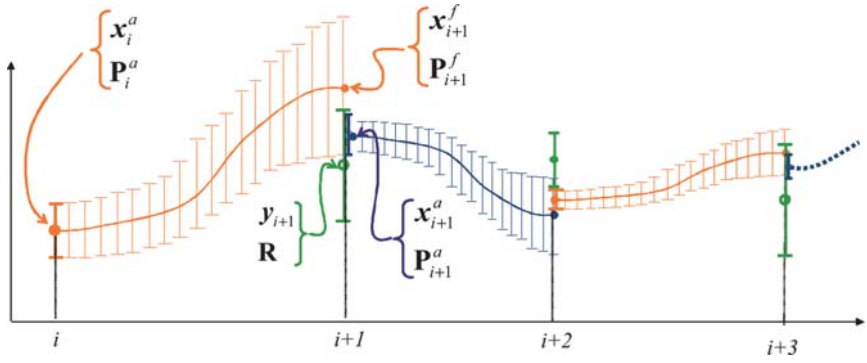

Figure 2. Conceptual representation of filtering with sequential assimilation.

3. From theory to real ocean applications

The Kalman Filter has been primarily developed in the context of ballistic applications, involving dynamical models of fairly low dimension. The more recent interest in the KF in Earth sciences (numerical weather prediction or oceanography) has raised new issues related to the huge number of degrees of freedom taken into account by the models, with consequences for the size of the discretized operators and the quantity of information to be manipulated. It is therefore important to examine at

this stage the practical representation of the mathematical objects introduced in the previous section, with examples taken from high-resolution circulation systems developed from an operational perspective.

3.1 The state vector and model operator

The state vector $\mathbf{x}$ is a discrete representation of the variables involved in the description of the system state. It characterizes information about the space variability of the physical or biological quantities, and about the multivariate relationships between the different dynamical variables. In a numerical model of the Primitive Equations (PE), $\mathbf{x}$ typically contains the 3D discretized temperature, salinity, zonal and meridional velocities, and the 2D sea-surface height or barotropic streamfunction field computed prognostically at every gridpoint of a finite-difference mesh. The typical size of a PE state vector n is approximately given by $N_X \times N_Y \times (N_Z \times 4 \text{ variables} + 1 \text{ variable})$, where N_X, N_Y and N_Z are the horizontal and vertical grid dimensions. The length n of the state vector is typically 10^6 to 10^8 in scientific or operational applications. A biological state vector would contain, for instance, the concentration distributions of nutrients, plankton, dissolved and particulate matter, etc. [*Carmillet et al.*, 2001].

Let us consider, for example, the $1/12^\circ$ model configuration of the North Atlantic ocean that has been developed using the HYbrid Coordinate Ocean Model (HYCOM) with a horizontal grid size of approximately 1400 x 1400, and 26 layers in the vertical direction. Figure 3 shows a snapshot extracted from a simulation, illustrating the space variability described by the multivariate HYCOM state vector.

Note that the vertical hybrid coordinate used in HYCOM is a combination of geometric and dynamic vertical coordinates that evolves dynamically with the state of the system itself [*Bleck*, 2002]. Due to the occurrence of outcropping layers at the base of the mixed layer, the number of discrete variables may be different from one timestep to another, and the dimension of the state vector is therefore dynamically-dependent. This feature slightly complicates the handling of the state vector [*Brankart et al.*, 2003; *Birol et al.*, 2004], compared with more conventional models based on static vertical coordinates such as OPA.

The size problems arise mainly from equations (18), (20) and (21) which involve the manipulation of $n \times n$ matrices. With a state dimension of 10^8, the storage of a full $n \times n$ matrix would require a memory of 10^5 gigabytes, representing about 1000 times the capacity of the largest computers available today. For this reason, it is necessary to by-pass the explicit representation of such matrices in the algorithms. For instance,

the model $M(t_i, t_j)$ will never be available as an explicit operator (even for linear dynamics) but the numerical code of the ocean model will be used as a routine to compute the time evolution of the model state and forecast error components, as required by (17) and (18).

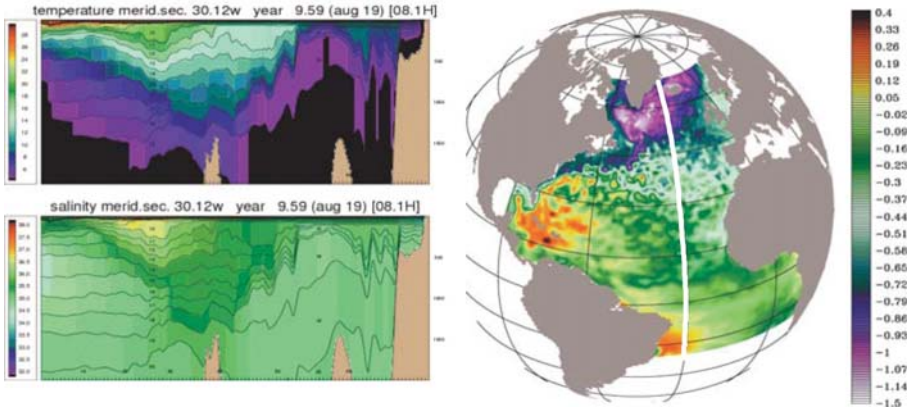

Figure 3. Physical representation of a portion of the HYCOM state vector in a North Atlantic model. The panel on the right shows a snapshot of the SSH field ; the panels on the left show a vertical section in the 3D temperature and salinity fields at 30°W, superimposed to the vertical grid oriented along the layer interfaces. The full state vector also includes the velocity components (not shown here).

3.2 The observation vector

The dimension p of the observation vector $\mathbf{y}$ depends on the capacity of the observation system and the frequency of data assimilation, but in general it is much smaller than the dimension of the state vector. The data sets available to control ocean circulation models in scientific or operational exercises can be categorized into measurements from space, which primarily reflect the surface signature of the ocean circulation and ocean-atmosphere interactions, and *in situ* measurements devoted to the monitoring of the ocean's interior.

Oceanic quantities measured from space include essentially Sea-Surface Temperature (SST), Sea-Level Anomalies (SLA) and ocean colour (which can be used to estimate the chlorophyll concentration in the upper ocean). Note that other important satellite data types will become available in the near future, such as surface salinity measurements from the SMOS (Soil Moisture and Ocean Salinity) mission and sea-ice observations from CRYOSAT. Unlike conventional measurements from field campaigns, satellite-based instruments are operated in routine over long

periods of time and provide considerable information about the ocean's horizontal and temporal variability. Figure 4 illustrates a typical SST picture obtained from composite AVHRR (Advanced Very High Resolution Radiometer) images, and a network of altimetric ground tracks covered by two satellites flying simultaneously over a 7-day period. Along the altimeter tracks, the measurements are available every 7 km and the separation between two tracks of Topex-Poseidon is about 300 km at the Equator. These data can be assimilated "along track" as shown in figure 4, providing in principle the best possible utilization of the observed information if the assimilation scheme is optimally tuned. Otherwise, the raw measurements can be transformed into gridded products using sub-optimal interpolation methods before assimilation.

The surface nature of satellite data poses specific challenges with regard to data assimilation because the surface information has to be projected downward to reconstruct the 3D-content of the ocean signal. As illustrated in the next section, the extrapolation process is achieved through the use of 3D, multivariate error covariances in the assimilation scheme. These surface data remain insufficient, however, for describing aspects of sub-surface variability, and other data available in the form of vertical temperature and salinity profiles from hydrographic casts, moorings or expandable bathythermographs (XBT) and profiling floats from the ARGO international program provide valuable information about the vertical stratification. It is therefore essential to assimilate both data types in a consistent manner.

Two additional operators related to the data must be introduced: the observation error covariance matrix $\mathbf{R}$ and the observation operator $\mathbf{H}$ (or a non-linear H). The observation operator converts the forecast state into "first guesses" of the observations. This conversion is needed because the observed variables may not be located on the model grid points so that horizontal or vertical interpolations are necessary. In addition, relationships of varying complexity may exist between the observed quantity and the model variables. One example is the ocean colour, which is related to the phytoplankton concentration through a fairly complex, and sometimes approximate, relationship. Another example is the relationship between sea-level (measured by altimetry) and the prognostic variables of a rigid-lid ocean model [*Pinardi et al.,* 1995]. In order to avoid too complex observation operators in the algorithms, it is sometimes advisable to augment the state vector with diagnostic variables (such as the surface pressure of a rigid-lid model which can be diagnosed as a function of the prognostic variables) that can be linked to the observed variables more easily.

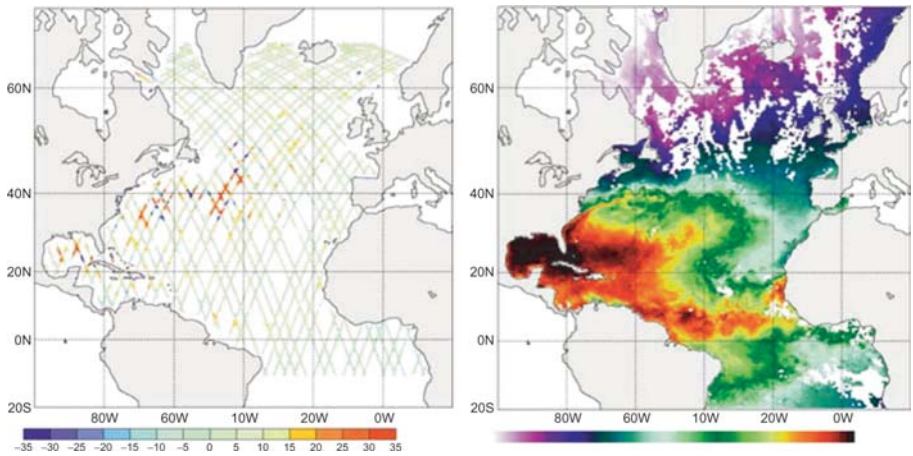

Figure 4. Multi-satellite data set collected during the week August 19-26, 1993, illustrating the different sampling properties. On the left: altimeter measurements (in cm) obtained by merging the Topex/Poseidon and ERS ground tracks; on the right: composite AVHRR picture of the sea-surface temperature.

The observation error accounted for by $\mathbf{R}$ has different possible sources. One source is the instrumental error which can often be considered as spatially uncorrelated. Another is the so-called "representativeness error", associated with variability described by the data at scales that cannot be faithfully represented by the model grid. A third source is the error associated with the mapping $\mathbf{H}$ between the model and the observation space. The spectrum of these errors is mainly concentrated on the short scales, and it is often a reasonable approximation to represent $\mathbf{R}$ by a diagonal matrix of error variances. Note, however, that the reduction in the dimensionality of the estimation problem introduced in Section 5 will possibly be a source of correlated observation error.

3.3 Error covariance matrices

As mentioned above, the assimilation sequence must be initialized with some initial guess for the state $\mathbf{x}_0$ and the associated error covariance $\mathbf{P}_0$. The initialization of $\mathbf{P}$ critically determines the functioning of the filter beyond the first assimilation stages through the process of error dynamics [*Ballabrera et al.*, 2001]. In order to understand the role played by error covariance matrices, let us consider the idealized case of an analysis step with one single observation, η_0. The observed variable, noted η, is assumed to be one of the discrete elements of the state vector

for simplicity, so that $\mathbf{y}$ is a scalar ($p = 1$) and $\mathbf{H}$ is a single-row vector of the form

$$\mathbf{H} = [0, ..., 0, 1, 0, ..., 0] \tag{23}$$

The Kalman gain (20) is then a single-column vector which simplifies as

$$\mathbf{K} = \frac{1}{p_{\eta\eta} + \epsilon^2} \{\mathbf{P}_0\}_\eta \tag{24}$$

where ϵ^2 is the error variance of the single observation, $p_{\eta\eta}$ is the error variance of the observed variable (the η-diagonal element of $\mathbf{P}_0$) and $\{\mathbf{P}_0\}_\eta$ is the η-column of the error covariance matrix. The correction to the initial guess is then proportional to $\{\mathbf{P}_0\}_\eta$ weighted by the prior model-data misfit:

$$\mathbf{x}^a - \mathbf{x}_0 = \frac{1}{p_{\eta\eta} + \epsilon^2} \{\mathbf{P}_0\}_\eta (\eta - \eta_0) \tag{25}$$

Thus, any row or column of the error covariance matrix can be interpreted as a multivariate influence function associated with the observed state variable. This explains the crucial role of the error structures specified in $\mathbf{P}_0$ and shows the importance of considering dynamically-balanced error covariance matrices.

The last operator to be prescribed is the model error covariance matrix $\mathbf{Q}$. This includes all the errors associated with the various physical parameterizations necessary in the model (mixing, diffusion, turbulent closure, hydrostatic approximation, etc.), the errors in the atmospheric forcings and more generally in the boundary conditions, and the errors due to the numerical discretization on the horizontal and vertical dimensions. Note that $\mathbf{Q}$ represents the model error statistics accumulated during an assimilation interval, and should not be confused with errors generated at every time step. Those errors are clearly distributed over a wide spectrum of space scales, and it would be extremely difficult to prescribe a full $\mathbf{Q}$ matrix. By considering the prior misfit between model simulations and observations, it is possible to derive some general properties of the model errors as manifested in the observation space. These difficulties promote the adoption of simplified parameterizations for $\mathbf{Q}$ following, for instance, the approach proposed by *Dee* [1991], and to adjust those parameterizations by sensitivity experiments. Note that systematic biases in the models often make the statistical assumption (4) inappropriate: such biases can be detected by examining the statistics of the innovation sequence, and in principle they should be removed from the forecast to preserve the essential properties of optimal analysis.

4. Simplified schemes based on static background errors

In the KF algorithm, the forecast error covariance is updated before every analysis step by using the model dynamics. Equation (18) can be rewritten as

$$\mathbf{P}_{i+1}^{f} = \mathbf{M}(\mathbf{M}\mathbf{P}_{i}^{a})^{T} + \mathbf{Q} \tag{26}$$

showing that the model code has to be used to propagate the n columns of $\mathbf{P}_i^a$ (in addition to propagate the model state itself). Given the huge size of realistic ocean models, this leads to computing requirements that even the largest computers in the world will not meet in a foreseeable future. In addition, the many imperfections inherent in the representation of the error covariance matrices $\mathbf{P}_i^a$ and $\mathbf{Q}$ make the explicit computation of this equation questionable.

If the flow of assimilated observations is fairly regular over time, it makes sense to assume that the errors in an assimilation sequence tend to fluctuate around some average level after a couple of cycles. This asymptotic behaviour reflects a balance between the increase of uncertainty during the forecast step and the error reduction during the analysis step. The existence of such an asymptotic limit for $\mathbf{P}_{i+1}^f$ provides justification for simply using a static error covariance, noted $\mathbf{B}$ for "background error" (a term usually adopted when the error statistics are not propagated from one assimilation cycle to the next), instead of explicitly computing the forecast error according to (18). Different techniques have been developed to prescribe the background error covariances in ocean and atmospheric assimilation schemes. These are discussed below.

4.1 Optimal Interpolation

Optimal Interpolation (OI) designates a wide range of statistical assimilation schemes in which the $\mathbf{B}$ matrix is pre-determined empirically. The main advantages of OI methods are their cost, their ease of use and the possibility of conducting sensitivity studies to test many different models of error covariances. Today, the vast majority of operational systems involved in GODAE (e.g., FOAM, MFS, HYCOM, MERCATOR) are based on OI schemes. In contradiction with its name however, OI is a sub-optimal assimilation process and it would actually be more correct to designate this class of methods as "statistical interpolation" [*Daley*, 1991].

Any covariance matrix can be normalized into a correlation matrix $\mathbf{C}$, dividing each component by the product of the error standard deviations.

Thus, $\mathbf{B}$ can still be written as

$$\mathbf{B} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2} \tag{27}$$

where $\mathbf{D}$ is the diagonal matrix of the variances. OI schemes commonly adopt further simplifications for modelling the background error covariances. Assuming that the correlations are functions of the distance only, analytical functions with open parameters such as correlation scales have been used for a long time to build $\mathbf{C}$ (e.g. *Thiebaux* [1985]). A significant property of well-posed correlation functions is their compact support which reflects the negligible influence of observations at large distances. It is also generally assumed that the correlations can be separated into a product of horizontal and vertical correlations. The concept of separability is related to the predominant role of stratification and the very different scales involved horizontally and vertically in the open ocean. However, different behaviours are expected near boundaries and coastal regions [*Echevin et al.*, 2000]. The SOFA (System for Ocean Forecasting and Analysis) scheme used in the first operational MERCATOR prototype performs OI in the horizontal direction to combine altimeter data with the model predictions [*De Mey and Benkiran*, 2002]. It offers the possibility of using different error covariance models, such as :

$$\mathbf{C}(\mathbf{x}_i, \mathbf{x}_j) = (1 + al + \frac{1}{3}a^2l^2)e^{-al} \quad \text{with} \quad l = dist(\mathbf{x}_i, \mathbf{x}_j) \tag{28}$$

where the parameter a determines the horizontal correlation scale [*Gavart et al.*, 1999]. Fine tuning of these correlation parameters can be achieved objectively via sensitivity studies or Monte-Carlo experiments [*Auclair et al.,* 2003]. In the MERCATOR system, the value of a is defined geographically, reflecting longer horizontal correlation scales in the tropics than at high latitudes.

It is worth noting that some applications of (27) allow for the error variances to be updated during the assimilation sequence using empirical prediction schemes, while still keeping the correlation matrix unchanged [*Rienecker and Miller*, 1991].

4.2 Asymptotic approximation

Other interesting approaches such as the time-asymptotic filter approximation have been proposed to determine a background error covariance matrix that is consistent with the error dynamics of the KF [*Fu et al.*, 1993; *Fukumori et al.*, 1993; *Fukumori and Malanotte-Rizzoli*, 1995]. In equations (18) , (20) and (21) above, operators $\mathbf{M}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$ have been assumed time-invariant for simplicity. Satellite altimeters,

for instance, can be considered as a time-invariant observation system given that the mission tracks are repeated in an exact manner. By combining (18) and (21), the so-called Riccati equation is obtained:

$$\mathbf{P}_{i+1}^{f} = \mathbf{M}\left[\mathbf{P}_{i}^{f} - \mathbf{P}_{i}^{f}\mathbf{H}^{T}[\mathbf{H}\mathbf{P}_{i}^{f}\mathbf{H}^{T} + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P}_{i}^{f}\right]\mathbf{M}^{T} + \mathbf{Q} \qquad (29)$$

which can be iterated prior to the assimilation sequence. It can be shown that the solution of the Riccati equation converges towards a steady-state solution, say $\mathbf{B}^{f}$, provided the $\mathbf{M}$, $\mathbf{H}$ and $\mathbf{Q}$ matrices have desirable properties. The details of these conditions are explained in *Fukumori et al.* [1993]. In addition, a fast convergence can be obtained by using appropriate recursion methods. The $\mathbf{B}^{f}$ solution results from a balance between three effects: evolution due to model dynamics, increase due to model errors and decrease due to the new information from assimilated data. Interestingly, the utility of a stationary $\mathbf{B}^{f}$ matrix in the Kalman gain was demonstrated even in the presence of evolving properties of the observation system [*Fukumori*, 1995].

The interest of steady-state filters has been illustrated in a number of oceanographic case studies, although additional approximations such as order reduction and model simplifications are generally needed to make the assimilation effective with realistic models.

5. Reduced-order Kalman filters

The previous section has shown that OI-based assimilation schemes over-simplify the propagation of errors by neglecting dynamical principles and statistical information. An alternative way to make the KF tractable with large-size models has been explored with the concept of reduced order, which aims at decreasing the computational burden of the algorithm while preserving the essential characteristics of error dynamics. Experiences from atmospheric re-analyses indicate the importance of the "errors of the day" which can be dominated by short time scales, and which are ignored when the forecast error is approximated by a constant as with OI [*Kalnay et al.,* 1997]. Similar behaviour is observed in the ocean, especially at scales dominated by instability mechanisms such as the scale of eddies, western boundary currents, etc [*Ballabrera et al.*, 2001].

Several arguments support the concept of order reduction. Firstly, the ocean can be considered as a driven/dissipative dynamical system governed by an "attractor" of finite dimension. The existence of a global attractor has been proved for the Navier-Stokes equations, with a dimension bounded by a function of the Reynolds number [*Lions et al.*, 1997]. Geostrophy, for instance, is one of the dominant properties of

the ocean attractor which cannot be ignored during the assimilation process. Therefore, it makes sense to renounce making corrections in the directions where the errors eventually die away because of the attractive nature of the system. Secondly, the state space as defined in Section 3 is determined by the number of degrees of freedom implied by the model discretization, which is much larger than the actual number of dynamical features of interest in the system. Thirdly, less is probably known about the errors than about the dynamics, and the lack of statistical information will make a full KF superfluous anyway [*Cane et al.*, 1996], especially in the ocean where the flow of observations accumulates at a fairly slow rate.

Reducing the dimensionality of the problem can be formulated in terms of state space or error space with quite different implications. A variety of approaches widely used in oceanography are discussed below.

5.1 Reducing the dimensionality of the state space

The reduced state can be defined explicitly using a transformation operator $\mathbf{T}$ to convert the full model state $\mathbf{x}$ (of dimension n) into a reduced state $\mathbf{w}$ of dimension $r < n$:

$$\mathbf{w} = \mathbf{T}\mathbf{x} \tag{30}$$

The statistical properties of $\mathbf{x}$ such as error covariance matrices can be easily transformed in the low-dimension space using (30). Dynamical equations can be derived for $\mathbf{w}$ using a pseudo-inverse of $\mathbf{T}$ and the KF can be entirely reformulated in the low-dimension space, with the condition that the null space associated with (30) be dynamically uncoupled from the reduced space [*Fukumori and Malanotte-Rizzoli*, 1995]. By introducing this transformation, the r elements of $\mathbf{w}$ become the actual degrees of freedom of the estimation problem.

Many possibilities exist for defining the transformation, such as a truncation of the model spectrum or a selection of multivariate modes of system variability [*Cane et al.*, 1996]. Another approach explored by *Dee* [1991] takes advantage of physical relationships between certain model variables. The reduction of the state space dimension can also be achieved simply by building the estimation vector from a selection of model state variables on a coarser grid [*Fukumori*, 1995] or with dynamical variables closely correlated with the observed signal. It is worth noting that, for sub-spaces which only preserve the scales larger than those of the observed signal, an extra term must be added to the observation error covariance matrix $\mathbf{R}$ to account for the truncated modes

[*Cane et al.,* 1996]. This kind of representativeness error typically corresponds to spatially correlated signals, and in this case off-diagonal elements must be included in **R**.

In the open ocean, dynamical considerations can justify the omission of some of the PE variables in the state vector, such as the horizontal velocity components which are expected to adjust very quickly to the density properties. Different choices of estimation space, motivated also by the faith in the error statistics to be prescribed for the selected variables, are discussed by *Brankart et al.* [2003] for primitive equation circulation models.

The question of assimilating satellite altimetric data into ocean circulation models has been approached many times by reducing the statistical estimation problem to the surface variables. The downward penetration of assimilated information is then achieved empirically, using a variety of statistical (e.g. *Hurlburt* [1986]; *Mellor and Ezer* [1991]; *De Mey and Benkiran* [2002]) or physical (e.g. *Cooper and Haines* [1996]; *Oschlies and Willebrand* [1996]) extrapolation schemes on the vertical direction. These different vertical projection methods, as well as the dynamical adjustment process described by *Brankart et al.* [2003] in the context of HYCOM, can be considered as approximations of the pseudo-inverse of **T** needed to convert the reduced state back to the full model space after statistical estimation.

5.2 Reducing the dimensionality of the error space

A number of other methods are based on approximations of the error with fewer degrees of freedom than the model itself. Error sub-spaces are built with the aim of selectively correcting the model state along the most representative directions of the forecast error. The concept fits well with statistical assimilation schemes because the analysis increment computed by (20) can only take place within the sub-space spanned by the forecast error covariance. The unknown are the coefficients of the correction projected along the multivariate error directions. The concept of Error Subspace Statistical Estimation (ESSE) introduced by *Lermusiaux* [1999] has been developed on the basis of this principle. Note that recent oceanographic studies have also started to explore the potential benefit of order reduction using variational assimilation methods [*Robert et al.*, 2005].

The sub-space can be prescribed as a time-invariant set of error directions. The success of assimilation depends essentially on the capacity of the sub-space to capture the observed variability of the system. A num-

ber of applications have been conducted successfully along these lines in the tropical oceans, where the dynamics is "slow" and nearly linear [*Verron et al.*, 1999]. Alternatively, the error sub-space can be allowed to evolve using deterministic or stochastic approaches, or a mixture of both. Evolving sub-spaces are in general more suitable to track the non-linear evolution of "fast" energetic error modes, or to dynamically adjust the unbalanced components of the sub-space [*Lermusiaux,* 2001]. Related concepts have been introduced in the context of atmospheric data assimilation by *Cohn and Todling* [1996], who proposed approximate schemes for error covariance propagation in the case of stable and unstable dynamics.

5.3 Low-rank error covariance matrix

By construction, a covariance matrix is symmetric and positive definite and can always be decomposed as $\mathbf{P} = \mathbf{N}\mathbf{\Lambda}\mathbf{N}^T$. The columns of $\mathbf{N}$ are formed by the orthonormalized eigenvectors $\mathbf{n}_k$ of $\mathbf{P}$, and $\mathbf{\Lambda}$ is a diagonal matrix formed with the corresponding eigenvalues λ_k. The inverse of $\mathbf{P}$ is then given by $\mathbf{P}^{-1} = \mathbf{N}\mathbf{\Lambda}^{-1}\mathbf{N}^T$. The pdf defined by (9) can be refomulated as follows:

$$\epsilon \rightarrow N(0,\mathbf{P}) \sim \prod_{k=1}^{n} \exp\left[-\frac{1}{2}\lambda_k^{-1}\epsilon_k^2\right] \tag{31}$$

where ϵ_k is the component of the error vector ϵ in the $\mathbf{n}_k$ direction.

A reduced-order Kalman filter can be implemented by approximating the error covariance matrix $\mathbf{P}$ with only the "leading" columns of $\mathbf{N}$ associated with the r largest eigenvalues. The pdf defined by such a low-rank matrix is obtained by taking the limit of (31) when $\lambda_k \rightarrow 0$ for $k > r$ (assuming that the eigenvalues have been sorted in decreasing order). Equation (31) shows that the probability tends to zero if $\epsilon_k \neq 0$ and the error vectors are confined in a sub-space of dimension r. From a stochastic point of view, the leading vectors describe the principal axes of the probability ellipsoid oriented along the dominant directions of uncertainty; from an algebraic point of view, they define the basis of a sub-space where the error is expected to lie.

5.4 Specification of error sub-spaces using EOFs

Several strategies can be adopted to determine the leading directions of the error sub-space. One of them is the Singular Value Decomposition (SVD) of a covariance matrix constructed with prescribed analytical functions. Other methods have been proposed in the literature which utilize singular, Lyapunov or breeding vectors of the transition matrix

(e.g., *Miller and Erhet* [2002]). An approach based on EOFs was put forward by *Cane et al.* [1996] to elaborate a reduced state Kalman filter, and by *Pham et al.* [1998] in the context of the Singular Evolutive Extended Kalman (SEEK) filter.

A practical way to estimate a low-rank error covariance matrix is to perform an EOF analysis of state vectors generated by a prior model simulation. The EOF analysis provides a compact description of the spatial and temporal variability of the model in terms of orthogonal functions. Usually, most of the variance of the time sequence is described by the first few orthogonal functions whose patterns may then be linked to dynamical mechanisms [*Emery and Thomson*, 1998]. In order to compute an EOF basis, a series of s model state vectors $\mathbf{x}(t_i)$ is extracted at regular intervals from a free model simulation, and the vectors

$$\frac{\mathbf{x}(t_i) - \overline{\mathbf{x}(t_i)}}{\sqrt{s-1}} \tag{32}$$

form the columns of a "scatter matrix" $\mathbf{X}$ of dimensions $n \times s$. The normalization factor $\frac{1}{\sqrt{s-1}}$ is introduced so that the unbiased estimation of the covariance matrix is given by the product $\mathbf{X}\mathbf{X}^T$. The size s of the sample is always much smaller than the dimension n of state vectors involved in realistic ocean models. The EOFs correspond to the orthonormalized eigenvectors $\mathbf{N}$ of the $n \times n$ matrix $\mathbf{X}\mathbf{X}^T$ which has a rank necessarily smaller than or equal to s. The explicit calculation of the matrix $\mathbf{X}\mathbf{X}^T$ is not required to compute the eigenmodes. Indeed, the matrix $\mathbf{X}^T\mathbf{X}$ of much lower dimensions $(s \times s)$ has the same eigenvalues $\mathbf{\Lambda}$, and its eigenvectors $\mathbf{V}$ allow the computation of the "large" eigenmodes $\mathbf{N}$ by simple multiplication, $\mathbf{N} = \mathbf{X}\mathbf{V}$. This provides a first practical method to calculate the EOFs. A second possible approach is to perform a singular value decomposition of the matrix $\mathbf{X}$ directly, using standard SVD algorithms [*Kelly*, 1988; *Emery and Thomson*, 1998]. This provides a decomposition of the form:

$$\mathbf{X} = \mathbf{N}\sqrt{\mathbf{\Lambda}}\mathbf{V}^T \tag{33}$$

The EOFs determined by the two methods are identical, but the SVD has the advantage of greater computational stability. One should keep in mind that the EOF decomposition of multivariate state vectors also depends on the units adopted to represent the different physical quantities, and the choice of a particular metric may be more critical than computational stability in determining the structure and ordering of the dominant EOFs.

The r dominant eigenvectors form the columns of an array noted $\mathbf{N}_0$ with dimensions $n \times r$, and the diagonal matrix of the associated

eigenvalues is noted $\mathbf{\Lambda}_0$. The covariance can be approximated by the matrix

$$\mathbf{X}\mathbf{X}^T \backsimeq \mathbf{N}_0\mathbf{\Lambda}_0{\mathbf{N}_0}^T = \mathbf{S}_0\mathbf{S}_0^T \tag{34}$$

of rank r, where $\mathbf{S}_0 = \mathbf{N}_0\sqrt{\mathbf{\Lambda}_0}$ contains dimensional quantities. Assimilation experiments can be initialized using $\mathbf{S}_0\mathbf{S}_0^T$ as a guess of the initial error covariance.

The minimum dimension of the error sub-space has to be determined by practical considerations. A condition of stability for a reduced-rank KF in idealized conditions (linear model and autonomous system) is shown to be that r should be larger than r^*, where r^* is the number of eigenvalues of $\mathbf{M}$ having an absolute value larger than or equal to 1 (*Pham et al.* [1998]; *Carme et al.*, [2001]). Since this number cannot be systematically evaluated with large ocean models, more pragmatic criteria must be considered to define the truncation threshold. For example, the statistical representativity of the r retained EOFs can be measured by the percentage

$$I = \frac{\sum_{k=1}^{r} \lambda_k}{\sum_{k=1}^{s} \lambda_k}$$

where the λ_k are the eigenvalues. The choice of an acceptable threshold of explained variance provides a means to determine the minimum dimension of the EOF basis for practical assimilation problems. This number typically ranges from a few tens to a few hundreds.

The assumption underlying the procedure described above is that the model is unbiased (which is a necessary condition for the Kalman filter to yield optimal estimations) and sufficiently "good" for its intrinsic variability to be statistically representative of real world variability. The hypothesis that a model provides an unbiased simulation implies the use of the mean vector $\overline{\mathbf{x}(t_i)}$ as a first guess. Sometimes, these ideal conditions are not verified with a free model simulation and more complex strategies must be developed. For example, an assimilation sequence may be carried out in a first stage with a simplified method which does not require EOFs (such as an OI or nudging scheme) so as to bring the model and observations together; an EOF analysis of the resulting fields is then performed, and a new assimilation sequence can be obtained in a second stage with a reduced-rank KF. The process can be iterated several times to provide an improved EOF basis which combines the variabilities of the model dynamics and of the observations.

5.5 Examples of sub-spaces based on EOFs

The EOFs calculated with primitive-equation models are fully multivariate and three-dimensional, and cover the whole model domain: all the variables of the state vector (SSH, temperature, salinity, zonal and meridional velocities) are considered together in a dynamically consistent manner. The extrapolation of the data from observed to non-observed variables is performed along the directions represented by these EOFs which connect all grid points of the numerical domain.

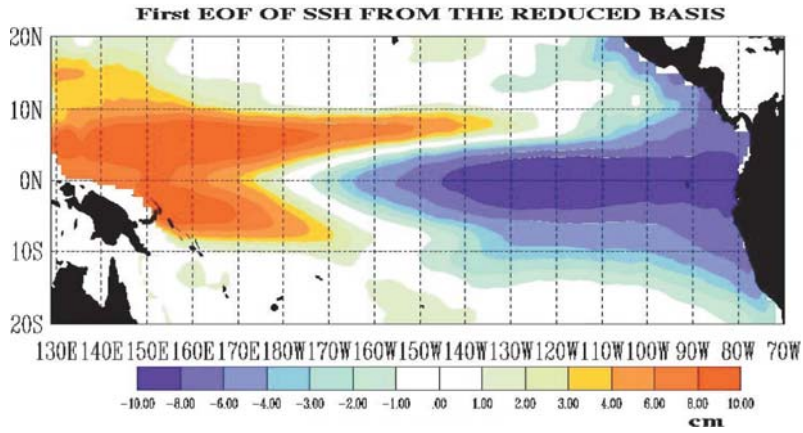

Figure 5. Sea-surface height component of the first EOF (in cm) of the interannual variability of the circulation in the Tropical Pacific (Parent [2000]).

The physical nature of an EOF basis is discussed by *Parent et al.* [2003], who studied the variability of the circulation in the Tropical Pacific ocean during the period 1994-1998 using a primitive-equation model of the Equatorial basin between 20°N and 20°S and a SEEK filter to assimilate satellite altimeter data. They computed an EOF decomposition of the simulated variability over the 5-year integration period and selected the first 15 EOFs to build a reduced basis for assimilation. The first dominant EOF illustrates the well-known west-east seesaw, which is the most important feature of the El Niño and La Niña phases of the ENSO phenomenon. In 1997, the warm pool of the western basin migrated towards the eastern basin and produced a positive SLA, whereas during the La Niña phase (second part of 1998), the opposite movement was observed (figure 5).

An issue of practical interest is the estimation of small correlations associated with distant variables, which is a well-known difficulty of finite

sampling procedures such as that described above. As pointed out by *Houtekamer and Mitchell* [1998], the accurate estimation of small correlation coefficients would necessitate a very large number of independent model samples to guarantee the statistical convergence of the computations. In order to keep the number of samples within tractable limits and prevent the data from exerting a spurious influence at remote distances through large-scale signatures in the EOFs, a technique based on EOFs with compact support has been setup by *Testut et al.* [2003] in which regional sub-domains of adjustable size are considered to characterize the error sub-space. From a theoretical point of view this approximation can be justified by the argument that, when the ocean surface is divided into local regions of moderate size, the background error in such regions tends to lie in a sub-space of much smaller dimension than the full ocean state. Similar hypotheses have been put forward to develop local ensemble Kalman filters for atmospheric data assimilation [*Ott et al.*, 2004].

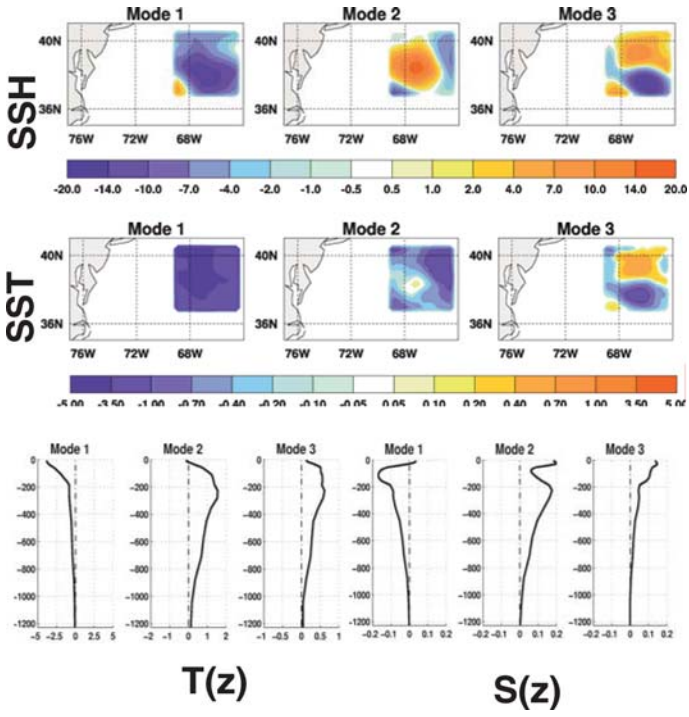

Figure 6. Multivariate local EOFs of the mesoscale ocean variability in the Gulf Stream region simulated by an eddy-permitting model: surface signature of temperature and sea-surface height of the first three dominant modes, and associated vertical extensions of the temperature and salinity structures (reproduced from Testut [2000]).

This technique has been developed more specifically for eddy-resolving ocean assimilation models in which the signal dominates in the small scales. Figure 6 illustrates the surface signature and vertical extension of the three dominant modes calculated with an eddy-permitting model in the Gulf Stream region. Theses modes reflect to some extent the anisotropic nature of the multivariate covariances associated with the local dynamics, with smaller scales represented by higher modes.

The local representation of the error sub-space has been shown to be particularly effective for capturing the mesoscale features of the turbulent ocean. In the example described by *Penduff et al.* [2002], an assimilative system based on local EOFs was implemented in a 1/3° South Atlantic OPA model to perform hindcast experiments for the 1993-1996 period. The assimilated data are similar to those shown in figure 4, consisting of composite AVHRR observations of SST and altimetric measurements of the sea-level anomalies. The results show that the assimilation is able to successfully reproduce the Agulhas Rings present at that time in the real ocean. In addition to correcting the variables observed at the surface, the three-dimensional multivariate properties of the EOFs also permitted a correction of the non-observed variables in the ocean's interior. The beneficial impact of the assimilation was particularly impressive on the mean salinity in the Confluence region down to about 1500 meters depth.

6. Low-rank Kalman filters

Several low-rank filters based on static or evolving error sub-spaces have been developed over the past ten years, such as the SEEK filter introduced by *Pham et al.* [1998] and the Reduced-Rank-SQuare-RooT (RRSQRT) formulation explored by *Verlaan and Heemink* [1997]. Many features of the Ensemble Kalman Filter (EnKF) put forward by *Evensen* [1994] can be discussed using a similar framework.

The Ensemble OI scheme [*Evensen,* 2003] and the SEEK filter with static error sub-space are two sub-optimal schemes which preserve a number of important properties of statistical estimation but require only a small fration of the computer resources needed by the model. In contrast to those simplified schemes, the EnKF or the SEEK filter with evolutive sub-space propagate the error statistics according to the model dynamics, but they need the simultaneous integration of model states perturbed along each direction of the error sub-space.

The consequences of using a low-rank error covariance matrix to compute the forecast and analysis steps of the KF are examined in the following sections.

6.1 Forecast error with a reduced basis

The reduced basis concept allows drastic simplifications to compute the evolution of the error statistics over the assimilation window. A hierarchy of algorithms of increasing sophistication and computer requirements has been proposed to compute the forecast step. Assuming that the analysis error covariance is represented as a low-rank matrix at time t_i

$$\mathbf{P}_i^a = \mathbf{S}_i^a \left(\mathbf{S}_i^a\right)^T, \tag{35}$$

where $\mathbf{S}_i^a$ (of dimension $n \times r$) defines the error sub-space associated with $\mathbf{x}_i^a$, Eq. (18) becomes

$$\mathbf{P}_{i+1}^f = \widetilde{\mathbf{S}}_{i+1}^f \left(\widetilde{\mathbf{S}}_{i+1}^f\right)^T + \mathbf{Q} \quad \text{with} \quad \widetilde{\mathbf{S}}_{i+1}^f = \mathbf{M}\mathbf{S}_i^a. \tag{36}$$

The computer load associated with Eq. (36) is primarily determined by the rank r of $\mathbf{P}_i^a$ which specifies the number of model integrations needed to evaluate the forecast error covariance matrix. As originally proposed by *Pham et al.* [1998], this algorithmic variant known as "Extended Evolutive Basis" requires the derivation of the tangent linear model $\mathbf{M}(t_i, t_{i+1})$ to update the error directions, i.e. the r columns of $\mathbf{S}_i^a$. The evolved sub-space $\widetilde{\mathbf{S}}_{i+1}^f$ reflects how the model dynamics affects uncertainty during the forecast.

An alternative scheme to Eq. (36) can be used for the calculation of the time evolution of the reduced basis as follows :

$$\left\{\widetilde{\mathbf{S}}_{i+1}^f\right\}_j = \frac{1}{\alpha}\left[M(t_i, t_{i+1})\left\{\mathbf{x}_i^a + \alpha\left\{\mathbf{S}_i^a\right\}_j\right\} - M(t_i, t_{i+1})\left\{\mathbf{x}_i^a\right\}\right] \tag{37}$$

where $\{\}_j$ is the jth column of the matrix, and α is an adjustable parameter that determines the size of the perturbations along each error direction. Equation (37) is a finite-difference approximation of the linear error evolution if α is small. However, the value of α is usually taken to be of the order of 1 to simulate the non-linear evolution of model perturbations that have an amplitude comparable to the error covariances. This algorithm known as "Interpolated Evolutive Basis" has a two-fold benefit: firstly, it avoids the computation of the tangent linear model which numerically can be a delicate task; and secondly, it seems more robust with regard to the model non-linearities because the finite difference takes into account the amplitude of the uncertainties, while the classic linearization does not.

Due to the recursive character of the Kalman filter, $\mathbf{P}_{i+1}^f$ should have the same rank as $\mathbf{P}_i^a$ in order to preserve the advantage of a low-dimension space for the subsequent assimilation cycles. The difficulty

arises from the fact that the rank of $\mathbf{P}^f_{i+1}$ intimately depends on the structure of $\mathbf{Q}$. A possible method is simply to project the error vector η_i and its associated covariance on the sub-space generated by the columns of $\widetilde{\mathbf{S}}^f_{i+1}$. As it is often impossible to specify the model error perfectly, for the reasons discussed above, a simple parameterization of system noise can be introduced, assuming for instance that

$$\mathbf{Q} = \frac{1-\rho}{\rho}\ \widetilde{\mathbf{S}}^f_{i+1}\left(\widetilde{\mathbf{S}}^f_{i+1}\right)^T \tag{38}$$

where ρ is a scalar quantity called the "forgetting factor" $(0 < \rho < 1)$, by analogy with the approach used in automatic control algorithms. This kind of model error parameterization leads to a forecast error covariance matrix of the form :

$$\mathbf{P}^f_{i+1} = \frac{1}{\rho}\widetilde{\mathbf{S}}^f_{i+1}\left(\widetilde{\mathbf{S}}^f_{i+1}\right)^T = \mathbf{S}^f_{i+1}\ \mathbf{S}^{fT}_{i+1} \text{ with } \quad \mathbf{S}^f_{i+1} = \frac{1}{\sqrt{\rho}}\widetilde{\mathbf{S}}^f_{i+1} \tag{39}$$

which is singular and has the same rank r as the previous analysis error covariance. The forgetting factor is one of the many possible options for accounting for some simple form of model error in the assimilation scheme. Other approaches may be implemented, however, such as using a perturbed model $\widetilde{M}(t_i, t_{i+1})$ instead of the original model to dynamically update the error modes through Eq. (37). For instance, in most EnKF implementations, stochastic perturbations are introduced in the surface forcings to update each ensemble member, accounting in this way for the uncertainty in the atmospheric fluxes.

With respect to Eq. (36), an even more drastic simplification of the forecast step can be obtained by simply neglecting the dynamical transformation of the error directions during the assimilation period (t_i, t_{i+1}), leading to the "Fixed Basis" algorithm:

$$\widetilde{\mathbf{S}}^f_{i+1} = \mathbf{M}\mathbf{S}^a_i \approx \mathbf{I}\mathbf{S}^a_i \tag{40}$$

where $\mathbf{I}$ is the identity matrix. As in the Ensemble OI scheme [*Evensen*, 2003], temporal persistence of the error sub-space basis is assumed in this variant. Static error sub-spaces have been successfully used in a variety of assimilation applications (*Verron et al.* [1999]; *Gourdeau et al.* [1999]; *Parent et al.* [2003]; *Penduff et al.* [2002]), being justified by two basic arguments. The first one is of practical interest: the cost of a Fixed Basis assimilation experiment is of the same order of magnitude as a model simulation, with only a few additional computations needed to perform the algebraic operations of the analysis step. This makes the algorithm extremely useful in evaluating the overall performance of the

system, testing new parameterizations, examining the impact of different observation configurations, etc. The second argument constitutes a more objective justification in terms of statistical estimation: it is often impossible to properly characterize the statistical structure of the model error originating from imperfections in the model forcings, discretization schemes, space/time resolution, etc. As a result, the approximation made by assuming persistence of the error sub-space directions is often negligible compared to the mis-specification of the systematic errors. When the amplitude of the model error increment $\mathbf{Q}$ dominates the difference between $\mathbf{P}^a$ and $\mathbf{M}\mathbf{P}^a\mathbf{M}^T$ in the error propagation Eq. (18), it is a waste of CPU resources to explicitly compute the error propagation with the model dynamics, because the result will eventually be polluted by mis-specified model error estimates. In such situations, it is sensible to by-pass the error propagation equation and concentrate on identifying the dominant forecast error directions using, for instance, adaptivity mechanisms as described in Section 7.

6.2 Analysis step with a reduced basis

The linear variance-minimizing analysis (19) can be re-formulated nicely when the forecast error covariance used to compute the Kalman gain is of low rank. Using equations (20), (39) and the matrix equality

$$\left[\mathbf{X}_1 + \mathbf{X}_{12}\mathbf{X}_2^{-1}\mathbf{X}_{21}\right]^{-1} = \mathbf{X}_1^{-1} - \mathbf{X}_1^{-1}\mathbf{X}_{12}\left[\mathbf{X}_2 + \mathbf{X}_{21}\mathbf{X}_1^{-1}\mathbf{X}_{12}\right]^{-1}\mathbf{X}_{21}\mathbf{X}_1^{-1}, \tag{41}$$

the expression for the Kalman gain, after some mathematical manipulations, can be transformed as follows :

$$\mathbf{K}_{i+1} = \mathbf{S}_{i+1}^f[\mathbf{I} + (\mathbf{H}\mathbf{S}_{i+1}^f)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{S}_{i+1}^f)]^{-1}(\mathbf{H}\mathbf{S}_{i+1}^f)^T\mathbf{R}^{-1}. \tag{42}$$

Equation (42) shows that the size of this inversion problem is determined by the error sub-space dimension, while the original Kalman gain (20) requires an inversion in the observation space. As the number of observations is usually much larger than the rank of the error sub-space, the inversion step becomes less expensive than the corresponding computation of the original Kalman gain. By combining equations (19) and (42), the correction of the forecast state can be written as a linear combination of the error modes

$$\mathbf{x}_{i+1}^a - \mathbf{x}_{i+1}^f = \mathbf{S}_{i+1}^f\mathbf{c}_{i+1} \tag{43}$$

with

$$\mathbf{c}_{i+1} = [\mathbf{I} + (\mathbf{H}\mathbf{S}_{i+1}^f)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{S}_{i+1}^f)]^{-1}(\mathbf{H}\mathbf{S}_{i+1}^f)^T\mathbf{R}^{-1}\left(\mathbf{y}_{i+1}^o - H\mathbf{x}_{i+1}^f\right) \tag{44}$$

The vector $\mathbf{c}_{i+1}$ contains the r amplitudes of the error modes that need to be estimated at each analysis step. Finally, the scheme evaluates the analysis error as follows

$$\mathbf{P}^a_{i+1} = [\mathbf{I} - \mathbf{K}_{i+1}\mathbf{H}]\mathbf{P}^f_{i+1} = \mathbf{S}^a_{i+1}\ \mathbf{S}^{aT}_{i+1} \tag{45}$$

and updates the vectors of the reduced basis according to

$$\mathbf{S}^a_{i+1} = \mathbf{S}^f_{i+1}[\mathbf{I} + (\mathbf{H}\mathbf{S}^f_{i+1})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{S}^f_{i+1})]^{-1/2}. \tag{46}$$

This shows that, if the rank of $\mathbf{P}^f_{i+1}$ is r, then the rank of $\mathbf{P}^a_{i+1}$ is equal to r also; therefore, recursivity in the forecast/analysis cycles is allowed. For better numerical conditioning, it is possible to re-orthonormalize the reduced basis by recomputing a SVD decomposition of the analysis error covariance matrix.

There is one additional consideration that should be mentioned concerning the practical computation of the Kalman gain in the reduced space (42). The robust estimation of small correlations associated with remote observations is a common difficulty that can be avoided by computing EOFs with compact support, as mentioned above. Instead of computing local EOFs, a simplification of the analysis scheme can be designed by enforcing to zero the error covariances between distant variables which are believed to be uncorrelated in the real ocean [*Houtekamer and Mitchell,* 1998].

This simplification is implemented by assuming that distant observations have negligible influence on the analysis. The global system is split into sub-systems, and for each of these the traditional analysis is computed. Only data points located within individual regions, centered on a sub-domain of one or several grid points to be updated, actually contribute to the gain in (42). This approach can be understood as a tuning of the observation operator according to the sub-domain in question. Intuitively, this approximation makes sense because only data points located in the neighborhood of a model grid point should objectively have an impact on the analysis for that grid point. The size of the regions is determined in such a way that the distribution of the observations available on the model domain always provides at least a few data points within each region of influence (if there were no data available in the region of influence, no correction would apply). This algorithmic simplification also improves the analysis because it enables a larger part of the estimation space to be spanned over a particular subdomain (see, for instance, *Brusdal et al.* [2003]).

Other interesting procedures exist for addressing the locality issue. One of these is the partitioning of the large estimation problem into

a series of separate smaller calculations using the partitioned Kalman filter and smoother proposed by *Fukumori* [2002]. The objective of the partition is to make global eddy-resolving data assimilation problems computationally viable. In their example, the reduced state consists of perturbations of the barotropic mode and the first five baroclinic modes defined in eight overlapping cells covering the globe.

6.3 Stochastic vs. deterministic filters

In a low-rank filter like SEEK, the error directions are determined at the initialization step and their evolution can be anticipated in a deterministic way. As these filters do not require any further randomization, they are termed deterministic filters in contrast to the EnKF which can be considered as a stochastic filter because ramdomization has to be repeated at each assimilation cycle.

Evensen [2003] reviews the theoretical formulation and practical implementation of the EnKF. The EnKF was introduced to avoid the occurrence of instabilities found with the Extended KF due to the non-linear evolution of the probability density functions [*Evensen* 1994]. However, the EnKF only solves half of the non-linearity problems because it still combines the model prediction and the data by using only the first two moments of the pdfs, assuming that the distributions are nearly Gaussian. The non-linear analysis equations would be more difficult to use in practical applications, as discussed by *Evensen and van Leeuwen* [2000].

The EnKF was initially proposed by *Evensen* [1994] based on the following general procedure: a sampling of the state space is achieved using Monte-Carlo methods to generate an ensemble of r model states $\mathbf{x}_i^{a,j}$ representing the spread of possible initial conditions at time t_i around the mean $\overline{\mathbf{x}_i^{a,j}}$. Each member is then propagated individually as

$$\mathbf{x}_{i+1}^{f,j} = M(t_i, t_{i+1}) \left\{ \mathbf{x}_i^{a,j} \right\} + \eta^j \quad , \quad j = 1, ..., r \tag{47}$$

using the non-linear model with stochastic perturbations. The vectors η^j have a covariance matrix $\mathbf{Q}$ and are generally introduced through the perturbation of the atmospheric forcings. The new ensemble $\mathbf{x}_{i+1}^{f,j}$provides implicitly the forecast error covariance matrix

$$\mathbf{P}_{i+1}^f = \frac{1}{r-1} \sum_{j=1}^{r} \left(\mathbf{x}_{i+1}^{f,j} - \overline{\mathbf{x}_{i+1}^{f,j}} \right) \left(\mathbf{x}_{i+1}^{f,j} - \overline{\mathbf{x}_{i+1}^{f,j}} \right)^T . \tag{48}$$

The statistical analysis equation of the KF is then applied to each individual member,

$$\mathbf{x}_{i+1}^{a,j} = \mathbf{x}_{i+1}^{f,j} + \mathbf{K}_{i+1}\left(\widetilde{\mathbf{y}}_{i+1} - H\left\{\mathbf{x}_{i+1}^{f,j}\right\}\right) \quad , \quad j = 1, ..., r \tag{49}$$

providing implicitly the analysis error covariance matrix

$$\mathbf{P}_{i+1}^{a} = \frac{1}{r-1}\sum_{j=1}^{r}\left(\mathbf{x}_{i+1}^{a,j} - \overline{\mathbf{x}_{i+1}^{a,j}}\right)\left(\mathbf{x}_{i+1}^{a,j} - \overline{\mathbf{x}_{i+1}^{a,j}}\right)^{T}. \tag{50}$$

In order to avoid the problem of systematic underestimation of the analysis covariance that occurs when the same data and the same gain are used in the set of analysis equations, an ensemble of perturbed observations $\widetilde{\mathbf{y}}_{i+1}$ is considered in (49) instead of the original $\mathbf{y}_{i+1}$ [*Houtekamer and Mitchell*, 1998; *Burgers et al.*, 1998]. The stochastic nature of the EnKF filter arises as a consequence of using perturbed observations [*Lawson and Hansen*, 2004]. Deterministic variants of the EnKF, which do not require perturbed observations, have been proposed recently involving square-root analysis schemes [e.g. *Whitaker and Hamill,* 2002].

An interpretation of the error propagation scheme in the SEEK filter has been proposed by *Ballabrera et al.* [2001] in terms of ensemble model integrations. Indeed, Eq. (37) depicts the natural dispersion of an ensemble of different model trajectories initialized in the vicinity of the initial guess; it represents the amplification of unstable error modes (or the damping of stable modes) inherent in the system's dynamics. A more exhaustive exploration of the similarities and differences between the SEEK and Ensemble Kalman Filter (EnKF) is discussed in *Brusdal et al.* [2003] and *Nerger* [2004].

7. Statistical consistency of assimilation schemes

A key issue concerning statistical assimilation systems relates to their capacity to produce reliable error statistics about the ocean state estimates, and to propagate those error statistics properly from one assimilation cycle to the next. In the linear KF, the specification of the system noise $\mathbf{Q}$, the observation error $\mathbf{R}$ and the background error covariance at the initial time $\mathbf{P}_0$ perfectly determines the subsequent evolution of the error statistics throughout the assimilation sequence. This is because the observations actually control the trajectory of the model state, but they have no impact on the evolution of the error statistics themselves (except through the observation network $\mathbf{H}$). For a KF to yield optimal performances, it is necessary to provide the correct *a priori* description of these error covariance matrices. As only guesses of these quantities

are available in real cases, it is necessary to verify the consistency of the underlying assumptions.

Comparisons with independent information provide, of course, the ideal way to assess the quality of the assimilation and to detect what has possibly gone wrong in the system. Testing the ability to produce a forecast from an analyzed state (i.e., using data that have yet to be assimilated) is also very helpful in checking whether the dynamical model and assimilation scheme are consistent with one another. The most common test is to compare the forecast to persistence [*De Mey et al.*, 2002]. Additionally, information acquired during system operation can be used to verify if the prior error statistics have been prescribed in a way which represents the actual errors in the model and the observations. The difference between the observations $\mathbf{y}$ used in the assimilation system and the forecast or analysis fields (called innovations $\mathbf{d}$ and residuals $\mathbf{r}$, respectively) provide two series of data that contain a wealth of important information about the consistency of the prior error estimates. It can be shown that examining diagnostics on the innovation vector is essentially equivalent to examining them on the residual vector, but in practice some features of the analysis may be easier to diagnose from one or the other vector [*Talagrand*, 1999].

In the section below, we will examine a number of useful criteria that can be diagnosed from the sequence of innovations and residuals to detect imperfections in the specification of the error statistics. Since such imperfections can be a source of drift in the assimilation system or, even more dramatically, can affect the stability of the filter, we will introduce the concept of adaptivity, the aim of which is to enforce consistency between the error statistics predicted by the filter and the observed misfits.

7.1 Verification of statistical consistency

Simple diagnostics can be implemented quite easily in assimilation systems, providing interesting tools to monitor overall performance and detect anomalies in system operations. A first-order check can be made by computing the mean innovation, which is expected to vanish over a sufficiently long assimilation sequence:

$$\overline{\mathbf{d}_i} = \overline{\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f} = \overline{\mathbf{H}\mathbf{x}_i^t + \epsilon_i^o - \mathbf{H}(\mathbf{x}_i^t + \epsilon_i^f)} = \overline{\epsilon_i^o} - \mathbf{H}\overline{\epsilon_i^f} = 0 \qquad (51)$$

A non-centered innovation sequence is the obvious indication of biases in the model and/or observations which in principle should be removed from the assimilation system. If the source of bias is in the model, $\overline{\mathbf{d}_i}$ only provides information about the projection of this bias in the observation

space, and the difficulty remains of inversing this information back to the model space.

Similarly, the expectation of the residual $\overline{\mathbf{r}_i} = \overline{\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^a}$ should be zero for an optimal system as well as the mean increment $\mathbf{H}(\overline{\mathbf{x}_i^a - \mathbf{x}_i^f}) = \overline{\mathbf{d}_i} - \overline{\mathbf{r}_i}$ [*Talagrand*, 1999]. Figure 7 illustrates the mean assimilation increment of sea-surface height computed from a 10-year analysis sequence of the MERCATOR global prototype (at $2^\circ \times 2^\circ$ horizontal resolution) assimilating altimeter data during the 1993-2003 period [*Ferry et al.*, 2005]. Over large portions of the ocean, the amplitude of the increment is indeed very small, but in some regions the assimilation system systematically corrects the predicted sea-surface pressure towards higher (e.g. in the southern oceans) or lower values (e.g. in the sub-polar gyre). *Ferry et al.* [2005] discuss how the detection of such biases can be used to improve the modelling or assimilation components of the MERCATOR system.

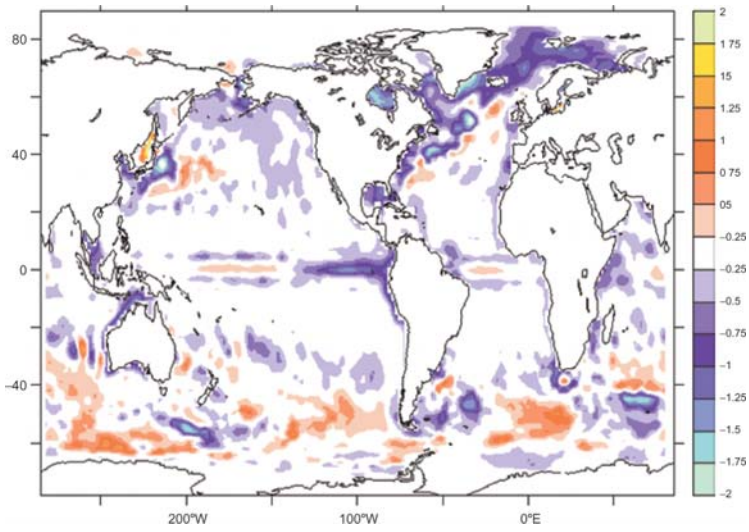

Figure 7. Mean SLA increment (in cm) diagnosed from the global MERCATOR prototype assimilating along-track altimeter data (according to Ferry et al. [2005]).

By taking the covariance of the innovation and remembering the assumptions made in Section 2 about observation and forecast errors, we obtain

$$\overline{\mathbf{d}_i \mathbf{d}_i^T} = \overline{(\epsilon_i^o - \mathbf{H}\epsilon_i^f)(\epsilon_i^o - \mathbf{H}\epsilon_i^f)^T} = \mathbf{R} + \mathbf{H}\mathbf{P}_i^f\mathbf{H}^T \tag{52}$$

if the error covariances are correctly specified. The comparison between the matrix $\overline{\mathbf{d}_i \mathbf{d}_i^T}$ and the sum of the observation and forecast error covariances used by the assimilation scheme indicates whether the forecast misfits "seen" by the filter are compatible with the prior information.

As all these quantities are full covariance matrices in the data space, the comparison is often made by looking only at a few selected diagnostics such as the trace of the matrix or its diagonal elements. After simple manipulations, it can be shown using the residuals that

$$\overline{\mathbf{r}_i \mathbf{r}_i^T} = \mathbf{R} - \mathbf{H}\mathbf{P}_i^a\mathbf{H}^T \tag{53}$$

This formula accounts for the fact that the same observation error affects both the data and the analysis. It shows that, for an asymptotically perfect estimation system (with $\mathbf{P}_i^a \sim 0$), the residual error covariance converges towards the observation error.

Another more synthetic diagnostic can be implemented in both statistical and variational assimilation systems. If the covariances are correctly estimated, the scalar quantity

$$J_i = \mathbf{d}_i^T(\mathbf{H}\mathbf{P}_i^f\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}_i \tag{54}$$

behaves as a chi-squared variable with as many degrees of freedom as there are independent data (say, p) [*Bennett*, 1992]. It can be regarded as a particular norm (the so-called Mahalanobis norm) of the innovation vector. Significance tests may therefore be used to accept or reject the covariance models. In the context of low-rank KFs based on the decomposition (39), the norm (54) can be written as

$$J_i = \mathbf{d}_i^T\left((\mathbf{H}\mathbf{S}_i^f)(\mathbf{H}\mathbf{S}_i^f)^T + \mathbf{R}\right)^{-1}\mathbf{d}_i \tag{55}$$

and can be used to objectively evaluate the suitability of the error subspace $\mathbf{S}_i^f$. Instead of testing the complete χ_p^2 behaviour of J_i, only the first and second statistical moments need to be examined. These should show that

$$\overline{J_i} = p \qquad \text{and} \qquad Var(J_i) = p \tag{56}$$

In particular, a too small (resp. large) value of $\overline{J_i}$ is symptomatic of an overestimated (resp. underestimated) innovation amplitude, which may be the consequence of too large (resp. small) observation or forecast error covariances.

7.2 Adaptivity

In most applied assimilation systems, large deviations may be observed with respect to the theoretical criteria (51), (52), (53) or (56), reflecting some flaws in the prior statistical assumptions. Feedback mechanisms can be implemented on-line or off-line to adjust the prior error statistics in such a way as to restore consistency between the errors diagnosed by the filter and the innovation or residual information. Inherent

in the family of techniques used for this purpose is the concept of adaptivity. A comparison between different sophisticated adaptive KFs is provided by *Blanchet et al.* [1997]. In this section, we examine more closely different examples of adaptive methods that have been explored in the literature to identify and correct model biases, to tune the parameterization of model errors and to build the error sub-space of low-rank KFs.

Research efforts aimed at improving error covariance modelling in assimilation systems are of limited interest if biases are left in the models or the observations, i.e. if criterion (51) is not verified. The correction for biases in operational systems is expected to have a very strong impact on assimilation performances. In the context of sequential assimilation, *Dee and Da Silva* [1998] proposed a rigorous method for estimating the forecast bias and correcting the forecast prior to the analysis, assuming unbiased observations. The algorithm is designed to perform on-line and its implementation does not require substantial modifications to the assimilation system. The basic idea consists in running a simplified KF-type algorithm to estimate $\overline{\mathbf{d}_i}$ in addition to the KF for the state itself. Assuming temporal persistence of the bias, the extra cost of the method is equivalent to one additional computation of the statistical analysis step.

The question of model error parameterization can been addressed by means of adaptive methods too. In many practical studies, the model error covariance $\mathbf{Q}$ is probably the least well known statistical quantity impacting the forecast error. *Mitchell and Houtekamer* [2000] developed an adaptive EnKF using a maximum likelihood method to estimate the parameterization of model errors from the innovation sequence. The approach recently developed by *Brankart et al.* [2003] can also be viewed as an adaptive parameterization of the model error. By adjusting the ρ coefficient of Eq. (38) according to the local innovation variance, the method is able to account for regional properties of the ocean dynamics. The estimation of the innovation variance is based on a weighted average of the latest innovations, using a weight that decreases exponentially with the age of the innovation.

This mechanism was explored in the context of hindcast experiments conducted for the 1993-1996 period. The SEEK filter was implemented in two different models : the 1/3° North Atlantic OPA model of the MERCATOR prototype system, and the Atlantic/Arctic MICOM model of the European DIADEM system. Sea-surface temperature from the NASA Pathfinder project and altimetric data from the Topex/Poseidon and ERS missions were assimilated in both systems every 10 days. *Testut* [2000] studied the distribution of the 10-day forecast error in the

OPA system once the assimilation experiment had reached an asymptotic regime (figure 8). The largest forecast errors are found along the Gulf Stream path between Cape Hatteras and 40°N, where the standard deviation exceeds 20 cm in some places. Local maxima are also detected in the North Atlantic Current extension and along the Azores Current at 35°N. Using the MICOM system, *Brankart et al.* [2003] compared the estimated innovation variance with the sum of the observation and forecast error variance diagnosed by the filter in a zonal section across the Gulf Stream region at 30°N. The comparison indicates a fairly good agreement between the 2 estimates in most regions, reflecting a successful adaptive procedure.

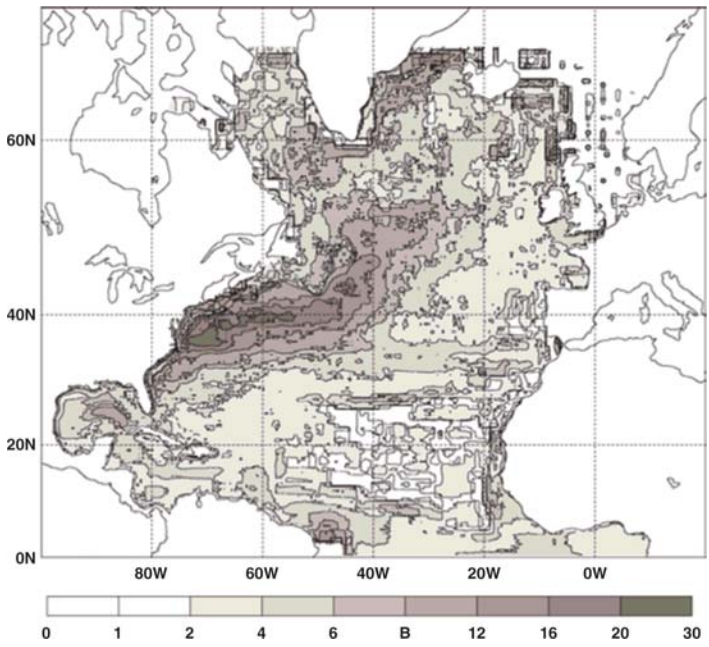

Figure 8. Distribution of the 10-day forecast error diagnosed with the adaptive SEEK filter implemented in the OPA model (reproduced from Testut [2000]).

Another interesting application of the adaptivity concept is the update of the initial structure of the error sub-spaces. Using a low-rank KF with an idealized primitive-equation model of the Gulf Stream, the assimilation error dynamics was investigated by *Ballabrera et al.* [2001]. They examined how the filter performances were affected by imperfect specifications of the initial error statistics. They observed that the truncation error does not impact the control of the solution when at least one

of the following three conditions is met: (i) the initial error is perfectly described by the reduced basis; (ii) the truncation error is dynamically uncoupled with the error components of the low-dimension space; (iii) the sub-space must contain all the components of the growing error during the forecast time period. As these three conditions are never perfectly verified in realistic assimilation systems, an adaptive procedure was developed in order to introduce some feedback between the data and the error sub-space used by the filter. The algorithm proposed by *Brasseur et al.* [1999] updates the error sub-space of the SEEK filter along the geostrophic attractor by extracting information left in the residual vector after each analysis step. This update of the reduced basis is performed in such a way as to attenuate the truncation error and to improve the projection of the next innovation onto the error sub-space: this leads to the concept of adaptive sub-space. An advantage of this variant of the SEEK filter is that it allows the evolution of the error sub-space without incurring the cost of propagating the whole set of error directions dynamically.

8. Intermittent vs. time-continuous filtering

In the basic assimilation problem introduced in Section 1, two major simplifications were considered: (i) the observations were available at discrete time intervals, and (ii) the analysis was performed at the exact time of the measurements. In oceanographic and atmospheric applications, the situation is quite different since the flow of observations can be considered as almost continuous in time (for instance, the acquisition of altimeter data). Therefore, ocean data assimilation with intermittent sequential filters necessarily involves approximations.

In principle, sequential filters could perform an analysis step as often as a new piece of information arrives. Time-continuous formulations of the KF exist [*Gelb* 1974] and have been applied to analogic signal processing. However, their application to oceanographic or atmospheric models is inappropriate: for practical reasons, it would be impossible to incorporate the data at their exact time of acquisition. Experience shows that it is necessary to accumulate a certain number of observations between two successive analysis steps to correct the ocean state with sufficient impact. Besides, operational assimilation systems must be scheduled on a regular temporal basis so as to avoid unnecessary algorithmic complications, to account for human intervention, delay in data delivery, etc.

8.1 Computing innovations using FGAT

Typical lengths of assimilation cycles are 3 to 7 days for mesoscale ocean current predictions, and 10 to 30 days for initialization of the oceanic component of seasonal climate predictions. In spite of the fact that the ocean cannot be considered as static over these time scales, intermittent sequential filters incorporate at one single instant the set of observations collected during the whole assimilation interval. This is a major difference with 4D-VAR algorithms, which can take full advantage of the temporal distribution of the data within an assimilation window.

Fairly simple solutions can be set up to alleviate these problems in the context of statistical filters. For example, the FGAT (First Guess at Appropriate Time) method initially introduced in meteorology can be used to evaluate the innovation vector more correctly: instead of computing the difference between the time-distributed data set and the model forecast at time t_{i+1} as in Eq. (19), the innovation is evaluated "on the flight" by cumulating the differences between each piece of observation and the corresponding element of the model forecast at the measurement time. This approach has also been used with 3D-VAR assimilation systems [*Weaver et al.*, 2003] to benefit from the temporal dimension. Due to the fast propagation of equatorial waves, the FGAT feature may be particularly important in the tropical oceans.

8.2 Incremental Analysis Update

A direct consequence of intermittency is the discontinuity of the forecast/analysis estimates, which is recognized as a major drawback of sequential methods. Two related problems, - shocks to the model and data rejection -, arise with intermittent corrections. It is found that observations assimilated into models may introduce transient waves excited by the impulsive insertion. These waves are often the result of imperfections in the corrected state associated with physically unbalanced error covariances. In the example illustrated in figure 9, six vertical profiles of temperature and salinity measurements are assimilated into the HYCOM model using the SEEK filter. The profiles are inserted in distant regions (Labrador Basin, Irminger Sea, Gulf Stream, Azores and North Brazil currents and Caribbean Sea) in order to avoid mutual interference. The corrected state is then integrated using the model with realistic forcings for one month after analysis time. Figure 9 depicts the SSH increment after 3 days of simulation, showing the occurrence of spurious transients (particularly in the Gulf Stream region) which the model generates to dynamically adjust the new state.

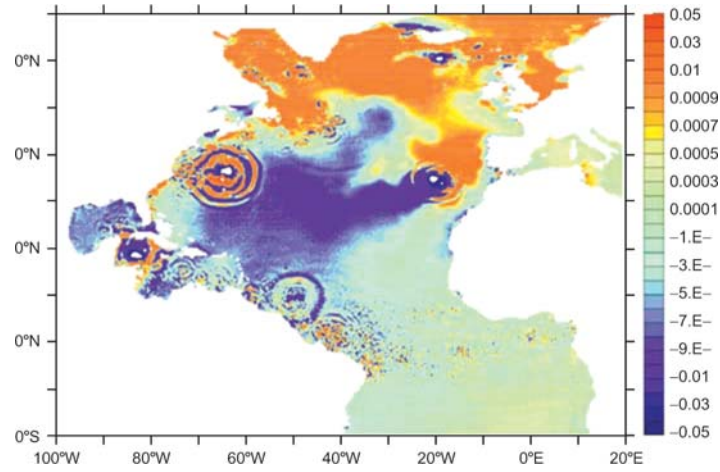

Figure 9. SSH increment (in m) obtained with the HYCOM model 3 days after assimilation of vertical T/S profiles.

In order to incorporate analysis increments in a more gradual manner, a new algorithm based on Incremental Analysis Updates (IAU) has been proposed by *Bloom et al.* [1996], which combines aspects of intermittent and continuous assimilation schemes. Using the regular KF equations, the IAU algorithm first computes the analysis correction $(\mathbf{x}^a_{i+1} - \mathbf{x}^f_{i+1})$. This correction is then distributed (uniformly or not) over the assimilation window and inserted gradually into the model. The state obtained at the end of the assimilation window can be used as the initial conditions for the next assimilation cycle, leading to time-continuous filtered trajectories. The concept of IAU can also be implemented by computing corrections to the atmospheric forcing or by introducing dynamical relaxation terms in the prognostic equations, according to the analysis increments. The IAU temporal strategy can be complemented by the FGAT scheme which computes the innovation "on the flight".

A possible implementation of the IAU scheme in primitive equation models may include the following steps: (i) first-guess model integration and evaluation of the innovation vector using FGAT; (ii) computation of the analysis increment using the Kalman gain; (iii) second integration on (t_i, t_{i+1}) using primitive equations modified by the increment. The additional cost incurred by this process will be one model integration or less, depending on the details of the algorithmic implementation. Noting by δT the increment of the temperature field, a possible modification of

the PE equation for the temperature may be written as follows:

$$\frac{\partial T}{\partial t} + \mathbf{u}.\nabla_h T + w\frac{\partial T}{\partial z} = D_h(T) + \frac{\partial}{\partial z}(\tilde{\lambda}\frac{\partial T}{\partial z}) + \frac{\delta T}{(t_{i+1} - t_i)} \qquad (57)$$

where $\delta T/(t_{i+1} - t_i)$ is a new term acting as a body force, in a similar way to the nudging assimilation technique [*Verron and Holland*, 1989; *Blayo et al.,* 1994]. The extra term is, however, different from newtonian relaxation since it does not explicitly involve the current value of T and has the advantage of being multivariate in general and weighted consistently with error statistics. A linear analysis of the IAU procedure shows that it has the properties of a low-pass temporal filter. Variants of the IAU procedure are currently being explored with the MERCATOR assimilation systems.

9. Conclusions

The Kalman filter provides a theoretical framework from which a hierarchy of algorithms of increasing sophistication and increasing computer requirements can be derived. These algorithms range from Optimal Interpolation schemes, which require only a few percent of the computer resources allocated to run the model, to non-linear Kalman filters such as the SEEK or the EnKF, which require the simultaneous integration of perturbed model trajectories in equal number to the dimension of the error sub-space. The research conducted in support of operational oceanography has shown that error statistics is at the heart of applied assimilation systems and remains a major challenge for ocean data assimilation. In this respect, the derivation of a hierarchy of simplified filters offers a unique possibility to test different approaches for specifying and calibrating the background, systematic and observation error statistics.

A salient feature of statistical estimation methods is their multivariate nature : observations related to several different model variables (e.g., SST, SLA, SSS, in situ temperature and salinity data) are used to correct the whole model state in a consistent manner. This opens promising avenues for effectively envisaging the simultaneous assimilation of *in situ* data with satellite data from different sensors. Error bars on state estimates can be computed by the algorithms on a rigorous basis, making the schemes useful in qualifying the reliability of different forecasts, or comparing the relevance of different observation systems. For practical implementations, the adjoint of the model code is not necessarily needed as with 4D-VAR, and the basic architecture of the algorithms is modular. The transition from one model version to another, or from one model code to another can be made smoothly without much recoding.

The large variety of sequential statistical methods developed in the context of scientific or operational applications is an indication that a given assimilation technique cannot be considered as a "plug-and-play" system, capable of solving universal problems. To design the best possible assimilation system, it is necessary to clearly define the purpose of data assimilation (forecast initialization, reanalysis, model error detection, etc.), the physical characteristics of the processes of interest, the sampling properties of the observation systems, and the limitations of the assimilation techniques. Moving towards hybrid sequential/variational methods is probably an astute way of taking the best from both approaches. Due to the complexity of models and algorithms, the success of an assimilation system also depends on the community of people working within a common framework. In the future, the sharing of generic assimilation tools between operational and research teams should accelerate the progress of the methods and provide feedback from more intensive utilizations.

A number of important issues for ocean data assimilation have yet to be fully resolved, such as the incorporation of inequality constraints in statistical estimation algorithms. Such constraints are needed, e.g. in the context of tracer data assimilation to maintain the positiveness of tracer concentrations, or for the proper handling of the physical properties of the water column such as static stability. Another important challenge will be to further develop assimilation systems for coupled physical-biological models, with the aim of demonstrating the capacity to estimate and forecast marine ecosystems and biogeochemical variables in the ocean on a routine basis. As a first step, the incorporation of biogeochemical models into the physical assimilative systems developed within GODAE will provide new "bio-metrics" to evaluate the representation of the physical processes of critical importance to biology. A further challenge will be to implement suitable methods to assimilate ocean colour data in the coupled assimilative systems.

Acknowledgements

The author gratefully acknowledges his colleagues from LEGI and MERCATOR Ocean for numerous discussions and valuable suggestions. He is supported by the Centre National de la Recherche Scientifique. Computations were carried out at the IDRIS/CNRS computing centre. The contribution of students from the GODAE school who commented an earlier version of these lecture notes has been much appreciated. This work has been partly supported by the MERSEA project of the European Commission under Contract SIP3-CT-2003-502885.

References

Auclair, F., P. Marsaleix and P. De Mey, 2003 : Space-time structure and dynamics of the forecast error in a coastal circulation model of the Gulf of Lions. *Dyn. Atmos. Oceans*, **36**, 309-346.

Ballabrera J., P. Brasseur and J. Verron, 2001: Dynamical evolution of the error statistics with the SEEK filter to assimilate altimetric data in eddy-resolving ocean models. *Quart. J. Roy. Meteor. Soc.*, **127**, 233-253.

Bennett A., 1992: *Inverse methods in physical oceanography.* Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press, 346 pp.

Birol F., J.M. Brankart, F. Castruccio, P. Brasseur and J. Verron, 2004: Impact of the Ocean Mean Dynamic Topography on Satellite Data Assimilation, *Mar. Geodesy*, **27**, 59-78.

Blanchet I., C. Frankignoul and M.A. Cane, 1997: A comparison of adaptative Kalman filters for a tropical Pacific ocean model. *Mon. Wea. Rev.*, **125**, 40-58.

Blayo E., J. Verron and J.M. Molines, 1994: Assimilation of Topex/Poseidon altimeter data into a circulation model of the North Atlantic. *J. Geophys. Res.*, **99** (C12), 24691-24705.

Bleck R., 2002: An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modelling*, **4**, 55-88.

Bloom S.C., L.L. Takacs, A.M. Da Silva and D. Ledvina, 1996: Data Assimilation Using Incremental Analysis Updates, *Mon. Wea. Rev.*, **124**, 1256-1271.

Brankart J.-M., C.E. Testut, P. Brasseur and J. Verron, 2003: Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: Application to a 1993-96 hindcast of the North Atlantic Ocean circulation, *J. Geophys. Res.*, **108**, 3074, doi: 10.1029/2001JC001198.

Brasseur, P., J. Ballabrera and J. Verron, 1999: Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model. *J. Mar. Syst.*, **22** (4), 269-294.

Brusdal K., J.M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P.J. van Leeuwen, E. Dombrowsky and J. Verron, 2003: A demonstration of ensemble-based assimilation methods with a layered OCGM from the perspective of operational ocean forecasting systems, *J. Mar. Syst.*, **40-41**, 253-289.

Burgers G., P. van Leeuwen and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter, *Mon. Wea. Rev.*,**126** , 1719-1724.

Cane M.A., A. Kaplan, R.N. Miller, B. Tang, E.C. Hackert and A.J. Busalacchi, 1996: Mapping tropical Pacific sea level: Data assimilation via a reduced state Kalman filter. *J. Geophys. Res.*, **101**, C10, 22,599-22,617.

Carme S., Pham D. T., et J. Verron, 2001: Improving the Singular Evolutive Extended Kalman filter for strongly non-linear models for use in ocean data assimilation. *Inverse Problems*, **17**, 1535-1559.

Carmillet V., J. M. Brankart, P. Brasseur, H. Drange,G. Evensen et J. Verron, 2001 : Assimilation of ocean color data into an ecosystem model of the North Atlantic ocean. *Ocean Modelling*, **3**, 167-192.

Cohn S.E. and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics, *J. Meteo. Soc. Japan*, **74**, 63-75.

Cooper M. and K. Haines, 1996: Altimetric assimilation with water property conservation. *J. Geophys. Res.*, **101** (C1), 1059-1077.

Courtier P., 1997: Variational Methods, *J. Met. Soc. Japan*, **75**, 211-218.

Daley R., 1991: *Atmospheric Data Analysis*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, 457 pp.

Dee D.P., 1991: Simplification of the Kalman filter for meteorological data assimilation, *Quart. J.Roy. Meteor. Soc.*, **117**, 365-384.

Dee D.P and Da Silva A., 1998: Data assimilation in the presence of forecast bias, *Quart.. J. Roy. Meteor. Soc.*, **124**, 269-295.

De Mey P. and M. Benkiran, 2002: A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation. In: *Ocean Forecasting: Conceptual basis and applications*, N. Pinardi and J.D. Woods, Eds, Springer-Verlag, Belin, Heidelberg, New York, 472 pp.

De Mey P., T. Awaji, M. Bell, A.F. Bennett, P. Brasseur, G. Evensen, K. Haines, I. Fukumori, O.M. Smedstad, D. Stammer and A. Weaver, 2002 : Approaches to Data Assimilation within GODAE. *Proceedings of the international symposium "En route to GODAE".*

Echevin V., P. De Mey and G. Evensen, 2000: Horizontal and Vertical Structure of the Representer Functions for Sea Surface Measurements in a Coastal Circulation Model. *J. Phys. Oceanogr.*, **30**, 2627-2635.

Emery W.J. and R.E. Thomson, 1998: *Data analysis methods in physical oceanography*, Pergamon, 634 pp.

Evensen G., 1994: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10,143-10,162.

Evensen G. and P.J. van Leeuwen, 2000: An Ensemble Kalman Smoother for nonlinear Dynamics, *Mon. Wea. Rev.*, **128**, 1852-1867.

Evensen G., 2003 : The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**(4), 343-367.

Ferry N., E. Remy, P. Brasseur and C. Maes, 2005: The MERCATOR global ocean operational analysis/forecast system: assessment and validation of a 11-year reanalysis, submitted to *J. Mar. Syst.*

Fu L.L., I. Fukumori and R.N. Miller, 1993: Fitting dynamic models to Geosat sea-level observations in the tropical Pacific ocean. Part II: A linear, wind-driven model. *J. Phys. Oceanogr.*, **23**, 2162-2181.

Fukumori I., J. Benveniste, C. Wunsch and D.B. Haidvogel, 1993: Assimilation of sea surface topography into an ocean circulation model using a steady state smoother. *J. Phys. Oceanogr.*, **23**, 1831-1855.

Fukumori I., 1995: Assimilation of Topex sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific ocean. *J. Geophys. Res.*, **100** (C12), 25,027-25,039.

Fukumori I. and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: an example with an idealized Gulf Stream model. *J. Geophys. Res.*, **100** (C4), 6,777-6,793.

Fukumori I., 2001: Data Assimilation by Models. In: *Satellite Altimetry and Earth Sciences*, L.L. Fu and A. Cazenave Eds., International Geophysics Series, Vol. **69**, Academic Press, 463 pp.

Fukumori I., 2002: A partitioned Kalman Filter and Smoother, *Mon. Wea. Rev.*, **130**, 1370-1383.

Gavart M., P. De Mey and G. Caniaux, 1999: Assimilation of satellite altimeter data in a primitive-equation model of the Azores-Madeira region, *Dyn. Atmos. Oceans*, **29**, 217-254.

Gelb A., 1974: *Applied optimal estimation.* MIT Press, Cambridge, MA, 374 pp.

Gourdeau L., J. Verron, T. Delcroix, A.J. Busalacchi et R. Murtugudde, 1999 : Assimilation of Topex/Poseidon altimetric data in a primitive equation model of the tropical Pacific ocean, 1992-1996. *J. Geophys. Res.*, **105** (C4), 8473-8488.

Hurlburt H.E., 1986: Dynamic transfer of simulated altimetric data into subsurface information by a numerical ocean model. *J. Geophys. Res.*, **91** (C2), 2372-2400.

Houtekamer P.L. and H.L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.*, **126**, 796-811.

Ide K., A.F. Bennett, P. Courtier, M. Ghil and A.C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential and variational. *J. Meteor. Soc. Japan*, **75**, 181-189.

Jazwinski A.H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.

Kalnay E., D.L.T. Anderson, A.F. Bennett, A.J. Busalacchi, S.E. Cohn, P. Courtier, J. Derber, A.C. Lorenc, D. Parrish, J. Purser, N. Sato and T. Schlatter, 1997: Data assimilation in the ocean and in the atmosphere: what should be next ? *J. Meteor. Soc. Japan*, **75**, 489-496.

Kalnay E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, 341 pp.

Kelly K., 1988: Comments on "Empirical Orthogonal Function Analysis of Advanced Very High Resolution Radiometer Surface Temperature Patterns in Santa Barbara Channel" by G.S.E. Lagerloef and R.L. Bernstein, *J. Geophys. Res.*, **93**, 15753-15754.

Lawson W.G. and J.A. Hansen, 2004: Implications of Stochastic and Deterministic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth, *Mon. Wea. Rev.*, **132**, 1966-1981.

Lermusiaux P.F.J., 1999: Data assimilation via error subspace statistical estimation: Part II. Middle Atlantic Bight shelfbreak front simulations and ESSE validation. *Mon. Wea. Rev.*, **127**, 1408-1432.

Lermusiaux P.F.J., 2001: Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay. *J. Mar. Syst.*, **29**, 385-422.

Lions J.L., O.P. Manley, R. Temam and S. Wang, 1997: Physical Interpretation of the Attractor Dimension for the Primitive Equations of Atmospheric Circulation, *J. Atmos. Sci.*, **54**, 1137-1143.

Mellor G.L. and T. Ezer, 1991: A Gulf Stream model and an Assimilation Scheme, *J. Geophys. Res.*, **96**, 8779-8795.

Miller R.N., 1989: A Kalman filter analysis of sea level height in the tropical Pacific, *J. Phys. Oceanogr.*, **19**, 773-790.

Miller R.N. and Ehret L., 2002: Ensemble Generation for Models of Multimodal Systems, *Mon. Wea. Rev.*, **130**, 2313-2333.

Mitchell H.L. and P.L. Houtekamer, 2000: An Adaptive Ensemble Kalman Filter, *Mon. Wea. Rev.*, **128**, 416-433.

Morrison D.F., 1988: *Multivariate Statistical Methods*, Probability and Statistics Series, McGraw-Hill, 415 pp.

Nerger L., 2004: Parallel Filter Algorithms for Data Assimilation in Oceanography, *Reports on Polar and Marine Research*, Alfred Wegener Institute for Polar and Marine Research, Germany, Volume **487**, 175 pp.

Oschlies A. and J. Willebrand, 1996: Assimilation of Geosat altimeter data into an eddy-resolving primitive equation model of the North Atlantic ocean. *J. Geophys. Res.*, **101**, 14175-14190.

Ott E., B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil and J.A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation, *Tellus*, **56A**, 415-428.

Parent L., 2000 : *Assimilation de données dans l'océan Pacifique Tropical sur la période 1994-1998*, Ph. D. dissertation, Univ. J. Fourier, Grenoble, 222 pp.

Parent L., C.E. Testut, J.M. Brankart, J. Verron, P. Brasseur and L. Gourdeau, 2003 : Comparative assimilation of Topex/poseidon and ERS altimetric data and of TAO temperature data in the tropical Pacific Ocean during 1994-1998, and the mean sea-surface height issue, *J. Mar. Syst.*, **40-41**, 381-401.

Penduff Th., P. Brasseur, C.E. Testut, B. Barnier and J. Verron, 2002: Assimilation of sea-surface temperature and altimetric data in the South Atlantic Ocean : impact on basin-scale properties, *J. Mar. Res.*, **60**, 805-833.

Pham D. T., J. Verron and M.C. Roubaud, 1998 : A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.*, **16**, 323-340.

Pinardi N., A. Rosati and R. Pacanowski, 1995: The sea-surface pressure formulation of rigid lid models. Implications for altimetric data assimilation studies. *J. Mar. Syst.*, **6**, 109-120.

Rienecker M. and R. Miller, 1991: Ocean Data Assimilation Using Optimal Interpolation with a Quasi-Geostrophic Model, *J. Geophys. Res.,* **96** (C8), 15093-15103.

Robert C., S. Durbiano, E. Blayo and J. Verron, 2005: A reduced order strategy for 4D-Var ocean data assimilation, *J. Mar. Syst.*, accepted.

Talagrand O., 1999: A posteriori verification of analysis and assimilation algorithms, *Diagnosis of Data Assimilation Systems Workshop Proceedings*, ECMWF, Reading U.K., November 1998, 17-28.

Testut C.E., 2000 : *Assimilation de données satellitales avec un filtre de Kalman de rang réduit dans un modèle aux Equations Primitives de l'océan Atlantique*, Ph. D. dissertation, Univ. J. Fourier, Grenoble, 228 pp.

Testut C.-E., P. Brasseur, J.M. Brankart and J. Verron, 2003 : Assimilation of sea-surface temperature and altimetric observations during 1992-1993 into an eddy-permitting primitive equation model of the North Atlantic Ocean, *J. Mar. Syst.*, **40-41**, 291-316.

Thiebaux H.J., 1985: On approximations to geopotential and wind-field correlation structures, *Tellus*, **37A**, 126-131.

van Leeuwen P.J. and G. Evensen, 1996: Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation, *Mon. Wea. Rev.*, **124**, 2898-2913.

Verlaan M. and A.W. Heemink, 1997: Tidal flow forecasting using reduced-rank square root filter, *Stochastic Hydro. Hydraul.*, **11**, 349-368.

Verron J. and W.R. Holland, 1989: Impact de données d'altimétrie satellitaire sur les simulations numériques des circulations générales océaniques aux latitudes moyennes, *Ann. Geophys.*, **7**, 31-46.

Verron J., L. Gourdeau, D.T. Pham, R. Murtugudde and A.J. Busalacchi, 1999: An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific Ocean: Method and validation, *J. Geophys. Res.* **104** (C3), 5441-5458.

Weaver A., J. Vialard and D.L.T. Anderson, 2003: Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks, *Mon. Wea. Rev.*, **131**, 1360-1378.

Whitaker J. and T.M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations, *Mon. Wea. Rev.*, **130**, 1913-1924.

Chapter 11

WHAT IS DATA ASSIMILATION REALLY SOLVING, AND HOW IS THE CALCULATION ACTUALLY DONE?

Ichiro Fukumori

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, USA

Abstract: Data assimilation is reviewed in the context of an inverse problem. The mathematical nature of the problem is examined and some of its common solutions are described, clarifying some of the implicit assumptions that underlie both problem and solution. For instance, Kalman filtering and Rauch-Tung-Striebel smoothing can be identified as recursive least-squares inversions of the assimilation problem but of different parts of the problem. The temporal evolution of a filtered solution is not physically consistent, but that of a smoothed solution is. Understanding these characteristics is essential in effectively assimilating observations as well as in utilizing and further improving the assimilated solution. Practical steps in implementing a filtering and smoothing algorithm are illustrated by examples from the Consortium for "Estimating the Circulation and Climate of the Ocean" (ECCO).

Keywords: Data assimilation, Kalman filter, smoother, consistency, ECCO.

1. Introduction

Data assimilation is a procedure in which observations are combined with models. The observations correct model errors on the one hand, and the models extrapolate the data information in space, time, and among different properties on the other. The result of assimilation is generally a more complete and more accurate description of the state of the modeled system than those obtained by either observations or model simulations alone. However, data assimilation is not a panacea for correcting every model error or for compensating all deficiencies of observations.

Because ocean models have finite degrees of freedom, model estimates are inherently different from observations regardless of errors in

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 317-342.

measurements. Moreover, most data assimilation schemes incorporate approximations and/or simplifications that dictate what is being solved and how the results could be utilized. For instance, because of data increments, budgets of heat and other properties cannot be closed in a physically consistent manner for many sequential data assimilation estimates while for other estimates budgets can be closed. Understanding what is being solved and how it is done so are fundamental to utilizing data assimilated estimates and to devising means of improving them further.

The nature of the data assimilation problem and some of its solutions are reviewed to clarify these underlying properties and to elucidate their implications. The data assimilation problem is mathematically identified in Section 2. In Section 3, the Kalman filter and Rauch-Tung-Striebel smoother are compared in the context of a least-squares solution to this mathematical problem. The nature of data and model errors that are utilized as weights in assimilation is reviewed in Section 4. Practical issues in implementing these solutions are described in Section 5, using examples from the near real-time data assimilation system of the Consortium for "Estimating the Circulation and Climate of the Ocean" (ECCO; Stammer *et al.* 2002.) The discussion is summarized in Section 6.

2. Data assimilation as an inverse problem

Mathematically, data assimilation can be identified as an inverse problem; the state of a dynamic system (model), $\mathbf{x}$, and its controls, $\mathbf{u}$, (non-state variables of the model) are estimated given a set of observations, $\mathbf{y}$, and a model; e.g.,

$$\begin{pmatrix} \vdots \\ \mathbf{H}\mathbf{x}_t \\ \mathbf{H}\mathbf{x}_{t+1} \\ \vdots \\ \mathbf{x}_{t+1} - \mathbf{A}\mathbf{x}_t - \mathbf{G}\mathbf{u}_t \\ \mathbf{x}_{t+2} - \mathbf{A}\mathbf{x}_{t+1} - \mathbf{G}\mathbf{u}_{t+1} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \mathbf{y}_t \\ \mathbf{y}_{t+1} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix} \tag{1}$$

where … denote similar equations at different instances, t, indicated by the subscripts. Variable $\mathbf{x}$ consists of all the model's prognostic variables and $\mathbf{u}$ includes forcing, boundary condition, and sources of model error. Terms that include quantities to be solved ($\mathbf{x}$ and $\mathbf{u}$) are on the left hand side and the rest are placed on the right hand side. The upper part of Eq (1) relates the model

state to the observations by the observation operator **H**. The lower part describes the model's temporal evolution by operators **A** and **G** that embody the model physics and dynamics. The right hand side of the model equations (lower part of Eq 1) is identically zero as all terms of the model are generally uncertain and are placed on the left hand side. (Sources of model error are included in **u**.)

For simplicity, we assume a linear model for most of this discussion. The problem above and the solutions discussed below can be extended to non-linear models with suitable linearization. Bold upper and lower case characters represent matrices and column vectors, respectively. The time increment from *t* to *t+1* above denotes an arbitrary increment, as opposed to a single model time-step, and corresponds to instances at which observations are available.

As in most geophysical inverse problems[1], Eq (1) is rank deficient. In particular, there are generally more unknowns than the number of constraints. For example, the dimension of **x**, excluding the temporal dimension, is of order several million for typical general circulation models, whereas there are only about 20,000 hydrographic profiles during the entire World Ocean Circulation Experiment. Consequently, there are an infinite number of solutions that could satisfy Eq (1). Different criteria are used to derive particular solutions. One such criterion is least-squares, and is reviewed below.

3. Kalman filter and Rauch-Tung-Striebel smoother as least squares inversions

The least squares solution (cf. Chapter 10) provides a general solution to inverse problems such as Eq (1). Namely, the least squares solution $\hat{\mathbf{a}}$ (^ denotes an estimate) and its error covariance matrix $\mathbf{R}_{\hat{a}\hat{a}}$ for a general linear inverse problem,

$$\mathbf{E}\mathbf{a} = \mathbf{b} \tag{2}$$

when the right hand side **b** is given (known), are,

$$\hat{\mathbf{a}} = \mathbf{a}_0 + \mathbf{R}_{\mathbf{aa}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R}_{\mathbf{aa}}\mathbf{E}^T + \mathbf{R}_{\mathbf{bb}}\right)^{-1}\left(\mathbf{b} - \mathbf{E}\mathbf{a}_0\right) \tag{3}$$

$$\mathbf{R}_{\hat{a}\hat{a}} = \mathbf{R}_{\mathbf{aa}} - \mathbf{R}_{\mathbf{aa}}\mathbf{E}^T\left(\mathbf{E}\mathbf{R}_{\mathbf{aa}}\mathbf{E}^T + \mathbf{R}_{\mathbf{bb}}\right)^{-1}\mathbf{E}\mathbf{R}_{\mathbf{aa}} \tag{4}$$

---

[1] Basic matrix algebra is fundamental to mathematical discussions below and data assimilation in general. See, for instance, Wunsch (1996) for a general discussion of inverse methods that includes a brief summary of matrix and vector algeba relevant to the subject.

where $\mathbf{a}_0$ is a prior estimate of **a,** and $\mathbf{R}_{\mathbf{aa}}$ and $\mathbf{R}_{\mathbf{bb}}$ are prior error covariance matrices of $\mathbf{a}_0$ and **b**, respectively. (The latter includes representation error for **E**. See Section 4.1 for further discussion.) Filtering and smoothing algorithms can be identified as such least squares inversions and are reviewed below focusing on what they respectively solve.

A least-squares solution is identical to a minimum variance estimate when weights used in least-squares are suitable inverse error covariance matrices. These solutions are optimal in the sense that they optimize a given criteria (function) and that the expected error variance of $\hat{\mathbf{a}}$ is minimum among all (linear) estimates. Least-squares, as well as filtering and smoothing described below, do not necessarily assume Gaussian statistics. When the statistical distribution of **a** is Gaussian, the least-squares estimate is also a maximum likelihood estimate. Otherwise, the least-squares solution and the maximum likelihood estimate are distinct.

3.1 What do Kalman filters solve?

The Kalman filter (e.g., Chapter 11) corrects model forecasts $\hat{\mathbf{x}}_t^f$ and its error covariance matrix $\mathbf{P}_t^f$ by,

$$\mathbf{x}\ddot{\mathbf{o}}_t^a = \mathbf{x}\ddot{\mathbf{o}}_t^f + \mathbf{P}_t^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_t^f \mathbf{H}^T + \mathbf{R}_t\right)^{-1} \left(\mathbf{y}_t - \mathbf{H}\mathbf{x}\ddot{\mathbf{o}}_t^f\right) \tag{5}$$

$$\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{P}_t^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_t^f \mathbf{H}^T + \mathbf{R}_t\right)^{-1} \mathbf{H}\mathbf{P}_t^f \tag{6}$$

using notation defined in Eq (1). **R** is the data error covariance matrix (cf Section 4.1). Superscripts *f* and *a* denote model forecast and filter analysis, and $\hat{\mathbf{x}}_t^a$ and $\mathbf{P}_t^a$ are the Kalman filter's state analysis and its corresponding error covariance matrix, respectively. The correspondence between Eqs (3) and (5) and between (4) and (6) shows that the Kalman filter can be regarded as a least squares inversion of operator **H**.

However, given that the data assimilation problem is a combined inversion of observations and model equations (Eq 1), the Kalman filter does not solve (invert) the entire data assimilation problem, in particular, the model equations (lower part of Eq 1). In fact, combining Eq (5) with the model forecasting step, $\hat{\mathbf{x}}_t^f = \mathbf{A}\hat{\mathbf{x}}_{t-1}^a + \mathbf{G}\hat{\mathbf{u}}_t^0$, where $\hat{\mathbf{u}}_t^0$ is the a priori estimate of the control, the temporal evolution of the Kalman filter analysis satisfies,

$$\hat{\mathbf{x}}_t^a = \mathbf{A}\hat{\mathbf{x}}_{t-1}^a + \mathbf{G}\hat{\mathbf{u}}_{t-1}^0 + \mathbf{P}_t^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_t^f \mathbf{H}^T + \mathbf{R}_t\right)^{-1} \left(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_t^f\right) \tag{7}$$

Eq (7) is different from the model equations (lower half of Eq 1) due to the filter's data increment (third term of Eq 7). As illustrated in Figure 1, the data increment (black line) is not ascribed to particular processes as are the first two terms of Eq (7) (dotted black curve), and thus the temporal

evolution between $\hat{\mathbf{x}}^a_{t-1}$ and $\hat{\mathbf{x}}^a_t$ is physically inconsistent. For instance, budgets of heat and other properties cannot be closed between the two instances.

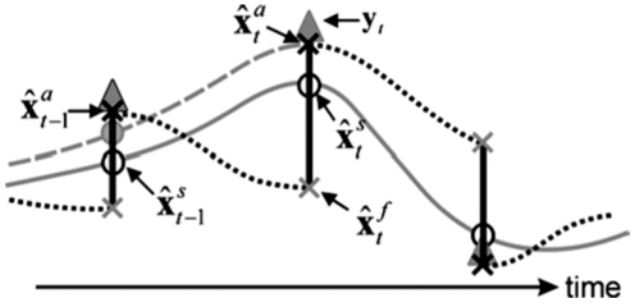

Figure 1. Schematic of a state element's temporal evolution in a typical sequential assimilation. Abscissa is time and ordinate is the state's value. Filtering progresses by a model forecasting step integrating the model along the dotted black curve from an analysis $\hat{\mathbf{x}}^a_{t-1}$ (black cross) to a forecast $\hat{\mathbf{x}}^f_t$ (gray cross). At time t the Kalman filter corrects the forecast $\hat{\mathbf{x}}^f_t$ to another analysis $\hat{\mathbf{x}}^a_t$ (black cross), bringing the model state closer to the observations $\mathbf{y}_t$ (gray triangle) along the solid black line. This filter correction is inverted by a smoother that corrects the model's prior evolution (dotted black curve) and the prior analysis $\hat{\mathbf{x}}^a_{t-1}$ (black cross) as depicted by the dashed gray curve and gray circle, respectively. In turn, differences at earlier times can be further inverted backwards in time. A general smoothed estimate and its temporal evolution initiated at some future instant is depicted by the white circles (e.g., $\hat{\mathbf{x}}^s_{t-1}$ and $\hat{\mathbf{x}}^s_t$) and the solid gray curve, respectively.

3.2 What is a smoother?

Whereas filters solve only the upper part of Eq (1), smoothers invert the entire data assimilation problem identified by Eq (1). The data increment in Eq (7) represents errors in the model that is being corrected by the assimilated data. These errors include those of the prior model evolution (dotted black curve in Figure 1) as well as those of the state at the previous assimilation instant from which the model forecasting step was taken (black crosses). The correspondence between the data increment and these errors can be recognized as another inverse problem defined by the lower half of Eq (1) that has not been solved by the Kalman filter (Eq 5). The sequential smoother described below employs the filtered solution to invert the model's temporal evolution that defines this lower half of Eq (1).

Namely, given the data assimilated analysis at time t, $\hat{\mathbf{x}}^a_t$, the model equations of Eq (1) define another inverse problem,

$$\hat{\mathbf{x}}_t^a = \mathbf{A}\hat{\mathbf{x}}_{t-1}^s + \mathbf{G}\hat{\mathbf{u}}_{t-1}^s = \left(\mathbf{A}\,\mathbf{G}\right)\begin{pmatrix}\hat{\mathbf{x}}_{t-1}^s \\ \hat{\mathbf{u}}_{t-1}^s\end{pmatrix} \tag{8}$$

to estimate the model state and control at time *t-1,* denoted by superscript *s*. Eq (8) can also be solved by least squares (Eq 3). In particular, as model error sources (process noise, See Section 4.2) are explicitly included in the formulation (**u**), an exact solution can be sought that would satisfy model constraints (e.g., closed heat budget, etc) by estimating these errors. This amounts to setting $\mathbf{R_{bb}} = 0$ in Eq (3). The filtered estimate $\hat{\mathbf{x}}_{t-1}^a$ and the a priori control $\hat{\mathbf{u}}_{t-1}^0$ provide the prior solutions in (3), and their error covariance matrix defines the equivalent of $\mathbf{R}_{aa}$;

$$\begin{pmatrix}\mathbf{P}_{t-1}^a & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{t-1}\end{pmatrix} \tag{9}$$

where **Q** denotes the error covariance of $\hat{\mathbf{u}}^0$. Standard Kalman filtering assumes temporally uncorrelated process noise that makes errors in $\hat{\mathbf{x}}_{t-1}^a$ and $\hat{\mathbf{u}}_{t-1}^0$ uncorrelated to each other, and thus off-diagonal blocks are zero in Eq 9.

Substitution of these elements in Eq (3) yields new estimates $\hat{\mathbf{x}}_{t-1}^s$ and $\hat{\mathbf{u}}_{t-1}^s$ such that,

$$\begin{aligned}\begin{pmatrix}\hat{\mathbf{x}}_{t-1}^s \\ \hat{\mathbf{u}}_{t-1}^s\end{pmatrix} &= \begin{pmatrix}\hat{\mathbf{x}}_{t-1}^a \\ \hat{\mathbf{u}}_{t-1}^0\end{pmatrix} \\ &+ \begin{pmatrix}\mathbf{P}_{t-1}^a\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{P}_{t-1}^a\mathbf{A}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\right)^{-1} \\ \mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}} + \mathbf{A}\mathbf{P}_{t-1}^a\mathbf{A}^{\mathrm{T}}\right)^{-1}\end{pmatrix}\left(\hat{\mathbf{x}}_t^a - \mathbf{A}\hat{\mathbf{x}}_{t-1}^a - \mathbf{G}\hat{\mathbf{u}}_{t-1}^0\right)\end{aligned} \tag{10}$$

Previous filtered estimates at time *t-2* can be improved and be made consistent with this estimate using these results ($\hat{\mathbf{x}}_{t-1}^s$ as opposed to the filter analysis $\hat{\mathbf{x}}_{t-1}^a$ in Eq 8) in another inversion. By induction, other filtered estimates at earlier instances can be improved by such inversion recursively back in time such that,

$$\begin{aligned}\begin{pmatrix}\hat{\mathbf{x}}_{t-1}^s \\ \hat{\mathbf{u}}_{t-1}^s\end{pmatrix} &= \begin{pmatrix}\hat{\mathbf{x}}_{t-1}^a \\ \hat{\mathbf{u}}_{t-1}^0\end{pmatrix} \\ &+ \begin{pmatrix}\mathbf{P}_{t-1}^a\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{P}_{t-1}^a\mathbf{A}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\right)^{-1} \\ \mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}} + \mathbf{A}\mathbf{P}_{t-1}^a\mathbf{A}^{\mathrm{T}}\right)^{-1}\end{pmatrix}\left(\hat{\mathbf{x}}_t^s - \mathbf{A}\hat{\mathbf{x}}_{t-1}^a - \mathbf{G}\hat{\mathbf{u}}_{t-1}^0\right)\end{aligned} \tag{11}$$

in general. (Note the use of $\hat{\mathbf{x}}_t^s$ in the last term instead of $\hat{\mathbf{x}}_t^a$, thus defining a recursion.)

The recursive relation Eq (11) can be recognized as the Rauch-Tung-Striebel (RTS) fixed-interval smoother. The RTS smoother can be shown to provide estimates of the state and control using all observations within a fixed time interval and is a general solution to the assimilation problem (Eq 1). (The smoother alters all filtered estimates except that at the end of the fixed time-interval; i.e., The Kalman filter estimate at the end of the time-interval is a least-squares solution of Eq (1) but not at intervening times.)

In Eq (11), past data information is contained in the Kalman filtered analysis $\hat{\mathbf{x}}_{t-1}^{a}$ while information of formally future observations is carried backward in time by the smoothed estimate $\hat{\mathbf{x}}_{t}^{s}$. Owing to the additional information from formally future observations, the smoothed estimates are generally more accurate (has smaller error) than corresponding filtered estimates. The error covariance matrix of the smoothed estimates $\hat{\mathbf{x}}_{t-1}^{s}$ and $\hat{\mathbf{u}}_{t-1}^{s}$ (Eq 11), $\mathbf{P}_{t-1}^{s}$ and $\mathbf{Q}_{t-1}^{s}$, respectively, is given by,

$$\begin{pmatrix} \mathbf{P}_{t-1}^{s} \\ \mathbf{Q}_{t-1}^{s} \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{t-1}^{a} \\ \mathbf{Q}_{t-1} \end{pmatrix} - \begin{pmatrix} \mathbf{P}_{t-1}^{a}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{P}_{t-1}^{a}\mathbf{A}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\right)^{-1}\mathbf{A}\mathbf{P}_{t-1}^{a} \\ \mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}} + \mathbf{A}\mathbf{P}_{t-1}^{a}\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{G}\mathbf{Q}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{L}_{t-1}\mathbf{P}_{t}^{a}\mathbf{L}_{t-1}{}^{T} \\ \mathbf{M}_{t-1}\mathbf{P}_{t}^{a}\mathbf{M}_{t-1}{}^{T} \end{pmatrix} \tag{12}$$

where,

$$\begin{aligned} \mathbf{L}_{t-1} &\equiv \mathbf{P}_{t-1}^{a}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{P}_{t-1}^{a}\mathbf{A}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\right)^{-1} = \mathbf{P}_{t-1}^{a}\mathbf{A}^{\mathrm{T}}\mathbf{P}_{t}^{f^{-1}} \\ \mathbf{M}_{t-1} &\equiv \mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}} + \mathbf{A}\mathbf{P}_{t-1}^{a}\mathbf{A}^{\mathrm{T}}\right)^{-1} = \mathbf{Q}_{t-1}\mathbf{G}^{\mathrm{T}}\mathbf{P}_{t}^{f^{-1}} \end{aligned} \tag{13}$$

are the coefficient matrices in Eq (11) (smoother gain matrices) introduced for shorthand notation.

The correspondence between Eqs (11) and (3) and between (12) and (4) shows that the RTS smoother is a recursive inversion of the model (Eq 8). In particular, the smoothed state estimate (upper part of Eq 11) and smoothed control estimate (lower part of Eq 11) can be identified as inversions of **A** and **G**, respectively. Moreover, as illustrated above, the smoother solution was derived to exactly satisfy the model equation, which can also be found by substituting results of Eq (11) to the right hand side of Eq (8) to yield,

$$\hat{\mathbf{x}}_{t}^{s} = \mathbf{A}\hat{\mathbf{x}}_{t-1}^{s} + \mathbf{G}\hat{\mathbf{u}}_{t-1}^{s} \tag{14}$$

The additional (last) term in Eq (12) relative to (4) reflects the uncertainties of the left hand side of Eq (8), and similar equations at other instances, while the smoother solves for such exact solution. The inversion and the physical consistency of the smoothed estimates are illustrated by the gray curves in Figure 1.

Although smoothed solutions satisfy model equations (Eq 14), smoothing should not be confused with the so-called "strong constraint" estimation (Sasaki, 1970) that assumes that models have no errors except in initial condition. In fact, the model solution by itself does not satisfy the model; viz., $\hat{\mathbf{x}}_t^s \neq \mathbf{A}\overline{\mathbf{x}}_{t-1}^s + \mathbf{G}\hat{\mathbf{u}}_{t-1}^0$. Smoothing is generally a "weak constraint" inversion that allows for model errors, but one that explicitly provides estimates of these inaccuracies. The explicit estimation of these model error sources as opposed to leaving them unknown ($\hat{\mathbf{u}}_t^s$ in Eq 14 instead of $\hat{\mathbf{u}}_t^0$), is what allows for the temporal evolution of the smoothed solution to be physically consistent.

While the discussion above has focused on sequential smoothing, there are other equally effective smoothing algorithms. In particular, when model error sources are made part of the estimate, the so-called adjoint method or 4dVAR (Chapter 10) is equivalent to the RTS smoother (Eq 11). The adjoint estimation directly solves for the smoothed solution ($\hat{\mathbf{x}}^s$ and $\hat{\mathbf{u}}^s$) without deriving intermediate filter estimates.

4. What are data errors and model errors?

"Data" error covariance **R** and "model" error covariance **Q** in effect define the solution to the data assimilation problem (e.g., Eqs 5 and 11). (**P** is a function of **R** and **Q**.) Their understanding and specification are, therefore, fundamental to assimilation and in utilizing their results. In fact, as described below, a part of what is commonly regarded as "model" error should in fact be considered "data" error. **R** and **Q** are better considered error covariances of the "data constraint" and the "model constraint", respectively.

4.1 "Data" Error

"Data" and "model" errors can be best understood by considering the true nature of the model vis-à-vis that of the observations and the ocean. The following discussion follows that of Cohn (1997). For instance, the model's true state $\overline{\mathbf{x}}_t$ (overbar denotes true solution) can be recognized as representing the ocean in finite dimension,

$$\overline{\mathbf{x}}_t \equiv \mathbf{\Pi}\mathbf{w}_t \tag{15}$$

Function $\mathbf{\Pi}$ defines the model state given the complete state of the ocean $\mathbf{w}_t$ (which has infinite degrees of freedom). Observations $\mathbf{y}_t$ are samples of this ocean $\mathbf{w}_t$ that could be written as,

$$\mathbf{y}_t = \mathbf{E}\mathbf{w}_t + \varepsilon \tag{16}$$

where **E** describes the sampling operation and ε denotes measurement errors (e.g., instrument error).

In terms of the model, Eq (16) can be rewritten as,

$$\mathbf{y}_t = \mathbf{H}\overline{\mathbf{x}}_t + \left\{\mathbf{E}\mathbf{w}_t - \mathbf{H}\Pi\mathbf{w}_t\right\} + \varepsilon \tag{17}$$

The last two terms of Eq (17),

$$\left\{\mathbf{E}\mathbf{w}_t - \mathbf{H}\Pi\mathbf{w}_t\right\} + \varepsilon \tag{18}$$

can be identified as the error of the observation equation $\mathbf{y}_t = \mathbf{H}\mathbf{x}_t$ that defines the data assimilation problem (Eq 1), i.e., the covariance of Eq (18) is **R**. The first part of (18) is the difference between error free observations ($\mathbf{E}\mathbf{w}_t$) and error-free equivalent of the model ($\mathbf{H}\Pi\mathbf{w}_t$). The two are generally different because the model does not simulate the entire spectrum of the ocean but only parts of it (Eq 15).

For instance, a coarse resolution model of 1º horizontal resolution does not simulate meso-scale variability, and a 1.5-layer reduced-gravity model does not simulate barotropic motion. What a model cannot simulate constitutes part of the errors of the observation equation as described by Eq (18) and is termed "representation error." Other common examples of representation error include,

- Baroclinic variability for a barotropic model
- External gravity waves for a rigid-lid model
- Skin temperature for most models with thick surface layers
- Micro-structure for most large-scale models

In numerical weather forecasting, meteorologists typically employ larger "data" error than the measurement accuracy of the observations so as to maximize the skill of their forecasts. Forcing models to agree with observations that the models cannot simulate, result in models propagating the data correction incorrectly in time, causing larger errors in the model evolution than otherwise.

Some observations are dominated by representation error, making them difficult to utilize. For instance, individual drifter and float trajectories can depend on small-scale variabilities of the ocean, such that two floats deployed a short distance away from each other have dramatically different trajectories (e.g., Paduan and Niiler, 1993). Such measurements (Lagrangian trajectory as opposed to Eulerian velocities along the trajectory) that are dominated by representation errors do not provide strong data constraints, and cannot be used effectively.

4.2 "Model" Error

The nature of model process noise **Q** can be deduced in a similar fashion as data error above. The model can be written in shorthand as,

$$\mathbf{x}_{t+1} = A\left(\mathbf{x}_t, \hat{\mathbf{u}}_t^0\right) \tag{19}$$

where A denotes a general function describing the evolution of the model state. $\hat{\mathbf{u}}^0$ denotes the model's particular control that includes its forcing, boundary condition, and parameters. For generality, $\hat{\mathbf{u}}^0$ also includes other sources of process noise as discussed below that are zero a priori. The ocean evolution can be thought of similarly as,

$$\mathbf{w}_{t+1} = L\left(\mathbf{w}_t, \mathbf{v}_t\right) \tag{20}$$

where L describes the evolution of the ocean and $\mathbf{v}_t$ is the forcing and boundary conditions of the ocean.

Then, the model evolution in terms of the true model state can be written as,

$$\begin{aligned} \overline{\mathbf{x}}_{t+1} &= \mathbf{\Pi}\mathbf{w}_{t+1} = \mathbf{\Pi}L\left(\mathbf{w}_t, \mathbf{v}_t\right) \\ &= A\left(\overline{\mathbf{x}}_t, \hat{\mathbf{u}}_t^0\right) + \left\{\mathbf{\Pi}L\left(\mathbf{w}_t, \mathbf{v}_t\right) - A\left(\mathbf{\Pi}\mathbf{w}_t, \hat{\mathbf{u}}_t^0\right)\right\} \end{aligned} \tag{21}$$

using Eqs (15), (19) and (20). The last term in {} mathematically describes what model error (process noise) is; process noise is the difference between the true evolution of the ocean projected to the model space $\mathbf{\Pi}L\left(\mathbf{w}_t, \mathbf{v}_t\right)$ and the model evolution given the true model state and its particular control $A\left(\mathbf{\Pi}\mathbf{w}_t, \hat{\mathbf{u}}_t^0\right)$.

As shown by Eq (21), process noise could be due to errors in the given control ($\mathbf{v}_t$ versus its equivalent in $\hat{\mathbf{u}}_t^0$) and to differences in model algorithm A and the true model evolution L and their interaction with operator $\mathbf{\Pi}$. The former includes, for example, errors in the particular external forcing, boundary condition, and model parameters used by the model. The latter includes errors due to finite differencing, truncation, and interaction with scales and processes ignored by the model. The two types of error sources could be considered external and internal errors of the model algorithm, respectively, and are both identified as elements of the model control vector.

4.3 Specification of Data Error and Model Error

While their principles are understood, the actual specification of data and model error covariances are not trivial. For instance, it is not entirely clear what operator $\mathbf{\Pi}$ that defines these errors is for different models, let alone the errors' statistical properties. However, there are some practical means of quantifying these errors prior to assimilation. Here we describe the so-called "covariance matching" method described by Fu et al. (1993).

Observations **y** and their model simulation's equivalent **m** could be written as the sum of the true signal $\mathbf{s}^{st}$(1 term on the right hand side of Eq 17) and their respective uncertainties **r** and **p**,

$$\begin{aligned} \mathbf{y} &= \mathbf{s} + \mathbf{r} \\ \mathbf{m} &= \mathbf{s} + \mathbf{p} \end{aligned} \tag{22}$$

To first approximation, we may assume **s**, **r**, and **p** to have zero means and to be uncorrelated with each other. (See section 5.2.5 for dealing with non-zero means.) Then, the covariance among these elements can be written as,

$$\begin{aligned} \left\langle \mathbf{y}\mathbf{y}^T \right\rangle &= \left\langle \mathbf{s}\mathbf{s}^T \right\rangle + \left\langle \mathbf{r}\mathbf{r}^T \right\rangle \\ \left\langle \mathbf{m}\mathbf{m}^T \right\rangle &= \left\langle \mathbf{s}\mathbf{s}^T \right\rangle + \left\langle \mathbf{p}\mathbf{p}^T \right\rangle \\ \left\langle \mathbf{y}\mathbf{m}^T \right\rangle &= \left\langle \mathbf{s}\mathbf{s}^T \right\rangle \end{aligned} \tag{23}$$

where brackets denote statistical expectation. Assuming ergodicity and stationarity, quantities on the left hand side could be estimated by averaging the data and model estimates in time. Then,

$$\begin{aligned} \left\langle \mathbf{r}\mathbf{r}^T \right\rangle &= \left\langle \mathbf{y}\mathbf{y}^T \right\rangle - \left\langle \mathbf{y}\mathbf{m}^T \right\rangle \\ \left\langle \mathbf{p}\mathbf{p}^T \right\rangle &= \left\langle \mathbf{m}\mathbf{m}^T \right\rangle - \left\langle \mathbf{y}\mathbf{m}^T \right\rangle \end{aligned} \tag{24}$$

The former is a direct estimate of data error covariance matrix **R**, while the latter provides an indirect estimate of process noise covariance **Q**. Namely, given a process noise model (**u** in Eq 1) and its covariance **Q**, the corresponding model simulation error can be estimated using standard methods. In particular, using the notation defined in Eq (1), the stationary limit of such error $\mathbf{P}^{sim}$ is the solution to the Lyapunov Equation,

$$\mathbf{P}^{sim} = \mathbf{A}\mathbf{P}^{sim}\mathbf{A}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T \tag{25}$$

which is related to the empirical estimate Eq (24) by,

$$\mathbf{H}\mathbf{P}^{sim}\mathbf{H}^T = \left\langle \mathbf{p}\mathbf{p}^T \right\rangle \tag{26}$$

Eqs (24), (25) and (26) provide a means to calibrate the process noise estimate **Q**.

Figure 2 illustrates an example of such estimate for assimilating altimetric sea level data with a coarse (1º) resolution model. Because of the model's limited spatial resolution, the data error estimate (a) is dominated by meso-scale variability that constitutes the model's representation error (Section 4.1), as evidenced by large values in western boundary regions. Wind error (c) is estimated to be the dominant source of model error for simulating large-scale sea level variability. Note the first order correspondence between the empirical (b) and theoretical (d) error estimates of model simulated sea level.

5. Examples of implementing assimilation; how is assimilation actually done?

While the theory of data assimilation is well understood, implementing assimilation is often nontrivial owing to its large computational requirements

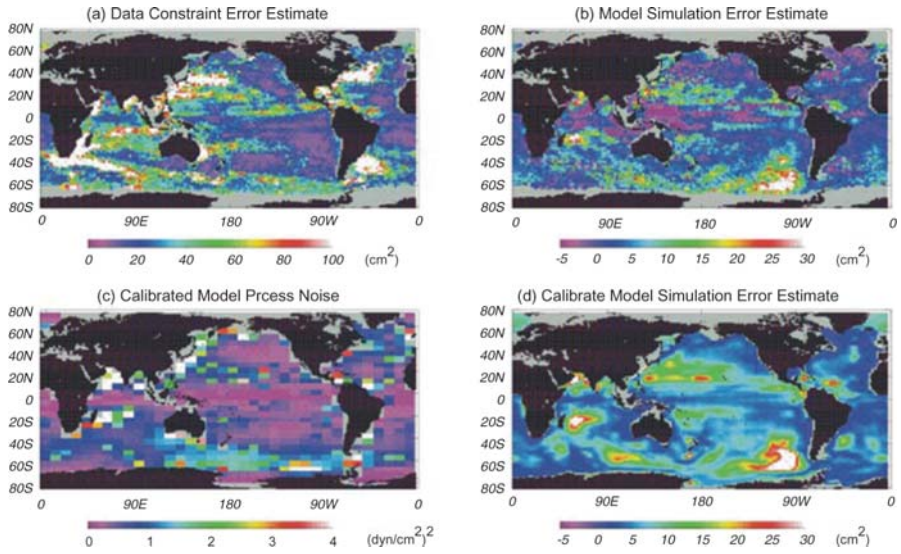

Figure 2. An example of prior error (variance) calibration; (a), (b) error estimates of altimetric sea level constraint and model simulated sea level, respectively, based on a model-data comparison (Eq 24), (c) calibrated wind stress error estimate (zonal component), (d) model simulated sea level error estimate based on (c). Note the first order consistency between (b) and (d) (Eq 26). The model is an ocean general circulation model with a 1º spatial resolution. (Adapted from Fukumori, *et al.*, 1999.)

and to a number of approximations and simplifications that are necessary to make the calculations tractable. An example specifying prior error estimates was described in section 4.3. Other examples of actually carrying out assimilation are described below to further elucidate some practical steps employed in an assimilation system. The examples are taken from the assimilation system of the Consortium for "Estimating the Circulation and Climate of the Ocean (ECCO)."

5.1 Consortium for "Estimating the Circulation and Climate of the Ocean" (ECCO)

The ECCO Consortium focuses on advancing data assimilation from an experimental tool to an operational means to study ocean circulation

(Stammer *et al.*, 2002.) ECCO estimates are characterized by their physical consistency (Section 3.2) owing to smoothing algorithms (RTS smoother and adjoint method). The estimates are based on a state-of-the-art primitive equation model (MITgcm; Marshall *et al.* 1997) and employ a diverse suite of in situ and satellite remote sensing observations including temperature and salinity profiles and sea level.

The ECCO estimates are available from its data server at http://www.ecco-group.org/las. In particular, ECCO has established a near real-time analysis producing estimates of large-scale global ocean circulation (73ºS~73ºN) on a regular basis (http://ecco.jpl.nasa.gov/external). The model employed is of moderate resolution (1º telescoping to 1/3º within 10º of the equator, 10m layers within 150m of the surface with a total of 46 vertical levels) with its parameters adjusted by a Green's function estimation (Menemenlis, *et al.*, 2004.) The near real-time analysis is conducted by an approximate Kalman filter and RTS smoother. Aspects of this near real-time assimilation are reviewed below.

5.2 ECCO near real-time analysis system

The recursive nature of the Kalman filter and RTS smoother is particularly suitable for near real-time computation. However, the computational requirements of evaluating the state error covariance matrix **P** make direct application of these methods impractical for most state-of-the-art ocean circulation models. Therefore, various methods have been put forth that approximate the derivation of **P** so as to make Kalman filtering and RTS smoothing feasible. In ECCO, three approximations are concurrently employed:

I. Time-asymptotic approximation (Fukumori *et al.,* 1993),
II. State reduction (Fukumori and Rizzoli, 1995),
III. Partitioning (Fukumori, 2002).

The *time-asymptotic* approximation evaluates and employs a time-invariant representative limit of **P**, thereby eliminating the computational cost associated with the continued model integration of the state error covariance matrix. Evaluation of this asymptotic limit is simplified by *partitioning* and *state reduction* where independent elements of **P** are evaluated separately from one another (*partition*) and within each partition only the dominant modes of the error are estimated (*state reduction*). A reduced-state model is derived for each partition to evaluate the errors while the original fully nonlinear unapproximated model is used to integrate the state. The smaller dimensionality of each partitioned-reduced-state model reduces the computational cost of evaluating **P**. Unlike global single-stage state reductions, the partitioning permits retaining many of the estimation

problem's degrees of freedom without incurring excessive computational requirements.

The reader is referred to Chapter 11 and to references above for further discussion on theoretical aspects of these and other approximations. Here we review examples of implementing the approximations and their implications.

5.2.1 Identifying process noise

To first approximation, different model errors sources could be considered independent of one another. Then different process noise and their consequent model state errors could be evaluated separately in the context of a partitioned estimation (Fukumori, 2002).

Different sources of process noise cause different errors in the modeled state. For instance, the response of a model to changes in large-scale wind can be effectively described in terms of the gravest few vertical dynamic modes (e.g., Cane, 1984). In comparison, a model's response to changes in air-sea heat flux is to first approximation confined to the sea surface. The modeled process noise dictates the most effective state approximation (e.g., state reduction and partitioning), and, therefore, its identification is the first step in designing an assimilation system.

The ongoing ECCO near real-time assimilation estimates uncertainties of wind forcing and its resulting model state errors. This estimate should not be confused as one that considers all model errors are due to errors in wind, but it is an estimate of only a part of the errors, as discussed above, albeit one of the dominant ones. The model's controllability (ability to uniquely solve $\mathbf{u}$ in Eq 1) limits aliasing of other model error sources to the particular process noise being estimated. The ECCO near real-time assimilation system described below is correspondingly designed to resolve the dominant response of the ocean to large-scale wind errors.

5.2.2 Regional partitions

Due to large differences in temporal and spatial scales, wind-driven barotropic errors could be considered independent of baroclinic errors and thus estimated separately. Having sub-basin length scales, the baroclinic components are estimated individually among seven different basins across the globe (Figure 3). These regions include three separate tropical basins (Indian, Pacific, Atlantic) and four mid- and high-latitude basins (North Pacific, North Atlantic, South Atlantic and Indian, South Pacific). The regions overlap each other to minimize edge effects caused by the regional approximation; errors in overlapping areas are considered to be split among

the different regions. The barotropic component, due to its large spatial scales, is estimated simultaneously over the entire model domain.

5.2.3 State reduction

Within each partition, additional vertical and horizontal approximations are defined to further reduce the errors' dimension. Vertically, state errors are expanded in terms of vertical dynamic modes of velocity and vertical

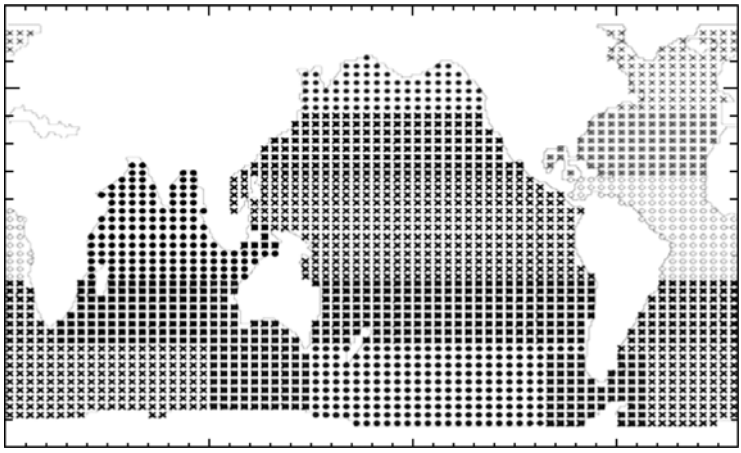

Figure 3. Coarse horizontal grid employed in ECCO partitioned reduced-state approximation. The different symbols denote different regional reduced-grid partitions used to estimate baroclinic errors of the model state.

displacement. For each baroclinic partition (Figure 3) the first five baroclinic modes are retained. Horizontally, large-scale errors are estimated by defining a coarse horizontal grid and an interpolation operator to map the coarse grid errors onto the model (fine) grid. The process noise (wind error) is reduced likewise, utilizing the same horizontal mapping operation.

The coarse grid is defined as a 5º-by-3º and 6º-by-6º (zonal and meridional resolution) grid for baroclinic and barotropic partitions, respectively. Objective mapping (Bretherton *et al.,* 1976) is employed as the coarse-to-fine grid interpolation operator, which can also be identified as a least-squares operator in itself (Eq 3). The interpolation assumes no underlying error and a Gaussian covariance function using the coarse grid dimensions as the correlation distance. To prevent spurious correlation across land (e.g., Pacific Ocean to Atlantic Ocean across the Isthmus of Panama), distances between model grid points used to define the mapping operation are computed around the model's land points.

The reduced state error thus consists of dynamic mode amplitudes $\mathbf{a}_u$, $\mathbf{a}_v$, $\mathbf{a}_\eta$ defined on a coarse grid for zonal and meridional velocity and vertical

displacement, respectively. The approximated control $\mathbf{a}_\tau$ is the magnitude of wind error defined on the same coarse horizontal grid. These approximated errors are related to those of the model state and model forcing by,

$$\delta\mathbf{u} = \mathbf{D}_{vel}\mathbf{O}\mathbf{a}_u, \quad \delta\mathbf{v} = \mathbf{D}_{vel}\mathbf{O}\mathbf{a}_v, \quad \delta\boldsymbol{\eta} = \mathbf{D}_\eta\mathbf{O}\mathbf{a}_\eta, \quad \delta\boldsymbol{\tau} = \mathbf{O}\mathbf{a}_\tau \tag{27}$$

where $\delta\mathbf{u}$, $\delta\mathbf{v}$, $\delta\boldsymbol{\eta}$, and $\delta\boldsymbol{\tau}$ are errors of model zonal and meridional velocity, vertical displacement, and wind stress, respectively. The **D**s consist of structures of vertical dynamic modes of respective variables that project the errors vertically to the model grid. **O** denotes the horizontal mapping operator from the coarse grid to the model grid. Errors of other state variables are diagnostically derived using estimates in Eq (27). For instance, errors of temperature **T** and salinity **S** are derived from those of displacement by,

$$\delta\mathbf{T} = \frac{\partial\mathbf{T}}{\partial z}\delta\boldsymbol{\eta}, \quad \delta\mathbf{S} = \frac{\partial\mathbf{S}}{\partial z}\delta\boldsymbol{\eta} \tag{28}$$

and errors of sea level can be defined as a function of $\delta\boldsymbol{\eta}$ and density (temperature and salinity). Errors from different partitions are summed together to form the overall model error estimate.

The total dimension of each partitioned-reduced-state is summarized in Table 1. The largest partition is the tropical Pacific cell consisting of nearly 12000 elements. In comparison, the total dimension of the model state (horizontal velocity, temperature, and salinity on the model grid) is 8 million.

Partition	Grid Points	Dimension
Tropical Indian	308	4620
Tropical Pacific	787	11805
Tropical Atlantic	350	5250
South Pacific	633	9495
South Atlantic & Indian	664	9960
North Pacific	271	4065
North Atlantic	198	2970
Global Barotropic	963	2889

Table 1. The reduced-state dimension of seven baroclinic partitions and the global barotropic partition. Each baroclinic partition employs the five gravest baroclinic modes. Each partition has three variables; zonal and meridional velocity and vertical displacement.

5.2.4 Derivation of state error covariance matrix

A time-invariant state error covariance matrix is derived for each separate partition by computing the asymptotic limit of the respective Riccati equation. (The Riccati equation describes the temporal evolution of the state error covariance matrix when integrating the model and assimilating

observations. See Chapter 10) The computation employs a representative approximation of the assimilation problem in which time-invariant system matrices **A**, **G**, **H**, **Q**, and **R**, (Eqs 1, 5 and 10) are derived and used. The so-called "doubling algorithm" provides an effective means to integrate the corresponding Riccati equation in increasing time-steps of powers of two (i.e., doubling) (Fukumori *et al*., 1993).

Model matrices **A**, **G**, and **H** are derived using a coarse grain Green's function of the corresponding partitioned reduced-state model and is computed by combining the state approximation and the original unapproximated model. For instance, a general state and control perturbation (error), $\delta\mathbf{x}$ and $\delta\mathbf{u}$, can be written as,

$$\begin{aligned} \delta\mathbf{x} &= \mathbf{B}\delta\mathbf{x}' + \mathbf{N}\mathbf{n} \\ \delta\mathbf{u} &= \breve{\mathbf{B}}\delta\mathbf{u}' + \breve{\mathbf{N}}\mathbf{m} \end{aligned} \tag{29}$$

where **B** and **N** define the range and null space of a particular partitioned reduced state (control) approximation described in Sections 5.2.2 and 5.2.3, respectively, and $\delta\mathbf{x}'$ and $\mathbf{n}$ are their amplitudes. $\breve{\mathbf{B}}$, $\breve{\mathbf{N}}$, $\delta\mathbf{u}'$ and $\mathbf{m}$ are corresponding counterparts for the control.

Then, given a general (nonlinear) model, Eq (19), the perturbations satisfy,

$$\delta\mathbf{x}_{t+1} = A\left(\tilde{\mathbf{x}} + \delta\mathbf{x}_t, \tilde{\mathbf{u}} + \delta\mathbf{u}_t\right) - A\left(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}\right) \tag{30}$$

where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{u}}$ are a representative state and control, respectively. (Time-means are used.) Substituting Eq (29) into (30) and multiplying both sides of the equation with the pseudo inverse of **B**, denoted $\mathbf{B}^*$, and noting the orthogonality between **B** and **N,** we have,

$$\delta\mathbf{x}'_{t+1} = \mathbf{B}^*\left(A\left(\tilde{\mathbf{x}} + \mathbf{B}\delta\mathbf{x}'_t + \mathbf{N}\mathbf{n}_t, \tilde{\mathbf{u}} + \breve{\mathbf{B}}\delta\mathbf{u}'_t + \breve{\mathbf{N}}\mathbf{m}_t\right) - A\left(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}\right)\right) \tag{31}$$

The approximation's dependence on the null space (**n** and **m**) is a source of error in defining a reduced-state model. However, because of their orthogonality, this dependency could be ignored if range and null space perturbations remain within their respective domain through the model integration, as $\mathbf{B}^*$ in Eq (31) will nullify any resulting null space perturbation. For example, to first approximation, a particular dynamic mode remains the same mode and large-scale perturbations remain large-scale. Then Eq (31) could be approximated in closed form in the reduced-space as,

$$\delta\mathbf{x}'_{t+1} = \mathbf{B}^*\left(A\left(\tilde{\mathbf{x}} + \mathbf{B}\delta\mathbf{x}'_t, \tilde{\mathbf{u}} + \breve{\mathbf{B}}\delta\mathbf{u}'_t\right) - A\left(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}\right)\right) \tag{32}$$

defining the partitioned reduced-state model.

Corresponding partitioned reduced-state matrices $\mathbf{A}'$ and $\mathbf{G}'$ that linearize Eq (32) around the representative state and control ($\tilde{\mathbf{x}}$ and $\tilde{\mathbf{u}}$),

$$\mathbf{B}^*\left(A\left(\tilde{\mathbf{x}} + \mathbf{B}\delta\mathbf{x}'_t, \tilde{\mathbf{u}} + \breve{\mathbf{B}}\delta\mathbf{u}'_t\right) - A\left(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}\right)\right) \approx \mathbf{A}'\delta\mathbf{x}'_t + \mathbf{G}'\delta\mathbf{u}'_t \tag{33}$$

are derived as coarse grain Green's functions. (The prime denotes the individual partitioned reduced-state equivalent.) Namely, an arbitrary column of the two matrices, $(\mathbf{A}')_i$ and $(\mathbf{G}')_i$, can be numerically derived as,

$$\begin{aligned}\left(\mathbf{A}'\right)_i &= \mathbf{A}'\mathbf{e}_i + \mathbf{G}'\mathbf{0} = \mathbf{B}^*\left(A\left(\tilde{\mathbf{x}} + \mathbf{B}\mathbf{e}_i, \tilde{\mathbf{u}}\right) - A\left(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}\right)\right) \\ \left(\mathbf{G}'\right)_i &= \mathbf{A}'\mathbf{0} + \mathbf{G}'\mathbf{e}_i = \mathbf{B}^*\left(A\left(\tilde{\mathbf{x}}, \tilde{\mathbf{u}} + \breve{\mathbf{B}}\mathbf{e}_i\right) - A\left(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}\right)\right)\end{aligned} \tag{34}$$

where $\mathbf{e}_i$ is the corresponding column of the identity matrix of appropriate dimension and $\mathbf{0}$ is a vector of zeroes.

Model implementation of Eq (34), and in particular, the pseudo inverse $\mathbf{B}^*$, requires some consideration. Since vertical displacement is not a variable in most models and inverting Eq (28) can be difficult where stratification is weak, vertical velocity is integrated in time in Eq (34) to diagnose $\delta\eta$ (cf. Section 5.2.3). Because of their orthogonality, implementing the pseudo inverse of the vertical transformation (the **D**s in Eq 27 that make up **B**) is trivial. However, the pseudo inverse of the horizontal operator **O** is not. The objective mapping operator is relatively sparse, and, therefore, **O** in Eq (27) is implemented as a sparse matrix multiplication retaining only significant elements of the matrix. However, the pseudo inverse of **O** tends to be a fairly large and dense matrix. An effective means of implementing the inversion of **O** is as,

$$\mathbf{O}^* = \left(\mathbf{O}^T\mathbf{O}\right)^{-1}\mathbf{O}^T \tag{35}$$

Matrix $\left(\mathbf{O}^T\mathbf{O}\right)^{-1}$ is a relatively small matrix that can be precomputed and stored. The left multiplication by **O** transpose can be achieved algorithmically given the sparse matrix **O** that is already available. (A multiplication by $\mathbf{O}^T$ is an adjoint of **O**.)

The partitioned reduced-state observation matrix $\mathbf{H}'$ can be numerically derived similarly to those in Eq (34):

$$\left(\mathbf{H}'\right)_i = \mathbf{H}'\mathbf{e}_i = H\left(\tilde{\mathbf{x}} + \mathbf{B}\mathbf{e}_i\right) - H\left(\tilde{\mathbf{x}}\right) \tag{36}$$

where H is a function defining the model equivalent of the observations.

The time-asymptotic approximation employs a time-invariant system in which not only the model ($\mathbf{A}'$ and $\mathbf{G}'$) but the observation matrix $\mathbf{H}'$ and the data and model error covariance matrices **R** and $\mathbf{Q}'$ are stationary. (Only the operators $\mathbf{A}'$, $\mathbf{G}'$, and $\mathbf{H}'$ and the statistics **R** and $\mathbf{Q}'$ are assumed stationary, not the state, control, or observation.) However, since in practice what is observed (H) varies in time, a representative set of observations is assumed to be available regularly in deriving the state error covariance matrix. For instance, to simulate the coverage and accuracy of satellite altimeter data, a three-day assimilation cycle is assumed during which all

satellite altimeter data within the 10-day repeat period is available but with 3-times the assumed data error.

The resulting system matrices are used to integrate the corresponding Riccati Equation to its asymptotic limit utilizing the doubling algorithm.

5.2.5 Implementation

Although the derivation of **P** assumed a 3-day assimilation interval, actual assimilation is performed every 6-hours (model time-step is 1 hour), assimilating all available observations within 3-hours of the assimilation instant. No observation is utilized more than once as dictated by standard estimation theory. The 6-hour assimilation interval is a compromise between computational requirements associated with applying the Kalman filter more frequently and the resolution of high frequency variability of the ocean (e.g., wind-driven barotropic motion).

For computational efficiency, the assimilation employs an alternate form of the Kalman gain matrix from the common formulation of Eq (5),

$$\mathbf{P}_t^f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_t^f \mathbf{H}^T + \mathbf{R}_t \right)^{-1} = \mathbf{P}_t^a \mathbf{H}^T \mathbf{R}_t^{-1} \tag{37}$$

The alternate form on the right hand side of Eq (37) employs the analysis error covariance instead of the forecast error covariance, and has fewer computational steps than the left hand side, when the respective state error covariance matrices are given.

In the partitioned reduced-state formulation, the filter (data) increment (the difference between analyzed state and forecast state, i.e., the third term of Eq 7), $\Delta \hat{\mathbf{x}}_t^a$, can be written as a sum of the increments in different partitions,

$$\Delta \hat{\mathbf{x}}_t^a = \sum_i \mathbf{B}_i \Delta \hat{\mathbf{x}}_{t,i}'^a \tag{38}$$

where,

$$\Delta \hat{\mathbf{x}}_{t,i}'^a = \mathbf{P}_i'^a \mathbf{H}_{t,i}'^T \mathbf{R}_t^{-1} \left(\mathbf{y}_t - H \left(\hat{\mathbf{x}}_t^f \right) \right) \tag{39}$$

is the filter increment of an individual partitioned reduced-state (subscript *i*). In Eq (39), the reduced state observation matrix $\mathbf{H}'$ (Eq 36) can be used. Alternatively, Eq (39) could be implemented as,

$$\Delta \hat{\mathbf{x}}_{t,i}'^a = \mathbf{P}_i'^a \mathbf{B}_i^T \mathbf{H}_t^T \mathbf{R}_t^{-1} \left(\mathbf{y}_t - H \left(\hat{\mathbf{x}}_t^f \right) \right) \tag{40}$$

using the adjoint of the model observation operator *H* as the left multiplication $\mathbf{H}^T$. Eq (40) involves less approximation and is of particular convenience when the observation operator is an implicit function of the state and its adjoint is available.

Due to inaccuracies in the marine geoid estimate, the analysis assimilates altimetric sea level anomaly relative to its temporal mean instead

of absolute sea surface height. For each partitioned reduced state, Eq (39) is computed by,

$$\Delta\hat{\mathbf{x}}'^{a}_{t,i} = \mathbf{P}'^{a}\mathbf{H}'^{T}_{t}\mathbf{R}_{t}^{-1}\left(\left(\mathbf{y}_{t}-\bar{\bar{\mathbf{y}}}\right)-\left(H\left(\hat{\mathbf{x}}^{f}_{t}\right)-\bar{\bar{\mathbf{m}}}\right)\right) \tag{41}$$

where $\bar{\bar{\mathbf{y}}}$ and $\bar{\bar{\mathbf{m}}}$ are time-mean altimetric sea level and its model simulation equivalent, respectively. This particular formulation corrects the model sea level variability without altering the model time-mean within the linearized time-asymptotic approximation. Such approximation is further sensible considering that errors in the time-mean state (bias) are due to time-correlated errors for linear models. Standard Kalman filtering and smoothing formulations assume temporally uncorrelated process noise and such correlated model errors require modification to the canonical estimation procedure. Thus, assimilation of other observations (e.g., temperature profiles) is similarly restricted to their temporal anomalies. Time-invariant process noise can alternatively be estimated separately from such temporally uncorrelated errors.

For computational efficiency, Eq (41) is carried out from the right as a series of left multiplications (or operations) of the innovation vector (i.e., data-model difference) and its products; i.e., no matrix-matrix multiplication is performed to compute the coefficient matrix in Eq (41). Contributions from different partitions are summed together (Eq 38) to correct the entire model forecast. The unapproximated fully nonlinear model is then integrated in time using the resulting analysis with all diagnostic variables updated consistently with these data increments.

In terms of the partitioned reduced-state formulation, the smoother increment (difference between analysis and smoother; second term on the right hand side of Eq 11) can also be written as a sum of smoother increments in the partitioned reduced state;

$$\begin{pmatrix}\Delta\hat{\mathbf{x}}^{s}_{t}\\ \Delta\hat{\mathbf{u}}^{s}_{t}\end{pmatrix} = \sum_{i}\begin{pmatrix}\mathbf{B}_{i}\Delta\hat{\mathbf{x}}'^{s}_{t,i}\\ \breve{\mathbf{B}}_{i}\Delta\hat{\mathbf{u}}'^{s}_{t,i}\end{pmatrix} \tag{42}$$

where,

$$\begin{pmatrix}\Delta\hat{\mathbf{x}}'^{s}_{t,i}\\ \Delta\hat{\mathbf{u}}'^{s}_{t,i}\end{pmatrix} = \begin{pmatrix}\mathbf{P}'^{a}_{i}\mathbf{A}'^{\mathrm{T}}_{i}\mathbf{P}'^{f\,-1}_{i}\\ \mathbf{Q}'_{i}\mathbf{G}'^{\mathrm{T}}_{i}\mathbf{P}'^{f\,-1}_{i}\end{pmatrix}\left(\Delta\hat{\mathbf{x}}'^{s}_{t+1,i}+\Delta\hat{\mathbf{x}}'^{a}_{t+1,i}\right) \tag{43}$$

are the smoother increments of state and control of a particular partitioned reduced state. The partitioned form of the smoother increment recursion, Eq (43), uses the second form of the smoother gain in Eq (13) and the definitions of $\Delta\hat{\mathbf{x}}^{s}$ and $\Delta\hat{\mathbf{x}}^{a}$ to rearrange the last term in Eq (11). Unlike the filter, elements of the approximate smoother gain matrix in Eq (43) are time-invariant and thus, for computational efficiency, the gain matrix can be explicitly derived and used in deriving the smoother increments. Smoother

increments of different partitions are summed together to correct the entire model state and control (Eq 42). However, because of the approximations, the resulting smoothed state and smoothed controls do not exactly satisfy the model equations and are, strictly speaking, physically inconsistent. Instead, estimates of a smoothed state that is fully consistent with the control are derived by re-integrating the model in time using the smoothed control estimates.

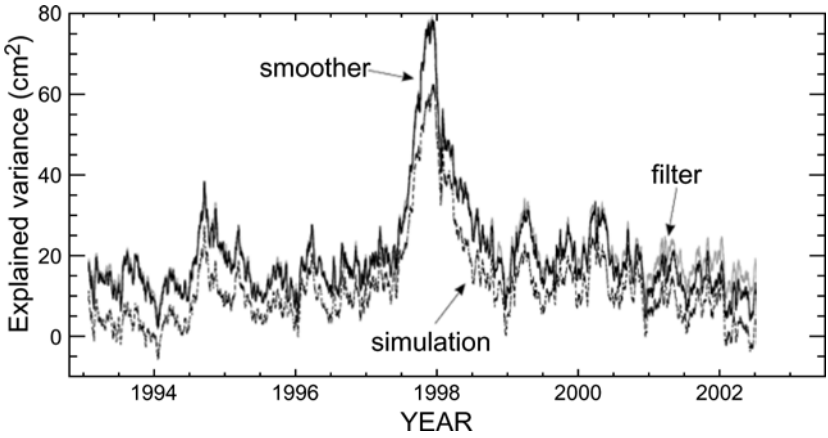

Figure 4. Model explained observed altimetric sea level anomaly variance; simulation (broken curve), Kalman filter (gray curve), smoother (solid curve). The explained variance is defined as the difference between data variance and model-data residual variance. Note the smoother results being nearly indistinguishable from the filter's (except near the end, 2001~2002) whereas the simulation's explained variance is substantially less than these throughout the experiment. Results are from the ECCO near real-time assimilation.

5.2.6 Assessment

The assimilation is assessed by examining its self-consistency and by comparisons with independent observations. Being a least-squares estimate, the estimates' errors are non-increasing functions of the amount and accuracy of the observations that are assimilated. An estimates' systematic degradation would indicate the assimilation's inaccurate assumption and/or errors in the implementation. Some examples of such assessment are briefly described below.

One of the useful and readily available measures for assessing assimilation is the difference between observations and their model equivalent, in particular, the innovation sequence (i.e., difference between observations and a filter's forecast). For instance, Figure 4 compares the amount of data variance (sea level) explained by the different model estimates. Explained variance is defined as,

$$\left\langle \mathbf{y}\mathbf{y}^T \right\rangle - \left\langle \left(\mathbf{y} - H(\mathbf{x})\right)\left(\mathbf{y} - H(\mathbf{x})\right)^T \right\rangle \tag{44}$$

The second term, the residual variance, is the variance of what the model cannot explain, and thus the difference with the data variance (first term) is a measure of what the model resolves. As the forecast does not yet utilize the particular observations, the innovation sequence also provides a measure of skill with respect to independent observations.

Figure 4 illustrates that the approximate Kalman filter explains significantly more data variance than does the model simulation without data

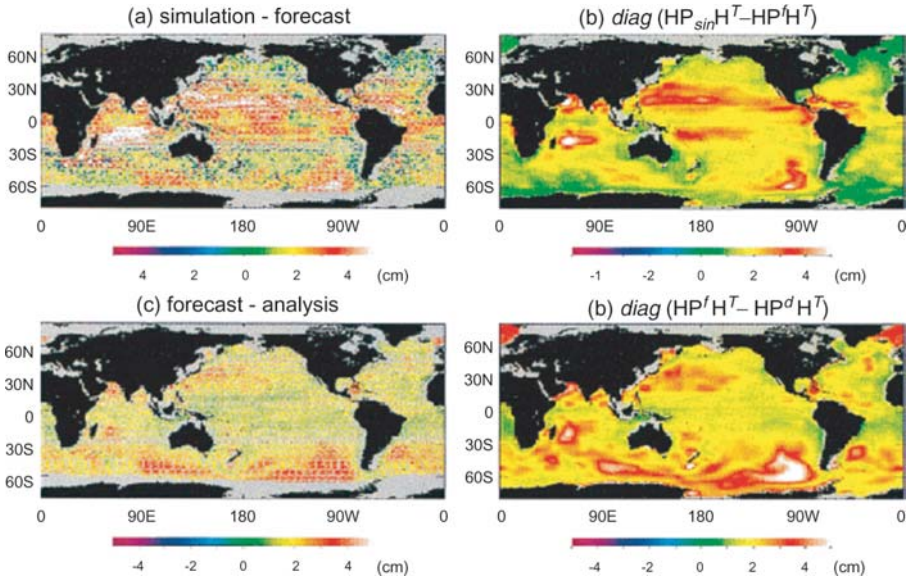

Figure 5. An assessment of model-data residuals with respect to their theoretical expectations. The panels show reductions in root-mean-square sea level residuals by assimilation of satellite altimeter data with a global ocean general circulation model. Panels (a) and (b) are differences between simulation and forecast and its theoretical expectation based on estimated errors, respectively; a positive value indicates an improvement by the latter model. Panels (c) and (d) are the same except between forecast and analysis. Note the first order consistency between (a) and (b) and between (c) and (d). Gray areas in (a) and (c) denote regions with no observations. Results correspond to assimilation using calibrated prior error estimates of Figure 2. (From Fukumori, *et al.,* 1999.)

constraints. Moreover, the figure shows that the smoothed estimate (smoothed-wind-driven model simulation) explains nearly as much variance as does the Kalman filter (model forecast), thus demonstrating the fidelity of the approximate smoother.

The assimilation's self-consistency can be assessed by comparing model-data differences with their theoretical expectations, i.e., formal error estimates computed and utilized by the Kalman filter algorithm. Figure 5

illustrates an example of such comparison. Differences between different model residuals are comparable to their respective theoretical expectations in both overall amplitude and spatial distribution. The absolute magnitudes of these estimates are statistically consistent, as the model-data difference of the simulation is also comparable with its theoretical expectation (Figure 2).

The fidelity of the assimilated analyses permits diverse studies and applications of not only ocean circulation (e.g., Fukumori *et al.,* 2004) but also of ocean biogeochemical processes (e.g., McKinley, 2002) and geodetic investigations (e.g., Dickey *et al.*, 2002.) For instance, Figure 6 illustrates such an application and an assessment of the data assimilated model estimate.

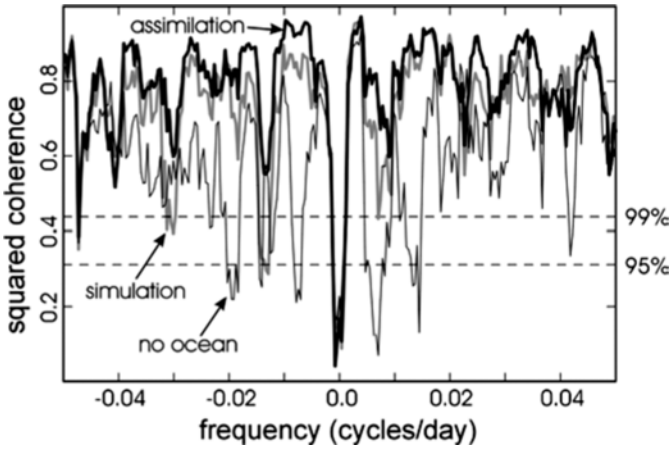

Figure 6. Coherence between observed and modeled excitation of Earth's wobble (polar motion); NCEP atmosphere reanalysis (thin black; "no ocean"), ECCO simulation plus NCEP atmosphere (gray; "simulation"), ECCO assimilation plus NCEP atmosphere (thick black; "assimilation"). Also shown are the 95% and 99% confidence levels. (Gross, 2003, personal communication. See Gross *et al.*, 2003, for related results.)

The figure shows coherence between observed excitation of Earth's polar motion (the wobble of Earth's rotation axis relative to the terrestrial frame) and that estimated by atmospheric (National Centers of Environmental Prediction (NCEP) Reanalysis, Kalnay *et al., 1996*) and oceanic models. While changes in atmospheric circulation (thin black curve) account for most of the observed polar motion, adding the ocean estimate (gray curve) significantly improves the coherence at almost all frequencies. Moreover, the ocean assimilation (thick black curve) further improves the coherence illustrating the impact of ocean data assimilation in improving the estimate of ocean circulation. Satellite navigation employs estimates of polar motion and thus would benefit from forecasts as well as near real-time ocean analysis systems.

6. Summary

Data assimilation concerns correcting models using observations. Although the concept is straightforward, there are various subtleties involved in both what data assimilation solves and how the computation is carried out. A careful understanding of these issues is helpful in assimilating observations, in utilizing their results, and in further improving their estimates.

Data assimilation can be considered a process of fitting models to observations. A solution that is consistent with both observations and model physics is sought. However, given that all models are in one way or another approximations of the real world (ocean), there are some, sometimes many, aspects of the observations that are real but inconsistent with the models. These aspects that models cannot inherently simulate (representation errors) therefore cannot be part of the assimilated solution and must be properly accounted for. Forcing models to agree with such measurements can lead to increased inaccuracies and inconsistencies. An assessment of what models do and do not simulate is important in carrying out the assimilation, and an understanding of what the assimilated estimates resolve is fundamental to utilizing the results.

Mathematically, data assimilation is an inverse problem. The temporally evolving state of the model and sources of model error are estimated by inverting model equations that consist of those relating the model state to the observations and those describing the model's temporal evolution.

The Kalman filter and other common filtering methods are inversions of the model equivalent of the observations but not of the model evolution, and, therefore, do not completely solve the assimilation problem. Smoothers additionally invert the model evolution completing the estimation, providing estimates of both model state and model error sources.

While state estimation is often used synonymously with data assimilation, it is in fact the estimation of the model error sources (process noise) that is most fundamental. Given smoothed model error corrections, and apart from corrections to the initial condition, the smoothed state can be derived by integrating the model in time, but not vice versa.

Because of model errors, data assimilated state estimates by themselves are not physically consistent, in the sense that the estimated states' temporal evolution cannot be physically accounted for. For instance, budgets of heat and other properties cannot be closed in terms of explicit physical processes. The smoother's explicit estimation of model error sources resolves the physical inconsistency, rendering the assimilated solution amenable to various process studies and applications.

Although methods of data assimilation are well known, their implementation is often hampered by the models' large dimension and their complex nonlinearities. Many approximations have been put forth that render their implementation feasible and practical. The near real-time assimilation system of the Consortium for "Estimating the Circulation and Climate of the Ocean" (ECCO) employs a hierarchy of such approximations to maximize utilization of observations.

The fidelity and scope of these and other analyses lend themselves to various studies in ocean circulation and their application. However, existing products are in certain respects yet incomplete. The present near real-time ECCO estimates utilize a simplification by only estimating errors resulting from uncertainties in wind forcing. Other ECCO estimates also estimate errors in diabatic forcing and uncertainties in some of the model parameters. However, there are many other model error sources that have not yet been addressed. Expanding the estimated suite of process noise remains a central task in further improving ECCO and other assimilation estimates.

For the approximate Kalman filter and RTS smoother, such extension requires an explicit modeling of the process noise that is physically sensible (identification of operator **G** in Eq 1) and in identifying an effective approximation (partition and state reduction operators and basis set **B** and $\breve{\mathbf{B}}$ in Eq 32) that would resolve the corresponding errors in the model state. An effective basis set not only has a small dimension but must also form a closed dynamic system (Eq 31 approximated as Eq 32). Understanding the nature of the modeled system is imperative to such design.

Acknowledgements

This study is a contribution of the Consortium for Estimating the Circulation and Climate of the Ocean (ECCO) funded by the National Oceanographic Partnership Program. This work was carried out in part at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

Bretherton, F. P., R. E. Davis, and C. B. Fandry, 1976: A technique for objective analysis and design of oceanographic experiments applied to MODE-73, *Deep-Sea Res.*, **23**, 559-582.

Cane, M. A., 1984: Modeling sea level during El Niño, *J. Phys. Oceanogr*, **14**, 1864-1874.

Cohn, S. E., 1997: An introduction to estimation theory, *J. Met. Soc. Japan*, **75**, 257-288.

Dickey, J. O., S. L. Marcus, O. de Viron, and I. Fukumori, 2002: Recent Earth oblateness variations: Unraveling climate and postglacial rebound effects, *Science*, **298**, 1975-1977.

Fu, L.-L., I. Fukumori and R. N. Miller, 1993: Fitting dynamic models to the Geosat sea level observations in the Tropical Pacific Ocean. Part II: A linear, wind-driven model, *J. Phys. Oceanogr.*, **23**, 2162-2181.

Fukumori, I., J. Benveniste, C. Wunsch, and D. B. Haidvogel, 1993: Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother, *J. Phys. Oceanogr.*, **23**, 1831-1855.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation; an example with an idealized Gulf Stream model, *J. Geophys. Res.*, **100**, 6777-6793.

Fukumori, I., R. Raghunath, L. Fu, and Y. Chao, 1999: Assimilation of TOPEX/POSEIDON data into a global ocean circulation model: How good are the results?, *J. Geophys. Res.*, **104**, 25,647-25,665.

Fukumori, I., 2002: A partitioned Kalman filter and smoother, *Mon. Weather Rev.,* **130**, 1370-1383.

Fukumori, I., T. Lee, B. Cheng, and D. Menemenlis, 2004: The origin, pathway, and destination of Niño3 water estimated by a simulated passive tracer and its adjoint, *J. Phys. Oceanogr.*, **34**, 582-604.

Gross, R. S., I. Fukumori, and D. Menemenlis, 2003: Atmospheric and oceanic excitation of the Earth's wobbles during 1980-2000, *J. Geophys. Res*., **108** (B8), 2370, doi:10.1029/2002JB002143.

Kalnay, E., and coauthors, 1996: The NCEP/NCAR 40-year reanalysis project, *Bull. Ameri., Meteorol. Soc.*, **77**, 437-471.

Marshall, J. C., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers, *J. Geophys. Res.*, **102**, 5753-5766.

McKinley, G. A., 2002: Interannual variability of air-sea fluxes of carbon dioxide and oxygen, Ph.D. thesis, Massachusetts Institute of Technology.

Menemenlis, D., I. Fukumori, and T. Lee, 2004: Using Green functions to calibrate an ocean general circulation model, *Mon. Weather Rev.,* (in press).

Paduan, J. D., and P. P. Niiler, 1993: Structure of velocity and temperature in the northeast Pacific as measured with Lagrangian drifters in Fall 1987, *J. Phys. Oceanogr.*, **23**, 585-600.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis, *Mon. Weather Rev.*, **98**, 875-883.

Stammer, D., C. Wunsch, I. Fukumori, and J. Marshall, 2002: State Estimation in Modern Oceanographic Research, EOS, Trans. Amer. Geophys. Union, 83(27), 289&294-295.

Wunsch, C., 1996: *"The Ocean Circulation Inverse Problem"*, Cambridge Univ. Press, New York, NY, 442pp.

Chapter 12

IMPORTANCE OF DATA: A METEOROLOGICAL PERSPECTIVE

Florence Rabier

Météo-France, CNRM, Toulouse, France

Abstract: The importance of data in meteorological data assimilation can be quantified in the context of re-analyses performed at Numerical Weather Prediction centres. The increasing quality and quantity of satellite data is seen to play a major role in the improvement of forecast performance, particularly in the Southern hemisphere. Further optimisation of the use of observations is possible through proper evaluation of the data impact, optimisation of the amount of data to be assimilated and of their error characteristics, and a relevant selection of data based on information content concepts. A more interactive forecasting system including an adaptive observation component is a new challenge to bring additional improvement in the forecasting of high-impact weather.

Keywords: Observations, Numerical Weather Prediction, data assimilation, observation targeting, data impact, optimisation, data selection, information content.

1. Introduction

Atmospheric data assimilation consists in combining information coming from a forecast model together with available observations. It is usually performed in a sequential way, with a time series of "assimilation cycles" including a model integration and a correction due to observations. As a new set of observations becomes available every six or twelve hours, a short-range forecast (so-called "background") is updated with the new set of data into a new "analysis" of the atmosphere. This analysis is then propagated in time with the forecast model to provide a new background field for the next "assimilation cycle". This series of steps in the data assimilation process shows that the atmospheric model is the basic ingredient which allows time continuity in our evaluation of the atmospheric flow. It also means that the observations are the crucial elements allowing to

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 343-360.

constantly re-adjust the model trajectory to produce a reasonable estimate of the true atmospheric state. From these analyses of the atmosphere, the model is run daily up to a few days to produce the forecast products which will guide the forecasters in their prediction of the weather. At the beginning of the 80's, data assimilation was a minor sub-discipline of numerical weather prediction, where the emphasis was mainly on the forecasting model itself. Simple correction methods were used to update the forecast such as nudging and linear optimal interpolation. Over the last two decades or so, this subject has expanded into quite a mature and motivating area of research and applications, with in particular the advent of variational methods. Such major scientific advances, combined with a large increase in available observations, has brought data assimilation to the forefront of operational weather forecasting. Its use is also spreading to climate applications through re-analyses and oceanography/chemistry applications. The experience gained in data assimilation in meteorology can be shared with scientists interested in other areas, such as oceanography. This paper will mainly address the issue of the importance of data in the assimilation process, in the context of global atmospheric modelling.

Firstly, the impact of observations on the forecast performance will be illustrated through the 40-year reanalysis performed at ECMWF (European Centre for Medium-range Weather Forecasts). Secondly, tools will be described which can help to perform an optimal use of observations, through data selection and error tuning. Finally, current developments towards an adaptive system will be described in the context of the THORPEX programme.

2. Impact of observations on forecast performance

Operational data used at Numerical Weather Prediction (NWP) centres are consisting of various data types provided by the global observing system. The backbone of this system is formed by surface observations from land and ship stations, and vertical soundings from radiosonde and pilot balloons. From the 1970s, other data types emerged such as drifting buoys, aircraft measurements, wind profilers, satellite radiances, satellite cloud-drift winds and scatterometers. On the one hand, observations such as land stations and radiosonde observations have been providing a stable source of information throughout the years, but their horizontal distribution is far from being homogeneous. On the other hand, satellite observations are blooming and becoming a major and horizontally homogeneous source of information in current systems. How did this increase in available observations translate into analysis and forecast quality? As a partial answer to this question, an illustration of the impact of observations is now provided in the context of the ERA-40 project (www.ecmwf.int/research/era). As summarized in

Simmons (2003), a re-analysis was conducted from September 1957 to August 2002 based on cycle 23r4 of ECMWF forecasting system operational from June 2001 to January 2002. It uses six-hourly variational analysis, a degraded version of the operational analysis scheme. The T159 horizontal resolution (~125km grid) is coarser than current operations which uses T511 (~39km grid). This re-analysis can then be seen as a cheaper version of the current ECMWF operational system. Figures 1 and 2 show anomaly correlations of 12UTC 500hPa height forecasts as a function of forecast range for the extratropical northern hemisphere (Figure 1) and for a smaller region encompassing Australia and New Zealand (Figure 2). These anomaly correlation scores quantify the quality of the forecasts, 100% being the maximum score and 60% the score below which the forecast is not generally considered useful. Results are shown for many of the ERA-40 years, verified against corresponding ERA-40 analyses, denoted by the colour scale shown in legend and for ECMWF operations (verified against the corresponding operational analyses, labelled Ops) for the calendar years of 1980 and 2001 and for the year ending 31 August 2003.

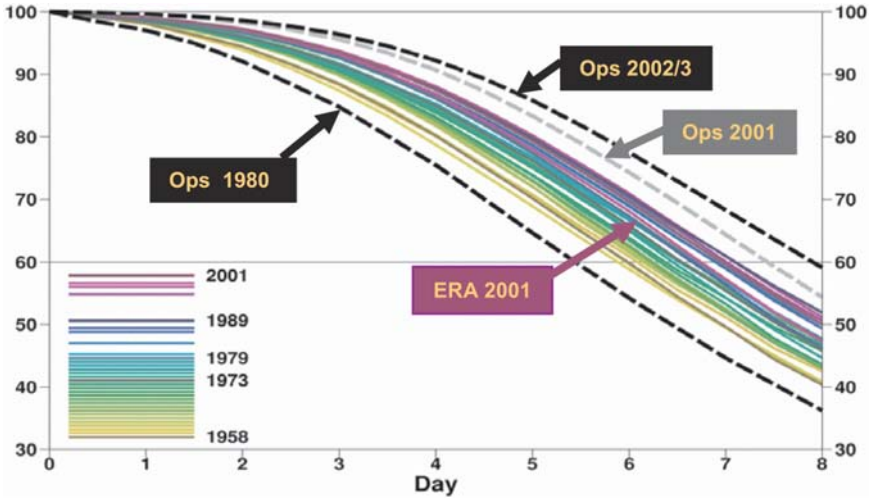

Figure 1. Anomaly correlations of 500hPa height forecasts over the Northern Hemisphere. From Simmons (2003).

These figures provide some evidence of the general improvement of the analyses over time, with interannual variations in predictability. The northern hemisphere results in Figure 1 show that whilst the observing system for medium-range prediction has improved over the years, a greater improvement in forecasts has been derived from the improvements in data assimilation and forecast models achieved since 1980. This can be seen by comparing the improvement in the coloured solid lines (same system, improvement entirely due to global observing system) and the larger

improvement in the dashed lines (both changes in the observations and in the NWP system). A different picture is seen for the southern hemisphere, where forecast performance is mainly driven by satellite data (Bouttier and Kelly, 2001). The area chosen for the score calculation in Figure 2 includes Australia and New Zealand where observational coverage is sufficient for some reliance to be placed on the quality of the verifying analyses throughout the period. Forecast quality is poor in the 1950s and 1960s. A dramatic jump in forecast quality occurs at the end of 1978 when the observing system was improved considerably with the introduction of radiances from the TOVS instruments and the addition of winds from geostationary satellites and many more data from drifting buoys and commercial aircraft. Observing-system improvements beyond 1979 have had larger impact on southern- than northern-hemisphere forecast accuracy, bringing forecast skill levels closer. In any case, observing system improvements have had a major impact on the forecast scores globally.

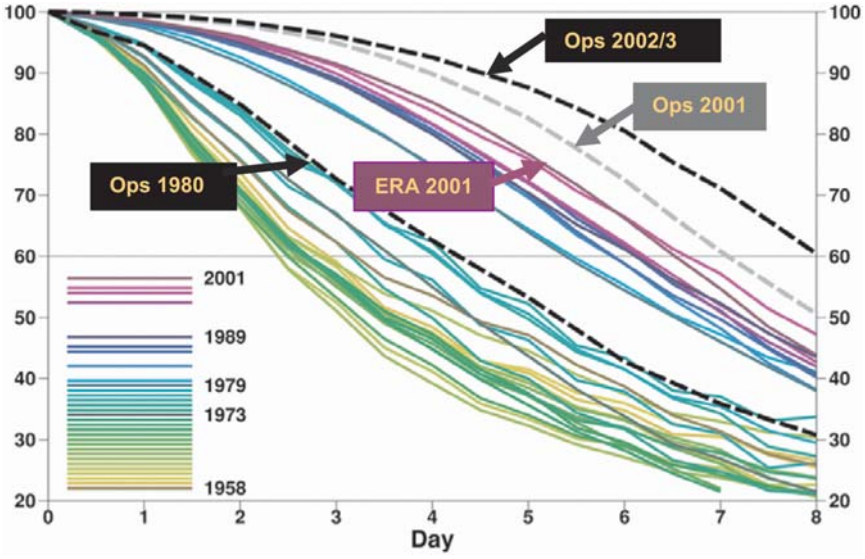

Figure 2. Anomaly correlations of 500hPa height forecasts over Australia-New Zealand. From Simmons (2003).

Such a large impact of observations has been made possible through the large increase in number and quality of observations combined with new approaches for their assimilation. One of the major advances has been the direct use of raw radiance observations in data assimilation in variational systems in the last few years in most NWP centres (and in the ERA-40 re-analysis). To explain this approach, let us start with the basics of satellite data assimilation. Radiometers provide a set of radiance measurements at various frequencies in the electromagnetic spectrum. Each of these radiance measurements provides information about temperature and/or humidity

integrated over a layer of the atmosphere. A set of a few of these measurements thus describes broad vertical structures in temperature and humidity. Data assimilation in some way or another converts these radiance measurements in temperature/moisture profiles. Different possibilities exist to process this information. One can use externally generated retrievals (profiles deduced from a set of radiances through regression typically), interactive retrievals using in-house information about short-range forecasts (e. g. 1D-Var retrievals), or the direct use of radiances (e.g. 3D-Var or 4D-Var). In NWP at least, the direct assimilation of satellite raw radiances has progressively replaced the assimilation of retrievals (Thépaut, 2003). This has been made possible because 3D and 4D-Var allow for some (weak) non linearities in the observation operator, and radiances are non-linearly linked with the atmospheric profiles. Retrievals always need prior background information, which either comes from independent statistics or from the short-range forecast. The direct assimilation of radiances has the advantage to avoid the contamination by such an external background information for which error characteristics are poorly known. Another advantage of global variational methods is that increments brought by satellite radiances are further constrained by many other observations/information. Finally, raw radiance observations exhibit less spatially correlated errors than processed retrieved information. In current data assimilation schemes, this allows to use observations with more spatial density, a subject which will be discussed further in the next section. Of course this use of raw data comes at the cost of developing the observation operator and the quality control appropriate for each observation for each data assimilation system in each NWP centre, but some of this effort is collaborative through EUMETSAT facilities for instance.

Zooming now on the period covering the most recent years, Figure 3 shows the number of data used in the ECMWF analysis between 1997 and 2003. This illustrates the tremendous increase in terms of observation numbers which took place lately, and most of this progression in data numbers comes from non-conventional asynoptic observations.

Such observation numbers have a significant impact, especially in an advanced data assimilation scheme such as 4D-Var which has been used since 1997 (Rabier et al, 2000). 4D-Var stands for Four-Dimensional Variational Data assimilation and it performs a global optimization of the model trajectory over a period of 6 to 12 hours typically. It performs an adjustment of the model trajectory with the observations taken explicitly at the precise time of the observation, thus allowing for a consistent use of data spread in time throughout the optimization period, such as satellite observations. In the linear approximation, 4D-Var is equivalent to a Kalman smoother: at any time in the assimilation window, information from past and future observations within this window will be taken into account to provide the best estimate of the flow (Rabier and Liu, 2003). It can also use the time-

tendency between various observations to adjust the model, which is particularly beneficial in the case of rapidly-developing weather systems (Järvinen et al, 1999). It is then particularly suited to the use of a large number of observations spread in time.

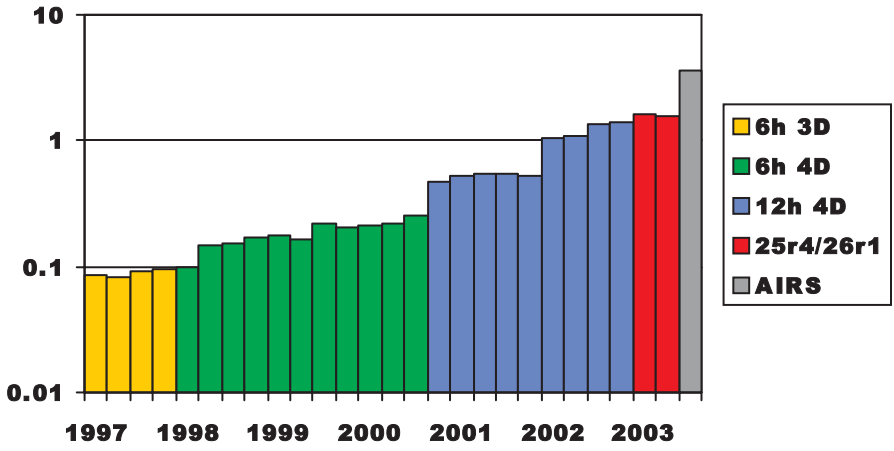

Figure 3. Number of observational data used in the ECMWF assimilation system, in millions. From Thépaut (2003).

What should be kept in mind after this introductory presentation is that data assimilation techniques now allow to make full use of observations, and in particular satellite measurements. These have become a major source of information in NWP systems, and their increase in number and quality is currently booming. It is then the right time to investigate their use in the view of optimally extracting the information contained in these data.

3. Optimal use of observations

3.1 Optimal resolution of observations

As already seen in the previous paragraph, the performance of current NWP systems benefits to a large extent from the increasing amount of globally available remotely-sensed observations used together with conventional observations to generate initial conditions for forecasts. Some of these data have fine horizontal resolution. The observation spacing can be smaller than the analysis grid of global NWP models. Not all of these observations are used in data assimilation systems because of various considerations. Firstly, current computing and storage power limits the use of all observations. Secondly, the errors affecting these observations may be horizontally correlated (instrument errors and/or representativeness errors);

current assimilation systems do not generally consider this correlation in the modelling of the observation-error covariance, because of a lack of accurate information on the correlation statistics and the technical difficulty of implementation. Alternatively, most NWP centres tend to use sub-optimal schemes for which the observation-error covariance matrix is designed to be diagonal. At the same time, horizontal thinning of remotely sensed observations is performed in order to reduce their effective error correlation.

Liu and Rabier (2002) have used a simple one-dimensional context to evaluate the optimal resolution of the observations leading to the best analysis. The framework is a 1D circle of a length of 8000km, with a grid-size of 100km. Background and observation errors have the same standard-deviation equal to 1 (arbitrary value).The background error correlation length-scale is taken equal to 200km. The analysis error covariance matrix is calculated for various observation spacings. Various scenarios were tested: uncorrelated observation errors and correlated observation errors with a correlation length of 100km. In the case of correlated observation errors, two analysis schemes were tested: the optimal one taking into account the proper observation error covariance matrix and a sub-optimal one neglecting the observation error correlations (similar to operational practice). Figure 4 shows the analysis error variance resulting from these combinations of observation density/observation correlation/analysis scheme. The main results are that, for uncorrelated observation errors, increasing the density always improves the analysis (dash-dotted line). This is the case even when the observation density is finer than the background error correlation length-scale and the analysis mesh. For correlated observation errors, increasing the observation density beyond a threshold can be harmful in a sub-optimal scheme for which no correlations are included in the observation error covariance matrix, as in current systems (dashed line). These results have been confirmed by a further study in a more realistic 4D context (Liu and Rabier, 2003) and might explain some of the results found in practical NWP experience.

It is also found that an optimal thinning of the dataset can extract most of the information contained in the data, and this approach is the pragmatic one used in most centres. The “optimal” observation density is usually found by trial and error. Another ad-hoc approach is to use most of the observations but to inflate artificially their errors to compensate for their correlations. More general solutions would of course be preferable. In particular, instead of performing a thinning of the observations, one might prefer to perform an averaging of neighbouring observations. The best theoretical framework might well be to model the correlations in the long term, if feasible.

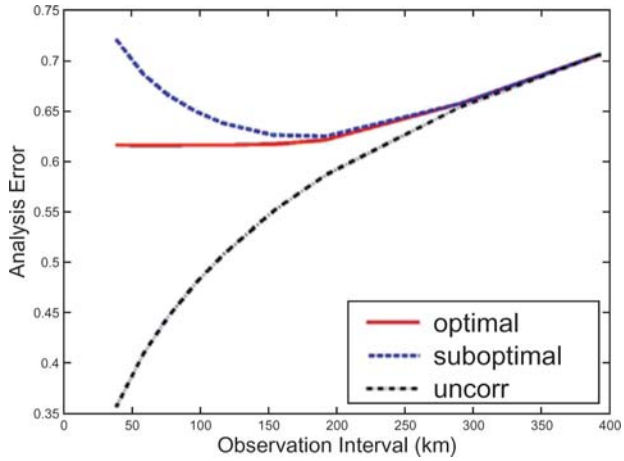

Figure 4. Analysis error standard-deviation as a function of the observation interval in a simple one-dimensional framework. The black dash-dotted line corresponds to uncorrelated observation errors. The solid red line corresponds to correlated observation errors, fully accounted for in the analysis. The dashed blue line corresponds to correlated observation errors, not accounted for in the analysis. From Liu and Rabier (2002).

3.2 Advanced diagnostics

Apart from the density issues explained in the previous section, another important question might arise in the use of observations, such as: what is the actual information content of the data? A simple data count might be misleading as not all observations are equal in what they measure and with what accuracy. In the perspective to diagnose the impact of observations on the data assimilation, some diagnostics were developed which are presented here.

Firstly, let us recall the equations relevant for statistical estimation, from the point of view of least squares. Let us assume that observations $\mathbf{y}$ are available, with a known observation operator $\mathbf{H}$ linking them to the atmospheric state vector $\mathbf{x}$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}_\mathbf{r} \tag{1}$$

together with a background vector (which usually comes from a short range forecast)

$$\mathbf{x}_\mathbf{b} = \mathbf{x} + \boldsymbol{\varepsilon}_\mathbf{b} \tag{2}$$

The least-squares method for estimating the analysed state $\mathbf{x}_\mathbf{a}$ is to minimize the cost-function

$$J(\mathbf{x})=1/2\ (\mathbf{x}-\mathbf{x}_\mathbf{b})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_\mathbf{b}) + 1/2\ (\mathbf{y}-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}) \tag{3}$$

where **B** is the covariance matrix of the background error $\boldsymbol{\varepsilon_b}$ and **R** the covariance matrix of the observation error $\boldsymbol{\varepsilon_r}$ which includes the instrument error and the representativeness error.

The solution in this linear case is given by

$$\mathbf{x_a} = \mathbf{x_b} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x_b}) \tag{4}$$

with the gain matrix **K**

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{5}$$

The corresponding analysis error is given by

$$\boldsymbol{\varepsilon_a} = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\boldsymbol{\varepsilon_b} + \mathbf{K}\boldsymbol{\varepsilon_r}. \tag{6}$$

The analysis error covariance is

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}. \tag{7}$$

This is the Optimal least-square estimator, which leads to the minimum variance for the analysis error, or BLUE = Best Linear Unbiased Estimator. If all errors are Gaussian, then it is also the maximum likelihood estimate.

When one wants to evaluate the gain brought by the observations, a pure data count can be misleading. If practically feasible, the computation of the analysis error covariance **A** and its comparison with the background error covariance matrix **B** will indicate how much benefit was brought by the observations, in terms of decreasing the error covariance of the estimation of the atmospheric state. Another approach is to compute the sensitivity of the analysis with respect to the observations. This leads to estimating the information content of data types. For example, one can compute the DFS = Degrees of Freedom for Signal (Rodgers, 2000)

$$\mathrm{DFS} = \mathrm{Tr}(\mathbf{K}\mathbf{H}) \tag{8}$$

where the trace of the matrix **KH** quantifies the gain in information brought by the observations. As shown by

$$\mathbf{H}\mathbf{x_a} = (\mathbf{I} - \mathbf{H}\mathbf{K})\,\mathbf{H}\mathbf{x_b} + \mathbf{H}\mathbf{K}\mathbf{y} \tag{9}$$

The **HK** matrix quantifies the sensitivity of the analysis to the observations

$$\partial_y \mathbf{H}x_a = (\mathbf{HK})^T \ . \qquad (10)$$

DFS=Tr($\partial_y \mathbf{H}x_a$)=Tr(**HK**) characterizes how the assimilation system uses the observations to pull the signal from the background. In the optimal case (i.e $\mathbf{K} = \mathbf{K}_{true}$), this is also the relative reduction of variance (Tr (**KH**) = Tr ((**B-A**)***B**$^{-1}$) = Tr(**I-AB**$^{-1}$)). It is only an upper bound in non-optimal cases. It indicates what the system does. One would need other information to give insight about what should be done to get the best analysis.

How to estimate Tr (**HK**)? This is not straightforward for large-dimension systems where the matrices are often implicitly known and not explicitly computed. Cardinali et al (2003, 2005) compute an estimate using the singular vectors of the hessian of the cost function provided by the Lanczos/Conjugate gradient minimizer. Another method was introduced in Desroziers and Ivanov (2001) and is used in Chapnik et al (2005), based on Girard (1987) for the evaluation of the trace of a matrix only known as an operator. Let us present the basis of this method, which is relatively easy to implement.

Let us consider a vector $\boldsymbol{\varepsilon}$ following a normal (Gaussian) distribution with mean **0** and covariance matrix the identity **I**. Girard (1987) proposed to use the following mathematical identity in order to evaluate the trace of a matrix **A** only known as an operator

$$E(\boldsymbol{\varepsilon}^T \mathbf{A} \boldsymbol{\varepsilon}) = \mathrm{Tr}(\mathbf{A}). \qquad (11)$$

The evaluation of the trace of **KH**, or equivalently the trace of **HK** can be perfomed based on this equation. The basic idea is to produce two analyses, one being deduced from the other by perturbing the observations. The difference between these two analyses will then be equal to the operator **K** applied to the perturbation, and applying **H** to this difference of analyses will give access to the **HK** operator needed. This method, as presented in Chapnik et al (2005) involves several steps

1. Perform a normal analysis from the information ($\mathbf{x_b}$, **y**) producing the analysis vector $\mathbf{x_a}$
2. Perform a perturbed analysis from the information ($\mathbf{x_b}$, **y***) with perturbed observations $\mathbf{y}^* = \mathbf{y} + \mathbf{R}^{1/2}\boldsymbol{\varepsilon}$ leading to $\mathbf{x_a}^*$. One notes that $\mathbf{x_a}^* - \mathbf{x_a} = \mathbf{K}(\mathbf{y}^* - \mathbf{y})$.
3. Then $(\mathbf{y}^* - \mathbf{y})^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x_a}^* - \mathbf{x_a})$ provides an approximation to Tr(**HK**).

If the number of observations is large, one sequence of such steps might be enough to get a reasonable estimation of the trace. Otherwise, one can use several realizations of the analysis, and concatenate the results.

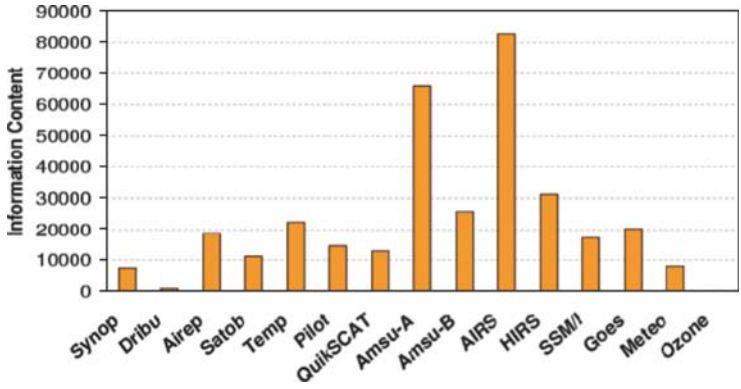

Figure 5. Information Content. Partition by observation type for the ECMWF system. Synop= surface observations, Dribu=drifting buoys, Airep=aircraft obs, Satob=winds from geostationary images, Temp=radiosonde obs, Pilot=wind profiler, QuikSCAT=scatterometer, Amsu-A +Amsu-B +AIRS +HIRS +SSM/I +Goes +Meteo =satellite radiances, Ozone=ozone information. From Cardinali et al (2003).

For non-linear cases, $\mathbf{H}(\mathbf{x_a}^*)-\mathbf{H}(\mathbf{x_a})= (\partial_{\mathbf{y}}\ \mathbf{H}(\mathbf{x_a}))\ (\mathbf{y}^*-\mathbf{y})$. Therefore,$(\mathbf{y}^*-\mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}(\mathbf{x_a}^*)-\mathbf{H}(\mathbf{x_a}))$ gives an approximation of $\mathrm{Tr}(\partial_{\mathbf{y}}\ \mathbf{H}(\mathbf{x_a}))$. This shows that this method allows the computation of DFS like quantities even for non-linear schemes.

An example of the use of the DFS as a diagnostic is shown in Figure 5. The partition by observation types allows to diagnose which observing system is pulling the analysis more or less than the other types. It can be seen that globally, the satellite observations have become a dominant source of information.

3.3 Observation error estimation

Apart from its use in pure diagnostic mode, it is possible to use DFS related quantities to improve specified covariance matrices. Following Desroziers and Ivanov (2001) and Chapnik et al (2004), suppose one can write the "true" perfect covariance matrices as a function of the ones actually used in the analysis

$$\mathbf{B}_{\mathrm{true}}= s_b\ \mathbf{B} \qquad (12)$$

$$\mathbf{R}_{\mathrm{true}}=s_o\ \mathbf{R} \qquad (13)$$

s_o and s_b being tuning coefficients. If

$$J(\mathbf{x})=1/2\ (\mathbf{x}-\mathbf{x_b})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x_b}) + 1/2\ (\mathbf{y}-\mathbf{Hx})^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Hx})= J_b +J_o \qquad (14)$$

is the cost function used in the sub-optimal system, with J_b the first term on the right-hand side and J_o the second term on the right-hand side, then

$$J_{true}(\mathbf{x}) = J_b/s_b + J_o/s_o \quad (15)$$

is the cost function using « true » matrices.
Let $\mathbf{x_a}$ be the minimizer of this cost function, then, following Talagrand (1999)

$$E(2J_o(\mathbf{x_a})/s_o) = Tr(\mathbf{I} - \mathbf{HK}) \quad (16)$$
$$E(2J_b(\mathbf{x_a})/s_b) = Tr(\mathbf{KH}) \quad (17)$$

yielding the following condition for the tuning coefficients

$$s_o = 2J_o(\mathbf{x_a})/ Tr(I - \mathbf{HK}) \quad (18)$$
$$s_b = 2J_b(\mathbf{x_a})/ Tr(\mathbf{KH}) \quad (19)$$

As **K** depends on s_o and s_b, this is a fixed-point relation, and a fixed point algorithm can be used to estimate the tuning coefficients. The denominator of those expressions can be computed using Girard's method, which is also used to compute the DFS.

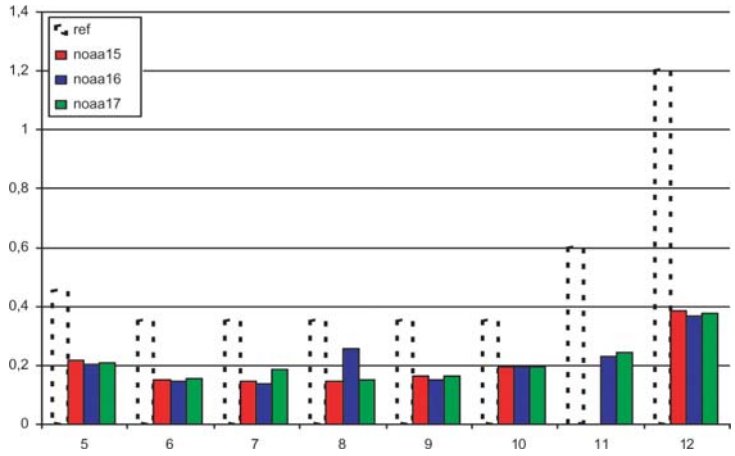

Figure 6. Observation errors, specified in the data assimilation system (dashed bars) and estimated by the optimisation method, for several channels of the AMSU-A instrument on board 3 NOAA satellites of the ATOVS series (coloured bars). From Chapnik et al (2005).

An example of the tuning of the error standard-deviations for satellite radiances of the ATOVS series is shown in Figure 6. One can see that the errors are generally over-estimated in the operational French global NWP model, and that the method can pick up small differences between the various satellites (which have been confirmed by the individual monitoring of the data). This type of information can be very valuable for a real-size NWP system, for which there are too many parameters to allow to perform a fine tuning on each by trial and error methods.

3.4 Channel selection for satellite sounders

Another optimisation of the use of observations can be the selection of the most valuable subset of data, if some of the global observing systems are providing too many pieces of information for the data processing capabilities. In particular, advanced infrared sounders provide thousands of radiance data at every observation location. The first instrument with kilo-channel data is the Atmospheric InfraRed Sounder (AIRS) on the Aqua satellite launched by the National Aeronautics and Space Administration (NASA) in 2002. On the European side, the French space agency Centre National d'Etudes Spatiales (CNES) and the European Meteorological Satellite organization (EUMETSAT) have developed the Infrared Atmospheric Sounding Interferometer (IASI) to be launched at the end 2005. For operational NWP systems, these data will provide temperature and humidity information with a fine vertical resolution. The number of individual pieces of information is not usable in an operational NWP context, and several possibilities are being investigated to choose an "optimal" subset of data. This would allow to extract the maximum information content from hyperspectral sounders, with a reduced number of individual data. Solutions proposed to solve this problem include the selection of relevant limited spectral bands (Aires et al, 2002), the grouping of highly correlated channels in the same spectral area into super-channels, the use of a partial eigen-decomposition of the radiance data (Joiner and Da Silva, 1998), and the selection of individual channels based on objective criteria (Rodgers, 1996).

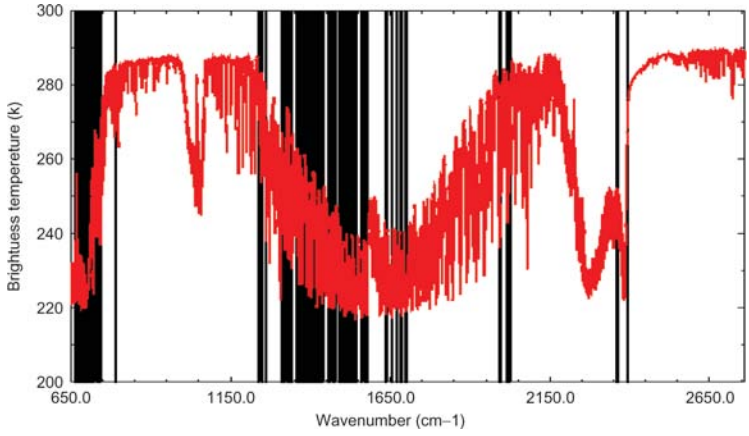

Figure 7. Location of 300 channels (black lines) selected by an iterative procedure based on information content for the retrieval of temperature and humidity for a IASI spectrum (red line). From Rabier et al (2002).

An example of such a channel selection procedure is presented in Rabier et al. (2002). The selection of individual channels is performed for simulated

IASI spectra (8461 radiance data). The procedure is iterative, based on Rodgers, (1996). At each step, one channel is picked. It is the most informative channel among those which have not been previously selected. The analysis error covariance matrix is then updated before proceeding to the next channel selection step. The choice of channels is based on information content with respect to the background or current analysis information. The selection criterion is the entropy reduction ER=-1/2 $\log_2\det(\mathbf{AB}^{-1})$ or the DFS= Tr ($\mathbf{I}$- $\mathbf{AB}^{-1}$). At each step, one optimises ER=-1/2 $\log_2\det(\mathbf{A_{i+1}}\ \mathbf{A_i}^{-1})$ or the DFS= Tr ($\mathbf{I}$- $\mathbf{A_{i+1}}\ \mathbf{A_i}^{-1}$) with $\mathbf{A_i}$ the analysis error covariance when using the first i selected channels, and $\mathbf{A_{i+1}}$ the analysis error covariance when using the first i+1 selected channels. Figure 7 shows the location of the "optimal" 300 channels selected for the retrieval of temperature and humidity information for a typical IASI spectrum. This type of work illustrates the benefit of using information content diagnostics for the benefit of the optimisation of data assimilation, through data selection.

4. Towards an adaptive system

Despite the notable increase in forecast skill over the past quarter-century, there is a necessity for further improvements, particularly in high-impact weather defined by their effect on society and the economy. The international programme THORPEX is a response to the challenge of improving the skill of high-impact weather forecasts. Its mission Statement is "Accelerating improvements in the accuracy of high-impact 1-14 day weather forecasts for the benefit of society and the economy". Information on this programme, and in particular the science plan, can be found on the World Meteorological Organisation web page (www.wmo.int). Research objectives are developed under four Sub-programmes: Predictability and Dynamical Processes, Observing Systems, Data Assimilation and Observing-Strategies, Societal and Economic Impacts. Among the core objectives, THORPEX plans to Contribute to the design and demonstration of interactive forecast systems which include the new concept of "targeted observations" and to perform THORPEX Observing-System Tests (TOSTs) and Regional field Campaigns (TReCs) to test and evaluate experimental remote-sensing and in-situ observing systems, and when feasible, demonstrate their impact on weather forecasts.

What is Targeting? In the last decade, strategies were developed that identify locations where additional observations would provide maximal improvements in the expected skill of forecasts. Targeting strategies are based on techniques that predict, prior to the actual measurements, the influence of an observation (or set of observations) on the uncertainty of a subsequent forecast. Different targeting techniques have been developed: some involve the adjoint of the linearized version of the forecast model or of

the assimilation scheme, others manipulate ensembles of forecasts. This concept is currently operational in the US and is called the Winter Storm Reconnaissance Program. The National Centers for Environmental Prediction uses the dispersion of the ensemble of forecasts run routinely and a set of pre-defined flight plans to evaluate which of the flight scenario would bring the maximum reduction in the dispersion of the forecasts. This flight scenario leads to a designated aircraft flying in the area and dropping dropsondes at regular intervals to provide additional observations. Majumdar et al (2002) provide a detailed comparison of various targeting techniques.

Apart from being used for selecting additional observations, targeting observing systems can be extended to other applications such as controlling the sampling rate of satellite sensors or the timing and location of mobile upper-air soundings. Targeting can also be used to determine which observations are to be discarded, i.e., to conduct effective thinning of the observations. This capability will become increasingly important, given the very large numbers of observations that will be available from next-generation satellites. Among the tools which can be useful for targeting, being able to quantify the impact of any observation on the analysis and the subsequent forecast is crucial. Such a tool has been developed in particular using the adjoint of the various operations involved (analysis step and model forecast) by Baker and Daley (2000) and Doerenbecher and Bergot (2001).This sensitivity to observations is illustrated in Figure 8. The "forward" step consists in the analysis represented by the Kalman gain matrix $\mathbf{K}$ and the forecast model $\mathbf{M}$. From the background $\mathbf{x_b}$ and the observations $\mathbf{y}$, it creates the analysis $\mathbf{x_a}$ and the subsequent forecast $\mathbf{x_f}$. The "adjoint" step consists in the adjoint of the forecast model $\mathbf{M}^T$ followed by the adjoint of the analysis process $\mathbf{K}^T$. It allows to compute the sensitivity of any aspect of the forecast with respect to the observations and/or the background.

This sort of tools initially developed mainly for targeting can also allow to compute the sensitivity of the analysis or forecast to various satellite sounder channels (eg Fourrié et al, 2002) and can also be used to select channels in an adaptive manner (Fourrié and Rabier, 2004).

Beyond these targeting issues, NWP is also expected to progress (within THORPEX and also independently) in flow-dependent specification of various parameters used in assimilation. The major of these parameters is the background error covariance matrix $\mathbf{B}$, where a lot of work is ongoing to incorporate more flow-dependence in the statistics through 4D-Var or ensemble methods mainly. The link to the observations is then the estimation of background errors in observation space ($\mathbf{HBH}^T$) to perform first-guess check (Andersson et al, 2000). There are also interesting developments in the context of flow-dependent tolerances for outlier observations for an adaptive buddy check (Dee et al, 2001).

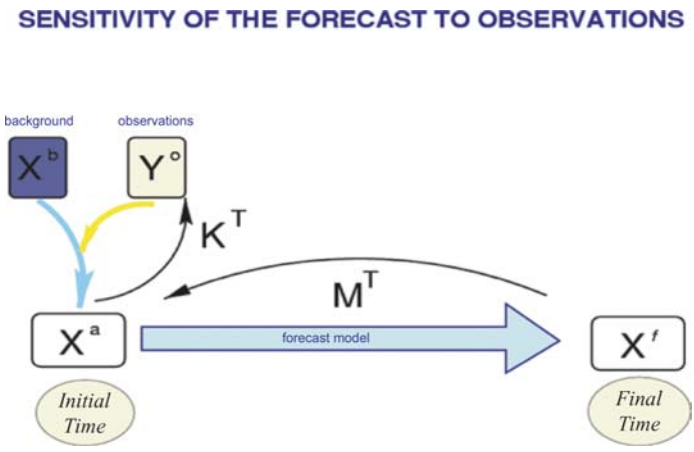

Figure 8. Schematic representation of the analysis and forecast steps K and M leading to the forecast and of its adjoints M^T and K^T leading to sensitivity computations.

5. Conclusion

In conclusion, there is ample evidence of the major improvements made in the last ten years or so in the context of Numerical Weather Prediction, and in particular in data assimilation. This can be seen in particular in the context of re-analyses programmes such a ERA-40, which highlight improvements coming from the increase in quantity and quality of data, mainly satellite observations, throughout the years. Satellite data are currently very successfully exploited by new data assimilation schemes (data assimilation schemes are now such that introducing additional well characterised satellite data generally improves the system). Variational methods have permitted to use such data in an innovative way, assimilating radiances directly in 3D/4D-Var, rather than using retrievals of temperature and humidity profiles obtained from the data. Furthermore, the proper inclusion of the time dimension in the assimilation period obtained in 4D-Var guarantees a near-optimal treatment of data which are not centred around the main synoptic times (0, 6, 12 and 18 UTC). In the future, the combined availability of new accurate satellite observations and improvement of models will allow an improved extraction of information content from these new data. In particular, observations related to the water cycle (clouds, rain…) will pose a great challenge to data assimilation. In general, the system can only cope with a small fraction of all available observations, and efficient tools have been built to evaluate observation impact. One of these tools is the Degrees of Freedom for Signal (DFS) quantity, which measures the sensitivity of the analysis with respect to observations. It can be used to investigate this sensitivity globally, data type

per data type, of by geographical areas, or even by parameters (Temperature, humidity, wind…). Such a diagnostic can also be used to perform optimal observation selection and error tuning. Looking forward to the future, we are now in a position to further optimise the use of observations, including more flow-dependency and a more interactive forecast system through the WMO programme THORPEX. The basic idea beyond this programme is to use the forecast system itself to predict where and when additional observations or a better treatment of planned observations would bring a major improvement in the forecasting of high-impact weather likely to have a high economic and societal impact. The NWP community has achieved a major improvement in average forecast scores in the last decade and is now mature enough to concentrate some of its efforts on the challenge of improving the forecast of rare events such as storms and floods.

Acknowledgements

This paper has benefited from contributions from Zhiquan Liu (Chinese Meteorological Administration), Nadia Fourrié (CNRS), Bernard Chapnik (CNRM), Adrian Simmons, Jean-Noël Thépaut and Carla Cardinali (ECMWF). Students and Gérald Desroziers (CNRM) are acknowledged for their review of the manuscript.

References

Aires, F., A. Chédin and N. A. Scott, 2002: A regularized neural net approach for retrieval of atmospheric and surface temperature with the IASI instrument. J. Appl. Meteorol, 41, pp 144-159.

Andersson, E., M. Fisher, R. Munro and A. McNally, 2000: Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. Quart. J. Roy. Meteor. Soc., 126, pp 1455-1472.

Baker, N. L, and Daley, R., 2000: Observation and background adjoint sensitivity in the adaptive observation targeting problem. Q. J. R. Meteor. Soc., 126, 1431-1454.

Bouttier, F. and G. Kelly, 2001: Observing-system experiments in the ECMWF 4D-Var data assimilation system. Quart. J. Roy. Meteor. Soc., 127, pp 1469-1488.

Cardinali, C., S. Pezzulli and E. Andersson, 2003: Influence matrix diagnostic of a data assimilation system. ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean.

Cardinali, C., S. Pezzulli and E. Andersson, 2005: Influence matrix diagnostic of a data assimilation system. Quart. J. Roy. Meteor. Soc., to appear.

Chapnik, B., G. Desroziers, F.Rabier, and O.Talagrand. 2004: Properties and first applications of an error statistic tuning method in variationnal assimilation. Quart. J. Roy. Meteor. Soc. 130, pp 2253-2276.

Chapnik, B., G. Desroziers, F.Rabier and O.Talagrand. 2005: Diagnosis and tuning of observational error statistics in a quasi operational data assimilation setting. Quart. J. Roy. Meteor. Soc. Accepted for publication.

Dee, D., L. Rukhovets, R. Todling, A. Da Silva and J. Larson, 2001: An adaptive buddy check for observational quality control. Quart. J. Roy. Meteor. Soc., 127, pp 2451-2471.

Desroziers, G. and S. Ivanov. 2001, Diagnosis and adaptive tuning of information error parameters in a variational assimilation. Quart. J. Roy. Meteor. Soc., 127, 1433-1452.

Doerenbecher A. and Bergot T., 2001: Sensitivity to observations applied to FASTEX cases. Non-linear Processes in Geophysics, 8(6), 467-481.

Fourrié, N., A. Doerenbecher, T. Bergot and A. Joly, 2002: Adjoint sensitivity of the forecast to TOVS observations. Quart. J. Roy. Meteor. Soc., 128, 2759-2777.

Fourrié, N. and F. Rabier, 2004: Cloud characteristics and channel selection for IASI radiances in meteorologically sensitive areas. Quart. J. Roy. Meteor. Soc., 130, 1839-1856.

Girard. D., 1987: A fast Monte Carlo cross-validation procedure for large leastsquares problems with noisy data. Technical Report 687-M,IMAG, Grenoble, France.

Järvinen, H., E. Andersson and F. Bouttier, 1999: Variational assimilation of time sequences of surface observations with serially correlated errors. Tellus, 51A, 469-488.

Joiner, J. and A. M. da Silva, 1998: Efficient methods to assimilate remotely sensed data based on information content. Quart. J. Roy. Meteor. Soc., 124, 1669-1694.

Liu,Z. and F. Rabier, 2002: The interaction between model resolution and observation resolution and density in data assimilation. Quart. J. Roy. Meteor. Soc., 128, 1367-1386.

Liu, Z. and F. Rabier, 2003: The potential of high-density observations for numerical weather prediction: A study with simulated observations. Quart. J. Roy. Meteor. Soc., 129, 3013-3035.

Majumdar,S. J, C. H. Bishop, R. Buizza and R. Gelaro, 2002: A comparison of ensemble-transform Kalman-filter targeting guidance with ECMWF and NRL total energy singular vector guidance. Quart. J. Roy. Meteor. Soc., 128, 2527.

Rabier, F., H. Jarvinen, E. Klinker, J-F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: experimental results with simplified physics. Quart. J. Roy. Meteor. Soc., 126, 1143-1170.

Rabier, F., N. Fourrié, D. Chafaï, and P. Prunet, 2002: Channel selection methods for infrared atmospheric sounding interferometer radiances. Quart. J. Roy. Meteor. Soc., 128, 1011-1027.

Rabier,F. and Z. Liu, 2003: Variational Data Assimilation: Theory and Overview. ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean.

Rodgers, C. D., 1996: Information content and optimisation of high spectral resolution measurements. Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II, SPIE Volume 2830, 136-147.

Rodgers, C. D., 2000: Inverse methods for atmospheres: Theory and practice. World Scientific Publishers, Singapore.

Simmons, A., 2003: Observations, assimilation and the improvement of global weather prediction- Some results from operational forecasting and ERA-40. ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean.

Talagrand., O., 1999: A posteriori verification of analysis and assimilation algorithms. In Proceedings of the ECMWF Workshop on Diagnosis of Data Assimilation Systems, 24 November pages 17--28, Reading.

Thépaut, J-N., 2003: Satellite data assimilation in Numerical Weather Prediction: an overview. ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean.

THORPEX International Science Plan: Mel Shapiro and Alan Thorpe (THORPEX Web site http://www.wmo.int)

Chapter 13

THE ECMWF PERSPECTIVE

David Anderson, Magdalena Balmaseda, and Arthur Vidard

ECMWF, Reading, UK

Abstract: ECMWF is involved in forecasting on time-scales from the medium range (days) to seasonal (6 months) ahead. This requires the preparation of initial conditions from which to start forecasts. For the medium range, a huge effort has been devoted to developing the most advanced assimilation strategies for analyzing the atmospheric state. Considerable effort has also been devoted to retrospectively analyzing the atmospheric state (ERA-40) using a more advanced assimilation system than was available at the time of the operational analysis. For forecasts beyond the medium range, coupled atmosphere-ocean models are used, requiring analyses of the ocean state. These in turn depend heavily on atmospheric analyses and reanalyses. Aspects of the atmospheric and oceanic assimilation systems used operationally are discussed, together with some limitations of current systems.

Keywords: ECMWF, forecasting, atmospheric analyses, reanalyses.

1. Introduction

In considering the merits of a data assimilation system, it is necessary to have the application in mind. That is after all what sets the measure by which one decides if the scheme works well or not. It is one thing to formulate an assimilation strategy theoretically but altogether a different matter to develop a practical implementation. No 'operational' assimilation system conforms to its theoretical optimum configuration.

At ECMWF (European Centre for Medium-Range Weather Forecasts), data assimilation is used in a number of applications. The key forecast applications are: deterministic forecasts for the medium range, ensemble prediction system (EPS) for the medium range, an ensemble of monthly forecasts and an ensemble of seasonal forecasts. The time ranges are 10 days for medium range, 31 days for monthly forecasts, 190 days for seasonal forecasts. As might be expected, the resolution of the model is not the same in all the applications. The model resolution decreases as one increases the

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 361-379.

forecast time or the ensemble size. Thus, for example, the deterministic forecast is made using a horizontal resolution of T511 (~40km), the EPS uses T255, the monthly uses T159, and the seasonal uses T95.

A further complication is that to make forecasts beyond a week or two requires information on the state of the ocean as well as that of the atmosphere. Forecasts of the monthly and seasonal range are made with coupled atmosphere-ocean models. The longer the range, the less important are the atmospheric initial conditions, but the more important the ocean initial conditions become (generally speaking). There might be a case for having an active ocean even for the short to medium range forecasts, as sea surface temperature changes associated with tropical cyclones and storm tracks might be important. This latter point requires confirmation, but regardless of whether the ocean impacts the atmosphere on the medium range, it is necessary to perform real-time ocean analyses for the monthly and seasonal forecast systems. At these longer forecast ranges, model error is sufficiently large that it cannot be ignored. One way to account for model error is to run the model for many realizations in the past to estimate the model climate and climate drift. This in turn brings in a need for extended analyses into the past as well as the present. ECMWF is thus heavily involved in reanalysis, both atmospheric and oceanic. It will be argued later that in fact reanalysis is an integral part of a forecast system, (at least for monthly and seasonal timescales and possibly also for medium range).

2. Atmospheric analyses

The atmospheric analysis is done using 4d-var (as discussed by F. Rabier in this volume). The 4 indicates the use of time as the 4th dimension. So the analysis is performed over a space-time box. It is performed over the highest resolution possible (this will be discussed later), producing an analysis which has the resolution of the first guess. The ECMWF model is formulated in spectral space rather than physical space i.e. in terms of spherical harmonics rather than grid points. In spectral space the resolution is T511 (about 20 km). The effective resolution is not this high, however. The T511 analysis is then used to provide initial conditions for the EPS, monthly and seasonal forecasts by truncating the analysis in spectral space to the appropriate resolution.

In addition to analyses made in near-real-time for the purpose of generating forecasts, analyses are made of past events as well, called reanalyses. Extended atmospheric reanalyses are major undertakings requiring years of dedicated work. So they are not undertaken lightly: in fact, at ECMWF, there have been only two to date. The first, denoted ERA-15 spanned the 15 years 1979-1993, and the second, denoted ERA-40, spanned the 40+ years Sept 1957- Aug 2002 (Uppala et al 2005). The analysis system used in these reanalyses should be the same throughout. For ERA-15, the scheme was OI, while for ERA-40 it was 3dvar. The expense

in performing the reanalyses means that it is not possible to use the latest scheme, nor the highest resolution. So for ERA-40, the resolution was T159.

3dvar does not try to fit the model trajectory throughout the data time window. It shares with Optimal Interpolation (OI) the approach of performing the analysis at discrete times. Account is taken of the time however, in the sense that the observations are compared with the model first guess at the time of the observation, rather than bunching all the observations in a given data window to the central time. This procedure is called FGAT (First Guess at Appropriate Time). Formally, if all assumptions are the same, OI and 3d-var are equivalent. In practice this is never so.

The analyses from ERA-40 have been used as initial conditions for medium-range weather hindcasts (hindcasts are made as if they were forecasts but are made for past events). Results from these hindcasts have been discussed by F. Rabier (this volume). The improvement of the forecasts as the observing system changes can be clearly seen. The improvements in forecast skill since the end of ERA-40 are also shown. They can result partly from improved data assimilation procedures (4d-var cf 3d-var) and partly from improvements in the quantity and quality of data.

An interesting finding is that the skill of forecasts for the southern hemisphere now rivals or exceeds that for the northern hemisphere, whereas in the first decades of ERA-40 reanalyses, the southern hemisphere skill was relatively poor compared to that of the northern hemisphere. Since the in situ measurements for the southern hemisphere are sparse, the implication is that the improvement results from increased satellite coverage. Fig 1, from Kelly et al 2004, shows results of various Observing System Experiments (OSE's) which clearly indicate the importance of satellite data for the southern hemisphere: when satellite data are withheld, the forecast skill is much reduced. By contrast, there is a much smaller degradation in the northern hemisphere. Other components of the observing system have a smaller impact.

Although the use of reanalyses to provide initial conditions for hindcasts is informative, it is not the sole or even primary purpose of performing a reanalysis. The analyses are used for diagnostic work to understand physical processes in the atmosphere. (The physical assumptions/simplifications that went into the analysis are important here). The application of most use to climate modelers is that they provide an estimate of the surface wind, heat and fresh water fluxes. As we shall later see, these are crucial in developing a monthly or seasonal forecast system, as they provide a means of dealing with model error. For monthly forecasting, the atmospheric initial conditions from the reanalysis are more important than the fluxes used to generate the ocean analysis. As the forecast range increases this balance shifts and for seasonal forecasts the fluxes are more important than the atmospheric initial conditions, as they play a major role in determining the ocean initial conditions.

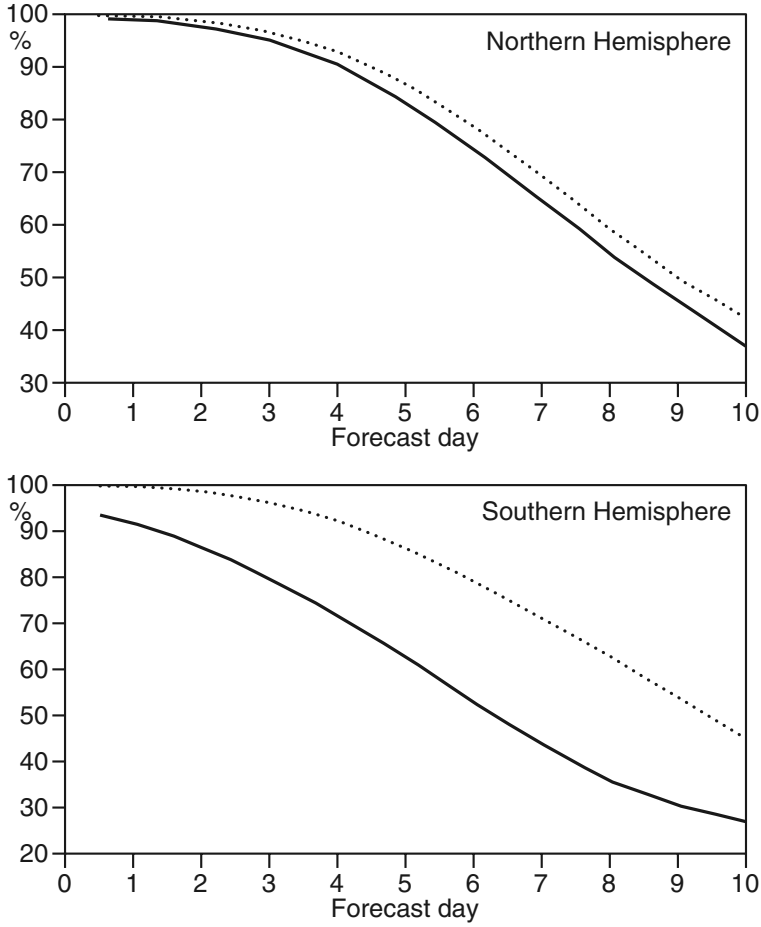

Figure 1. This plot shows the anomaly correlation between the predicted and analysed 500 hPa height field as a function of lead time out to 10 days for both the northern and southern hemispheres. The dotted line shows the skill when the full observing system is used: the solid line shows the skill when satellite data are withheld. Comparing upper and lower dotted curves shows that the skill of the SH is now commensurate with that of the NH. If one withdraws the satellite data, then the skill of the SH drops markedly whereas that for the NH drops only slightly. This OSE confirms the importance of satellite data, especially for the SH. From Kelly et al 2004.

The 4d-var assimilation system, seeks to minimise a cost function measuring the departure of the model trajectory from the data, subject to certain side-constraints, for example, that the departure from the first guess shouldn't be too large. The variables in the cost function are typically those governing the initial conditions. A data window (i.e. a time period over which the fit to data is minimized) is chosen. This is typically 6 to 12 hours for large-scale atmospheric models (Courtier et al 1994, Fisher 2005) but 30 days in the case of the ocean (Weaver et al 2003).

4d-var is cutting-edge technology and only a few operational weather forecasts centres have been able to adopt it. First was ECMWF, then Meteo France. The UK Met Office has recently introduced it. It is expensive, as one has to integrate the model forwards and its adjoint backwards through the data window (12 hours) of order 100 times. Simplifications are needed to make it feasible. First, an incremental approach is used (Courtier et al 1994). In this approach one assumes that only relatively small departures from the first guess will be made. The forward model is simplified, sometimes adjusting the physics to be more linear, so helping the derivation of the tangent linear (TL) and its adjoint. Although the TL can be a simplification of the forward model, the adjoint must normally be the exact adjoint of the tangent linear. The cost function, based on the tangent linear and its adjoint, is quadratic and convergence is faster than for a more general function. Further acceleration is achieved by using a lower resolution for the TL and its adjoint. At ECMWF, the resolution drops from T511 to T95 for the first 70 iterations of the inner loop. It is necessary to keep track of the full nonlinear model: so the trajectory is recalculated using the full nonlinear model at full resolution. This is followed by another 30 iterations of the inner loop at the slightly higher resolution of T159.

Because the ECMWF model is formulated in spectral space, rather than in gridpoint space, it is harder to impose covariances which vary geographically, e.g. to have different scales for the tropics to those for the extratropics and to have different degrees of isotropy. It can be done partially by using a variable such as vorticity, combined with linear balance constraints (Derber and Boutier 1998). Later, the implementation in an ocean context will be discussed and contrasted with the atmospheric case.

Although the same assimilation approach is used throughout the reanalysis, one can still get spurious low-frequency variability. This is because the observing system is not stationary. Although the number of observations generally increases as one approaches the present time, this increase is not monotonic. Some observing systems decline while other new systems come on stream.

An important component of any assimilation system is quality control: deciding which observations to keep and which to reject. In Rabier (op.cit), there is discussion of how to handle correlated observation error and how to thin potentially too-high-density observations such as those which can come from high-resolution satellites. Without getting into detail, as this is a difficult area (See Andersson and Jarvinen 1999), one counterintuitive example will be shown.

The late summer of 2004 has seen several intense hurricanes, several of which have caused substantial damage. When a hurricane is seen to be developing or tracking towards the Caribbean, additional meteorological observations are taken by flying aircraft into the storm and releasing dropsondes. These measure temperature, humidity and wind on their descent. In principle they should provide useful information on the location

of the eye, of the central pressure and of the structure both of the near and far fields. In addition scatterometers can sense the near-surface winds. Fig 2 shows the location of the extra observations along the tracks of hurricanes Frances and Ivan. Grey indicates the measurements that were accepted, black those which were rejected. One can see that almost all the measurements close to the centre of the hurricanes were rejected. As indicated by central pressure differences, there were big discrepancies between the model estimates and the measurements which was why the data were rejected.

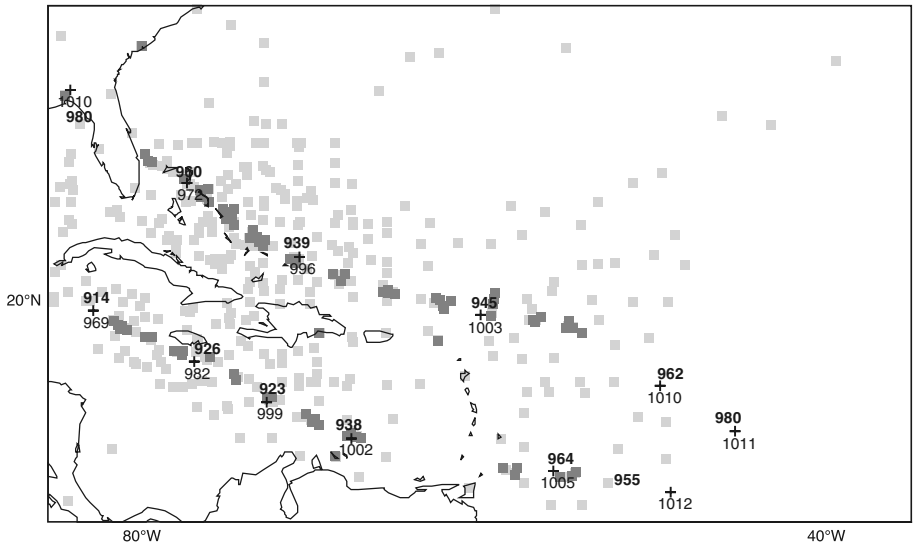

Figure 2. Dropsonde coverage for two tropical storms, Hurricanes Frances and Ivan in August September 2004. Bold numbers indicate the observed central pressure, normal numbers the analysed central pressure. Black squares indicate wind observations rejected by the analysis system as departing too far from the first guess. Light squares indicate observations which were accepted. Rejected data are mainly those close to the centres of the two storms. From Federico Grazzini, ECMWF.

Of course it is not clear that we are comparing like with like. The observation is a spot measurement, representative of only a small area of the storm. The model on the other hand represents an area average that is much larger. At best the grid is 20 km square but this is not the real resolution of the analysis system. First, the inner loops are run at T95/T159 (~180-120km). Secondly, the effective resolution of the analysis is considerably less than that as the structure functions are broader. In principle one could have adaptive structure functions with smaller scales around a storm but this is not done.

Often the analysis actually weakens the first guess. The model, when run forward, has a resolution of T511 and can develop i.e. intensify a tropical

storm over the next 12 hours. When the analysis kicks in, with its lower effective resolution, the storm can get weakened. So although the observations say deepen this storm significantly, the analyses says fill it in. This conflict in analysing small-scale, but important, features is generic.

3. The ocean analysis

3.1 Introduction

Ocean analyses are needed to provide initial conditions for both monthly and seasonal forecasts. For the monthly and seasonal forecast systems we use essentially the same ocean analysis systems. They are not exactly the same, as the monthly system needs a faster analysis than the seasonal and also has different cut-off times for the receipt of data. The resolution of the ocean model is 1 degree except near the equator where it is 1/3 degree in the meridional direction to better resolve the equatorial waves which are important in processes such as El Nino. In future we will explore the need for even higher resolution models coupled to the medium range weather forecast model, with potential impact on tropical cyclone prediction, but that will not be discussed further here.

3.2 The observing system

Ocean observations are mainly of thermal data and mainly in the upper 500m. Some salinity measurements are now available from ARGO but only in the last few years. Other measurements of salinity e.g. from CTD are sparse and have not generally been available in real-time. There are almost no measurements of velocity, except in the surface layer. Fig 3 shows the thermal data coverage in a typical 10-day window. Several features are worthy of note. First, there is quite a substantial coverage in real-time. In fact most data are now received within a day or so at operational centres such as ECMWF through a network called the GTS (Global Telecommunications System). When data are received, each individual observation is checked. It is compared against the model first guess and also with an analysis performed without the datum being checked. The actual quality control is quite complex and will not be gone into in detail. See Smith et al, 1991 for further discussion. Data profiles on fig 3 which are rejected are in grey. These are mainly located near the coast since coastal points are not well represented in the ocean model. Some profiles are partially accepted; at some depth the model and data differ by too great an amount and the datum is rejected but data above and below this level might be accepted.

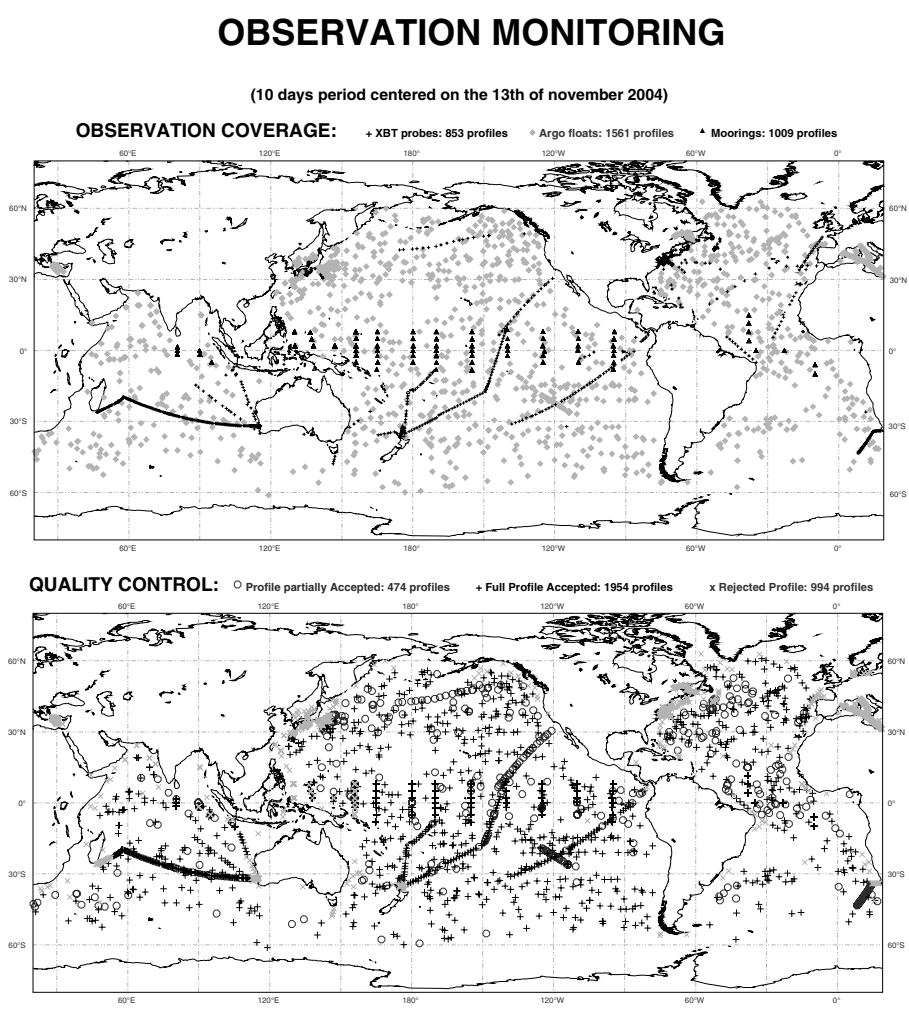

Figure 3. a) Data coverage for a 10 day period 9th-18th Nov 2004 inc. showing the TRITON/TAO/PIRATA mooring array in the Pacific/Atlantic oceans (Black triangles), the XBT network (black crosses) and the ARGO floats (grey diamonds). The moorings report daily in the Pacific to the east of the dateline and in the Atlantic. Hourly reports are received from moorings west of the dateline. b) This figure shows the profile data which have been fully accepted (black crosses), partly accepted (black circles) and completely rejected (grey crosses).

The array of triangles in the tropical Pacific and Atlantic oceans, seen most clearly in fig 3 is the TRITON/TAO/PIRATA. These report daily-mean values, so in a 10-day period one would expect ~10 measurements. For the TRITON array in the west Pacific the reporting is hourly. The straight or slightly curved lines of data are from merchant ships (called VOS or Voluntary Observing Ships) making XBTs measurements. XBTs are instruments which measure temperature to a depth of ~500m. A few

measure salinity. The diamonds are ARGO float measurements from buoys which drift at ~1000m, but every 10 days pop up to the surface, measuring temperature and often salinity, and relay the information via satellite to a ground station where it is put on the GTS. The mooring and XBT data are also available in real-time. If there are many data points in close proximity in space and time, they can be 'super-obbed', a process by which they are combined into one 'super-observation' and given an increased weight. See Smith et al 1991 for discussion on super-obbing and quality control issues.

3.3 The ocean analysis system

The scheme currently in operational use is OI (Optimal Interpolation) using a time window of 10 days. FGAT has not been implemented in this scheme: rather all observations in the 10 day window are taken to apply at the centre of the window. It differs from a standard OI, however, in that the correction to the FG following an OI is not applied instantaneously; rather the increment is divided by the number of timesteps in a 10 day window and then that fraction of the increment is applied every timestep. The idea behind this was to allow the model to generate its own circulation field following an OI in which only thermal data were assimilated. Although this worked to some degree, a better circulation field can be produced. In principle, one can update the velocity field and salinity field even though only temperature data are being assimilated. This can be done through having multi-variate covariances. However, at the time of implementation we did not have these well tuned. Rather, we sought to improve the univariate assimilation of T by building in some corrections, done in physical space rather than through multi-variate covariances. So a geostrophic velocity field is calculated following the T-assimilation and then the velocity increments are also spread over 10 days. See Burgers et al 2002 and Balmaseda 2003 for more details.

Salinity is also corrected by applying an S correction such that the T-S relationship is preserved. In the ocean when T changes, so does S in such a way as to preserve the T-S relationship. There are regions in which this approximation doesn't work- for example in the surface layers where heat and fresh-water fluxes will change T-S. The true T-S relationship is not known at any time, so we use the model T-S. Unfortunately the model T-S can drift from truth; for example, if the model has too much or too little mixing or if the precipitation is wrong and the surface layer starts to interact with the layers beneath. Nonetheless, although not a perfect solution, it does give beneficial impact, especially in the equatorial Atlantic and Pacific. See Troccoli et al 2002 for further details on the strategy and its impacts.

A novel feature of the ocean analysis system is that not just a single analysis but multiple analyses are performed. The purpose of the analysis is to provide initial conditions for monthly and seasonal forecasts. Such forecasts must be probabilistic. This implies that an ensemble of forecasts

must be made. In the case of the monthly forecast system, the ensemble size is 51, while for the seasonal forecast system, it is 40. The ensemble is there to sample uncertainty arising from the chaotic nature of the atmosphere. However, it should also take into account uncertainty in the ocean initial conditions. One method of representing this uncertainty is through running an ensemble of ocean analyses. In our case the ensemble size is 5. This ensemble is not to be confused with the EnKF in which the size might be ~100. Experiments are underway to assess the EnKF strategy as part of the EU project ENACT but results will not be presented here.

In the case of the atmosphere, almost all the information on which an analysis is based comes from observations of the atmosphere. In the case of the ocean, a substantial amount of information on the ocean state can be obtained not through ocean observations but through atmospheric observations - in fact all the observations that are involved directly or indirectly in defining the surface wind, heat and fresh water fluxes. For seasonal forecasting the most important of the surface forcings is the wind. The wind field that is used to force the ocean has uncertainty. We estimate that uncertainty and then force 5 ocean analysis streams with wind fields that are perturbed commensurate with the estimated uncertainty in the wind. In addition, the SST field is not known sufficiently accurately either. So perturbations to it are also applied. This is discussed more fully in Vialard et al 2005, who show the spread generated by different ensemble generation strategies.

Although the data coverage might look reasonable in fig 3 where the symbols are quite large, it is probably barely adequate even for today. Ten or twenty years ago the coverage was much worse. It is thanks in large measure to a major international programme called TOGA (Tropical Ocean Global Atmosphere) that the real-time coverage is as good as it is. Starting in 1985, this programme steered oceanography towards a free exchange of data in near real-time.

3.4 The value of data assimilation

There are surprisingly few clean sets of experiments to show that data assimilation improves the skill of seasonal forecasts. Any results probably apply only to the system being tested, as improvements in either the ocean model or the forcing fields through improved atmospheric analyses or reanalyses, could change the results. One clean set of experiments was performed by Alves et al 2004. Four sets of analyses were performed and four ensembles of forecasts were run from these analyses. Two different wind products were used and for each, experiments with and without data assimilation were performed. Fig 4 shows the growth of error (upper panel) and the anomaly correlation (lower panel) from these forecasts in the tropical Pacific, a key region for seasonal forecasts. The forecasts based on analyses with data assimilation are clearly better than those without: the rms

error is smaller and the anomaly correlation is higher. The lower curves on fig 4a give a measure of the spread of the ensemble.

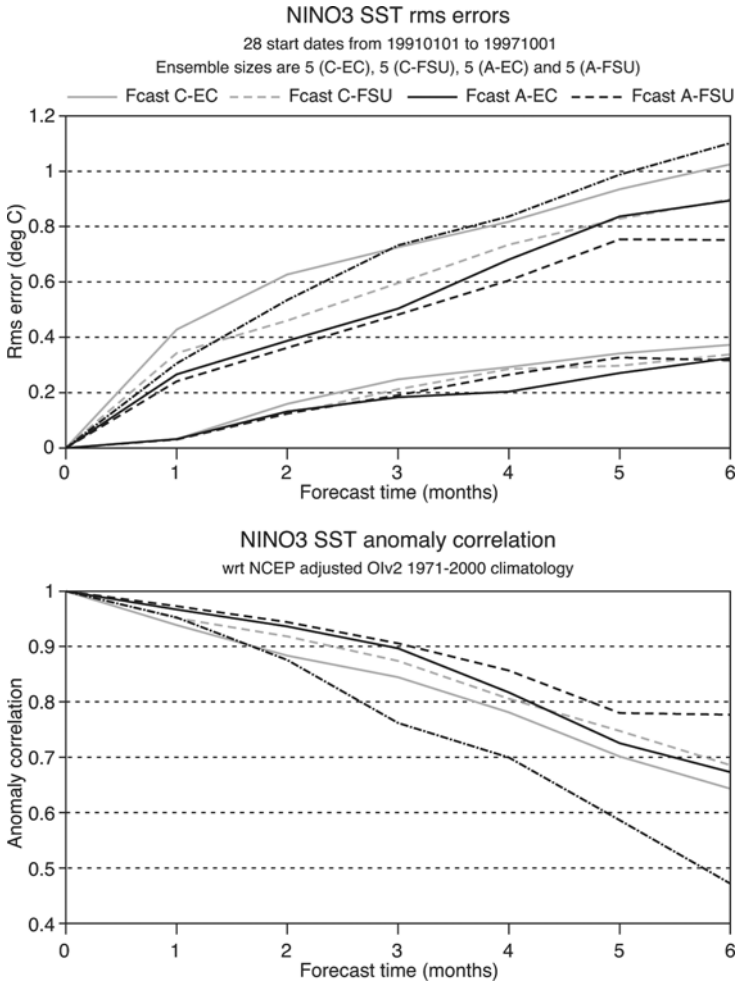

Figure 4. Upper panel: RMS error as a function of forecast lead time out to 6 months for the region NINO3 in the central east equatorial Pacific Ocean (upper set of curves). The lower set of curves on this panel shows the spread in the ensemble of forecasts. Lower panel: The anomaly correlation. Results for four experiments are shown. Two use ocean conditions with data assimilation (denoted A-); two do not assimilate sub-surface ocean data (denoted C-). Two different wind fields (FSU and ECMWF) are used to force the ocean during the assimilation process. This figure shows that the two runs with data assimilation have higher skill (smaller rms error and higher correlation) than those without, and that differences arising from differences in the wind field are reduced in the case of data assimilation. The dash dotted curve indicates the skill of persistence. From Alves et al 2004.

Vialard et al 2005 also show the growth of error in the forecasts. As for fig 4 which used earlier versions of both the ocean and atmospheric analysis

systems and the coupled forecast model than that used by Vialard et al, the spread is considerably smaller than the rms error, indicating a problem in the analysis/forecast system. At this stage we do not know if this is due to the analysis or the forecast model. One way of interpreting the difference in the growth of error versus the growth of spread in the ensemble is that the forecasts are too confident, indicating that all the uncertainty in the forecasts is not being accounted for. There is another more optimistic interpretation: that the spread is the theoretical measure of predictability. If the model error were small and the initial conditions were correct, then this is how the forecast error should grow. This estimate can be model dependent so it is not a hard argument. In practice, it is likely that by improving the models and the initial conditions, the error growth can be reduced and by improving the ensemble generation the spread can be increased. At ECMWF we are developing a multi-model forecasting facility in collaboration with the UK and French Met Offices and in this multi-model system the separation between error growth and spread is reduced. See also Palmer et al 2004.

3.5 Problems with the winds

In section 1, we considered the importance of atmospheric reanalyses. There we showed that the medium-range forecasts from say 20 years ago are now much better than they were then, partly because of improvements in the current analysis system compared with what was done then. That is encouraging but is not particularly useful in its own right, unless one wants to use these past forecasts for calibration. Calibration on past events has not really taken hold in the medium range community, though some moves in this direction are afoot (Lalaurette 2003). But calibration on past events is a major feature of seasonal and monthly forecast systems. This is described in Stockdale et al 1998. When making forecasts to a few days ahead, there are plenty of cases on which to test the model forecasts. When making forecasts to 6 months ahead, there are very few cases of events such as El Nino that can be tested in real-time. To evaluate such a system, one has to go back in time and to make hindcasts from as far back as one can reasonably go. As mentioned earlier, there are insufficient ocean observations to make ocean analyses directly. But by using the forcing fields from the atmospheric reanalyses one can produce ocean initial conditions back say 15 or 20 years. There is a further reason for using these past hindcasts. All models have errors. For forecasts out to a few days, these have largely been ignored though realization that the model would benefit from calibration on past events is growing. For monthly and seasonal forecasts, the error is sufficiently large that it can not be ignored. To first order it is estimated from the past hindcasts and this information is used to calibrate the real-time forecasts. The atmospheric reanalyses are therefore very important in enabling ocean reanalyses to be performed. It seems that the ERA40 reanalyses are considerably better than the earlier ERA-15 reanalyses in that

the ocean reanalyses using ERA-40 match the independent data set of sea-level from satellite altimeters such as TOPEX and Poseidon more closely and lead to better forecasts.

In fig 5, we show the zonal wind stress anomalies in the equatorial Atlantic averaged between 5S to 5N from both ERA15 and ERA40 as a function of time from 1987 to 2002 (ERA15 ended in 1993 so from then on we use the operational equivalent). Potential improvements in the assimilation system will therefore be present in the post 1992 era for ERA15. In contrast, ERA40 used the same 3d-var system throughout. One can see that the winds are substantially different between the two products and even post 2000 the differences are large even though both are using good (though not the same) assimilation methods and models. It is also clear that there were major differences in 1996 for reasons unknown, but possibly the ERA15/OPS product is better than ERA40 in this case.

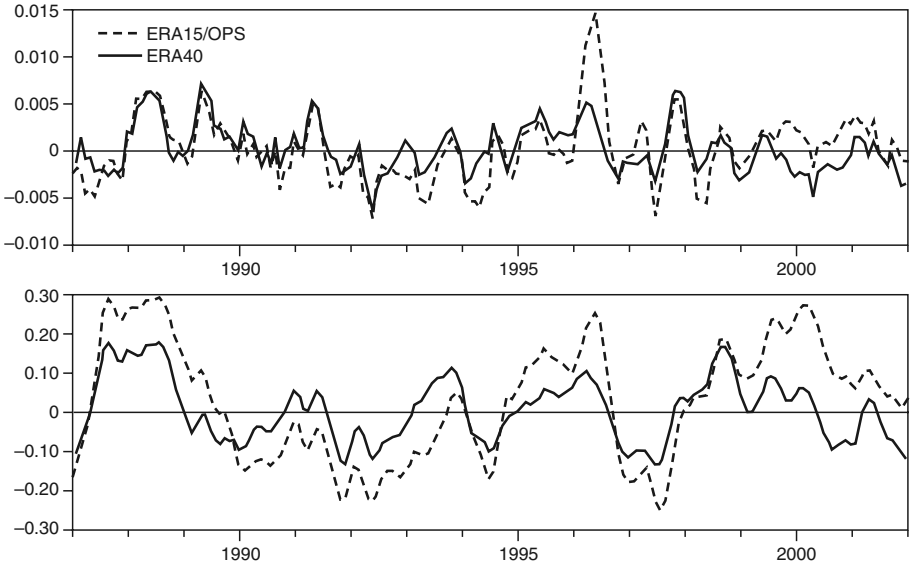

Figure 5. Upper panel: Plot of the wind stress anomalies in the equatorial Atlantic averaged from 5S to 5N. Two different reanalysis products are shown: ERA-40, solid line and ERA-15 dashed line. These wind fields differ considerably throughout the period but the differences are especially large in 1996 for reasons that are unknown. Lower panel shows the average temperature anomalies in the upper 300m for the same region. The ocean acts as a filter and integrator of noise. Thus the signal is redder than that of the wind. The differences are large throughout the integration, and not just in 1996.

What effect do these different winds have on the ocean? The lower panel shows the effect on the temperature averaged vertically over the top 300m and between 5S and 5N in the equatorial Atlantic. One can see that the effect is substantial; in fact the difference in the curves is nearly as large as the interannual signal itself. The ocean can act as an integrator of noise and in fact the signal is quite large in the late 80's early 90's even though this is

not the time of greatest wind error. Data assimilation acts to reduce these differences (not shown). Further illustration of the extent to which data assimilation can act to reduce the impact of wind error is shown in Vialard et al 2005.

4. Weaknesses in the ocean assimilation strategy

The hypothesis underlying the assimilation strategy is that the system is unbiased. This hypothesis is definitely not true in the case of ocean data assimilation. This can be seen by evaluating the mean increment applied in the assimilation. Fig 6 shows this for a section along the equator in the upper 400m. The upper panel shows the mean increment in temperature, averaged over eight years. Far from this being zero, one can see that the assimilation acts to warm the ocean in the west and to cool it in the east Pacific i.e. to strengthen the gradient along the equator. There is a strong systematic effect in the Atlantic too but not much in the Indian ocean (which might just reflect the fact that there are few observations in the Indian ocean).

If one makes the velocity correction mentioned earlier, this bias is reduced somewhat. Including a salinity correction also acts to reduce the bias. Nonetheless, regardless of these changes the bias remains substantial. It could result from error in the wind, in the ocean model physics, in the way that momentum is transferred from atmosphere to ocean, or in the assimilation system itself. Whatever the reason, the assimilation system will operate at reduced efficiency since it is not designed to deal with bias. In fact most assimilation schemes assume the system is unbiased. Can we somehow adapt the assimilation system to take account of the bias? This has been considered by Dee and Da Silva 1998, Dee and Todling 2000, Vidard et al 2004, and Bell et al 2004. One approach is to correct the pressure gradient as suggested by Bell et al 2004. One might think that it would be better to correct the temperature since that is the field which seems to be biased, but this is not the case. The vertical velocity along the equator is distorted (not shown). It is not possible to show what this field should really look like since it is hard to measure, but a descending circulation in the east Pacific is generated which looks very unlikely. It is thought that this arises from the assimilation cycle itself. (See also Vialard et al 2003, Huddleston et al 2004). Correcting bias in the pressure gradient greatly reduces this spurious circulation and the mean temperature increment (not shown). On the other hand correcting the bias in T actually aggravates the situation and leads to an enhanced spurious vertical circulation, although the T increment is reduced.

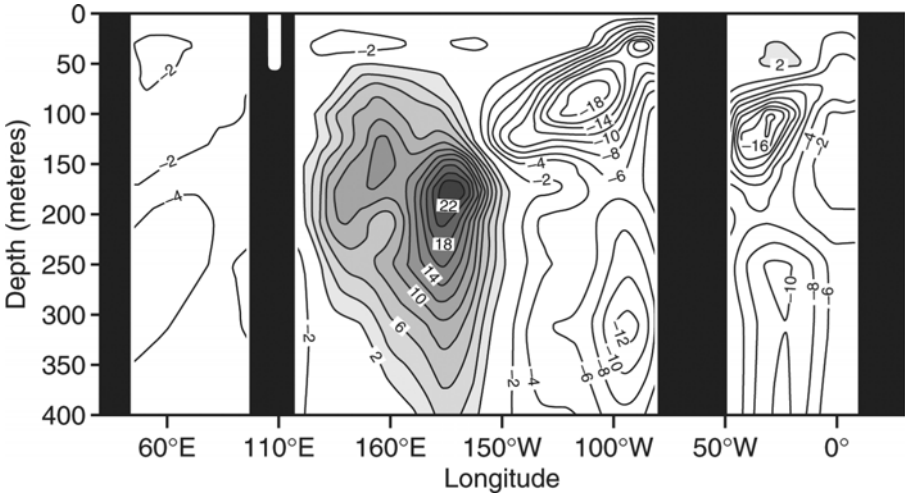

Figure 6. Assimilation increment averaged over an 8-year period in a vertical section along the equator. The left panel is the Indian ocean, the middle the Pacific and the right is the Atlantic. In a well-balanced system the mean increment should be close to zero. In practice most ocean assimilation systems have considerable bias, as shown here. The assimilation is acting to warm the west Pacific but to cool the east. Some of the bias is caused by the assimilation system itself. Warming regions, i.e. where the assimilation increment is positive, are shaded. Contour interval is 2K/year.

5. Ocean observing system experiments

As funding is always limited, the question of the relative merit of each observational system arises. This can be estimated through observation system experiments (OSEs), well known to meteorologists and described in fig 1. See Daley 1992 for general discussion and Anderson et al 1991 and Isaksen and Stoffelen 2000 for discussion relevant to surface wind field from scatterometers. In oceanography, this is a relatively new field, as observations have always been sparse.

The fairest way to evaluate the potential of an observing system is to selectively withdraw components of the system starting with the full array. So we take the system with TAO/PIRATA/TRITON moorings, the XBT network and the ARGO array as the standard, and then remove, TAO/PIRATA/TRITON. The impact of this component of the observing system can then be assessed. The XBTs can be assessed by removing them and comparing with the standard system. This strategy is different to starting with a zero system and then adding components. The approach used here allows for redundancy. So, it might be that a component of the array can be withdrawn at little overall effect because more or less the same information is available from another component of the observing system. Some redundancy is not necessarily bad, however, as it allows an evaluation of the different parts of the array. It should also be remembered that even though there is some redundancy in some average sense, there might be

occasions when this is not so. There are good examples of this in weather forecasting though we are not aware of this in ocean analyses for climate forecasting. Evaluation of observing systems through their impact on forecasts is standard practice in meteorology and should generally be so in oceanography, even if it has not been so in the past.

It is important to realise that results from OSEs are dependent on the analysis system used and on the weight given to the data. The results will be application-dependent. We are interested mainly in seasonal forecasts; this emphasizes the tropics over middle latitudes. For other forecast horizons, or other objectives, different areas may be important and different conclusions might be drawn.

Various experiments to assess the impact of the TAO and XBT networks are given in Vidard et al 2005. To assess the importance of the observing systems on forecasts, 200 six-month forecasts are made, spanning the period 1993-2002 using the ocean analyses as initial conditions. The importance of the equatorial moorings is demonstrated. The exciting new aspect of the observing system is the development of the ARGO network. Can this system replace the XBT network? To consider this, the impact that can be derived from ARGO is discussed. The ARGO experiments cover only the two years 2002-3, but show considerable impact even when the array is only partially developed and optimum techniques for using the data have not been developed (for example, salinity data from the ARGO floats are not yet assimilated).

In addition to OSEs, one can conduct Observing System Simulation Experiments (OSSEs) to assess the potential impact of a proposed observing system, or to assess the relative merits of a given array design. It is difficult to gauge the error characteristics of such observing systems and the experience from meteorology is that results are often too optimistic because errors or difficulties in using the data are underestimated.

6. Summary and conclusions

In this paper we have considered data assimilation methods and issues arising, as related to operational analysis of the atmosphere and ocean. At ECMWF, the primary purpose is, and has been, to provide initial conditions for various forecasts made. These range from forecasts out to 10 days at resolution of ~20km to seasonal forecasts at atmospheric resolution of ~200km. Forecasts for the monthly and seasonal timescales are made with coupled atmosphere-ocean models and so require initial conditions for the coupled system. In real-time, these are made by taking the atmospheric analysis performed for the highest resolution forecast and truncating to the appropriate resolution needed for the forecast. In the case of the ocean, they are obtained by running the ocean analysis system, assimilating all in situ thermal data and relaxing strongly to observed sea surface temperature. The ocean analysis systems for the monthly and seasonal forecast systems are

essentially the same, though they do differ slightly as the analysis for the monthly forecast has to be available within one day of real-time, whereas the seasonal forecasts start 11 days behind real-time. The analysis for the monthly system starts from the last analysis for the seasonal system but then is accelerated to real time using the surface forcing from the atmospheric analyses and assimilating what ocean data are available.

Because model error is significant at longer forecast range, it is necessary to take this into account when preparing forecast products. This is done by calculating the model drift and climate over a calibration period which is currently 15 years but will be longer in future. The forecast products are anomalies relative to the model climate. Hindcasts over the period 1987-2004 are possible, largely because ocean forcing fields are available from the reanalyses ERA-15, ERA-40. These atmospheric reanalyses are not without their problems, as shown in the paper, but mark a significant improvement over previously available products, such as the analyses carried out at the time. It is hoped that atmospheric reanalyses will continue to be performed, perhaps every 5 to 10 years, and covering perhaps the last 50 years or so. These are major undertakings using advances in model development and assimilation techniques and some recovery of old data. They should be viewed as part of the effort to produce monthly, seasonal and even multi-annual forecasts. Ocean reanalyses spanning the period of the atmospheric reanalyses are also made on a routine basis. These are currently done every year or so but as model resolution increases and/or assimilation techniques become more sophisticated, they will be undertaken less frequently. Currently, several ocean groups are active in this area, in contrast to meteorology where only two or three are performed.

At present it is unclear how best to initialize the coupled model. Initialising the atmosphere and ocean separately, as is done currently, may not be the best. One would like to do a more coupled assimilation but this is some way off at present. It is not straight forward because of the disparate time scales of the atmosphere and ocean.

The ocean observing system has advanced rapidly over the last decade. In the tropics it is clear that there are systematic differences between the ocean model state and the observations. This could be because of deficient forcing fields, ocean model physics or the way the surface fluxes are transferred from atmosphere to ocean. ECMWF uses a wave model as part of all its forecast systems, but this model is not fully tied into the ocean as yet. Most other groups using coupled models do not even include a wave model and do not pass the fluxes through the wave field. The assimilation system itself may contribute to the bias increments noted here. Finally, we are just beginning to evaluate the observing system and full use is not yet made of all the data available. For example, salinity from the ARGO floats is only just beginning to be used (Haines et al 2005). The combination of an expanding ocean observing system, new strategies for assimilation, different

techniques for initializing coupled models combined with better models suggest a busy time ahead.

Acknowledgements

The authors thank the members of the ECMWF analysis and medium range and seasonal forecast groups for constructive discussions, and Graeme Kelly and Federico Grazzini for providing figs 1 and 2 respectively. This work was supported in part by ENACT EU-contract EVK2-2001-00077.

References

Anderson D.L.T., A. Hollingsworth, S. Uppala, and P. Worceshyn, 1991: A study of the use of scatterometer data in the ECMWF Operational Analysis - Forecast model. Part II Data impact. J. Geophys. Res., 96, C2, 2635-2649.

Andersson E., and H. Jarvinen, 1999: Variational quality control. Quart. J. Roy. Met. Soc., 125, 697-722.

Alves O., M.A. Balmaseda, D.L.T. Anderson, and T. Stockdale, 2004: Sensitivity of dynamical seasonal forecasts to ocean initial conditions. Quart. J. Roy. Met. Soc., 130, 647-668.

Balmaseda M.A., 2003: Ocean data assimilation for seasonal forecasts. In Recent developments in data assimilation for atmosphere and ocean. Seminar Proceedings ECMWF.

Bell M., M.J. Martin, and N.K.Nichols, 2004: Assimilation of data into an ocean model with systematic errors near the equator. Quart. J. Roy. Met. Soc., 130, 873-893.

Burgers G., M.A. Balmaseda, F. Vossepoel, G.J van Oldenburgh, and P.J. van Leeuwen, 2002: Balanced ocean data assimilation near the equator. J. Phys. Ocean., 32, 2509-2519.

Courtier P., J-N Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4DVAR, using an incremental approach. Quart. J. Roy. Met. Soc., 120, 1367-1388.

Daley R., 1991: *Atmospheric Data Analysis*, Cambridge University Press, 457pp

Dee D.P., and A. Da Silva, 1998: Data assimilation in the presence of forecast bias. Quart. J. Roy. Met. Soc., 124, A, 269-295.

Dee D.P., and R Todling, 2000: Data assimilation in the presence of forecast bias: the GEOS moisture analysis. Mon. Wea. Rev., 128, 3268-3282.

Derber J., and F. Bouttier, 1998: A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus 51A, 195-221.

Fisher M., 2005: On the equivalence between Kalman Smoothing and weak-constraint four-dimensional variational assimilation. Quart. J. Roy. Met. Soc., to appear. See also ECMWF Tech Memo 447 available on line.

Haines K., J. Blower, J-P Drecourt, C. Liu, A. Vidard, I. Astin, and X. Zhou, 2005: Salinity assimilation using S(T) relationships. Part 1: theory. Mon Wea. Rev., to appear.

Huddleston M.R., M.J. Bell, M.J. Martin, and N.K.Nichols, 2004: Assessment of wind stress errors using bias-corrected ocean data assimilation. Quart. J. Roy. Met. Soc., 130, 853-871.

Isaksen L., and A.C.M. Stoffelen, 2000: ERS scatterometer wind data impact on ECMWF's tropical cyclone forecasts. IEEE Trans. Geosci. Remote Sens., 38, 4, 1,885-1,892.

Kelly G., 2004: The proceedings of the third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction (Alpabach, Austria, 9-12 March 2004).

Lalaurette F., 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. Quart J. Roy. Met. Soc., 129, 3037-3057.

Palmer T., F.J. Doblas-Reyes, R. Hagedorn, and 22 others, 2004: Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). BAMS, 85, 853-872.

Smith N.R., J.E. Blomley, and G. Meyers, 1991: A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. Prog. Oceanogr., 28, 219-256.

Stockdale T., D.L.T. Anderson, M.A. Balmaseda, and O. Alves, 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. Nature 392, 370-373.

Troccoli A., M. Balmaseda, J. Segschneider, J. Vialard, D. Anderson, K. Haines, T. Stockdale, F. Vitart, and A Fox, 2002: Salinity adjustments in the presence of temperature data assimilation. Mon. Wea. Rev., 130, 89-102.

Uppala S.M., P. Kallberg, A.J. Simmons, and forty-three others, 2005: The ERA-40 reanalysis. Submitted. See also ECMWF Newsletter 101. Available from ECMWF or from the web: http://www.ecmwf.int/publications/newsletters/

Vialard J., A. Weaver, D.L.T. Anderson, and P. Delecluse, 2003: Three and four dimensional variational data assimilation with a general circulation model of the tropical Pacific ocean. Part 2: Physical validation. Mon. Wea. Rev., 131, 1379-1395.

Vialard J., F. Vitart, M.A. Balmaseda, T. Stockdale, and D.L.T. Anderson, 2005: An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. Mon. Wea. Rev., in press.

Vidard A., A. Piacentini, and F.-X. Le Dimet, 2004: Variational data analysis with control of the forecast bias. Tellus A, 56, 177-188.

Vidard A., D.L.T. Anderson, and M.A. Balmaseda, 2005: Impact of ocean observation systems on seasonal forecasts. ECMWF Techical Memorandum 460.

Weaver A., J. Vialard, and D.L.T. Anderson, 2003: Three and four dimensional variational data assimilation with a general circulation model of the tropical Pacific ocean. Part 1: Formulation, internal diagnostics and consistency checks. Mon. Wea. Rev., 131, 1379-1395.

Chapter 14

MERCATOR OCEAN GLOBAL TO REGIONAL OCEAN MONITORING AND FORECASTING

Pierre Bahurel and the MERCATOR Project Team

MERCATOR OCEAN, Ramonville St Agne, France

Abstract : The MERCATOR OCEAN monitoring and forecasting system has been routinely operated in Toulouse in near-real-time since early 2001. MERCATOR OCEAN service is aiming at providing estimates of the ocean circulation and thermodynamics at high resolution at the global scale. Products are already used by more than 150 referenced users from various communities: public bodies such as met services and agencies dealing with the ocean and its environment, as well as private bodies that are directly linked with the customers operating in the marine environment.

The system is based on high resolution ocean general circulation models (OGCM), real-time processing of remotely sensed and in situ observations, and data assimilation techniques. Its has been regularly upgraded, expanding the geographical coverage from regional to global ocean, improving models and assimilation schemes, adding new data and building new products. Three prototypes of the MERCATOR system are currently running: one global coarse resolution (2°) configuration, one in the north and equatorial Atlantic at medium resolution (1/3°), and one with high resolution (1/15°) in the north Atlantic and Mediterranean. The goal is to build a high resolution O(1/12°) global unique system by the end of 2008. This service is one component of GODAE in Europe, and is one of the key components of the GMES/MERSEA European integrated project. A brief overview of MERCATOR OCEAN, the project and the systems, some recent upgrades and some examples of application using MERCATOR inputs are presented.

Keywords : MERCATOR, operational oceanography, ocean monitoring and forecasting, data assimilation, scientific assessment, ocean services.

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 381-395.

1. MERCATOR OCEAN, an assimilation center for ocean monitoring and forecasting

MERCATOR is the French ocean monitoring and forecasting center. Objectives are to:

- simulate the global ocean with a primitive-equation high resolution model, assimilating satellite and in situ data, to provide hindcasts and near-real time nowcasts and forecasts of the global ocean circulation,
- be operated on an operational mode (ie routine and near-real-time) to provide continuous and well-assessed global/regional ocean monitoring and forecasting information
- through a new ocean service, serving (1) Institutional Operational applications; (2) Research; (3) Private sector Operational Recreational and Commercial applications and (4) Environment Policy Makers end-user needs,

The project was launched in 1995 by the six major French agencies involved in oceanography (namely: Centre National d'Etudes Spatiales (CNES), Centre National de Recherche Scientifique (CNRS), Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Institut de Recherche pour le Développement (IRD), Météo-France, Service Hydrographique et Océanographique de la Marine (SHOM), with involvement of their subsidiaries CLS and CERFACS). The project is lead today by the MERCATOR OCEAN public company, created in 2002 to develop this joint operational capacity for global high resolution ocean monitoring and forecasting, with commitments to prepare transition to an operational centre.

The MERCATOR OCEAN monitoring and forecasting center is in Ramonville St Agne near Toulouse (France), with a team of around 50 people gathering the R&D, Integration, Assessment, Operation & Services requested skills.

Since 2001, the team has been providing weekly ocean bulletins without any service interruption to a wide range of users, and went through 3 major releases of the forecasting system. Real-time outputs, as well as validation reports, are available at http://www.MERCATOR.eu.org.

MERCATOR is a key partner of the international Global Ocean Data Assimilation Experiment "GODAE" and the European MERSEA project for operational oceanography.

2. The MERCATOR OCEAN monitoring and forecasting system

The MERCATOR system provides a full 3D depiction of the ocean dynamics and thermohaline circulation (T, S, currents, mixed layer depth,), with a priority given to high resolution (eddy resolving) scales. Information

is available on a near-real-time and routine basis, by providing weekly Near-Real-Time Analysis and 2-week Forecasts; and on a Reanalysis mode, with data assimilation.

Three prototypes of the MERCATOR system are currently running: one global low resolution (2°) configuration, one in the north and equatorial Atlantic at medium resolution (1/3°) and one with high resolution (1/15°) in the north Atlantic and Mediterranean. An upgrade in the resolution will soon be achieved for the global configuration from 2 to 1/4°

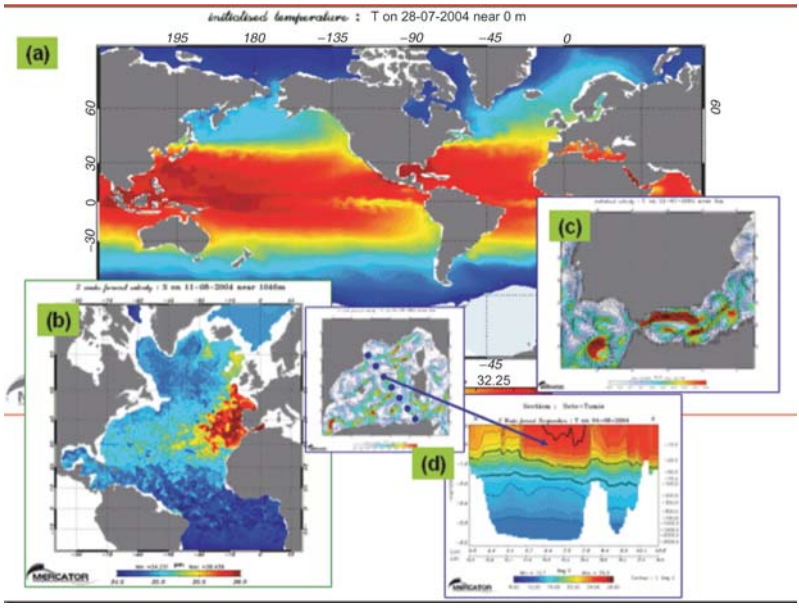

Figure 1. Examples of MERCATOR OCEAN system outputs. (a) Global Ocean (2° model) Sea Surface Temperature field, Real-Time Analysis 28 July 2004 computed 28 July 2004; (b) North & Tropical Atlantic (1/3° model) 1000 m depth Salinity, 2-week forecast 11 August 2004 computed 28 July 2004 ;(c) Gibraltar straight (5-7 km model, ~1/16°) surface currents; Real-Time Analysis 21 July 2004 computed 21 July 2004 ; (d) Mediterranean sea(5-7 km model, 1/16°) Temperature vertical section between Sète and Tunis (and surface current map) 2-week forecast 4 August 2004 computed 21 July 2004.

2.1 Input data

MERCATOR relies on existing data assembly centres to collect, process and validates its input real time and delayed data mode. Input data for MERCATOR include several in situ observations and remote sensing ones. Data are used for several applications: forcing, assimilation and model validation. The *Table 1* lists the current datasets used.

DATA	CENTER	RETRIEVAL FREQUENCY	FORCING	ASSIMILATION	VALIDATION
***In situ* data** : high resolution XBT and CTD, low resolution TESAC and BATHY, ARGO profiling floats, moorings (TAO/PIRATA/TRITON/), drifters, ...	**Coriolis Data Center**	weekly		**X**	**X**
Altimetry : Jason-1, GFO, Envisat, Topex-Poseidon	**SSALTO/DUACS**	weekly		**X**	
Reynolds **Sea Surface Temperature**	**NOAA**	weekly	**X**	**X**	
High Resolution SAF Ocean&Ice Atlantic **Sea Surface Temperature** (10 km, daily product)	**Eumetsat/ Météo-France**	daily			**X**
Climatologic **Sea Surface Salinity**	*Reynaud & al. (1998)*		**X**	**X**	
6-hour analyses and predictions of winds, heat fluxes, Evaporation-Precipitation, cloudiness, air surface temperature, air surface humidity, surface wind	**ECMWF for operational forcing**	weekly	**X**		
Monthly Climatologic **Runoffs**	**Unesco database** *J. D. Milliman and R. H. Meade, 1983* *G. L. Russell and J. R. Miller, 1990* *F. Van D Leeden & al*		**X**		
Real-time scatterometry **winds**	**Cersat**				**X**
Mean Sea Surface Height combining gravity/*in situ* data	*M.H. Rio & al. (2004)*	Regularly upgraded		**X**	
High resolution daily **sea ice concentration** (12 km) and **drift** (60 km). Real-time.	**Cersat** *R. Ezraty and J.-F. Piollé (2004)* *R. Ezraty & al. (2004)*				**X**

Table 1. Input MERCATOR data used for assimilation, forcing and model assessment.

2.2 Model

MERCATOR uses the OPA-NEMO primitive equation ocean code developed at LODyC, Paris [*Madec et al.*, 1998]. The four MERCATOR configurations have a lot of common points like using the rigid lid assumption, a vertical z-coordinates and a turbulent kinetic energy mixing parameterisation (1.5 closure scheme). The bathymetry is processed from the Smith et Sandwell, data base completed in the Antarctic region, the

initialisation temperature and salinity fields come from Levitus Climatology [*Levitus and Boyer*, 1994] completed with the Reynaud Climatology *[Reynaud et al.*, 1998] in the Atlantic and the Medatlas climatology [*MEDAR/MEDATLAS*, 2002] in the Mediterranean Sea. At this time the surface forcing function uses daily stress of wind, evaporation, precipitation, net heat and solar fluxes provided by the European Center for Medium-range Weather Forecast (ECMWF) analyses and forecast. The surface forcing includes a retroaction term in the net heat flux, based on the difference between the model SST and the weekly Reynolds Sea Surface Temperature [Reynolds], with the constant value of $40 W/m^2$ [*Barnier et al.*, 1995]. A relaxation term to the Sea Surface Salinity from climatological data is also added to the E-P flux with the constant value of 5.10^{-6} $m.s^{-1}$. The next version of all the configurations will using Bulk Formulae which allow a better coherence between ocean surface and atmospheric fields and also is the good way to forced the sea ice model. The main rivers are represented by an input of fresh water at the river mouth given by the climatological monthly data base from UNESCO [*UNESCO*, 1996]. A fresh water flux is also added along Antarctica to simulate the melting of the continental ice.

Two configurations are MERCATOR target configurations for the GODAE intensive phase, so that to cover global ocean with an eddy-permitting resolution, and North Atlantic and Mediterranean Sea basins with an eddy-resolving resolution:

- A high resolution basin configuration (5 to 7 km horizontal resolution, 43 vertical levels from 6 m at the surface to respectively 200 m and 300 m at the bottom of the Mediterranean Sea and the Atlantic) covering North Atlantic from 9°N to 70°N and Mediterranean sea [*Siefridt et al.*, 2002]; this configuration focuses on mesoscale processes [*Drillet et al.*, 2004] and links with coastal modelling in European seas. The first version of this configuration used a bilaplacien operator for the horizontal viscosity and diffusivity, a free slip lateral boundary condition and climatological buffer zone at the south and north boundaries of the domain. Several improvements will be soon used in the operational version of this model like laplacien isopycnal diffusivity, a partial slip lateral boundary condition, a bottom boundary layer and a south open boundary allowing a coupling between global and regional configuration. The next version of the high resolution Atlantic model will design the future global high resolution model (1/12°) with a free surface, a partial step vertical coordinate, the atmospheric bulk formulae and a sea ice model.

- A middle resolution global configuration (1/4° horizontal resolution, 46 vertical levels from 6 m at the surface to 250 m at the bottom) covering global ocean; this configuration aims at providing the best ocean state estimates for global ocean analysis and

boundary conditions for regional models worldwide. This configuration use the same parameterisation and physics than the high resolution model describe upper and the improvements (free surface, partial step vertical coordinate, atmospheric bulk formulae and sea ice model) will be implemented .

Two lighter configurations are also implemented in real-time by MERCATOR for demonstration and testing of new algorithms:

- A middle resolution basin configuration developed during the clipper project [*Treguier et al.*, 2001] (1/3° horizontal resolution, 43 vertical levels from 12 m at the surface to 200 m at the bottom) covering North and Tropical Atlantic.
- A low resolution global configuration (2° horizontal resolution, 30 vertical levels) covering global ocean [*Madec and Imbard*, 1996].

In September 2004, the two basin configurations (middle and high resolution) as well as the low resolution global ocean one are operational; the ¼° global ocean model is under development and will integrate the operational chain in the coming months.

2.3 Assimilation method

An ocean monitoring and forecasting system is based on two integrated components: the remotely sensed (e.g. SST, altimetric data) and in situ (e.g. temperature and salinity profiles) observations and the thermodynamical ocean model that are combined to give the best possible description of the real ocean. The way to optimally combine the information given by each system component is called “data assimilation”. This optimal combination (i.e. data assimilation) is achieved by taking advantage both from the information contained in the observations of the real ocean and from the constraints imposed by the ocean model physics. Data assimilation allows for instance eddy permitting ocean models to have meso-scale structures that are in phase with what can be observed and to provide a description of the ocean closer to the reality.

MERCATOR assimilation deals with altimeter sea level anomaly, sea surface temperature and temperature and salinity in situ profiles data into its basin and global scale models. Assimilation is considered for routine near-real-time nowcasts and forecasts issues, but also long term reanalysis products.

MERCATOR is developing a suite of assimilation tools (called “SAM” for MERCATOR Assimilation System) of increasing complexity, from sequential to variational method: the first release, SAM-1, is based on optimal interpolation ; the second release, SAM-2, will consider a Singular Extended Evolutive Kalman (SEEK) filtering analysis method ; and the third

one, SAM-3, will consider advanced methods such as 4D variational method.

The SAM suite uses the PALM coupler [Lagarde et al., 2001] which performs explicit communications between different heterogeneous units such as between model and assimilation operator units, and simplifies or makes easier the transition from one SAM version to the other.

In September 2004, SAM1 is used on operational real time basis since early days of year 2001, SAM2 is under development and will be integrated in the operational chain, and SAM3 is studied on a research mode.

2.3.1 SAM-1

The SAM-1 is based on a Reduced Order Optimal Interpolation method (De Mey and Benkiran, 2002). Two different assimilation versions of this system are used in the MERCATOR prototypes: the SAM1v1 and SAM1v2 techniques.

- Version 1 (SAM1v1) is based on the Cooper and Haines (1996) lifting-lowering of isopycnal approach and allows assimilating SLA data only. The SAM1v1 algorithm starts by calculating a global sea level anomaly increment from innovations (i.e. model-data differences) of along tracks sea level anomaly. It is the reduced order optimal interpolation tool which calculates this increment $\delta\eta$. Based on the statistics for the 3 last months of the ocean model simulation, $\delta\eta$ is partitioned into a baroclinic and a barotropic contribution. The baroclinic part is used to build the temperature increment and the salinity increment by vertically shifting the isopycnals for each water column of the model. Geostrophic velocities are diagnostically adjusted to the new mass field. The barotropic contribution to sea level is used to build a barotropic current stream function increment from which one can deduce a barotropic velocity increment. More details can be found in Ferry et al. (2005).

- Version 2 (SAM1v2) is MERCATOR's first multivariate assimilation method that allows assimilating simultaneously temperature and salinity profiles, SST and SLA data. This system uses fully multivariate 1D Empirical Orthogonal Functions (EOFs of T(z), S(z) and the barotropic stream function) to perform a ROOI (Benkiran et al., 2005). This assimilation system works as follows: first, the differences between the observed and forecast SLA, SST data and in situ measurements are computed over a one week model integration. These differences are then analyzed using a fully multivariate Optimal Interpolation (OI). It is worth noting that the SST product assimilated is the daily Reynolds SST at the analysis time (note that it is already used to correct the flux through a

restoring term). The model equivalent to this SST is the temperature at the first level of the model (6m for the 1/3° North Atlantic system). The Reynolds product is used with a non-Gaussian error so as not to damp the model's meso-scale features but rather to correct locally the model's large drifts. As a consequence, only innovations larger than 2σ (σ~0.6°C r.m.s) are taken into account in the analysis. The model state is updated by the sum of the contribution of each selected EOF to the gain multiplied by the innovation. This scheme is fully multivariate, since the covariances between the errors of different variables are taken into account in the EOFs calculation. The effect is that the assimilation of a single observation has an impact on all the variables in the state vector through the multivariate statistics.

2.3.2 SAM-2

The next generation of multivariate assimilation system, referred as SAM-2, is being developed from Reduced Order Kalman Filters using 3D multivariate modal decomposition of the forecast error covariance (See Brasseur [2005] for a complete description of these methods). The use of 3D modal representation for the error statistics is intended to overcome some of the limitations of SAM1v2 in highly inhomogeneous, anisotropic, and non separable regions of the world ocean such as shallow areas, as well as in the surface layer. SAM-2 includes two versions differing by the formulation of the analysis kernel. The final scheme derived from the SEEK (Singular Evolutive Extended Kalman) algorithm [Pham et al., 1998] (LEGI, Grenoble) has an inversion in the modal space, whereas the intermediate kernel has an inversion in the observation space.

2.3.3 SAM-3

Advanced data assimilation techniques, such as the variational approach, are also investigated in MERCATOR. The variational approach is based on the minimization of the model-observation square misfit over a time period with their given a priori errors. The solution given is the closest trajectory to the observations dynamically consistent with the model equations in a least square mean sense. At given time, the solution is constrained by past and future observations available on the assimilation window. This technique will be used first with the coarse resolution (~2°) global ocean configuration, ORCA2, to assimilate both in-situ and altimetric data.

2.4 Products dissemination

Dissemination of MERCATOR products is made through www and FTP automated tools, both for real-time and archived products. Dissemination

tools guaranty access to multi-year reference simulations (and reanalysis to come) with different extraction tools on one hand and real-time ocean bulletins and numerical fields (recent hindcasts, real-time nowcats and 2-week forecasts) on the other hand. For this purpose, a Live Access Server (LAS) tool is implemented and has been defined to be the standard interface for the GODAE community, and the MERSEA group in Europe.

MERCATOR outputs are freely available for Research and Educational applications; a user identification procedure (login/password) enables a precise knowledge of our user community and two-way dialogue with it.

A User Bureau is taking care of any request and can be contacted at mailto://products@mercator-ocean.fr.

3. System assessment

For operational activities, system assessment is a critical issue to guaranty high-level products to the system end-users; MERCATOR OCEAN organized four assessment loops with different scopes, time scales and team involved to ensure a perfect validation of any output of the system.

Each assessment result is widely communicated (see MERCATOR website) in a quaterly newsletter, so that interested users can access to this information. So far, 13 newsletters have been published: they give a good overview of the MERCATOR system products validation activity.

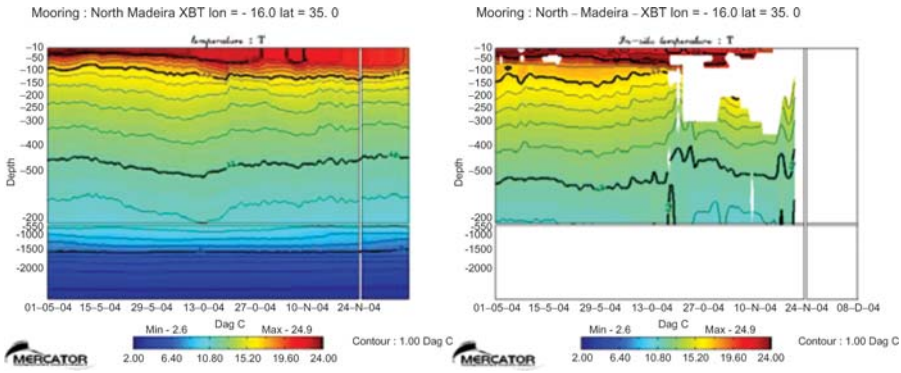

Figure 2. Temperature time series at North Madeira XBT mooring for model (left) and interpolated in situ data (right). See the MERCATOR Newsletter n°6 for details (*http://www.mercator-ocean.fr/html/lettre/presentation_lettre_en.html*).

3.1 Short loop

The first loop of validation is under the responsibility of MERCATOR OCEAN forecasters and considers real-time validation associated to the weekly bulletin. Before allowing any new Ocean Bulletin diffusion, Ocean Forecasters validate the new information processed during Tuesday/Wednesday night. Input data entering the system are first assessed

through the Armor tool (data 3D optimal interpolation tool for global ocean) to detect any strong anomalies in the data sets (combination and cross-correlation). Model outputs are compared to independent in situ data set where available, and simple scores (e.g. Analysis / Forecast comparison) are computed. The systematic comparisons with in situ observations, as illustrated on *Figure 2*, are weekly computed and are available on the MERCATOR Web site, as well as internal diagnostics of assimilation.

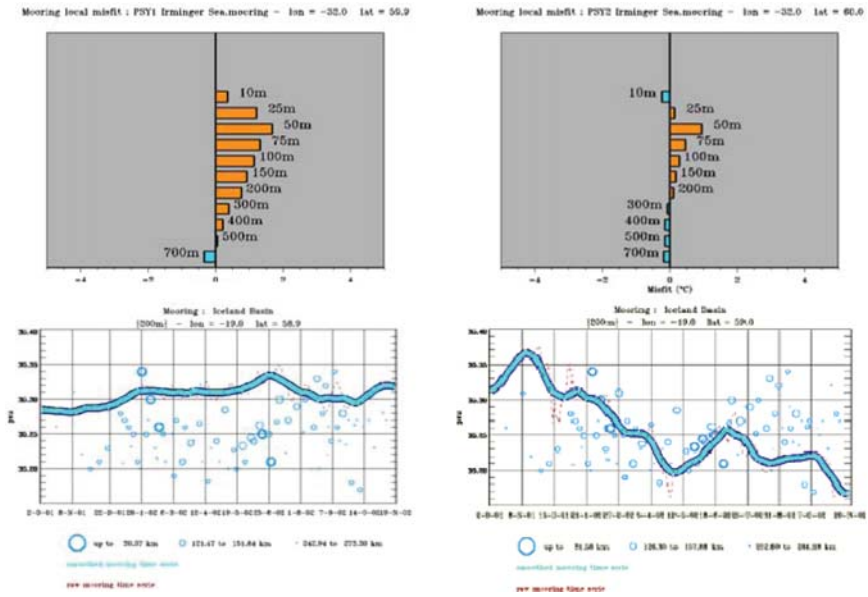

Figure 3. Performance comparison between middle and high resolution operational models (PSY1v1 and PSY2v1). In situ observations (Coriolis database – Ifremer) are used to measure the improvement between the two systems. The upper graphs show the temperature differences between PSY1v1 (left) and PSY2v1 (right) and interpolated observations for several depths and for the Irminger Sea mooring (32W-60N). A positive value indicates a colder model than observations. The lower curves represent the salinity PSY1v1 (left) and PSY2v1 (right) time series at 200 meters depth and for the Iceland mooring (19W-58.9N). The blue lines are 30-days-smoothed time series. Circles represent in situ observations located around the mooring point. The bigger the circle is, the nearer the observation is from the model mooring point. Both analyses show a significant improvement from PSY1v1 to PSY2v1.

3.2 R&D loop

The second loop involves the R&D teams inside MERCATOR and its main objective is to assess precisely impacts of any R&D improvements in the operational system. Change in the model parameterization, upgrades of the assimilation systems, integration of a new data set in the assimilation system could have minor or major impact and could have positive or negative impact. But in any case, this impact has to be perfectly known to measure the progress made to give to the user all the information needed to

appreciate the impact for its own application. This loop also permits to analyse more precisely the model performances and diagnoses the model ability to represent physical processes. As examples of such analysis, the Figure 3 illustrates the quality jump realised at the time of the transition from middle (PSY1v1) to high (PSY2v1) resolution model.

3.3 Internal metrics and intercomparison plans

This third loop entered recently ocean operational centres assessment activity thanks to the effort undergone in the GMES/EC MERSEA Strand 1 project and the GODAE experiment at international level.

An open collaboration with FOAM (UK), MFS (Italy), TOPAZ (Norway) and HYCOM-Miami (US) teams enabled to define standard "metrics" adopted for the North Atlantic and Mediterranean Seas and to be extended to the global ocean (see the chapter by L. Crosnier in this volume and http://www.mersea.eu.org/html/strand1/intercomparison.html#metrics). This is now a core component of the assessment activity of the MERCATOR OCEAN centre. Systematic intercomparison has already allowed clear improvements in the different systems.

3.4 Scientific and user feedback

The fourth loop considers the feedback of the user community, with a clear focus on the scientific community using MERCATOR outputs in its own research activity. The MERCATOR Science Working Team is composed of around 100 researchers, directly associated to this scientific assessment activity and improvements of MERCATOR algorithms. Coupling with ecosystem models, with coastal models, statistical analysis, etc., any research initiative conducted with MERCATOR outputs conveys direct opportunities to assess quality of the MERCATOR system.

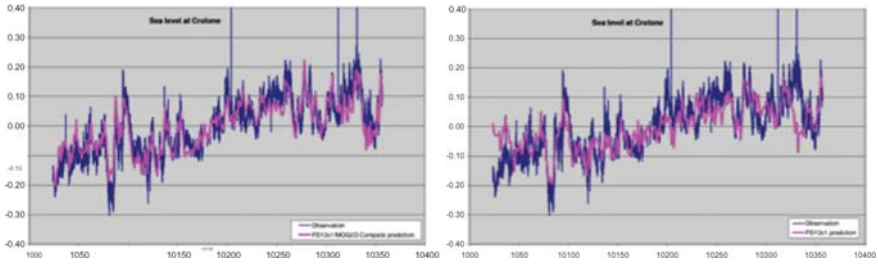

Figure 4. Crotone tide gauge: observed sea level versus MERCATOR (left), versus composite (Mog2D plus low-pass filtered MERCATOR) sea level (right). Series start February the first 2002 (by Florent Lyard and Laurent Roblou, July 2003 – MERCATOR Newsletter 10).

Extension to a wide user community is clearly considered and first feedbacks are already collected. To illustrate this last point, we show in Figure 4 a result of a combination of two dynamical models. The first one,

MERCATOR PSY2v1, is designed to solve the ocean circulation and the second one, Mog2D, is designed to solve the tide and storm surge processes. In the figure, the MERCATOR alone solution and the composite solution are compared with tide gauge measurements (for details, see F. Lyard and L. Roblou, July 2003 – MERCATOR Newsletter n°10).

4. Serving ocean services

Developing this new generation of ocean service, able to provide a fruitful operational, 4D space & time consistent and accurate "general ocean information" to specialized ocean services, is the key objective of MERCATOR operational oceanography.

After more than 3 years of continuous operations, ie more than 180 ocean bulletins for ocean forecasters, and more than 150 referenced users in continuous interaction, the MERCATOR team has been studying user feedbacks in a wide range of application sectors where MERCATOR OCEAN inputs were.

Four categories are considered: (1) institutional operational applications; (2) research; (3) private sector operational recreational and commercial applications and (4) environment policy makers.

Under category 1, MERCATOR has been involved in various experiments concerning: Oil Spill drifts experiments (Météo-France, Met.No, CEDRE), Navy operations (SHOM), Ocean inputs for Seasonal Forecasting system (Météo-France) and Education (schools, user training sessions, individual requests,...). Under category 2, MERCATOR has been requested to provide boundary conditions to coastal models, ocean inputs for biogeochemical models and seasonal forecasting systems, and involved in various Research Sea campaign (IRD, Ifremer, CNRS, IFM Kiel,...). MERCATOR has been serving under category 3 commercial activities (offshore and fisheries) and many recreational marine activities (sailing races, rowing races,...), and a growing range of activities enter today category 4 amongst with: assessment on observation network (satellite & in situ) for decision makers, monitoring and expertise on extreme ocean climate events (2003 hot summer event, Bay of Biscay ocean synthesis bulletin) or new indicators for ocean pollution risk.

Figure 5 is an illustration of MERCATOR role in serving institutional operational applications: here oil spill fate activity conducted by Météo-France. The impact of MERCATOR OCEAN currents in Météo-France's pollution dispersion model (called MOTHY) in assessing the consequences of the Prestige shipwreck were studied with great attention. As a preliminary test case to assess MERCATOR impact on his operational oilspill model, P.Daniel (Météo-France) used MERCATOR 103 meters depth currents to force his model with large-scale time/space information and showed the usefulness of such ocean currents forcings for long simulations.

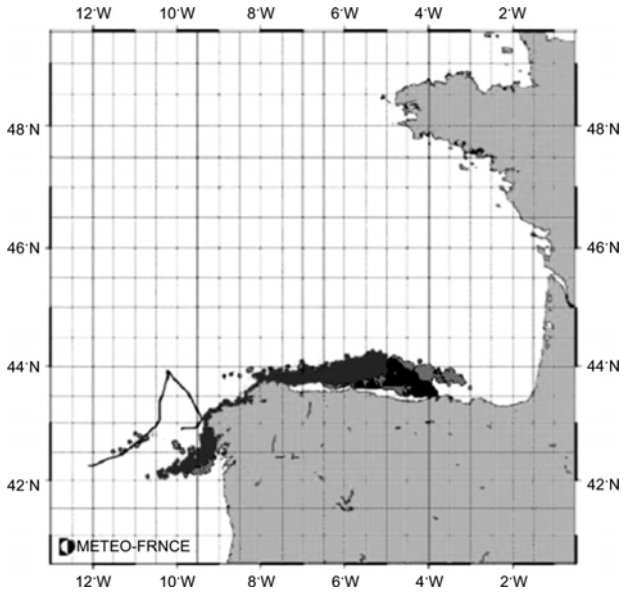

Figure 5. Impact of MERCATOR currents inputs on operational oil drift forecast. Courtesy P.Daniel (Météo-France). Oil spill position (05/12/2002 analysis) forecasted by Météo-France Mothy model alone (dark grey),with MERCATOR inputs (light grey) compared to observations (black).

5. Conclusion

The MERCATOR OCEAN assimilation centre has been running in Toulouse a pre-operational ocean monitoring and forecasting activity for more than 3 years, covering today European basins at 5-7 km, global ocean at 2° and coming to global ¼°; the team developed a strong experience in this matter, and is one of the core team of European GMES program actions to build an Integrated capacity at European level for global ocean monitoring and forecasting.

The MERCATOR OCEAN forecasters team has progressively developed an ocean service for services capacity, serving today more than 150 referenced users, from different application sectors (institutional to private sectors, operational to research, ocean maps to assessment), and bring to them ocean expertise and assimilation capacity.

This ocean monitoring and forecasting is now mature enough to consider the challenge of high resolution and global ocean coverage raised by strong user demands and the transition to a strong operational activity.

But this activity is also fully dependant on real-time, continuous and accurate ocean observation networks, and associated actions for in situ (Argo, ...) and satellite (satellite altimetry, ...) sustained network are critically required.

Acknowledgements

In memory of Christian Le Provost. Many thanks to the MERCATOR project team for the work presented here, with special thanks to N. Verbrugge (CLS), Y. Drillet (Cerfacs), N. Ferry and V. Toumazou (MERCATOR OCEAN) for their contribution to this paper.

References

Barnier, B., L. Siefridt, and P. Marchesiello, 1995: Thermal forcing for a global ocean circulation model using a three-year climatology of ECMWF analyses, J. Mar. Sys., 6, 363-380.

Benkiran, M., E. Greiner, and E. Dombrowsky, 2005: The Multi variate multi data assimilation in the MERCATOR project. J. Marine Sys., submitted.

Brasseur, P., 2005: Sequential methods based on the Kalman Filter for Ocean Data Assimilation: From theory to practical implementations. This volume.

Cooper, M, and K. Haines, 1996: Data assimilation with water property conservation. J. Geophys. Res., 101, 1059-1077.

De Mey, P., and M. Benkiran, 2002: A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation. In : Ocean Forecasting : Conceptual basis and applications, N. Pinardi and J.D. Woods, Eds, Springer Verlag, Berlin, Heidelberg, New York, 472 pp.

Drillet, Y., R. Bourdalle-Badie, L. Siefridt, and C. Le Provost, 2004: The MEDDIES in the MERCATOR North Atlantic and Mediterranean Sea eddy-resolving model. J. Geophys. Res., in press.

Ezraty R., J.-F. Piollé, L. Kaleschke, and G. Heygster, 2004: Sea-ice concentration and drift in the Central Arctic estimated from Special Sensor Microwave Imager data, User's Manual, Version 1.0, Ref. CERSAT C2-MUT-W-07-IF, CERSAT.

Ezraty R., and J.-F. Piollé, 2004: Sea-ice drift in the Central Arctic combining QuikSCAT and SSM/I sea ice drift data. User's Manual, Version 0.1, Ref. CERSAT C2-MUT-W-05-IF, CERSAT.

Ferry, N., E. Rémy, P. Brasseur, and C. Maes, 2005: The MERCATOR global ocean operational analysis / forecast system: assessment and validation of an 11-year reanalysis. J. Mar. Sys., submitted.

Lagarde Th., A. Piacentini, and O. Thual. 2001: A new representation of data assimilation methods: the PALM flow charting approach, Q.J.R.Meteorol.Soc., 127, 189-207.

Levitus, S., and T.P. Boyer, 1994: World ocean atlas 1994 volume 4 : Temperature, National Ocean and Atmosphere Administration.

Madec G., P. Delecluse,, M. Imbard, and C. Lévy, 1998: OPA 8.1 ocean general circulation model reference manual, Notes du pôle de modélisation IPSL, France, 91 pp. WWW Page http://www.lodyc.jussieu.fr/opa/

Madec, G., and M. Imbard, 1996: A global ocean mesh to overcome the north pole singularity, Clim Dyn, 12, 381-388.

MEDAR/MEDATLAS Group, MEDAR/MEDATLAS 2002 database. Cruise inventory, observed and analysed data of temperature and bio-chemical parameters (4 CD-Roms), 2002.

Milliman J.D. and R.H. Meade, 1983: World-wide delivery of river sediment to the oceans, J. Geology, 91, 1-21.

Pham, D., J. Verron, and M. Roubaud, 1998: A Singular Evolutive Extended Kalman filter for data assimilation in oceanography. J. Mar. Sys., 16, 323-340.

Reynaud, T., P. Le Grand, H. Mercier, and B. Barnier, 1998: A new analysis of hydrographic data in the Atlantic and its application to an inverse modeling study, *International WOCE Newsletter*, *32*, 29-31.

Reynolds, R.W. and T.M. Smith, 1994: Improved global sea surface temperature analyses.

Rio, M.-H., and F. Hernandez, 2004: A Mean Dynamic Topography computed over the world ocean from altimetry, in-situ measurements and a geoid model. J. Climate, 7, 929-948.

Russell G.L. and J.R. Miller, 1990: global river runoff calculated from a global atmospheric general circulation model, J. Hydrology, 117, pp 241-254.

Siefridt, L., Y. Drillet, R. Bourdallé-Badie, K. Béranger, C. Talandier, and E. Greiner, 2002: Mise en oeuvre du modèle MERCATOR à haute résolution sur l'Atlantique nord et la Méditerranée, la lettre trimestrielle de Mercator, n°5, pp 1.

Treguier, A.-M., B. Barnier, A.P. de Miranda, N. Grima, M. Imbard, C. Le Provost, G. Madec, C. Messager, J.-M. Molines, S. Michel, and T. Reynaud, 2001: An eddy permitting model of the Atlantic circulation : evaluating open boundary conditions., J. Geophys. Res, 106, 22115-22129.

UNESCO, 1996: Global River Discharge Data Base. Vol. I: Africa, Vol. II Asia and Vol. III: Europe, UNESCO.

Van Der Leeden F., F.L. Troise, and D.K. Todd, 1990: The water encyclopedia, Second edition, Lewis publishers.

Weatherly J.W., and J. E. Walsh, 1996: The effects of precipitation and river runoff in a coupled ice-ocean model of Arctic Clim. Dyn., 12, 785-798.

Chapter 15

THE FORECASTING OCEAN ASSIMILATION MODEL (FOAM) SYSTEM

Michael J. Bell, Rosa Barciela, Adrian Hines, Matt Martin, Alistair Sellar, and David Storkey

Met Office, Exeter, UK

Abstract: We present a detailed technical description of the present FOAM system and discuss some representative examples of the scientific investigations we undertake to track-down problems within the system and to understand the importance ("impact") of the various inputs to it. We also provide an historical perspective on the development of the system and the changing demands for it, and describe the way in which we are adapting to meet these demands.

Keywords: Operational ocean forecasting, data assimilation, assessments.

1. Introduction

The Forecasting Ocean Assimilation Model (FOAM) is a system for assimilating oceanographic measurements into a coupled dynamical model of the deep ocean and sea-ice. It is used on a routine daily basis to make forecasts out to five days ahead representing/resolving the ocean's mesoscale structure in selected regions. The system has been developed with funding from the Royal Navy and is used to support their operations. It also provides boundary conditions for a shelf-seas forecasting system operated by the Met Office. We aim to demonstrate in the near future that its analyses and forecasts of ocean currents are sufficiently skilful to be useful for search and rescue, oil spill drift prediction, and the deep-ocean oil and gas industry. We are also exploring the application of the system to monitoring of open ocean ecosystems and air-sea CO_2 fluxes and management of fisheries and are likely to explore its application to short-range, coupled, atmosphere-ocean forecasts.

The second section of this chapter provides an overview of the FOAM system and a technical summary of its inputs, dynamical model and assimilation methods as they stood in the operational suite in August 2004. The third section attempts to give some insight into the intellectual

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 397-411.

challenges inherent in developing these systems. It describes some representative examples of scientific trouble-shooting and some investigations of the impact of new observations and changes to assimilation methods on the performance of the system. The final section attempts to give an historical perspective. It summarises first the 20-year history of the FOAM project and then the changing world context in which it has been developing. Finally four major changes in the direction of the FOAM project are described and related to this changing context.

2. Description of the FOAM system

2.1 Overview and present configurations

The FOAM system produces 5-day forecasts of three-dimensional ocean temperatures, salinities and currents and sea-ice properties on a routine daily basis. It assimilates temperature profile data, surface height data from satellite-borne altimeters and satellite and in situ surface temperature data and is driven by 6-hourly surface fluxes from the Met Office's Numerical Weather Prediction (NWP) system. High resolution model configurations are nested inside the global configuration. Statistics on the differences between the model forecasts and observations are routinely produced and reanalyses can be generated from 1997 onwards.

The FOAM configurations that are presently run on a routine daily basis within the operational suite at the Met Office cover the globe with a 1^{o} grid; the Atlantic and Arctic Oceans and the Indian Ocean with 35 km grids; and the North Atlantic, the Mediterranean Sea and the Arabian Sea with 12 km grids. An Antarctic configuration with a 27 km grid is also run on a daily basis and is to be transferred into the operational suite in the first half of 2005. All of these configurations have 20 vertical levels. The global, Atlantic and Arctic, and N Atlantic configurations are illustrated in figure 1.

2.2 Inputs

Six-hourly full-resolution **surface-flux** fields from the global forecasts by the Met Office's NWP system to 5-days ahead are currently used to drive all the FOAM configurations (in future fluxes from limited-area forecasts will drive some configurations). The flux fields used are the wind stress (vector with two components), wind mixing energy, penetrating heat flux, non-penetrating heat flux and precipitation minus evaporation. The NWP system calculates fluxes over sea-ice and open water ("leads") separately and combines them using sea-ice concentration analyses generated by NCEP. The surface temperature and salinity fields are also weakly relaxed towards the monthly Levitus et al. (1998) climatologies.

The global configuration is also driven by climatological monthly **river inflow** data from the Global Runoff Data Centre (GRDC) with the outflows from the largest 20 rivers adjusted to accord with Baumgartner & Reichel (1975).

Temperature and **salinity profile** data are assimilated at all depths (see section 3.1). In the operational system the data are obtained from BATHY, TESAC and BUOY messages distributed by the Global Telecommunications System (GTS). These message formats are used to report expendable bathythermograph (XBT) data reported by Voluntary Observing Ships (VOS), and data from the Argo profiling floats and TAO/Triton equatorial moorings respectively. Quality control checks on these data include track, stability, background and buddy checks (Ingleby & Huddleston 2004).

Altimeter data from the Jason-1, Envisat and Geosat Follow-On (GFO) satellites are assimilated in all but the global configuration using products supplied twice a week by Collecte Localisation Spatiale (CLS) in Toulouse.

In situ **surface temperature** data from ships and drifting and moored buoys are assimilated. At present only advanced high resolution radiometer (AVHRR) data on a coarse grid (2.5^{o} spacing) are assimilated. All these data are distributed by the Global Telecommunications System (GTS). We will upgrade to using GODAE High Resolution SST (GHRSST) satellite data products when they become available.

Sea-ice concentration fields supplied by the Canadian Met Centre (CMC) on a daily basis are also assimilated. These fields are based on SSM/I (special sensor microwave imager) data processed using the York/AES algorithm (Ramseier et al. 1988).

2.3 Dynamical model

Storkey (2004) provides an excellent summary of the formulation of the physical ocean and sea-ice models used by FOAM in July 2004. The ocean model code, which originated from the Bryan-Cox code (Bryan 1969, Cox 1984), is developed jointly with groups in the Hadley Centre who use it for climate prediction. The FOAM formulation is quite close to that used by HadCM3 (Gordon et al. 2000).

Various bathymetries (Smith & Sandwell 1997, DBDB2 and GEBCO) have been used in building the present configurations. The bathymetry after interpolation to the model's grid is smoothed twice using a 1-4-1 filter. Grid-scale holes are filled to avoid an instability (Pacanowski & Griffies 1999) which appears to be associated witih the B-grid staggering of variables and the depth and width of important channels are adjusted using Thompson (1996) as a reference. At open boundaries of nested models the bathymetry in the relaxation zone (see below) is reset to be as similar as possible to the model providing its boundary data. Tests of the impact of code to achieve a

smoother bathymetry using partial bottom cells (Pacanowski & Gnanadesikan 1998) are in progress.

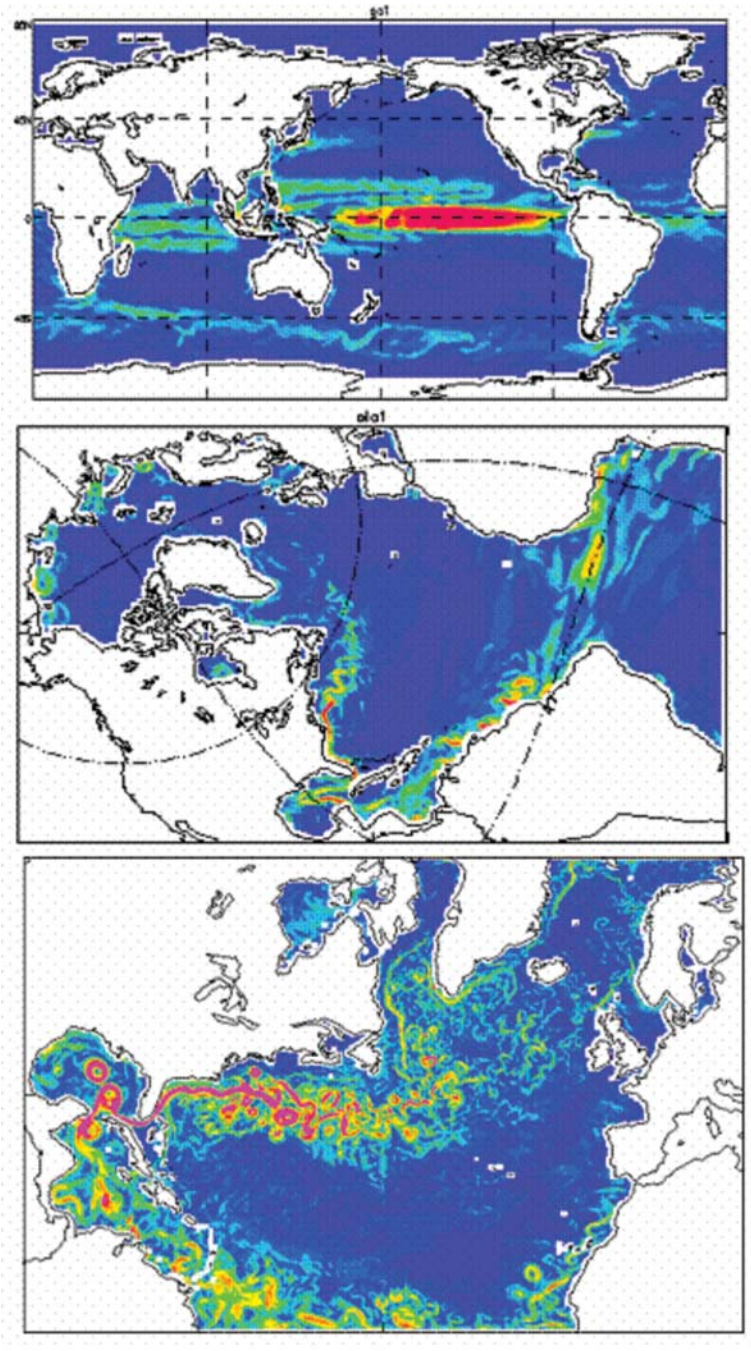

Figure 1. Surface current speeds in the FOAM global, Atlantic and Arctic, and North Atlantic configurations.

The limited area models use the Flow Relaxation Scheme (FRS) (Davies 1983, McDonald 1997) as boundary conditions for all prognostic variables (including temperature, salinity and horizontal velocity components). This relaxes the model fields in the inner model towards those in the outer model over a relaxation zone typically 4-8 gridpoints wide, the strength of the relaxation increasing as the outer edge of the inner model is approached. (When FOAM transitions to a free-surface we will transition to Flather conditions for the external modes.) Most of the limited area models use a rotated latitude-longitude grid to achieve for given resolution the largest minimum grid-spacing Δx. This allows a longer time-step to be used. At present resolutions it has been found that the maximum model timestep Δt is limited by the CFL criterion $2(c+u)\Delta t < \Delta x$ in which c is the speed of fastest internal waves (about 3 m/s), u is the fastest advecting velocity in the model and the factor of 2 arises from the use of the leapfrog scheme.

The prognostic equation for horizontal momentum is similar to that in the Bryan-Cox code except that the advection of momentum uses the Webb (1995) scheme and a simple quadratic bottom friction with $C_D = 0.00125$ is used to crudely parametrise tidal mixing. To increase the timestep that can be used the Coriolis term is calculated semi-implicitly in coarser resolution configurations, and the pressure gradient is averaged across timesteps in higher resolution configurations (Brown & Campana 1978). A combination of harmonic and biharmonic viscosities is used to damp gridscale noise and westward migrating eddies. The choice of parameters has a significant impact on the model simulation (Chassignet & Garraffo 2001). The barotropic flow is represented by a streamfunction using the rigid-lid approximation (see Storkey 2004 for details).

The prognostic equation for tracers presently uses a form of third-order upwind advection similar to Holland et al. (1998). A combination of a less diffusive advection scheme and the thickness diffusion scheme of Gent & McWilliams (1990) is being trialled as an alternative. The Griffies et al. (1998) formulation of isopycnal diffusion is employed.

The formulation of vertical mixing is explained by Gordon et al. (2000) and Storkey (2004). Momentum and tracers are mixed using the Pacanowski & Philander (1981) scheme and a simplified form of the Large et al. (1994) scheme. In addition tracers are mixed using a mixed-layer energetics scheme based on Kraus & Turner (1967) and Davis et al. (1981). Convective adjustment of tracers is performed by applying the Roussenov scheme (Roether et al.1994) followed by the Rahmstorf (1993) scheme.

The thermodynamic component of the **sea-ice** model uses the zero-layer model of Semtner (1976) and Hibler's (1979) formulation for leads processes. The dynamic component is based on Bryan et al. (1975): the ice concentration is advected using the top-level ocean currents and smoothed using Laplacian diffusion. The EVP formulation of Hunke & Dukowicz

(1997) and ice thickness distribution scheme of Lipscomb et al. (2001) are being trialled.

2.4 Assimilation methods

Data assimilation is based on a new version of the analysis correction (a/c) scheme. The a/c scheme was originally devised by Lorenc et al. (1991) and implemented for FOAM by Bell et al. (2000a). The new version (Bell et al. 2003) provides a sub-optimal approximation to a variant of 4D variational assimilation. Analysis steps are performed once per day. Each observation makes its full impact on the model on the day it arrives and on subsequent days is taken into account by giving additional weight to the model at the observation's location. Each analysis step consists of a number of iterations. On each iteration the observations are separated into groups which are easily related (thermal profiles, saline profiles, surface temperature, surface height). For each group of observations (e.g. the temperature profile data), increments are calculated first for the directly related model variables (e.g. the temperature fields). These increment fields are then used to calculate increments for less directly related model variables (e.g. the velocity fields) using hydrostatic and geostrophic balance relationships, water property conservation or statistical relationships. These balancing increments make the analysis multivariate. Increments are also made to the observations (Bratseth 1986) so that the iterations converge towards the statistically optimal analysis. The univariate components of the model error covariance are specified as the sum of two 3D error covariances, one describing the ocean mesoscale, the other large scales including atmospheric synoptic scales (Martin et al. 2002). These and the observation error covariances are estimated from statistics of observation minus model values obtained from hindcast assimilations. Altimeter data are assimilated by displacement of isopycnal surfaces (an extension of the Cooper & Haines 1996 scheme). A pressure correction technique (Bell et al. 2004) is employed to improve the dynamical balance near the equator (see section 3.1) and analyses performed with large correlation scales are used to attempt to remove large-scale biases in the AVHRR surface temperature data.

3. Trouble-shooting, assessments of impact and developments

3.1 Trouble-shooting

Bell et al. (2004) report a serious problem encountered assimilating thermal profile data into the global FOAM configuration in the equatorial Pacific region where the TAO moorings provide good observational

coverage. Figure 2 shows the annual mean temperature increments applied by the assimilation scheme along the equator. The units are ^{o}C per month. Just below 100 m depth, between 150^{o}W and 120^{o}W, over the course of a year the assimilation scheme is decreasing the temperature by as much as 30 ^{o}C! Since the change in temperature over the course of a year is a small fraction of this, it is clear that in this integration the ocean model must be increasing the temperature at this location by a similar amount. Since internal sources and sinks of heat are relatively small, the change in temperature is due to advection. Diagnostics of the vertical velocities confirm that they are much stronger at and below 100 m depth when the model is assimilating data than when it is not assimilating data. Bell et al. (2004) propose a dynamical explanation for these spurious over-turning circulations and suggest that the problem arises from inaccurate parametrisation of the downward mixing of momentum input by the wind stresses acting on the ocean surface. Assuming that these inaccuracies result in a slowly varying bias in the momentum equation they propose a scheme to estimate the bias using the observational data. Huddleston et al. (2004) show that the scheme is quite effective in reducing the vertical circulations and the net heat input by the assimilation scheme and improves the zonal currents along the equator in integrations using a number of wind stress products.

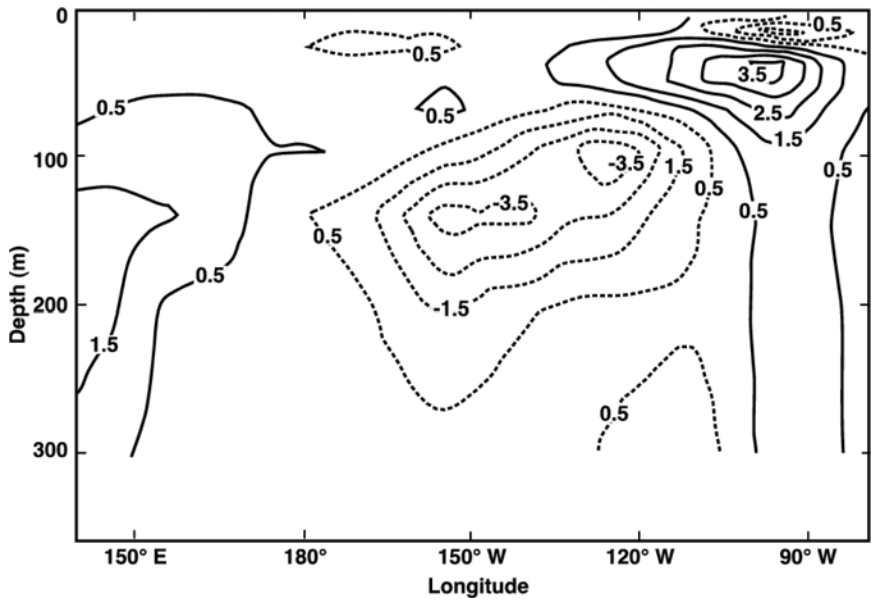

Figure 2. Annual mean potential temperature increments (^{o}C per month) for a cross-section along the equator between 140^{o}E and 90^{o}W. Negative contours are dashed.

It has been found that the assimilation of temperature and salinity data below 1000 metres depth can have major impacts on the barotropic flow and the meridional overturning in the FOAM system. The version of the FOAM system implemented in 1997 deliberately excluded observational data below 1000 metres depth because of the deleterious impact of occasional deep observations in the Gulf Stream region (Bell 1994). With the advent of the Argo system it is highly desirable to assimilate both temperature and salinity data at all depths (see next sub-section). It is important for the quality control of the Argo data to detect suspect observations, particularly those with depth independent offsets.

Small-scale noise in ocean forecasts is undesirable for several reasons. Storkey (2004) describes how biases in model integrations can develop from small-scale noise when upwind vertical advection schemes are employed. He found that for these advection schemes reducing the horizontal viscosity below a certain limit led to significant biases within the thermocline.

3.2 Assessments of impact

In order to prioritise developments one would like to be able to predict what the impact of a given development is likely to be. Should one give highest priority to the use of additional observational data, to improvements to the assimilation scheme or to the dynamical model?

Figure 3 shows the impact on verification statistics of assimilating Argo profile data into the FOAM 1^{o} model. The statistics are root mean square differences between profile observations and model fields valid the day before the observations (i.e. fields in which the observations have not been assimilated). All the model integrations covered the period January to May 2003, were forced by 6-hourly NWP fluxes and were started from the operational analysis for 1st January 2003. A "control" integration assimilated no data; a second integration assimilated only salinity profile data from Argo; a third assimilated only temperature profile data from Argo and the final integration assimilated both temperature and salinity profile data. No other data were assimilated and an early version of the new assimilation scheme was used.

It is clear that the Argo data have a major beneficial impact on the model fields but several points of detail are worth noting. First much of the impact in the deeper temperature and salinity fields arises from corrections to biases in the model fields which had accumulated in the operational model since 1995 when the integrations were initialised. Second the assimilation scheme does not attempt to conserve T/S properties (Troccoli & Haines 1999). It is likely that assimilation of temperature data only using T/S conservation ideas could produce better salinity analyses. Third, assimilation of salinity data only degrades the temperature fields compared to the control and similarly assimilation of temperature data only degrades the salinity fields.

Assimilation of temperature and salinity data produces markedly better salinity statistics than assimilation of just salinity data and slightly better temperature statistics than assimilation of just temperature data. These results may be explained by the impact of the assimilation on the advecting velocity field but this hypothesis has not been verified in detail.

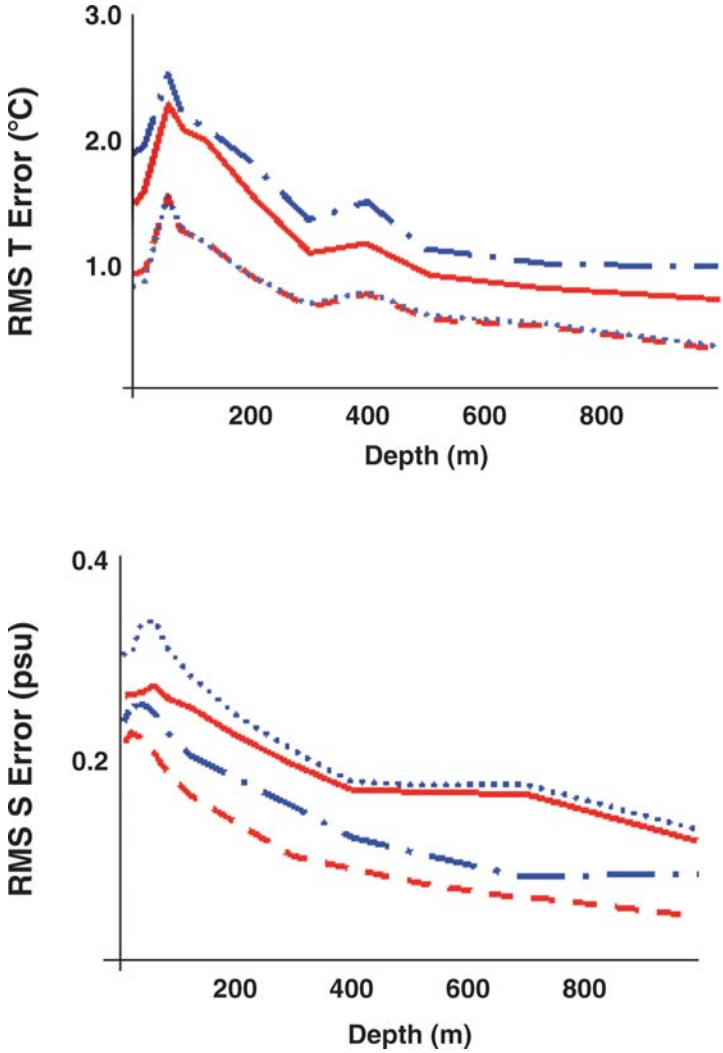

Figure 3. Impact of assimilating Argo data on FOAM global model: no assimilation - full line; temperature data only – dotted; salinity data only - dot dash; temperature and salinity data – dashed.

The impact of the representation of the background error covariance on the effectiveness of the data assimilation is also of considerable interest as differences between most assimilation schemes can be interpreted as differences in the representation of the background error covariance. Figure

4 gives a simple indication of the importance of the background error covariance for assimilation of SST data into the global FOAM configuration.

The r.m.s. difference between the model and AVHRR satellite observations just before they are assimilated is again taken as an indication of the effectiveness of the assimilation. The original FOAM scheme with a 300 km correlation scale produced the rms scores indicated by triangles and a version of the scheme using two iterations, one with a 500 km correlation scale and one with a 100 km correlation scale produced the scores indicated by crosses. A number of other tests (including ones in which biases in the satellite data are estimated) indicate that it is important to cover a range of scales in analysis of SST data but that the results are not particularly sensitive to the details of how this is done.

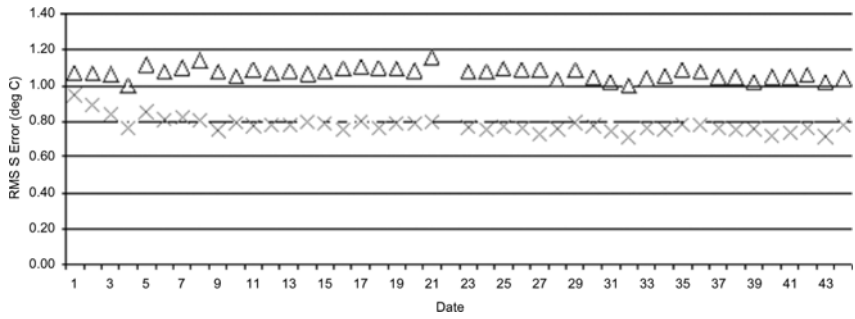

Figure 4. Time-series of global average root mean square differences between model fields and satellite SST observations prior to their assimilation. Assimilation used one scale of 300 km for triangles and two scales of 500 km and 100 km for crosses. The ordinate is the day number. Day 1 is 22 January 2001. See text for more details.

4. Historical perspective

4.1 A history of the development of the FOAM system

The first proposals for the development of a FOAM system to make daily forecasts of the three-dimensional temperature, salinity and current structure of the ocean and of sea-ice were written in 1985 by Howard Cattle & Adrian Gill. Developments started in 1988 and by the end of 1994 a global configuration of FOAM on a 1^{o} grid with 20 vertical levels, driven by six-hourly surface fluxes and assimilating temperature profile data, was being run on a daily basis (Alves et al. 1995). This system was adapted for introduction into the operational suite used for numerical weather prediction in November 1996. Its assimilation scheme was described in some detail by Bell et al. (2000a). A simple sea-ice assimilation scheme and improvements to the representation of atmospheric surface fluxes over water partially

covered by sea-ice were introduced in July 1999 (Bell et al. 2000b). Significant amendments to the advection and diffusion within the ocean model were introduced in September 2000. A configuration of FOAM covering the North Atlantic and Arctic oceans with a $1/3^{o}$ grid, nested inside the global model, was introduced into the operational suite in January 2001 and assimilation of altimeter data based on the Cooper-Haines scheme (Cooper & Haines 1996 and Forbes 1996) was introduced in November 2001. The capability to implement quickly high resolution configurations for new areas was also demonstrated in 2001. Daily pre-operational running of a North Atlantic configuration with a $1/9^{o}$ grid and distribution of the output via a Live Access Server first for GODAE and then for the MERSEA intercomparison project started in April 2002. The major changes to the assimilation scheme described in section 2.3 were introduced into the operational suite in November 2003 and July 2004.

4.2 Changing priorities

The FOAM project was devised towards the end of the Cold War and the development of FOAM has been supported largely by Navy funding. By 1995 the Navy's requirement for high-resolution open ocean forecasts had weakened and the main value of deep ocean forecasts was seen to be in the provision of boundary conditions for forecasts of shelf-seas and coastal waters. The Met Office started to collaborate with Proudman Oceanography Laboratory (POL) to implement a shelf-seas system for the North-West European continental shelf. Configurations of POLCOMS (POL Coastal Ocean Model System) have been run in the operational suite at the Met Office since June 2000.

Proposals for a Global Ocean Data Assimilation Experiment (GODAE) and an Argo system of autonomous profiling floats emerged during 1997. Since then the Argo system has revolutionised the sub-surface in situ observing system. GODAE is motivated by the need to demonstrate the value of the oceanographic observational networks; both space-based (e.g. altimeter, scatterometer and surface temperature) and in situ. This is essential to justify their transition from research to operational funding and urgent in view of the extended planning periods and expense of satellite programs. In order to accelerate this demonstration, the International GODAE Steering Team (IGST) has championed open scientific collaboration. This includes open access to the inputs to the systems (e.g. surface fluxes) and forecasts from them, shared documentation of the input and output data sets, development of shared tools for serving of data and products, and detailed intercomparisons of methods and results.

During 2001, the European Commission (EC) and European Space Agency (ESA) started to implement the GMES (Global Monitoring for Environment and Security) program which aims to supply the environmental

information needed to formulate and monitor policies for sustainable management of the environment. In the oceans there is a particular emphasis on coastal regions, management of pollution and the health of biological ecosystems.

Freely distributed "community" ocean models (such as MOM, HYCOM and ROMS) have been developed in the USA for two decades. In the last 5 years, flexible software tools (such as PRISM) for generating complex earth system models from component models (e.g. of the atmosphere, ocean and sea-ice) have started to gain maturity. Within Europe collaborative projects (e.g. EnAct) have started to compare data assimilation methods for ocean models.

4.3 Response to Changing Priorities

The Met Office is adapting its program for ocean forecasting in response to these changing priorities and the improved climate for collaboration and coordination.

First, it is starting to transition all of its ocean modelling activities (including seasonal forecasting and climate prediction) to use the NEMO (Nucleus for European Modelling of the Ocean) code. NEMO will be jointly owned by a consortium (including CNRS, Mercator-Ocean and the Met Office) who undertake to maintain and develop it. It will be freeware, the aim being to encourage a wide range of ocean modellers to use it and to contribute to its development. NEMO will be based on the OPA code and developed for use in shelf and coastal waters in addition to the deep ocean. It will be coupled to other models at the Met Office through a Flexible Unified Model Environment (FLUME) which will build on experience gained within PRISM.

Second, the Met Office is actively engaged in the Mersea (Marine Environment and Security in the European Area) project, which is building the open ocean component of the monitoring and forecasting system required for GMES. Mersea aims to improve the collaboration and coordination between European ocean forecasting systems and the Met Office is actively supporting this aim.

Third, a National Centre for Ocean Forecasting is being established at the Met Office in association with four of the NERC (Natural Environment Research Council) institutes (namely Proudman Oceanography Laboratory (POL), Plymouth Marine Laboratory (POL), Southampton Oceanography Centre (SOC) and Environmental Systems Science Centre (ESSC)).

The fourth initiative aims to strengthen the awareness within relevant UK government departments and the offshore industry of our growing capabilities in ocean forecasting and to work together to assess their needs and to develop the capabilities to meet them. The Ocean Customer Group formed to do this includes government departments with responsibilities for

marine pollution and search and rescue (Maritime Coastguard Agency), water quality, fisheries and coastal flooding (Environment Agency and Dept for Environment, Food and Rural Affairs), environmental impact assessments and offshore wind farms (Dept for Trade and Industry), and representatives of the oil industry.

Together it is hoped that these initiatives will consolidate the UK contribution to operational oceanography, strengthening the scientific input into it, the coordination and collaboration with European colleagues, and enabling it to evolve to meet the UK needs for environmental management.

References

Alves, J. O. S., M. J. Bell, N. P. J. Brooks, A. L. Cooper, S. J. Foreman, R. M. Forbes, and C. G. Sherlock, 1995: Performance review of the prototype FOAM system. Forecasting Research Division Technical Report 159. Available from Met Office, UK.

Baumgartner, A. and E. Reichel, 1975: The world water balance. Mean annual global, continental and maritime precipitation, evaporation and run-off. 179 pp. Elsevier, Amsterdam.

Bell, M. J., 1994: Results from initial experiments with the assimilation of thermal profiles into a dynamical model of the Atlantic Ocean. Forecasting Research Division Technical Report Tech Rep 98. Available from Met Office, UK

Bell, M. J., R. M. Forbes, and A. Hines, 2000a: Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Sys*., **25**, 1-22.

Bell, M. J., T. Allen, J. O. S. Alves, and A. Hines 2000b: The FOAM system and the use of satellite information. Proceedings of SAF Training Workshop, Ocean and Sea Ice. EUMETSAT Hosted by Meteo-France, Perros-Guirec, France 30 November - 2 December 1999.

Bell, M. J., A. Hines, and M. J. Martin, 2003: Variational assimilation evolving individual observations and their error estimates. Ocean Applications Tech Note 32. Available from Met Office, UK.

Bell, M. J., M. J. Martin, and N. K. Nichols, 2004: Assimilation of data into an ocean model with systematic errors near the equator. Quart. J. Roy. Metero. Soc., 130, 853-871.

Bratseth, A. M., 1986: Statistical interpolation by means of successive corrections. *Tellus,* 38A, 439-447.

Brown, J. A., and K. A. Campana, 1978: An economical time-differencing system for numerical weather prediction. Mon. Wea. Rev., 106, 1125-1135.

Bryan, K., 1969: A numerical study of the circulation of the world ocean. J. Comp. Phys., 4, 347-376.

Bryan, K., S. Manabe, and R. C. Pacanowski, 1975: A global ocean-atmosphere climate model. Part II: the ocean circulation. J. Phys. Ocean. 5, 30-46.

Chassignet, E. P. and Garraffo, Z. D., 2001: Viscosity parameterization and the Gulf Stream separation". In "From Stirring to Mixing in a Stratified Ocean". Proceedings 'Aha Huliko'a Hawaiian Winter Workshop. U. of Hawaii. January 15-19, Eds. P. Muller and D. Henderson., 37-41.

Cooper, M., and K. Haines, 1996: Altimetric assimilation with water property conservation. J. Geophys. Res., 101, C1, 1059-1077.

Cox, M. D., 1984: A primitive equation, 3-dimensional model of the ocean. GFDL Ocean Group Tech. Rep. No. 1, Princeton, New Jersey, Geophysical Fluid Dynamics Laboratory.

Davies, H. C., 1983: Limitations of some common lateral boundary schemes used in regional NWP models. Mon. Wea. Rev., 111, 1002-1012.

Davis, R. E., R. de Szoeke and P. Niiler, 1981: Variability in the upper ocean during MILE. Part II. Modelling the mixed layer response. Deep-Sea Res., 28A, 1453-1475.

Forbes, R. M., 1996: Initial results from experiments assimilating satellite altimeter sea surface height data into a tropical Pacific ocean model. Ocean Applications Technical Note 12. Available from Ocean Applications, Met Office, Bracknell, UK.

Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. J. Phys. Oceanogr., 20**,** 150-155.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B. and Wood, R.A., 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics, 16, 147–168.

Griffies, S. M., A. Gnanadesikan, R. C. Pacanowski, V. D. Larichev, J. K. Dukowicz, and R. D. Smith, 1998: Isoneutral diffusion in a z-coordinate ocean model. *J. Phys. Oceanogr*., 28, 805-830.

Hibler, W. D. III, 1979: A dynamic thermodynamic sea ice model. J. Phys. Oceanog., 9,815-846.

Holland, W. R., J. C. Chow and F. O. Bryan, 1998: Application of a third-order upwind scheme in the NCAR ocean model. J. Climate**,** 11, 1487-1493.

Huddleston M. R., M. J. Bell, M. J. Martin, and N. K. Nichols, 2004: Assessment of wind stress errors using bias corrected ocean data assimilation. Quart. J. Roy. Meterol. Soc., 130, 853-871.

Hunke, E. C., and J. K. Dukowicz,, 1997: An elastic-viscous-plastic model for sea ice dynamics. J. Phys. Oceanogr., 27, 1849-1867.

Ingleby, B., and M. R. Huddleston, 2004: Quality control of ocean profiles – historical and real-time data. To be submitted to J. Mar. Sys.

Kraus, E. B., and Turner, J., 1967: A one dimensional model of the seasonal thermocline. II The general theory and its consequences. Tellus, 19, 98-106.

Large, W. G., J. C. McWilliams, and S. C. Doney, 1994: Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev. Geophys. 32, 4, 363-403.

Levitus, S., T. P. Boyer, M. E. Conkright, T. O'Brien, J. Antonov, C. Stephens, L. Stathoplos, D. Johnson, and R. Gelfeld, 1998: World Ocean Database 1998. Volume 1: Introduction. NOAA Atlas NESDIS 18, 346 pp.

Lipscomb, W. H,. 2001: Remapping the thickness distribution in sea ice models. J. Geophys. Res., 106, C7, 13,989-14,000.

Lorenc, A. C., Bell, R. S. and MacPherson, B., 1991: The Meteorological Office analysis correction data assimilation scheme. Quart. J. Roy. Meteor. Soc., 117, 59-89.

Martin, M. J., M. J. Bell, and A. Hines, 2002: Estimation of three-dimensional error covariance statistics for an ocean assimilation system. Ocean Applications Technical Note 30. Available from Met Office, UK.

McDonald, 1997: Lateral boundary conditions for operational regional forecast models : a review. HIRLAM Tech. Report, 32, HIRLAM-5 Project, c/o Per Unden, SMHI, S-601 76 Norrkoping, Sweden.

Pacanowski R. C., and A. Gnanadesikan, 1998: Transient response in a z-level ocean model that seolves topography with partial ceslls. Mon. Weath. Rev., 126, 3248-3270.

Pacanowski, R. C., and S. M. Griffies, 1999: *MOM Manual 3.0*. Draft version. Geophysical Fluid Dynamics Laboratory/NOAA. [Available from GFDL, Princeton University, Princeton, NJ 08542.]

Pacanowski, R. C., and S. G. H. Philander, 1981: Parametrisation of vertical mixing in numerical models of tropical oceans. J. Phys. Oceanogr., 11, 1,443–1,451.

Rahmstorf, S., 1993: A fast and complete convection scheme for ocean models. Ocean Modelling, 101, 9-11.

Ramseier, R. O, I. G. Runbinstein, and A. F. Davies, 1988: Operational evaluation of special sensor microwave/imager. North York, Ont., York University. Centre for Research in Experimental Space Science. Atmospheric Environment Service.

Roether, W., V. M. Roussenov, and R. Well, 1994: A tracer study of the thermohaline circulation of the eastern Mediterranean. In Malanotte-Rizzoli, P. and Robinson, A. R., (eds) Ocean processes in climate dynamics: global and Mediterranean example. Kluwer Academic Press, Dordrecht, 371-394.

Semtner, A. J. Jr., 1976: A model for the thermodynamic growth of sea ice in numerical investigations of climate. J. Phys. Oceanogr., 6, 379-389.

Smith, W. H. F., and Sandwell, D.T., 1997: Global sea floor topography from satellite altimetry and ship depth soundings. Science, 277, 1956-1962.

Storkey D., 2004: Initial tuning of FOAM high-resolution models. Ocean Applications Tech. Note, 34. Available from Met Office, UK.

Troccoli A., and K. Haines, 1999: Use of temperature-salinity relation in a data assimilation context. J. Atmos. Ocean. Tech., 16, 2011-2025.

Thompson, S. R., 1996: Sills of the global ocean: a compilation, http://www.soc.soton.ac.uk/JRD/OCCAM/sills.html

Webb, D. J., 1995: The vertical advection of momentum in Bryan-Cox-Semtner ocean general circulation models. J. Phys. Oceanog., 25, 3186-3194.

Chapter 16

OCEAN PREDICTION WITH THE HYBRID COORDINATE OCEAN MODEL (HYCOM)

Eric P. Chassignet[1], Harley E. Hurlburt[2], Ole Martin Smedstad[3], George R. Halliwell[1], Patrick J. Hogan[2], Alan J. Wallcraft[2], and Rainer Bleck[4]

[1]RSMAS/MPO, University of Miami, Miami, Florida, USA; [2]Naval Research Laboratory, Stennis Space Center, Mississippi, USA; [3]Planning Systems Inc., Stennis Space Center, Mississippi,USA; [4]Los Alamos National Laboratory, Los Alamos, New Mexico, USA

Abstract: This chapter provides an overview of the effort centered on the HYbrid Coordinate Ocean Model (HYCOM) to develop an eddy-resolving, real-time global and basin-scale ocean prediction system in the context of the Global Ocean Data Assimilation Experiment (GODAE).

Keywords: HYCOM, GODAE, LAS, data assimilation, metrics.

1. Introduction

A broad partnership of institutions[1] is presently collaborating in developing and demonstrating the performance and application of eddy-resolving, real-time global and basin-scale ocean prediction systems using the HYbrid Coordinate Ocean Model (HYCOM). The plan is to transition these systems for operational use by the U.S. Navy at both the Naval Oceanographic Office (NAVOCEANO), Stennis Space Center, MS, and the Fleet Numerical Meteorology and Oceanography Center (FNMOC), Monterey, CA, and by NOAA at the National Centers for Environmental Prediction (NCEP), Washington, D.C. The partnership is also the eddy-resolving global ocean prediction system development effort that is sponsored by the U.S. component of the Global Ocean Data Assimilation Experiment (GODAE). GODAE is a coordinated international effort envisioning *"a global system of observations, communications, modeling, and assimilation that will deliver regular, comprehensive information on the state of the oceans, in a way that will promote and engender wide utility and availability of this resource for maximum benefit to the community"*. Three

[1] U. of Miami, NRL, Los Alamos, NOAA/NCEP, NOAA/AOML, NOAA/PMEL, PSI, FNMOC, NAVOCEANO, SHOM, LEGI, OPeNDAP, U. of North Carolina, Rutgers, U. of South Florida, Fugro-GEOS, ROFFS, Orbimage, Shell, ExxonMobil

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 413-426.

of the GODAE specific objectives are to apply state-of-the-art models and assimilation methods to produce short-range open ocean forecasts, boundary conditions to extend predictability of coastal and regional subsystems, and initial conditions for climate forecast models (GODAE Strategic Plan, 2000). HYCOM development is the result of collaborative efforts among the University of Miami, the Naval Research Laboratory (NRL), and the Los Alamos National Laboratory (LANL), as part of the multi-institutional HYCOM Consortium for Data-Assimilative Ocean Modeling funded by the National Ocean Partnership Program (NOPP) in 1999 to develop and evaluate a data-assimilative hybrid isopycnal-sigma-pressure (generalized) coordinate ocean model (Bleck, 2002; Chassignet et al., 2003; Halliwell, 2004).

Traditional ocean models use a single coordinate type to represent the vertical, but recent model comparison exercises performed in Europe (DYnamics of North Atlantic MOdels - DYNAMO) (Willebrand et al., 2001) and in the U.S. (Data Assimilation and Model Evaluation Experiment - DAMÉE) (Chassignet et al., 2000) have shown that no single vertical coordinate – depth, density, or terrain-following sigma − can by itself be optimal everywhere in the ocean. These and earlier comparison studies (Chassignet et al., 1996; Roberts et al., 1996, Marsh et al., 1996) have shown that the models considered are able to simulate the large-scale characteristics of the oceanic circulation reasonably well, but that the interior water mass distribution and associated thermohaline circulation are strongly influenced by localized processes that are not represented equally by each model's vertical discretization. The choice of the vertical coordinate system is one of the most important aspects of an ocean model's design and practical issues of representation and parameterization are often directly linked to the vertical coordinate choice (Griffies et al., 2000). Currently, there are three main vertical coordinates in use, none of which provides universal utility. Hence, many developers have been motivated to pursue research into hybrid approaches. Isopycnal (density tracking) layers are best in the deep stratified ocean, z-levels (constant fixed depths) are best used to provide high vertical resolution near the surface within the mixed layer, and σ-levels (terrain-following) are often the best choice in shallow coastal regions. HYCOM combines all three approaches and the optimal distribution is chosen at every time step. The model makes a dynamically smooth transition between the coordinate types via the layered continuity equation.

This chapter describes the various components of the HYCOM data assimilative system and is organized as follows: an overview of the main HYCOM characteristics is presented in section 2, the performance of the present near real time Atlantic forecasting system is discussed in section 3 and section 4 provides an outlook.

2. The ocean model

HYCOM is designed to provide a significant improvement over the existing global operational ocean products, since it overcomes design limitations of present systems as well as limitations in vertical resolution. The ultimate goal is a more streamlined system with improved performance and an extended range of applicability (*e.g.*, the present U.S. NAVY systems are seriously limited in shallow water and in handling the transition from deep to shallow water). The generalized coordinate (hybrid) ocean model HYCOM retains many of the characteristics of its predecessor, the isopycnic coordinate model MICOM (Miami Isopycnic Coordinate Model), while allowing coordinate surfaces to locally deviate from isopycnals wherever the latter may fold, outcrop, or generally provide inadequate vertical resolution in portions of the model domain. The freedom to adjust the vertical spacing of the coordinate surfaces in HYCOM simplifies the numerical implementation of several physical processes (mixed layer detrainment, convective adjustment, sea ice modeling,…) without robbing the model of the basic and numerically efficient resolution of the vertical that is characteristic of isopycnic models throughout most of the ocean's volume.

The implementation of the generalized coordinate in HYCOM follows the theoretical foundation set forth in Bleck and Boudra (1981) and Bleck and Benjamin (1993): i.e., each coordinate surface is assigned a reference isopycnal. The model continually checks whether or not grid points lie on their reference isopycnals and, if not, attempts to move them vertically toward the reference position. However, the grid points are not allowed to migrate when this would lead to excessive crowding of coordinate surfaces. Thus, vertical grid points can be geometrically constrained to remain at a fixed depth while being allowed to join and follow their reference isopycnals in adjacent areas (Bleck, 2002). The default configuration in HYCOM is one that is isopycnal in the open stratified ocean, but smoothly reverts to a terrain-following (σ) coordinate in shallow coastal regions and to fixed pressure-level coordinates (hereafter referred to as p) in the surface mixed layer and/or unstratified seas (Figure 1). In doing so, the model combines the advantages of the different types of coordinates in optimally simulating coastal and open-ocean circulation features. It is left to the user to define the coordinate separation constraints that control regional transitions among the three coordinate choices. Figure 1 illustrates the transition that occurs between p/σ and isopycnic (ρ) coordinates in the fall and spring in the upper 400 meters and over the shelf in the East China and Yellow Seas. In the fall, the water column is stratified and can be represented with isopycnals; in the spring, the water column is homogenized over the shelf and is represented by a mixture of p and σ coordinates. A particular advantage of ρ coordinates

is illustrated by the density front formed by the Kuroshio above the peak of the sharp (lip) topography at the shelfbreak in Fig. 1a. Since the lip topography is only a few grid points wide, this topography and the associated front is best represented in ρ coordinates.

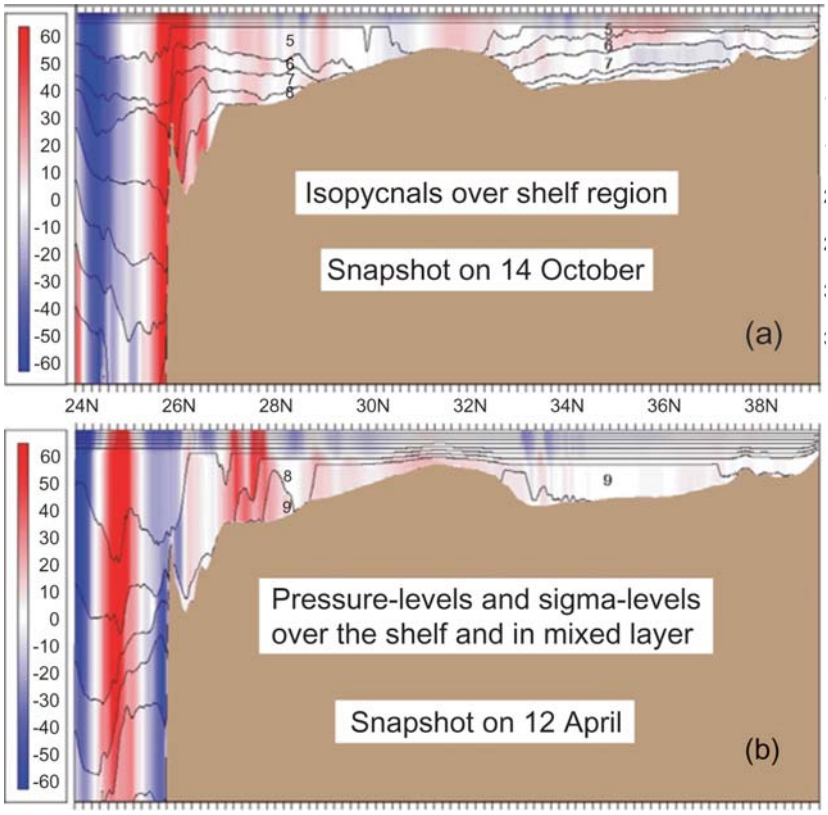

Figure 1. Upper 400 meters north-south velocity cross-section along 124.5°E in the East China and Yellow Seas: (a) Fall; (b) Spring.

The algorithm that maintains the hybrid vertical coordinates is T/S conservative and monotonicity-preserving (i.e., no new T/S extrema during re-gridding). It is referred to as the "grid generator" (Bleck, 2002) and is the final algorithm executed during each model time step. The grid generator relocates vertical interfaces to restore isopycnic conditions in the ocean interior to the greatest extent possible while enforcing the minimum thickness requirements. The minimum thickness is enforced by a "cushion" function (Bleck, 2002) that produces a smooth transition from the isopycnic to the p-domain. The grid generator first attempts to restore the density of a given layer to its isopycnic reference density if necessary. If a layer is less

dense that its isopycnic reference density, the generator attempts to move the bottom interface downward so that the flux of denser water across this interface increases density. If the layer is denser than its isopycnic reference density, the generator attempts to move the upper interface upward to decrease density. In both cases, the generator first calculates the vertical distance that the interface must be relocated so that volume-weighted density of the original plus new water in the layer equals the reference density. Repeated execution of this algorithm at every time step does maintain layer density very close to its reference value as long as a minimum thickness does not have to be maintained. To insure that a permanent p-coordinate domain exists near the surface year round at all model grid points, the uppermost layers are initialized with reference densities smaller than values found anywhere in the model domain. The minimum thickness constraint is not enforced at the bottom in the open ocean, permitting model layers to collapse to zero thickness there as in MICOM.

The capability of assigning additional coordinate surfaces to the HYCOM mixed layer allows the option of implementing sophisticated vertical mixing turbulence closure schemes [see Halliwell (2004) for a review]. The full set of vertical mixing options contained in the latest version of HYCOM (http://hycom.rsmas.miami.edu) includes five primary vertical mixing submodels, of which three are vertically "continuous" models and two are predominantly or totally bulk models. The three continuous models, which govern vertical mixing throughout the water column, are: K-Profile Parameterization of Large et al. (1994) (KPP), the level 2.5 turbulence closure of Mellor and Yamada (1982) (MY), and the Goddard Institute for Space Studies (GISS) level 2 turbulence closure of Canuto et al. (2001,2002). The other two are the quasi-bulk dynamical instability submodel of Price et al. (1986) (PWP) and the bulk Kraus-Turner (1967) submodel (KT).

The following procedure is used to implement the three continuous vertical mixing submodels. Velocity components are interpolated to the *p* grid points from their native *u* and *v* points. The one-dimensional submodels are then run at each *p* point to calculate profiles of viscosity coefficients along with T and S diffusion coefficients on model interfaces. The one-dimensional vertical diffusion equation is then solved at each *p* point to mix T, S, and tracer variables, which involves the formulation and solution of a tri-diagonal matrix system using the algorithm provided with the KPP submodel (Large et al., 1994). To mix momentum components, viscosity profiles stored on interfaces at *p* grid points are horizontally interpolated to interfaces at *u* and *v* grid points. Then the vertical diffusion equation is solved on both sets of points. For more details on the implementation of the various mixing schemes, the reader is referred to Halliwell (2004).

3. The North Atlantic prototype ocean prediction system

While HYCOM is a highly sophisticated model, including a large suite of physical processes and incorporating numerical techniques that are optimal for dynamically different regions of the ocean, data assimilation is still essential for ocean prediction a) because many ocean phenomena are due to flow instabilities and thus are not a deterministic response to atmospheric forcing, b) because of errors in the atmospheric forcing, and c) because of ocean model imperfections, including limitations in resolution. One large body of data is obtained remotely from instruments aboard satellites. They provide substantial information about the ocean's space-time variability at the surface, but they are insufficient by themselves for specifying the subsurface variability. Another significant body of data is in the form of vertical profiles from XBTs, CTDs, and profiling floats *(e.g.,* ARGO). While these are too sparse to characterize the horizontal variability, they provide valuable information about the vertical stratification. Even together, these data sets are insufficient to determine the state of the ocean completely, so it is necessary to exploit prior knowledge in the form of statistics determined from past observations as well as our understanding of ocean dynamics. By combining all of these observations through data assimilation into an ocean model it is possible to produce a dynamically consistent depiction of the ocean. It is important that the ocean model component of the forecast system has skill in predicting the ocean features of interest. Then the model can act as an efficient dynamical interpolator of the observations.

Performance of HYCOM in the North and Equatorial Atlantic has been documented by Chassignet et al. (2003) within the framework of the Community Modeling Experiment (CME). The near real time 1/12º (~7 km mid-latitude resolution) HYCOM Atlantic Ocean prediction system (http://hycom.rsmas.miami.edu/ocean_prediction.html) spans from 28^{o}S to 70^{o}N, including the Mediterranean Sea and has been running since July 2002. The vertical resolution consists of 26 hybrid layers, with the top layer typically at its minimum thickness of 3 m (*i.e.,* in fixed coordinate mode to provide near surface values). In coastal waters, there are up to 15 sigma-levels, and the coastline is at the 10 m isobath. The northern and southern boundaries are treated as closed, but are outfitted with 3 ^{o} buffer zones in which temperature, salinity, and pressure are linearly relaxed toward their seasonally varying climatological values. Three-hourly wind and daily thermal forcing (interpolated to three hours) are presently provided by the FNMOC Navy Operational Global Atmospheric Prediction System (NOGAPS) (Rosmond et al., 2002), available from NAVOCEANO and the U.S. GODAE data server in Monterey. The HYCOM prediction system uses

surface wind stress, air temperature, and specific humidity (from dewpoint temperature and sea level pressure) in addition to shortwave and longwave radiation. Surface heat flux is calculated using NOGAPS fields and the Kara et al. (2002) bulk parameterization of latent and sensible heat flux, which uses model SST.

Mostly because of its simplicity, robustness, and low computational costs, operational ocean prediction systems around the world (NLOM, MERCATOR, FOAM, etc.) are presently using Optimal Interpolation (OI) based data assimilation techniques. For the current $1/12^{o}$ Atlantic HYCOM ocean forecasting system, we have adopted a similar approach by selecting an OI technique with Cooper and Haines (1996) for downward projection of SSH from altimetry [see Chassignet et al. (2005) for details]. Real time satellite altimeter data (Geosat-Follow-On (GFO), ENVISAT, and Jason-1) are provided via the Altimeter Data Fusion Center (ADFC) at NAVOCEANO to generate the two-dimensional Modular Ocean Data Assimilation System (MODAS) SSH ($1/4^{\circ}$) analysis (Fox et al., 2002) that is assimilated daily. The MODAS analysis is an OI technique which is using a complex covariance function that includes spatially varying length and time scales as well as propagation terms derived from many years of altimetry (Jacobs et al., 2001). The model sea surface temperature is relaxed to the daily MODAS $1/8^{\circ}$ SST analysis which uses daily Multi-Channel Sea Surface Temperature (MCSST) data derived from the 5-channel Advanced Very High Resolution Radiometers (AVHRR) – globally at 8.8 km resolution and at 2 km in selected regions. The system runs once a week every Wednesday and consists of a 10-day hindcast and a 14-day forecast.

At the present time, evaluation of the model outputs relies on systematic verification of key parameters and computation of statistical indexes by reference to both climatological and real time data, and, in a delayed mode, to quality controlled observations. The accuracy of data assimilative model products is theoretically a non-decreasing function of the amount of data that is assimilated. A degradation caused by assimilation generally indicates inaccurate assumptions in the assimilation scheme. While models can be forced to agree with observations (*e.g.*, by replacing equivalent model fields with data), improvements with respect to independent observations are not trivial. An assessment of model improvement (or lack of degradation) with respect to unassimilated, independent measurements is therefore an effective means of assessing the performance of an assimilation system. Variances of these model-data differences serve as common measures of the estimation accuracy. For the evaluation of flow accuracy and water mass characteristics, we follow the guidelines put forward by the international GODAE metrics group as well as the validation tests commonly used at the operational centers before official transition to operational use. In the

remainder of this section, we outline some of these metrics and provide examples for the HYCOM Atlantic forecasting system.

Large-scale circulation features: These tests evaluate whether the global and basin-scale models correctly place the large-scale features of ocean circulation, such as gyres, strong fronts, and currents. It is indeed necessary to know the oceanic mean SSH over the time period of the altimeter observations before one can assimilate the SSH anomalies determined from satellite altimeter data. Furthermore, at the scales of interest (tens of kilometres), it is also necessary to have the mean of major ocean currents and associated SSH fronts sharply defined. This is not feasible from coarse hydrographic climatologies (~1º horizontal resolution) and from present geoid measurements since the geoid is not yet known accurately on the mesoscale. The approach taken by the HYCOM-based system is to use a model mean generated by a previous 1/12º North Atlantic simulation performed with MICOM (Chassignet and Garraffo, 2001).

Eddy kinetic energy/SSH variability: These tests evaluate whether the models have a realistic level and distribution of energy (mean and variability) at depths where observations are available.

Sea Surface Height (analysis, forecast): Provide an assessment of the models' ability to represent observed sea surface heights.

Sea Surface Temperature (analysis, forecast): These tests evaluate whether the models are producing acceptable nowcasts and forecasts of sea surface temperature. The near real-time system is routinely compared to buoy observations of SST.

Vertical profiles, time series of profiles and vertical cross sections (analysis, forecast): Since the present forecasting system assimilates only surface quantities (SSH, SST), quantitative comparisons of model temperature and salinity to unassimilated profile data from XBTs, CTDs, and ARGO floats, and moored buoys can be used to assess the model's performance in the ocean interior. In Figure 2, model temperature sections are compared to XBT measurements obtained from the Marine Environmental Data Service (MEDS) dataset. A quantitative assessment using the RMS difference between the model and data profiles is shown in Figure 3. With assimilation of surface data only, the Atlantic HYCOM forecasting system has, overall, larger RMS error than climatology or MODAS-3D. MODAS-3D (Fox et al., 2002) uses the statistics of the historical hydrographic data base to downward project the same MODAS SSH anomaly and SST analyses assimilated by HYCOM, indicating superior performance for a data-based method of downward projection than the Cooper and Haines (1996) technique used in HYCOM, at least in this application.

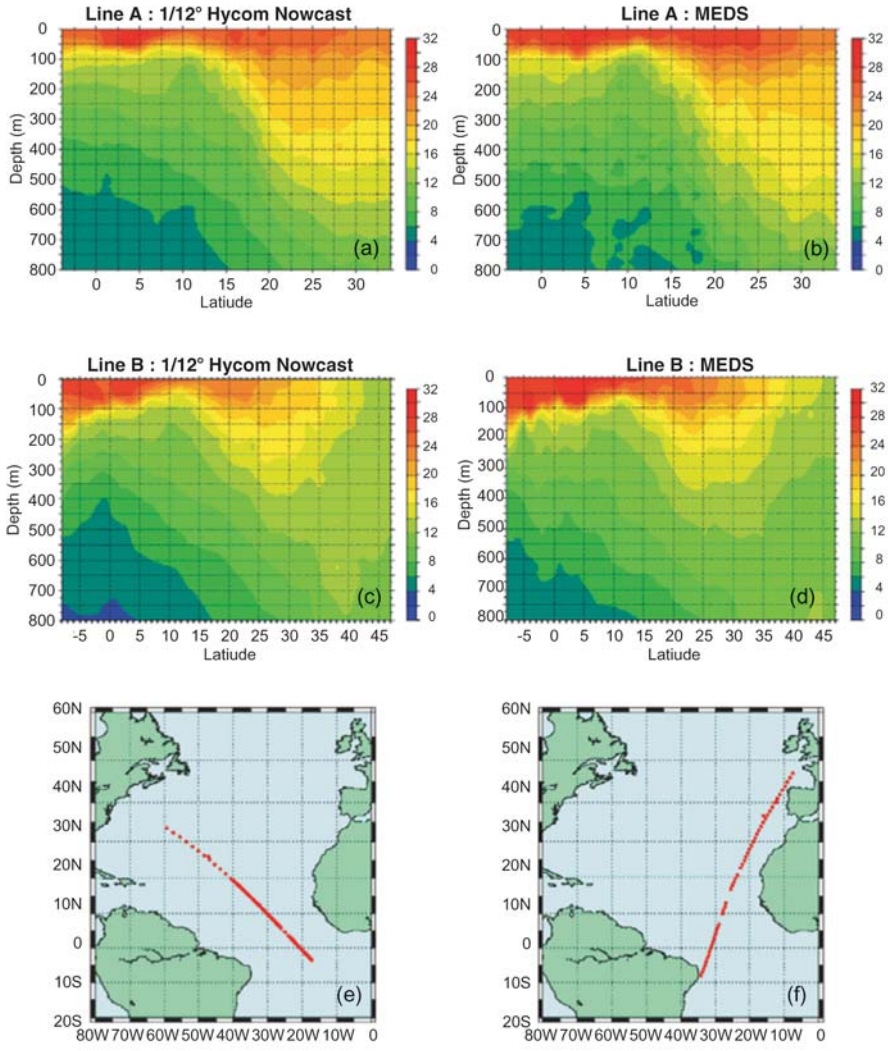

Figure 2. (a) Temperature section along line A from the 1/12°near real-time Atlantic system, (b) corresponding section from the MEDS data, (c) temperature section along line B from the 1/12°near real-time Atlantic system, (d) corresponding section from the MEDS data.

Current cross sections: These tests evaluate model velocity cross-sections through qualitative and quantitative comparisons of biases when data are available. When observations are available, transport time series provide an excellent measure of the model's ability to represent daily to seasonal variability (see example shown in Figure 4 for the Florida Straits).

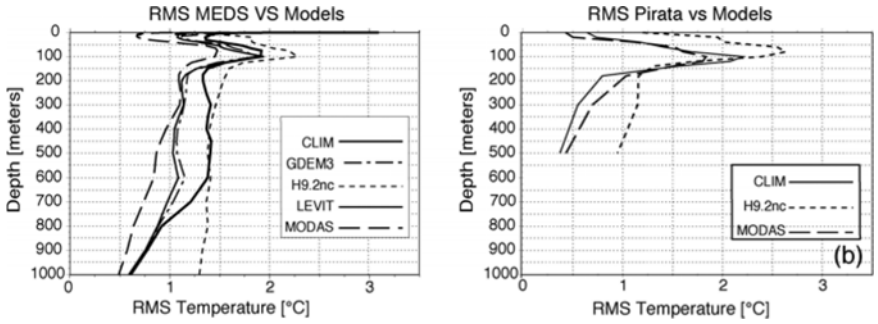

Figure 3. (a) Statistics for the month of February between the 1/12° HYCOM system and available Marine Environmental Data Service (MEDS) profile observations. The RMS difference between the MEDS data, MODAS3D (MODAS), and different climatologies (MODAS (CLIM), Levitus (LEVIT), and the Generalized Digital Environmental Model (GDEM3)) is also shown. (b) Statistics for the month of May between the 1/12° HYCOM system and available PIRATA profile observations. The RMS between the PIRATA data, MODAS3D, and MODAS climatology (CLIM) is also shown.

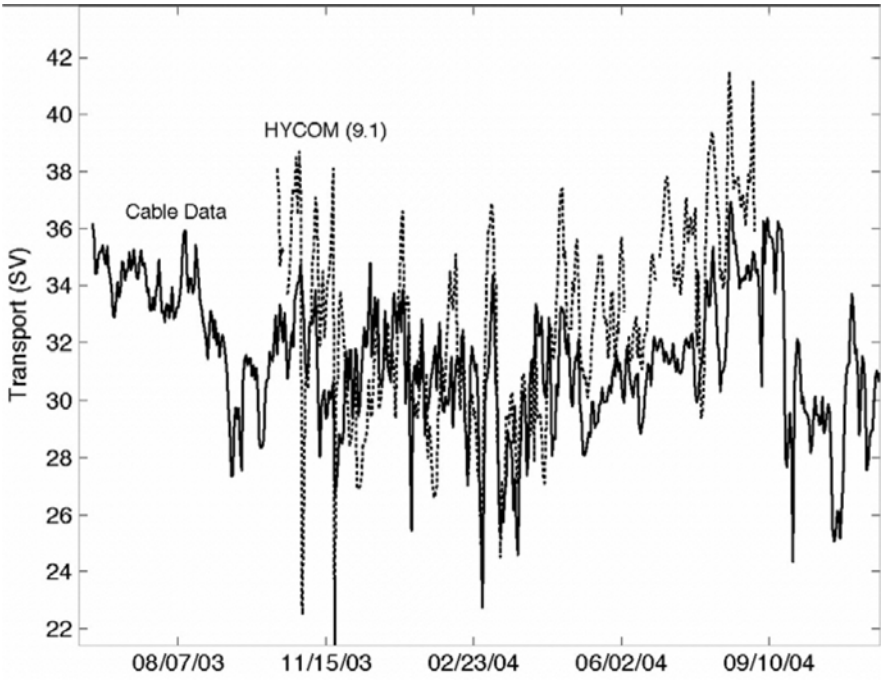

Figure 4. The transport in the Florida Current at 27°N from the 1/12° Atlantic near real-time system are shown with dotted lines. Observations from the cable data are shown in solid black.

Comparison with drifting buoys: These tests will evaluate the models' ability to produce ocean currents that yield drifter and ARGO floats trajectories similar to observations.

Mixed Layer Depth (MLD) (analysis, forecast, simulation without ocean data assimilation): Model analyses, forecasts, and simulations will be compared to mixed layer depths from profile data (e.g. XBTs, ARGO floats, CTDs, and moored buoys) and to an MLD climatology.

Event comparisons: Independent data are used for qualitative and quantitative evaluation of prediction system skill in nowcasting and forecasting specific oceanic events and features. A classical example is the impact of hurricanes on the ocean circulation (Zamudio et al., 2002). Comparisons of surface height and temperature with ocean color imagery can at times provide clear and dramatic qualitative model assessment (Chassignet et al, 2005).

The near real-time North Atlantic basin model outputs are made available to the community at large within 24 hours via the Miami Live Access Server (LAS) (http://hycom.rsmas.miami.edu/las). Specifically, the LAS supports model-data and model-model comparisons; provides HYCOM subsets to coastal or regional nowcast/forecast partners as boundary conditions, and increases the usability of HYCOM results by "application providers".

4. Outlook

The long term goal is an eddy-resolving, fully global ocean prediction system with data assimilation based on HYCOM to be transitioned to the Naval Oceanographic Office at 1/12° equatorial (~7 km mid-latitude) resolution in 2007 and 1/25° resolution by 2011. This paper summarizes the present status of the HYCOM effort and illustrates its capabilities. The present systems are a first step towards the fully global 1/12° HYCOM prediction system. The size of the problem makes it very difficult to use sophisticated assimilation techniques. Some of these methods can increase the cost of running the model by a factor of 100. It is, however, important to evaluate the performance of these advanced data assimilation techniques. Several additional techniques for assimilating data into HYCOM are already in place or are in the process of being implemented. These techniques vary in sophistication and computational requirements and include: NRL Coupled Ocean Data Assimilation (NCODA), Singular Evolutive Extended Kalman (SEEK) filter, Reduced Order Information Filter (ROIF), Ensemble Kalman Filter (EnKF), Reduced Order Adaptive Filter (ROAF) (including adjoint), and the 4D-VAR Representer method.

NCODA is an oceanographic version of the multivariate optimum interpolation (MVOI) technique widely used in operational atmospheric forecasting systems. A description of the MVOI technique can be found in Daley, (1991). The ocean analysis variables in NCODA are temperature, salinity, geopotential (dynamic height), and velocity. The horizontal correlations are multivariate in geopotential and velocity, thereby permitting adjustments to the mass field to be correlated with adjustments to the flow field. NCODA assimilates all available operational sources of ocean observations. This includes along track satellite altimeter observations, MCSST and in situ observations of SST and SSS, subsurface temperature and salinity profiles from BT's and profiling floats, and sea ice concentration.

Both the SEEK filter (Pham et al.,1998) and ROIF (Chin et al., 1999) are sequential in nature, implying that only past observations can influence the current estimate of the ocean state and are especially well suited for large dimensional problems. The ROIF assumes a tangent linear approximation to the system dynamics, while the SEEK filter can use the non-linear model to propagate the error statistics forward in time (Ballabrera et al., 2001). Besides the NCODA, SEEK and ROIF methods, other techniques such as the EnKF and the ROAF are also being evaluated. Because of their cost, they are presently being evaluated mostly within coastal HYCOM configurations or in specific limited areas of high interest. The NCODA and SEEK techniques are being considered as the next generation data assimilation to be used in the near real-time system.

Development of the global HYCOM prediction system is presently taking place and includes model development, data assimilation, and ice model embedment. The model configuration is fully global with the Los Alamos CICE ice model embedded and will run at three resolutions: ~60 km, ~20 km and ~7 km at mid-latitudes with the NCODA data assimilation. As stated above, some of the more expensive data assimilation techniques, while impractical over a high resolution global domain, can be used in subregions of the global model domain where there is special interest or where they provide particular value added.

Acknowledgements

This work was sponsored by the National Ocean Partnership Program (NOPP), the Office of Naval Research (ONR), and the Operational Effects Programs (OEP) Program Office, PMW 150 through the following projects: NOPP HYCOM Consortium for Data-Assimilative Ocean Modeling, NOPP U.S. GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean

Model (HYCOM), 6.1 Global Remote Littoral Forcing via Deep Water Pathways (ONR), 6.4 Large Scale Ocean Models, and 6.4 Ocean Data assimilation (all the 6.4 projects sponsored by PMW-150).

References

Ballabrera-Poy J., Brasseur P. and Verron J., 2001: Dynamical evolution of the error statistics with the SEEK filter to assimilate altimetric data in eddy-resolving ocean models, *Q. J. R. Met. Soc*., **127**, 233-253.

Bleck, R., and D. Boudra, 1981: Initial testing of a numerical ocean circulation model using a hybrid (quasi-isopycnic) vertical coordinate. *J. Phys. Oceanogr.*, **11**, 755-770.

Bleck, R., and S. Benjamin, 1993: Regional weather prediction with a model combining terrain-following and isentropic coordinates. Part I: Model description. *Mon. Wea. Rev.*, **121**, 1770-1785.

Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modelling*, **4**, 55-88.

Canuto, V.M., A. Howard, Y. Cheng, and M.S. Dubovikov, 2001: Ocean turbulence. Part I: One-point closure model-momentum and heat vertical diffusivities. *J. Phys. Oceanogr*., **31**, 1413-1426.

Canuto, V.M., A. Howard, Y. Cheng, and M.S. Dubovikov, 2002: Ocean turbulence. Part II: Vertical diffusivities of momentum, heat, salt, mass, and passive scalars. *J. Phys. Oceanog,.* **32**, 240-264.

Chassignet, E.P., L.T. Smith, R. Bleck, and F.O. Bryan, 1996: A Model Comparison: Numerical Simulations of the North and Equatorial Atlantic Ocean Circulation in Depth and Isopycnic Coordinates. *J. Phys. Oceanogr.,* **26**, 1849-1867.

Chassignet, E.P., H. Arango, D. Dietrich, T. Ezer, M. Ghil, D.B. Haidvogel, C.-C. Ma, A. Mehra, A.M. Paiva, and Z. Sirkes, 2000: DAMEE-NAB: the base experiments. *Dyn. Atmos. Oceans.*, **32**, 155-184.

Chassignet, E.P., and Z.D. Garraffo, 2001: Viscosity parameterization and the Gulf Stream separation. In "From Stirring to Mixing in a Stratified Ocean". Proceedings 'Aha Huliko'a Hawaiian Winter Workshop. U. Hawaii. January 15-19, 2001. P. Muller and D. Henderson, Eds., 37-41.

Chassignet, E.P., L.T. Smith, G.R. Halliwell, and R. Bleck, 2003: North Atlantic simulations with the HYbrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference density, and thermobaricity. *J. Phys. Oceanogr.*, **33**, 2504-2526.

Chassignet, E.P., H.E. Hurlburt, O.M. Smedstad, C.N. Barron, D.S. Ko, R.C. Rhodes, J.F. Shriver, A.J. Wallcraft, and R.A. Arnone, 2005: Assessment of ocean data assimilative systems in the Gulf of Mexico using ocean color. In AGU Monograph "*New developments in the circulation of the Gulf of Mexico*", in press.

Chassignet, E.P., H.E. Hurlburt, O.M. Smedstad, G.R. Halliwell, P.J. Hogan, A.J. Wallcraft, R. Baraille, and R. Bleck, 2005: Data assimilative ocean modeling with the HYbrid Coordinate Ocean Model (HYCOM), *J. Mar. Sys*., in press.

Chin, T.M., A.J. Mariano, and E.P. Chassignet, 1999: Spatial regression with Markov random fields for Kalman filter approximation in least-squares solution of oceanic data assimilation problems. *J. Geophys. Res*., **104**, 7991-8014.

Cooper, M., and K. Haines, 1996. Altimetric assimilation with water property conservation. *J. Geophys. Res*., **101**, 1059-1078.

Griffies, S.M., C. Böning, F.O. Bryan, E.P. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A.-M. Treguier, and D. Webb, 2000: Developments in ocean climate modelling. *Ocean Modelling*, **2**, 123-192.

Halliwell, G., 2004: Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid Coordinate Ocean Model (HYCOM). *Ocean Modelling*, **7**, 285-322.

Hoang, S., R. Baraille, O. Tallagrand, X. Carton, and P. De Mey, 1997: Adaptive flitering: Application to satellite data assimilation in oceanography. *Dyn. Atmos. Oceans*, **27**, 257-281.

Jacobs, G.A., C.N. Barron and R.C. Rhodes, 2001: Mesoscale characteristics, *J. Geophys. Res*., **1006**, 19, 581-19, 595.

Kara, A.B., P.A. Rochford, and H.E. Hurlburt, 2002: Air-sea flux estimates and the 1997-1998 ENSO event, *Boundary-Layer Meteorology*, **103**, 439-458.

Kraus, E.B. and J.S. Turner, 1967. A one-dimensional model of the seasonal thermocline: II The general theory and its consequences. *Tellus*, **19**, 98-106.

Large, W.G., G. Danabasoglu, S.C. Doney and J.C. McWilliams, 1997: Sensitivity to surface forcing and boundary layer mixing in a global ocean model: Annual-mean climatology. *J. Phys. Oceanogr.,* **27**, 2418-2447.

Marsh, R., M.J. Roberts, R.A. Wood, and A.L. New, 1996: An intercomparison of a Bryan-Cox type ocean model and an isopycnic ocean model. Part II: The subtropical gyre and heat balances. *J. Phys. Oceanogr*., **26**, 1528-1551.

Mellor, G.L. and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Geophys. and Space Phys.,* **20,** 851-875.

Pham D.T., J. Verron and M.C. Roubaud, 1998. Singular evolutive extended Kalman Filter with EOF initialization for data assimilation in oceanography, *J. Mar. Syst*, **16** (3-4), 323-340**.**

Price, J.F., R.A. Weller and R. Pinkel, 1986: Diurnal cycling: Observations and models of the upper ocean response to diurnal heating, cooling and wind mixing. *J. Geophys. Res.,* **91,** 8411-8427.

Roberts, M.J., R. Marsh, A.L. New, and R.A. Wood, 1996: An intercomparison of a Bryan-Cox type ocean model and an isopycnic ocean model. Part I: The subpolar gyre and high-latitude processes. *J. Phys. Oceanogr*., **26**, 1495-1527.

Rosmond, T.E., J. Teixeira, M. Peng, T.F. Hogan, and R. Pauley, 2002: Navy Operational Global Atmospheric Prediction System (NOGAPS): Forcing for ocean models, *Oceanography*, **15**(1), 99-108.

Schott F. A., M. Dengler, P. Brandt, K. Affler, J. Fischer, B. Bourles, Y. Gouriou, R.L. Molinari, and M. Rhein, 2003: The zonal currents and transports at 35°W in the tropical Atlantic, *Geophys. Res. Lett.,* **30** (7), 1349, doi:10.1029/2002GL016849

Willebrand, J., B. Barnier, C. Böning, C. Dieterich, P.D. Killworth, C. LeProvost, Y. Jia, J.- M. Molines and A.L. New, 2001: Circulation characteristics in three eddy-permitting models of the North Atlantic. *Prog. Oceanogr.,* **48,** 123-161.

Zamudio, L., H.E. Hurlburt, E.J. Metzger, and O.M. Smedstad, 2002: On the evolution of coastally trapped waves generated by hurricane Juliette along the Mexican West Coast. *Geophys. Res. Lett.,* **29** (56), 1-4.

Chapter 17

BLUELINK: LARGE-TO-COASTAL SCALE OPERATIONAL OCEANOGRAPHY IN THE SOUTHERN HEMISPHERE

Andreas Schiller[1] and Neville Smith[2]

[1]CSIRO Marine Research, Hobart, Tasmania, Australia; [2]Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

Abstract: In 2003, the Australian Government, through the Bureau of Meteorology (BoM), Royal Australian Navy and CSIRO, initiated BLUElink - *Ocean Forecasting Australia*, a project to deliver operational short-range ocean forecasts for the Asian-Australian region by 2006. Global advances in technologies necessary to observe and simulate the oceans have provided scientists at CSIRO and the Bureau of Meteorology with the tools to deliver near real-time information on ocean behaviour. Central to BLUElink is the development of a global and nested ocean prediction system. The BLUElink initiative centres on ocean prediction and analysis, and forecasting of day-to-day variations in ocean currents and temperatures around Australia. Ocean forecasts will be updated to include the latest changes in the ocean state and weather systems, particularly extreme conditions such as from tropical cyclones. The aim of the project is to generate ocean charts for marine users similar to weather forecast charts available to the rest of the community. The BLUElink system will provide information on coastal and open-ocean currents, surface and subsurface ocean properties, products that impact and are linked to maritime and commercial operations, defence applications, safety-at-sea, marine environmental sustainability, and regional and global climate. CSIRO is currently developing a high-resolution, coupled atmosphere-ocean model predicting out to 3 days which has been specifically designed for coastal and shelf applications. Standard products of the coastal forecasting system will be surface winds and sea-surface height, and 3-dimensional fields of ocean temperature, salinity and currents.

Keywords: Operational ocean forecasting, South-East Asian-Australian region, Indonesian throughflow, nested modeling, ocean reanalysis.

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 427-439.

1. Introduction

Ocean forecasting has become achievable because of rapid developments in computing power, remote observations and autonomous ocean instruments. Technical developments and intensive research have given the global marine community a broad understanding of the behavior and circulation of the oceans and seas. Although there is still a long way to go, scientists today have a good understanding of the "inner-workings" of the ocean. The state of the ocean can now be monitored and predicted. In line with international trends, operational oceanography in Australia is evolving in support of the community and its marine industries, as well as defense, environmental protection, and improved climate prediction.

BLUElink is a three-year project to develop the ocean model, analysis and assimilation systems required to provide timely information on, and forecasts of, the oceans around Australia. It will build on Australia's weather and climate forecasting systems, extending its capability to the marine environment, through an initial investment of $US 10M. BLUElink forecasts will provide information on coastal and open-ocean currents and eddies, and surface and sub-surface ocean properties; information that is useful for maritime and commercial operations, defense applications, safety-at-sea, sustainability of the marine environment, and regional and global climate. BLUElink is Australia's contribution to the international Global Ocean Data Assimilation Experiment (GODAE), an international experiment in ocean prediction. With new computer technologies and a greater understanding of the ocean and "how it works", scientists can start to run models that emulate the ocean environment and predict its evolution.

The Australian Bureau of Meteorology already provides operational weather services for marine users. This service focuses on the state of the sea surface and forecasts of wind conditions, storm surge heights, and sea-surface temperatures. BLUElink will provide the core of an extended oceanographic service at the Bureau of Meteorology in which the ocean model and assimilation prediction system will maintain up-to-date analyses of the state of the ocean. High-resolution analyses will be generated each day and twice-a-week forecasts will be generated out to 28 days incorporating the latest changes in the marine weather and wind systems. These extensions will complement the Bureau's existing weather and climate services and provide an integrated suite of products for the public and private sectors. The service will provide essential infrastructure for industry and the ocean community.

2. Data and product management

The international marine community shares a common desire to synthesize and assimilate global, real-time data of winds, ocean temperature, salinity and sea level. These data are collected by nations participating in the Global Ocean Observing System (GOOS) and stream in hourly to shared data networks. European and US satellite missions and the new wave of ocean-borne measuring instruments, such as the *Argo* profiling floats, are the centerpieces of the ocean observing system. BLUElink has instituted a sophisticated arrangement for reception, management and quality control of these data. The Bureau of Meteorology has installed the Meteorological Archive and Retrieval System (MARS) developed by ECMWF and this will be used to manage the entire ocean and product data streams for the operational ocean systems. State-of-the-art information technology will be deployed to ensure data are provided swiftly to models and to ensure products are available to participants and the wider community as soon as possible, e.g. via Live Access Servers. A preliminary web page with some background information about the project is available at http://www.marine.csiro.au/bluelink/.

3. Analysis and modelling systems

Just as zones of high and low pressure drive atmospheric changes and our weather, similar forces are at work in the ocean. The ocean responds directly to exchanges of energy with the atmosphere, through wind and surface heating. However, the ocean also delays its reactions to these forces, to later generate its own “weather” (ocean eddies) and fronts, as well as longer-term climate variations such as El Niño. The topography of the sea floor and the presence of the bordering land also have profound effects on the ocean, in a way that is unique to the ocean and its coastal subsystems. The BLUElink ocean models aim to capture key regional ocean currents, such as the East Australian and Leeuwin Currents, the Indonesian Throughflow and the Antarctic Circumpolar Current (Fig. 1). All of these currents are subject to strong short-term variability, which is superimposed to variability on seasonal and climate change timescales (for further details see Tomczak and Godfrey, 1994). Furthermore, basin-scale wave dynamics in the Pacific, Indian and Southern Oceans on seasonal and longer timescales has a direct impact on regional ocean variability around Australia (Wijffels and Meyers, 2004). All of these features need to be properly simulated in an ocean model which has a regional focus on the oceans around Australia.

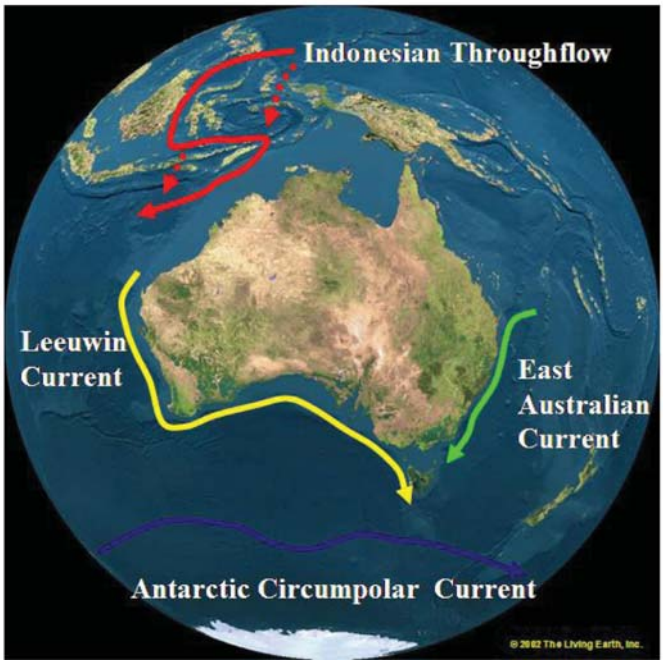

Figure 1. Schematic of major currents in the Asian-Australian region. The background image is © Copyright 1996 The Living Earth, Inc./Earth Imaging by permission of the publisher.

3.1 Analysis system

Modern weather prediction relies on access to accurate observations of the atmosphere. BLUElink will depend upon access to a range of surface and sub-surface observations of the ocean. To be useful, the data must be communicated and managed in a way that maximizes its contribution to the forecasts. The information required includes surface winds and temperature, sea level, ocean currents, and sub-surface temperature and salinity. Primary data types and sources are:

- Historical information obtained since about 1900, but mostly from recent decades (T, S);
- *Argo* profiling floats (T, S);
- Expendable bathythermographs (XBTs) deployed from "Ships of Opportunity" (T);
- Surface weather and ocean observations from voluntary observing ships (T, S, **u**);
- Surface drifting instruments (**u**);
- Coastal sea level from tide gauges (η);

- Open ocean sea level from the *Envisat* (ESA), *Jason* (NASA-CNES) and *GFO* (U.S. Navy) satellite altimetry (η);
- Sea surface temperatures from polar orbiting platforms and geostationary platforms, using both infrared and microwave instruments (SST);
- Surface wind estimates from satellite-borne instruments and numerical weather models (τ);

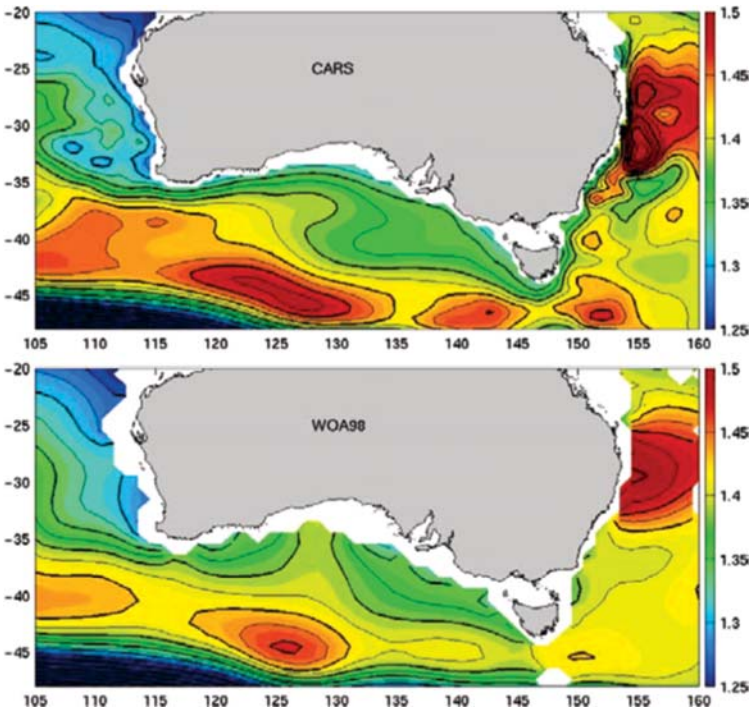

Figure 2. Dynamic height 400/2000 dbar. Top: CSIRO Atlas of Regional Seas (CARS) (Ridgway et. al., 2002); bottom: World Ocean Atlas (Levitus and Boyer, 1994).

CSIRO has already developed and maintains the CSIRO Atlas of Regional Seas – a regional (10ºN-60ºS: 90ºE-180ºE) high resolution, quality controlled, three-dimensional ocean climatology of temperature and salinity fields. The Atlas has been enhanced with additional historical temperature and salinity information for the Indonesian and New Zealand regions. Fig. 2 shows improvements in the fine-scale structures of CARS compared to the World Ocean Atlas (WOA). In CARS, a continuing coastal current is visible (Leeuwin and South Australian Current), which is missing in the WOA. The atlas uses a methodology which explicitly allows for the influence of

bathymetry and island barriers which resolves both the large-scale structure as well as narrow boundary features.

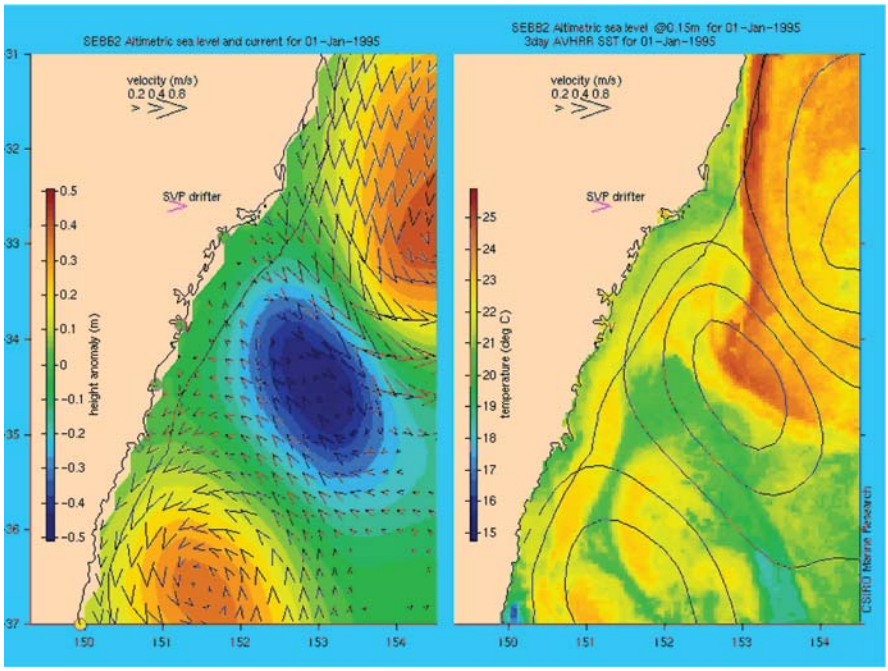

Figure 3. Example of satellite-based data analysis (ERS-2, Topex Poseidon and NOAA AVHRR). Date: 1 January 1995. Left panel: altimetric sea-level height (background) and vectors of geostrophic currents. Right panel: sea-surface temperature (background) with contours of sea level overlayed (contour interval 0.15 m). Shown are a cold-core eddy at the centre of each panel with two adjacent warm-core eddies to the north and south. Courtesy David Griffin, CSIRO Marine Research.

Near-real-time satellite altimeter data from platforms such as the *Jason-1* and *Envisat* satellites and coastal tide-gauge sea-level estimates are used to map sea-level (Fig. 3). These maps will be used in conjunction with a statistical projection method, the Regional Atlas, and available ocean *in situ* data (e.g. *Argo* profiling floats, XBTs, research and navy vessels) to infer three-dimensional estimates of the ocean's state. An enhanced, high-resolution real-time sea-surface temperature (SST) analysis will also be developed, with around 2–5 km spatial resolution in the Australian region. The analysis will take advantage of geostationary satellite data to partially resolve the diurnal cycle and of microwave SST data to eliminate gaps arising from cloud cover. BLUElink will contribute Level 2 data from the Australian region to the GODAE High Resolution SST Pilot Project (GHRSST) and develop dedicated data sites to support the validation of SST

products. Surface currents will be estimated from images of remotely-sensed sea-surface temperatures, using a method that tracks the movements of small-scale features. A high-quality enhanced regional wind speed and direction product will also be developed.

3.2 Modelling and assimilation system

3.2.1 Global model and data assimilation system

BLUElink has developed a global ocean model based on the GFDL MOM4 code (Griffies et al., 2004), called the Ocean Forecasting Australia Model (OFAM), with a resolution telescoping from 2º in the North Atlantic and North Pacific to (1/10)º in the Asian-Australian region. The Asian-Australian region is defined as extending from longitudes 90º E to 180º E, and from latitude 16º N to 75º S which is very similar in extent to the part of the earth shown in Fig. 1. OFAM has 47 vertical levels with 35 levels in the top 1000 m and a thickness of the uppermost levels of 10 m. The model uses the third-order "quicker" scheme for tracer advection (Leonard, 1979), a hybrid mixed-layer model based on Chen et al. (1994), viscosity mixing based on Smagorinsky (1993) and isopycnal mixing with the Gent-McWilliams parameterisation of eddy-induced stirring (Gent and McWilliams, 1990). Due to the model's variable grid size anisotropic options have been chosen for the latter two parameterisations.

OFAM will form the backbone of the operational ocean forecasting system run by the Bureau of Meteorology. A multivariate Optimum Interpolation scheme (BODAS) is used to assimilate ocean observations in near-real time into the model to improve the initial conditions of the forecasts. A schematic diagram of this forecast system is presented in Fig. 5. BODAS is a stand-alone system that is intended to be adaptable to other ocean models. BODAS calculates a global analyses of surface height η, temperature T, salinity S and the horizontal components of current (u, v) by combining a forecast from OFAM with observations of sea-level anomaly (SLA) from along-track altimetry and temperature and salinity from various sources including Argo profiles, CTD surveys and moored arrays. BODAS performs this calculation through a series of steps that are summarized as follows:

1. Calculate analysis of η, T and S
2. Convectively adjust T and S
3. Spatially filter increments of η, T and S
4. Calculate geostrophic increments for u and v
   (or use statistical approach).

The analyses for η, T and S are produced using ensemble-based multivariate optimal interpolation (EMOI), which is based on simulated ensembles produced by the model's spin-up run. Modules from MOM4.0 are used to convectively adjust gravitationally unstable temperature and salinity profiles. The spatial filter is applied to eliminate near-grid-scale features in the analysis increments. These modifications to the EMOI- based analyses are made before the geostrophic adjustments are computed for u and v in an attempt to construct dynamically balanced analyses. Further details about BODAS can be found in Oke (2004, unpublished manuscript).

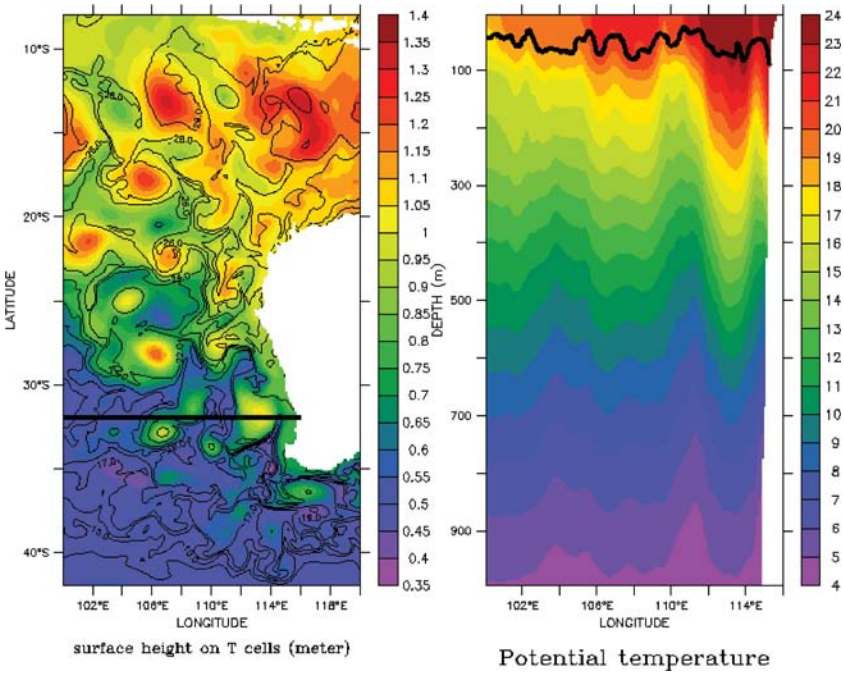

Figure 4. Example of daily-averaged eddy structures in OFAM forced by ERA-40 winds. Left Panel: sea-surface height (colour) and sea-surface temperature (contours) along Leeuwin Current. Right panel: Potential temperature and mixed-layer depth along a section indicated by the horizontal line in the left panel.

The model will be run daily in analysis mode. Twice a week forecasts will be issued with a 7-day short-range forecast and 28-day long-range forecast, respectively.

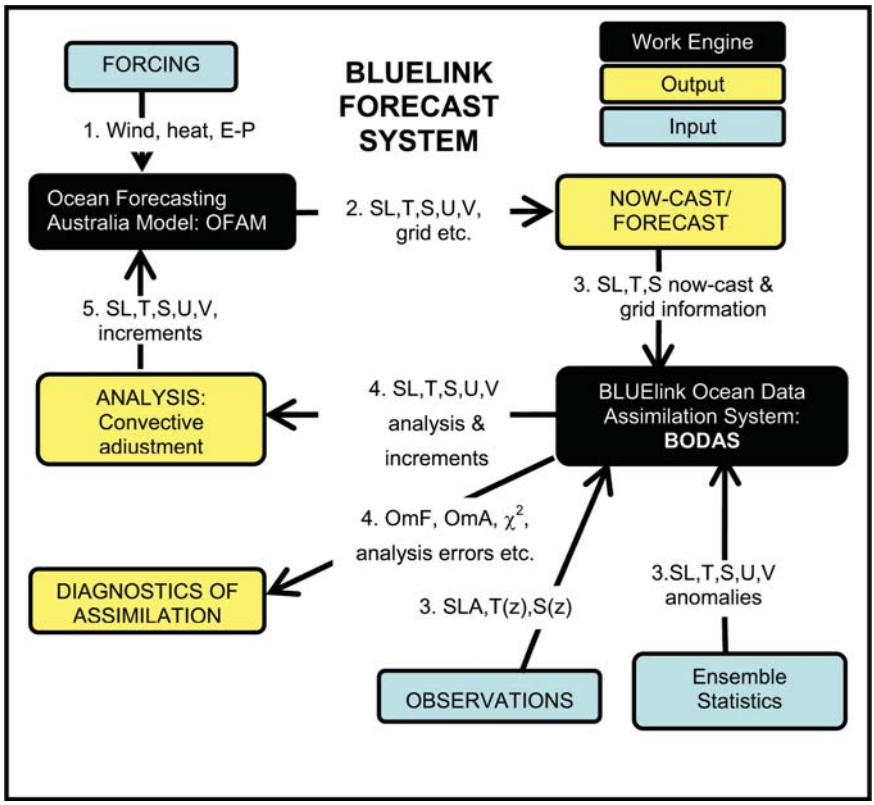

Figure 5. Schematic of BLUElink Ocean Data Assimilation System (BODAS). The three main components are the ocean model and the data assimilation component ("work engines"), the observational input data (surface forcing, satellite and in situ observations), and the analysed output fields which are combined with the model to provide now- and forecasts. The numbers denote the sequence of steps performed during an assimilation cycle. Courtesy Peter Oke, CSIRO Marine Research.

3.2.2 Relocatable ocean atmosphere model

A relocatable, high-resolution, coupled oceanic-atmospheric model called the Relocatable Ocean-Atmosphere Model (ROAM) is under development. The ocean component of ROAM is based on a primitive equations model developed by Walker and Waring (1998) and the atmospheric component has been adopted from the Colorado State University (Regional Atmospheric Modelling System; Pielcke et al., 1992). ROAM will be embedded within the global OFAM model, bringing the resolution down to 2 km for domains of the order of 100 km x 100 km.

ROAM will enable high-resolution forecasts of the water column, of the atmosphere and of conditions at the ocean-atmosphere interface out to 7-10 days (Fig. 6). In this way, scientists can account for the effect of large-scale winds and currents on the Australian region. With ROAM, an operator at a computer terminal will be able to specify a particular geographic region. This might be Bass Strait or the Timor Sea, for example. ROAM will then automatically interrogate the latest output from BLUElink's global ocean and the Bureau's global atmospheric models, and "nest" more detailed atmosphere and ocean models inside it. It will then produce a forecast of the meteorological and ocean conditions for the next several days. ROAM will predict wind, rain and temperature in the atmosphere, and currents, temperature and salinity in the ocean. The model is being developed for the Navy to calculate the behavior of radar and sonar in particular operational locations for their next few days of activity. As implemented in BLUElink, ROAM will typically cover the continental shelf. ROAM will facilitate the implementation of higher resolution nested circulation models for coastal embayments and estuaries.

4. Conclusions

By 2006, the BLUElink project will be the first operationally implemented ocean forecasting system in the Southern Hemisphere covering large-to-coastal scales. BLUElink participates in the international GODAE project and results will be made available to the scientific community and to commercial users of ocean forecasts. Potential benefits from BLUElink will include:

- Optimum ship routing to achieve fuel savings and shipping schedules.
- Enhanced environmental information for onboard naval tactical response systems, improved sonar performance.
- Predictions of changes in ocean conditions with design and cost implications for floating (oil exploration, oceanic and coastal fish farms, marinas) and permanent structures (jetties, breakwaters, pipelines, cables).
- Enhanced coastal and wildlife protection through modelling the dispersion of accidental oil and pollutant spills.
- Improved capabilities for maritime rescue and safety authorities in support of offshore maritime operations and recreational activities.
- Sustainable fisheries and fisheries management, through better understanding of how changes in the currents and the temperature and salinity of the water influence the fisheries resource.

- Identification of changes in coastal water temperatures, tides, sea level and currents that influence, for example, coral reefs and aquaculture management.
- Longer lead-times in climate predictions providing agriculture, tourism, energy generators, disaster response and insurance sectors with greater planning flexibility to respond where possible to the impacts of climate.
- Access for researchers to enhanced ocean data (observed and modelled) with the potential to generate value-added applications for industry and the community.

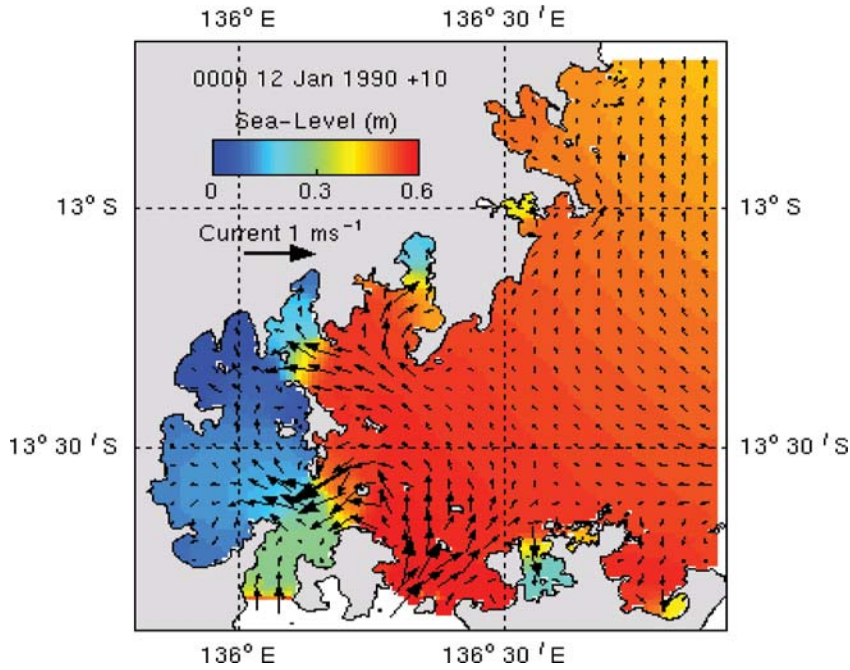

Figure 6. Example of high-resolution ROAM ocean simulation showing sea level and currents in the Gulf of Carpentaria. Courtesy Mike Herzfeld and Scott Condie, CSIRO Marine Research.

At the time of writing this chapter, the work focussed on the first global reanalysis run over the last decade which also serves as a first test of the combined global model, data assimilation system and observations (see section 3.1 for observations assimilated in the model). In hindcast/ reanalysis mode the model is forced by 6-hourly ERA-40 data from the ECMWF (plus

weak restoring towards observation-based estimates of sea-surface temperature and sea-surface salinity), whereas in forecast mode atmospheric forecasts from the Bureau of Meteorology's NWP system will be used as surface forcing for OFAM. This first experiment will also provide preliminary but important information about the statistical behaviour of the system (the box labelled "Diagnostics of Assimilation" in Fig. 5).

The next step will be the operational implementation of the forecasting system at the Australian Bureau of Meteorology. Observations received in near-real time from satellites (altimetry and SST) and in situ platforms such as Argo floats will be analysed and ingested daily and into the forecasting system. Short-range forecasts will be issued twice a week.

Future developments beyond the timescale of the current project will focus on an improved data assimilation system (e.g. Ensemble Kalman Filter) and on the implementation of a sea-ice model.

Acknowledgements

Parts of this paper are based on a brochure developed by the BLUElink team and the CSIRO Marine Research Communications Group. Among others, the contributions by Rick Bailey, John Parslow, Craig Macaulay, Lea Crosswell, Peter Craig, David Griffin, Peter Oke, Ken Ridgway and Scott Condie are particularly appreciated. This work was partly funded by the Bureau of Meteorology, the CSIRO and a grant from the Royal Australian Navy in support of the BLUElink ocean forecasting partnership project.

References

Chen, D., L. M. Rothstein and A. J. Busalacchi, 1994: A hybrid vertical mixing scheme and its application to tropical ocean model. J. Phys. Oceanogr., 24, 2156-2179.

Gent, P. R. and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models, *J. Phys. Oceanogr*., 20, 150-155.

Griffies, S. M., M. J. Harrison, R. C. Pacanowski, and A. Rosati, 2004: A technical guide to MOM4. GFDL Ocean Group Technical Report No. 5. NOAA/Geophysical Fluid Dynamics Laboratory, 371 pp.

Levitus, S., and T. P. Boyer, 1994: World Ocean Atlas 1994. Volume 4: Temperature. NOAA Atlas NESDIS 4, 117pp.

Leonard. B. P., 1979: A stable and accurate convective modelling procedure based on quadratic upstream interpolation. *Computer Methods in Applied Mechanics and Engineering,* 19, 59-98.

Oke, P. R., 2004: BLUElink Ocean Data Assimilation System: BODAS. User's Manual. CSIRO Marine Research, 24 pp.

Pielke, R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, W. A. Lyons, L. D. Grasso, M. E. Nicholls, M.D. Moran, D.A. Wesley, T.J. Lee and J.H. Copeland, 1992: A comprehensive meteorological modelling system - RAMS., *Meteorol. Atmos. Phys.,* 49, 69-91.

Ridgway, K. R., J. R. Dunn and J. L. Wilkin, 2002: Ocean interpolation by four-dimensional weighted least squares – Application to the waters around Australia, *J. Atmos. and Oceanic Technology*, 19, pp. 1357-1375.

Smagorinsky, J., 1993: Some historical remarks on the use of nonlinear viscosities. B. Galperin and S. A. Orszag, Eds., *Large Eddy Simulation of Complex Engineering and Geophysical Flows*, Cambridge University Press.

Tomczak, M. and J. S. Godfrey, 1994: Regional oceanography: an introduction, Pergamon Press, Oxford, 422pp.

Walker, S. J. and J. R. Waring, 1998: A multiple grid, 3-dimensional, non-linear, variable-density hydrodynamic model with curvilinear horizontal coordinates and level (z) vertical coordinates, CSIRO Marine Research, Report OMR-118/120.

Wijffels, S. and G. A. Meyers, 2004: An intersection of oceanic wave guides: variability in the Indonesian Throughflow region, *J. Phys. Oceanogr*., 34, 1232-1253.

Chapter 18

OPERATIONAL OCEANOGRAPHY: A EUROPEAN PERSPECTIVE

Jean-François Minster

IFREMER, Paris, France

1. Introduction

Operational oceanography aims at providing services and products to users of the ocean, and to help in sustainable exploitation of its resources. Indeed, one observes worldwide an increasing perception of ocean issues, such as its environment and climate, resource exploitation, or usage of the ocean space.

There are a lot of international legislations, treaties and declarations on environmental protection, marine resources, transport, fisheries, policies to take into account. These include for example KYOTO, UNFCCC, UNCLOS, AMAP, OSPARCOM, MARPOL, BARCELONE, and HELCOM. This variety induces public policies which are among the main drivers for operational oceanography.

The huge range of users interested in operational prediction systems, include among others:

- European Environmental Agency (EEA)
- Meteorological services
- Coastal protection agencies
- National and international environment administrators
- Water basin authorities
- Climate and environmental research organisations and communities

Knowledge and technologies developed during the last decades allows us to propose new services for all intermediate and end users, including monitoring activities, systems allowing impact studies, and prediction systems, all three aspects being complementary. In general, operational oceanography is used to describe prediction systems for ocean currents and its ecosystem: indeed, it is the new development which can potentially encompass observation and impact study systems. Many operational and

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 441-447.

preoperational prediction systems are being implemented. This note discusses some general issues of this development.

2. Open ocean operational oceanography in France

The CNES, NASA Jason-1, and its Jason-2 extension: Jason-1 was launched on December 7, 2001, on the same orbit as Topex/Poseidon. It carries an altimetry payload derived from Topex/Poseidon. The Jason-1 commissioning phase was successfully completed on 4 March 2002. Jason Operational Sensor Data Records are delivered within three hours.

An international effort is being made in the operational high precision satellite altimetry programme, with the recent decision of the National Oceanic and Atmospheric Administration (NOAA) and the European Organisation for the Exploitation of Meteorological Satellites (Eumetsat) to contribute to Jason-2. Furthermore, new technological developments, such as Altika, are on their way.

Coriolis: Coriolis is the French contribution to ARGO (an international programme that calls for the deployment of 3,000 free drifting profiling floats distributed over the global oceans by 2006). It includes instrument development, deployment and a near real-time global scale data centre. In the year 2001 there were data sets from XBT, CTD, and moorings providing global coverage with 165 190 profiles (T, S).

Prediction systems: MERCATOR is the preoperational open ocean current-prediction system which has been implemented in France. The partners in this project are CNRS (the French national scientific research centre), METEO-FRANCE (the French national weather service), CNES (the French space agency), SHOM (the French Navy's hydrography & oceanography department), Ifremer, and IRD (the French development research institute), and their goal is to progressively develop an operational capacity to analyse and predict the global ocean, through assimilation of near real-time satellite and *in situ* data into an ocean model. Customers of the system will be public services, civil security, defence, and commercial applications of oceanography. MERCATOR already provides weekly bulletins predicting the state of the ocean for the next two weeks, including outputs on the temperature and salinity of ocean currents at various depths, and sea surface height at a 5 km resolution in the North Atlantic and the Mediterranean Sea and at a lower resolution globally. It is a contribution to the Global Ocean Data Assimilation Experiment (GODAE).

The next steps: Existing preoperational systems need to be transformed into operational systems. This requires:

- Satellite recurrent systems, which is potentially feasible because of the lower cost of new technologies and the involvment of operational agencies;
- *In situ* observation operational teams and tools which requires long term coordinated commitment by national agencies;
- Permanent modelling and prediction systems;
- Services to intermediate and end users though public or private service companies.

3. GMES and GEOSS

It is clear that such an implementation requires an European or global perspective. GMES (Global Monitoring for Environment and Security) is a joint initiative of the European Commission and the European Space Agency (ESA) to provide a sound basis to European policies related to environment and security. GMES will provide Europe with an independent access to global information useful, for example, for international conventions such as the Kyoto Protocol. GMES will also develop applications for global change, environmental security, and natural hazards.

GMES will include an extended partnership between the European commission space agencies, industry and science. Development of GMES is based on the ESA GMES services element (funded €83 million, 2002-06), the 5th FP call for proposals, and the 6th FP Integrated Projects and Networks of Excellence. GMES should become the European contribution to a GEOSS (Global Earth Observation System of Systems) to be developed on the world scale.

4. The MERSEA concept

General objectives: MERSEA intends to produce, assess, and deliver real-time and continuous observations of the ocean's three-dimensional structure and associated biochemical components. MERSEA will also produce, assess, and deliver in real-time hindcasts and forecasts of the three-dimensional ocean variability at the highest resolution possible for the short time scales (a few weeks). MERSEA intends to deliver a global scale operational system. This will include support for shelf sea systems and interconnection with coastal zone systems.

A "European Centre for Ocean Monitoring and Forecasting" (ECOMF): The ECOMF concept aims to operate a global system, with a short to medium-term (e.g., one month) prediction capacity at high resolution. The ECOMF should have strong research connections and build partnerships with national centres.

Some European countries are already running national global ocean systems for practical or political reasons (e.g., defence needs). These national systems should enter into partnerships with the ECOMF, so as to share services and products for their mutual benefits. It is expected that several services currently operated under national systems will be transferred to the ECOMF.

Most practical issues and applications are regional and require very high resolution monitoring and modelling. Such systems can be best managed in a distributed manner. Regional "outcentres" can be run as integrated systems, carrying out observations, modelling and assimilation, real-time and off-line operation, validation, analysis, distribution of products, and regional services. For example, they could use ECOMF outputs as boundary conditions and will contribute to data acquisition and model development useful for ECOMF. Outcentres will generally develop in an open and competitive manner, but some can be part of an institutional network (e.g. the Mediterranean, Baltic, and Artic Seas).

MERSEA data processing modules: MERSEA does not contribute to infrastructures (e.g., ships, satellites, computers), but includes modules necessary to ensure that ocean observations are adequately processed.

MERSEA *in situ* observations: A global *in situ* observation system is required, such as continuing ARGO and time series observations. MERSEA includes part of the European contribution to this world scale system. MERSEA will aim at creating real-time access to environmental ocean monitoring data, and at implementing their assimilation into numerical models to improve the value of this environmental monitoring.

The Biogeochemistry component: There are a number of contrasting requirements for ecosystem prediction. Global requirements include CO_2 fluxes (for climate change) and primary production. These are of interest to end-users and decision makers.

Regional requirements include trophic interactions to zooplankton and predators, and harmful algal blooms. These are interesting to intermediate and end-users such as fisheries, aquaculture, and tourism.

Local (coastal) requirements include complex ecosystems with benthic, pollutants and suspended matter. These are important to intermediate and end-users, such as tourism managers and local policymakers.

This is a complex problem: preoperational systems are less developed and there is a lack of data. We need global primary production with ocean colour data assimilation. Models should have regional and local very high resolution with operational dispersion modelling, requiring preoperational complex systems.

We are talking about an evolutionary system, which will have to include both continuous improvments and occasional decisions implement new generations.

MERSEA interfaces: MERSEA will have to establish interfaces with many other organisations, including:

- Marine science activities of relevance, including European networks of excellence and integrated projects;
- Other operational systems, more particularly those dealing with meteorology, climate prediction, fisheries, and marine environment monitoring;
- National agencies and private companies involved in the development of operational oceanography, whether at global or regional scales (likely to become members or associates of ECOMF).

The MERSEA general principles: MERSEA will build on incremental developments of ongoing science and technology. It will aim at establishing a European operational system. MERSEA will identify and help organise a set of European agencies to implement and fund a long-term operational system as a component of a worldwide organisation by 2008. It should have the capacity to adjust to new requirements, research results and technologies.

5. Operational oceanography and research

Operational Oceanography must maintain connections with research, as:

- Evolution of the systems will benefit from open research and technological developments;
- Evolution of requirements will lead to define research and development questions to be addressed;
- Research activities devoted to outputs is an essential component of their validation, beyond the operational, in-house validation;
- Research will be a customer of services and products

It is necessary to organise this link. It is simpler to think of it at various levels, including:

- The general research and technology development in laboratory, with the need to identify the progress of internet for operational oceanography through science advisory committees ;
- The targeted research and technology developments, dealing with subjects identified by operational oceanography services; such developments can be supported by programs oriented by the operational services; this should also include validation exercises. Beside ecosystem issues, developments are very much needed in three directions : higher vertical resolutions, including near bathymetric features; higher horizontal resolutions, for example in frontal regions or near continental slopes and margins; higher frequencies, in particular in the mixed layer;
- Developments of the operational service, such as for example adaptation of new codes; it is more easily done inside the operational service, or through contracts to research terms.

In addition to these, outputs from operational services are necessary for research teams. They should be made available at the minimum cost, as their use by research is clearly at the benefit of all.

6. Services using operational oceanography

The success of operational oceanography will be measured by the number of public or private services which will be derived from the outputs. At the present stage, it is necessary to facilitate their implementation, through specific developments, demonstrations and access to appropriate outputs.

In particular, one should not underestimate the additional developments required by such services. It is first necessary to adjust the core production of the operational center to take into account their requirements: for example, high frequency outputs applications in fishery science or for Defence needs. Secondly, transformations of the outputs should be made possible for the service: for example, the precise position of high energy currents for oil platform operation.

7. Conclusion

Operational oceanography should be considered as one of the most important strategic axes of oceanographic research. It will not only

contribute to sustainable exploitation of the ocean, as it is already concentrating and orienting a very significant fraction of the research and technology effort in the discipline. It is very strongly focused by the need to implement systems; yet, it is a fascinating adventure with many opportunities.

Chapter 19

MERSEA: DEVELOPMENT OF A EUROPEAN OCEAN MONITORING AND FORECASTING SYSTEM

Ocean and Marine Applications for GMES

Yves Desaubies

Laboratoire d'Océanographie Spatiale, IFREMER, Plouzané, France

Abstract: The strategic objective of the MERSEA Integrated Project is to provide an integrated service of global and regional ocean monitoring and forecasting to intermediate users and policy makers in support of safe and efficient offshore activities, environmental management, security, and sustainable use of marine resources. Thus MERSEA objectives are very close to those of GODAE, to which the project federates the European contribution. As it is still in its initial phase, this chapter does not present results yet, but aims to outline the scope and work plan of the project.

Keywords: Mersea, GMES, global monitoring, operational oceanography

1. Introduction

The MERSEA Integrated Project is funded by the EU Commission to develop the ocean and marine applications component of the future GMES system[1]. The GMES system is being developed to fulfill the requirements of the EU for independent global information services in support of policy making and sustainable economic development. GMES addresses also such issues as risk and conflict prevention, or the negotiation and monitoring of treaties and conventions.

The concept of GMES is to provide monitoring and operational services for the global environment. The best established and developed systems are the National Weather Services (NWS) for the atmosphere, which are truly

[1] GMES: Global Monitoring for the Environment and Security, a joint initiative of the EU and the European Space Agency

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 449-453.

operational and global in their coverage, and issue weather forecasts regularly. NWS are also responsible for observations, analysis and forecasts in marine meteorology (wave forecasts, sea-ice conditions, storm surges), and marine safety (oil spills and drifting objects, search and rescue). Global ocean analyses and forecasts are also produced by some NWS for seasonal or climate applications, or in support of naval operations.

However, improvements in the ocean components of those systems are necessary and feasible. Progresses in high resolution numerical ocean modeling and data assimilation techniques, together with the availability of global data sets, allow more accurate representation of the three dimensional currents distribution – in contrast to many models that consider only the wind-driven component, for instance. Such advances have been implemented in several global systems, described in this volume. In particular, the FOAM, MERCATOR, and Mediterranean Forecasting systems are key components of MERSEA, complemented by major regional alliances around the North-West European shelves, the Baltic and North seas, and the Arctic and northern basins.

Specific ocean analysis and forecasting systems are needed to extend the range of services presently provided by NWS – and to improve them – as well as to respond to increased demand in several sectors, most notably ocean research, global ocean climate monitoring, ecosystem modelling, support for commercial applications in the offshore industry and maritime transport, or coastal management. These fields of application are the motivation for several international programmes, e.g. GODAE, or the Global Ocean Observing System (GOOS) and its regional alliances.

The four year MERSEA project started in April 2004, with the goal to deliver a working pan-European ocean system by early 2008. Some thirty eight agencies, institutes or universities in sixteen countries collaborate on the research and development effort.

Development of operational oceanography in Europe has been conducted under national programmes or research projects funded by the EU or the ESA. Consequently, there is a wide range of practices, formats, technical development, or institutional implementation. Some systems are fully operational, others are research projects, while some regions are not covered. It is one of the objectives of MERSEA to pool resources for the development of a high resolution system, to integrate regional systems by implementing standards and promoting best practices, and to fully validate the systems.

A preliminary project, MERSEA Strand 1, was conducted from January 2003 to June 2004, with participation of most of the partners of the present project. The objective was to assess the strengths and weaknesses of present systems, to conduct selected demonstrations (ecosystem modelling, and oil spill drift predictions), and to report on the lessons learned and make recommendations for future progress. The chapter by L. Crosnier and C. Le Provost in this volume describes one aspect of the project. It has proven

quite effective in building the MERSEA consortium, and in providing a baseline and starting point for the Integrated Project.

2. The MERSEA work plan

At the core of the system under development is the collection, validation and assimilation of remote sensed and in situ data into ocean circulation models that allow for the self consistent merging of the data types, interpolation in time and space for uniform coverage, nowcasting (i.e., data synthesis in real-time), forecasting, and hind-casting, and delivery of information products. Those products are aimed at intermediate users, such as NWS, or marine institutes and agencies in charge of ocean monitoring, or private service providers, who will use them in turn to improve their services or develop customized applications.

Accordingly, the work plan is structured in a number of work packages, that can be grouped in three main modules: one concerned with input data (remote sensed, in situ, and atmospheric forcing); the second with system design and development, production, research and validation, information management; the third with the development of specific applications. Other issues of overall assessment, outreach and training, and communication are addressed in the management and coordination work packages.

The various tasks and topics addressed in the project are summarized in a set of key specific objectives:

- Develop the systems needed, and use them, to provide real-time, high quality, validated, merged products from satellite data for surface height, surface temperature, ocean color, sea-ice and surface velocity.
- Implement and assess the value of moorings and gliders for bio-geochemical data. Initiate routine provision of real-time data from specific research vessels. Make cost-effective contributions to the in situ system for the Mediterranean and the global Argo system. Collect in situ data and make them available in real-time through a unique server.
- Implement and test a high resolution global ocean model with assimilation of remote sensed and in situ data; develop a deep ocean model for use in shelf seas and evaluate its performance on NW European shelf; evaluate methods for nesting of models ; develop and assess bio-geochemical models ;
- Progressively implement a coordinated ocean and sea-ice monitoring and forecasting system for the global ocean and European seas, assimilating in situ and satellite data and providing high-resolution forecasts on a daily or weekly basis for physical and bio-geochemical variables.

- Make systematic assessments of the performance of the analyses and forecasts of the physical and bio-geochemical variables.
- Assess the value of boundary data for regional forecasts and implement data assimilation methodologies for regional seas and test their impact.
- Develop an information management system that will deliver both real-time and delayed mode information to users, allowing them to exploit multiple data sets from many different sources with user-friendly internet-based interfaces. Provide full documentation and meta-data.
- Develop and demonstrate applications in support of the safety of marine operations including improved wave-forecasting, and forecasts for ship routing, the offshore oil and gas industry, oil spill drift.
- Implement research results and perform specific experiments to improve now-casting (analysis) and forecasting for ecosystems and seasonal weather forecasting.
- Help to organize appropriate teams and agencies to establish the proper framework for the ocean component of GMES by 2008.
- Develop a communication, promotion, education, and outreach program towards the general public, policy makers, end-users, and specialists to increase awareness and knowledge on the global ocean environment.

3. The MERSEA system

At the end of the project, in mid 2008, a working prototype system will be delivered. Its exact architecture is not defined at this early time, since the design is an essential task of the project. The present concept is that of a set of thematic centres: satellite products, in situ data, forcing fields, and the main ocean forecasting systems: global and regional. Those centres must be integrated by the adoption of common formats, practices, and quality standards. The institutional framework under which the system will be operated is an open question; the participating institutions (Met agencies and marine institutes) will continue to play a major role. The possibility of establishing a European Centre for Ocean Monitoring and Forecasting, is considered. Such a Centre would provide a focal point of excellence to conduct research and development, for training, to pool resources and operate the global system, and to distribute basic products to participating members. The major regional systems would be out-centres, with identified operators, and close ties to the global system. Considerations of funding, administrative status and international agreements are complex, and go beyond the scope of the project.

4. Conclusions

Several initiatives, such as the Group on Earth Observations and its System of Systems concept, and the GMES, provide a strong impetus towards the establishment of operational oceanography services in Europe. MERSEA intends to play a major role towards that end. There is certainly a strong public and political demand and interest in monitoring and understanding our global environment. There is perhaps also a somewhat utopian view in some quarters that everything can be observed and forecast, for the benefit of economic growth, sustainable development, public safety, and risk prevention. The recent devastating Indian Ocean tsunami must humble us: the terrible toll is due not only to the rare physical event and the difficulty to predict it; it is compounded by socio-economic factors (e.g. patterns of land use, settlement, and population growth), inadequate public education on the phenomenon, and improper communications for warnings.

Our models are imperfect, but significant progress is being made, and their realism and forecasting skill are constantly improving. These improvements result from advances in ocean modelling and assimilation, and from the availability of global data sets from satellites and in situ observing systems. One of the challenges is to sustain them in the long term and to transition them from research projects to operational systems.

MERSEA intends not only to develop the backbone of a future European ocean monitoring and forecasting system, but also to demonstrate its value for practical applications of interest to society.

Acknowledgements

The MERSEA Integrated Project involves many participants who share the credit and responsibility of its success. Special thanks go to the members of the Executive Committee, who are particularly active in steering the project and managing its work packages: P. Bahurel, M. Bell, E. Buch, J. Johannessen, P.-Y. Le Traon, G. Manzella, N. Pinardi, S. Pouliquen, R. Rayner; H. Roquet, U. Send, and J. Verron. The project is partially funded by the EU under FP6, MERSEA contract SIP3 – CT- 2003- 502885.

Chapter 20

INTERNAL METRICS DEFINITION FOR OPERATIONAL FORECAST SYSTEMS INTER-COMPARISON: EXAMPLE IN THE NORTH ATLANTIC AND MEDITERRANEAN SEA

Laurence Crosnier, Christian Le Provost, and the MERSEA-strand1 team

MERCATOR OCEAN, Ramonville St Agne, France.

Abstract : The European MERSEA and international GODAE projects have built a common methodology for real-time inter-comparison of forecast systems. Internal metrics, i.e. a mathematical definition of chosen diagnostics, are defined and aim at testing the consistency, quality and performance of each system. They are sorted into four classes (Class 1 to 4) and described here for the North Atlantic basin and the Mediterranean Sea. Possible use of such metrics and comparison to existing litterature is also briefly described.

Keywords: MERSEA, GODAE, internal metrics, inter-comparison, North Atlantic, Mediterranean Sea.

1. Introduction: MERSEA framework

The European project MERSEA conducts a real time inter-comparison of 5 operational forecast systems for the North Atlantic and Mediterranean basins, gathering in alphabetical order: FOAM from UK, HYCOM-US from USA, MERCATOR from France, MFS from Italy and TOPAZ from Norway. MERSEA project has developped a web site: http://www.mersea.eu.org. Its final aim is to build a European GMES (Global Monitoring for Environment and Security) operational system in 2008. MERSEA teams have built a common methodology, defining a common grid on to interpolate their outputs and internal metrics which aim at testing the consistency, quality and performance of each system.

E. P. Chassigne and J. Verron (eds.), Ocean Weather Forecasting, 455-465.

- 'Consistency' means that operational systems outputs have to be consistent with the current knowledge of the ocean circulation and climatologies.
- 'Quality' means that operational systems outputs have to be in agreement with independent observations (i.e. not assimilated).
- 'Performance' means that internal metrics should quantify the capacity of each system to provide accurate short term forecast.

Following those criteria, internal metrics are sorted into different classes. Class 1, 2, and 3 metrics allow testing the consistency and quality of the systems. Class 4 are diagnostics to check the performance of the system. Definitions for the North Atlantic and Mediterranean basins are given in this paper. Some complementary metrics (Class 1 to 4) are currently being defined in the context of GODAE for the Pacific Ocean (Masa Kamachi, personal communication), the Arctic and Antarctic Oceans with metrics for the ice (Gilles Garric, personal communication) and the Indian and Southern Oceans (Neville Smith, personal communication).

Standardized output fields and diagnostics are distributed via OPeNDAP servers and can be visualized through a Live Access Server (LAS) or with DODS clients.

2. OPeNDAP and LAS servers

2.1. OPeNDAP server

OPeNDAP allows to access remote data of interest over the internet, using familiar data analysis and visualization packages like: Matlab, Ferret and others, without worrying about data storage formats. More information about OPeNDAP/DODS can be found on the web:

- http://opendap.org/faq/what_is_OPeNDAP_software.html
- http://ferret.pmel.noaa.gov/Ferret/DODS/ferret_dods.html

A non exhaustive list of all datasets available via OPeNDAP servers is indicated on the OPeNDAP homepage: http://www.opendap.org/data/

2.2. LAS server

The Live Access Server (LAS) is a highly configurable Web server designed to provide flexible access to geo-referenced scientific data. LAS enables the Web user to visualize data with on-the-fly graphics, request subsets of variables in a choice of file formats. A user can quickly obtain products such as plots, images, and formatted files with any t, z, y, x combination. LAS has a comparison mode which allows the user to select any data sets distributed on Internet via OPeNDAP, and then compute

differences (with automated re-gridding), overlay them graphically and view them as side-by-side plots.

More information about LAS can be found on the web:

- http://ferret.pmel.noaa.gov/Ferret/LAS/ferret_LAS.html
- http://ferret.pmel.noaa.gov/Ferret/LAS/LAS_forInSituData.pdf

2.3. MERSEA servers

Mersea **OPeNDAP server** URL addresses are (Password and user name are available upon request):

http://user:password@www.nerc-essc.ac.uk:9090/dodsC/ for FOAM
http://hycom.rsmas.miami.edu/dodsC/ for HYCOM-US
http://user:password@opendap.mercator-ocean.fr/dodsC/ for MERCATOR
http://thredds.sincem.unibo.it:8080/thredds/dodsC/ for MFS
http://mersea.nersc.no/dodsC/ for TOPAZ

MERSEA LAS URL is the following:

http://las.mersea.eu.org (restricted access)

2.4. MERSEA forecast systems main characteristics

Ocean Model? Vertical Coordinate? Ice Model? Spin Up Length? Mixing Parameterization? Which Basin?	**Horizontal and Vertical Resolution? Relaxation to Mediterranean Water?**	**Which Heat and Momentum Forcing? Relaxation?**	**Which Data Assimilation Method? Which Data Assimilated? Which MSSH used (Mean Sea Level used as a reference for data assimilation)?**
MERCATOR FR OPA 8.1 Z coord., Rigid Lid. Simple diagnostic sea ice. SPIN UP : 15days TKE . ATL + MED.	Hori.1/15°(5/7km) . Verti. 43 levels. Relaxation to Medatlas (T,S) in Gulf of Cadiz below 500m.	Daily ECMWF forcing. Relaxation to Reynolds SST and Reynaud SSS.	OI SOFA. SLA SSALTO-DUACS along track once a week. MSSH from Rio et al.(2004) in the Atlantic and blend of previous runs in Mediterranean basin.
TOPAZ NO HYCOM 1.0 Hybrid coord., Free surface. Dyn/thermodynanic sea ice. SPIN UP: 20years . KPP mixing. ATL.	Hori. 20 to 30km. Verti. 22 hybrid layers. Closed boundary without relaxation.	6 hourly ECMWF forcing (Bulk formulae momentum and heat). Precipitation Clim + Relaxation to Levitus SSS.	EnKF. SLA SSALTO-DUACS maps once a week. SST from CLS AVHRR. Maps of ice concentration. MSSH from previous OCCAM run.
FOAM UK Brian-Cox Hadley Center.	Hori. 1/9° (12km). Verti. 20 levels.	6 Hourly UK-Met-Office forcing.	OI Cooper&Haines. SLA SSALTO-DUACS along track once a week.

Z coord., Rigid Lid. Dyn/thermodynanic sea ice. SPIN UP: 5months. Kraus-Turner. ATL + MED.	No Med. Water relaxation.	Relaxation to Levitus SST and SSS.	SST 2.5° gridded (ARGO). (T+S) vertical profiles. Maps of ice concentration. MSSH from previous run.
MFS IT MOM 1.0 Z coord. , Rigid Lid. No ice model. SPIN UP: 7 years . Constant vertical mixing + vertical adjustment. MED.	Hori. 1/8°. Verti. 31 levels. Transport through Gibraltar parameterized.	6 Hourly ECMWF forcing (Bulk formulae momentum and heat). Relaxation to satellite night time SST and SSS climatologies.	OI SOFA. SLA SSALTO-DUACS along track once a week. SST and T vertical profiles along track once a week. MSSH from previous run with 1993-99 forcing.
HYCOM US HYCOM *2.1* Hybrid coord., Free surface. No ice model. SPIN UP: 15years. KPP mixing. ATL.	Hori1/12°(6.5km). Verti. 26 hybrid layers. Entrainment parameterization of Med Water Outflow.	3 hourly NOGAPS forcing (Bulk formulae for heat). SSS=50%(E-P) +50% relaxation to Levitus SSS. Relaxation to MODAS SST analysis.	OI Cooper&Haines. SLA MODAS Maps once a week. MSSH from previous 1/12° MICOM run with perpetual ECMWF forcing.

3. Definition of internal metrics

3.1. Common grid

All the systems interpolate their outputs on the so called MERSEA grid with:

- A horizontal resolution of 1/8°.
- A vertical resolution with 8 vertical levels (at 5, 30, 50, 100, 200, 500, 1000, and 2000 meters) in the Mediterranean basin.
- And 12 vertical levels (at 5, 30, 50, 100, 200, 400, 700, 1000, 1500, 2000, 2500, and 3000 meters) in the North Atlantic.

The common geographical domain extends from 10°N to 68°N for the North Atlantic and covers the whole Mediterranean Sea excluding the Black Sea from 6°W eastward.

Class 1 to 3 diagnostics are provided on a real time basis by all teams through their OPeNDAP server for the daily mean (or snapshots for HYCOM-US) best estimates fields (the best estimate corresponds to the best analysis field that each system can produce), as well as for the sixth day forecast.

3.2. Class 1 metrics

Class 1 diagnostics gathers 2-D and 3-D fields interpolated on the MERSEA grid. Two dimensions fields are:

- **The zonal and meridional wind stress (Pa)**,
- **The total net heat flux including relaxation term (W/m^2),**
- **The freshwater flux including relaxation term ($kg/m^2/s$),**
- **The Barotropic Streamfunction (henceforth BSF) (Sverdrup=$10^6 m^3/s$).** The BSF characterizes the wind-driven circulation established in response to wind forcing. One year mean Sverdrup BSF can also be computed using the provided Class 1 wind stress fields. At 25°N, it is commonly assumed that the vertically integrated transport is governed by a flat-bottom Sverdrup balance (Towsend et al. 2000; Boning et al. 1991) at least in the eastern basin. The DYNAMO (Willebrand et al, 2001, their figure 15) five years mean numerical simulations without data assimilation at this latitude showed a good agreement in the eastern basin between models and Sverdrup for all models except the Sigma coordinates models.

- **The Mixed Layer Depth (henceforth MLD) (m).** Two kinds of MLD diagnostics are provided in the Atlantic basin: MLD(θ) with a 1°C criteria and MLD(ρ) with a 0.05kg/m^3 surface potential density criteria. In the Mediterranean Sea, one MLD(ρ) with a 0.011kg/m^3 surface potential density criteria is provided. Hovmuller plots of the MLD behaviour in chosen regions can show convection time periods. In the mediterranean basin for example, in convection areas as the Levantine Basin, the MLD maximum depth could be plotted. Volume of newly formed Levantine Intermediate Water could be estimated (Castellari et al. 2000).

- **The Sea Surface Height (SSH) (m).** For instance, a zonal section at 48°N of the one year mean SSH can show whether systems have a realistic North Atlantic Current (Willebrand et al. 2001, their figure 13). The path of major currents can also be derived from SSH averaged over several months using the Le Provost and Bremond (2003) algorithm which allows to display path location associated with geostrophic currents. True current position which are well known from compilation of in situ data and remote sensing observations (Auer, 1987) can also be displayed. SSH time series at some moorings locations can also be compared to available observed tide gauge measurements (Smedstad et al. 2002; Tokmakian and McClean, 2003).

- **The Mean Sea Surface Height (henceforth MSSH) (m)** used as a reference sea level during the assimilation procedure. Each system is using a different MSSH field (cf section 2.4). Differences in the MSSH fields between the systems can be large is some areas and are shown to have major influence on the system behaviour (Birol et al. 2004).

The three dimensional fields are:

- **The potential temperature (°C) and salinity (psu).** Those Class 1 metrics allow to test the consistency and quality of the systems. For example, the comparison of the monthly mean Class 1 fields to available climatologies (Levitus 1998; Reynaud et al. 1998; Medatlas 2002) can put in light drifts in the systems away from initial climatological conditions at depths because of long spin-up. Such tests have been used in recent inter-comparison experiments such as DYNAMO (Meincke et al. 2001) and DAMÉE (Chassignet and Malanotte-Rizzoli, 2000).
- **The zonal and meridional velocity fields (m/s).** A derived Class 1 diagnostic is the surface Eddy Kinetic Energy (EKE) which can be computed using surface Class 1 velocities and compared to EKE observations derived from satellite altimetry (Ducet and Le Traon, 2001).

3.3. Class 2 metrics

Class 2 diagnostics are interpolated on high 10km resolution vertical sections (Figures 1 and 2). Class 2 fields are:

- **The potential temperature (°C) and salinity (psu).** The vertical Class 2 sections can be compared, when possible, to historical WOCE synoptic sections. This brings relevant insights on the systems water masses characteristics, for example the 18°C Mode Water (Worthington, 1959), Madeira Mode Water (Siedler et al. 1987), or the Mediterranean Water outflow (Bryden and Kinder, 1991). Class 2 model sections can also be compared to observed XBT MEDS sections gathered within the SOOP program.
- **The zonal and meridional velocity fields (m/s).** Those Class 2 bring information on the vertical structure of currents, as for example, the Deep Western Boundary Current below the Gulf Stream transporting North Atlantic Deep Water (Willebrand et al. 2001; Lee et al. 1996), the Azores current (Paillet and Mercier, 1997; Sy, 1988) and the North Brazil Current (Johns et al. 1998; Schott et al. 1998; Stramma and Schott, 1996). Class 2 model velocities can also be compared to observed ADCP data.

Another Class 2 diagnostic using the velocity fields is the EKE which can be computed for instance along the 48°N zonal section in the various systems and compared to EKE observations from Colin de Verdiere et al. (1989). EKE distribution through the Gulf Stream at 55°W has also been documented from current meter, drifter and float data by Richardson (1983; 1985). Abyssal eddy kinetic energy can also be referred to estimates given by Schmitz (1984; 1996).

3.4. Class 3 metrics

Class 3 diagnostics are integrated quantities (integration done on the original grid) such as:

- **Volume transports (Sverdrup=$10^6 m^3/s$)** across chosen sections (Figures 1 and 2). Depending on the section considered, one has to provide the total volume transport or the volume transport per defined potential temperature classes or density classes. For example, the water flowing through the Florida Strait comes from different Caribbean passages. The knowledge of the flow distribution through these passages appears to be a significant test for the North Atlantic model simulations (Böning et al. 1991; Maltrud et al. 1998). Class 3 model volume transport across the Florida Straits can be compared to real time Cable Voltage measurements (Larsen, 1992). These measurements indicate an annual mean mass transport of approximately 30 Sv, modulated by a seasonal cycle in transport of roughly 6 to 10 Sv.

  In the Mediterranean basin, the volume transport seasonal variablility across several straits in the models can be compared to observations gathered in Astraldi et al. (1999).

- **The Overturning Streamfunction (OSF) (Sverdrup=$10^6 m^3/s$)** as a function of latitude and depth (m) or potential temperature (°C) or potential density (kg/m^3). The OSF characterizes the thermohaline circulation established in response to external forcings (wind, heat and freshwater fluxes) and to the water masses conservation taking place in the buffer zones. The large scale overturning is not directly observable, but an annual mean maximum OSF from 16 to 20 Sv between 20°N and 40°N in the depth range 1000m to 1500m is consistent with the estimates of the corresponding heat transport. At 24°N, repeated transoceanic sections contributed to get a remarkably stable estimate of 17-18 Sv (Hall and Bryden, 1982; Roemmich and Wunsch, 1985).

- **The Meridional Heat Transport (MHT) (PW=10^{15}Watt).** The MHT is a variable of high climatological interest. The MHT is strongly linked to the OSF and mostly reflects the North Atlantic Deep Water (NADW) overturning cell behaviour: the stronger the NADW cell, the stronger the northward MHT. The canonical value is 1.2 +/- 0.3 PW at 24°N (Hall and Bryden, 1982). The OSF and MHT Class 3 diagnostics provide a significant index of the thermodynamic behaviour of the model.

3.5. Class 4 metrics

Class 4 metrics are root mean square statistics (equation 1) in the model and observation space to assess data assimilation performance and forecast skill.

$$rms\,(o - Hs) = \sqrt{\frac{1}{N}\sum (o - Hs)^2} \qquad (1)$$

"o" is the observation vector available in the [T0-7; T0] daily temporal window.

"s" is a hindcast, a forecast, an analysis or a persistence at day T0.

"H" is an operator that converts the "s" vector into the space in which the observation "o" is expressed, i.e. horizontal or vertical interpolation.

The state variables used are sea level anomaly, potential temperature and salinity. The results are given in the form of spatial averages over agreed regions and depth classes. Each team is using the exact same set of independent observations in order the diagnostics to be coherent and meaningful.

4. Conclusion

The methodology based on metrics definition and distribution of outputs through OPeNDAP servers has been applied during the MERSEA-strand1 project and allows a successful demonstration of real time inter-comparison of basin-scale systems in the North Atlantic and Mediterranean basins (Crosnier and Le Provost, 2004). The inter-comparison exercise is being pursued during the European MERSEA Integrated Project (2004-2008). The methodology developped provides a forum to share experience and discuss the areas where progress is needed. It allows identify required characteristics to build a performing operational system. Recommendations for improvement can regularly be addressed to system's team. The methodology allows a continuous and comprehensive assessment of the performances of each system including all components as the observing system, the modeling, assimilation and product distribution components. The framework

built during the MERSEA-strand1 project helps setting up the GMES Ocean Component. It has also been adopted by the GODAE partners, who are defining more metrics adapted to the global Ocean.

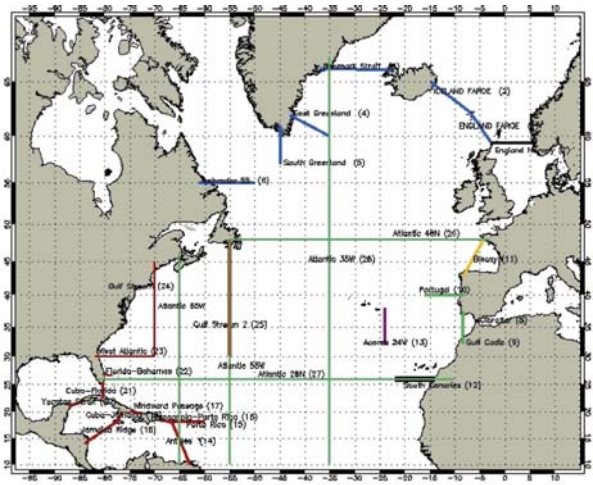

Figure 1. Class 2 sections in the Atlantic.

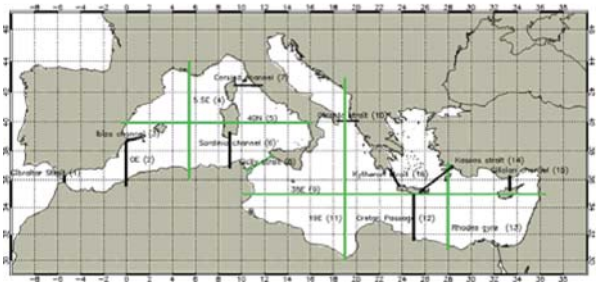

Figure 2. Class 2 sections in the Mediterranean Sea.

References

Astraldi M., S. Balopoulos, J. Candela, J. Font, M. Gacic, G.P. Gasparini, B. Manca, A. Theocharis and J. Tintoré, 1999: The role of straits and channels in understanding the characteristics of Mediterranean circulation. Progress in Oceanography 44, 65-108.

Auer S.J., 1987: Five-year climatological survey of the Gulf Stream system and its associated rings. J. Geophys. Res., 92, 11709-11726.

Birol F., J.-M Brankart., F. Castruccio, P. Brasseur, and J. Verron, 2004: Impact of the Ocean Mean Dynamic Topography on Satellite Data Assimilation, Mar. Geodesy, 27, 59-78.

Böning C.W., R. Döscher, and H.J. Isemer, 1991: Monthly mean wind stress and Sverdrup transports in the North Atlantic: A comparison of Hellerman-Rosenstein and Isemer-Hasse climatologies. J. Phys. Oceanog., 21, 221-235.

Bryden H.L. and T.H. Kinder, 1991: Steady two-layer exchange through the Straits of Gibraltar. Deep Sea Res., 38, 445-463.

Castellari S., N. Pinardi, and K. Leaman, 2000: Simulation of water mass formation processes in the Mediterranean Sea: Influence of the time frequency of the atmospheric forcing, J. Geophys. Res. 105, 24157-24181.

Chassignet E.P. and P. Malanotte-Rizzoli, 2000: Ocean circulation model evaluation experiments for the North Atlantic Basin, Dyn. Atmos. Oceans, 32, 155-431.

Colin de Verdière A., H. Mercier, and M. Arhan, 1989, Mesoscale variability from the western to the eastern Atlantic along 48°N, J. Phys. Oceanog, 19, 1149-1170.

Crosnier L. and C. Le Provost, 2004: Inter-comparing five forecast operational systems in the North Atlantic and Mediterranean basins: The MERSEA-strand1 Methodology, submitted to J. Mar. System.

Ducet N. and P.Y. Le Traon, 2001: A comparison of surface kinetic energy and Reynolds stresses in the Gulf Stream and the KuroShio current systems from merged T/P and ERS-1/2 altimetric data, J. Geophys. Res., 106, 16603-16622.

Hall, M.M and H.L. Bryden, 1982: Direct estimates and mechanisms of ocean heat transport, Deep Sea Research, 29, 339-359.

Johns W.E., T.N. Lee, R.C. Beardsley, J. Candela, R. Limeburner, and B. Castro, 1998: Annual cycle and variability of the North Brazil Current, J. Phys. Oceanog, 28, 103-128.

Larsen J.C, 1992: Transport and heat flux of the Florida Current at 27°N derived from cross-stream voltages and profiling data: theory and observations. Philosophical Transactions of the Royal Society of London A, 338, 169-236.

Lee T.N., W. Johns, R. Zantopp and E. Fillenbaum, 1996: Moored observations of western boundary current variability and thermohaline circulation at 26.5°N in the subtropical North Atlantic, J. Phys. Oceanog., 26, 962-983.

Le Provost C., and M. Bremond, 2003: Resolution needed for an adequate determination of the mean ocean circulation from altimetry and an improved geoid, Space Science Reviews, 108, 163-178, 2003, In Earth Gravity Field from Space - From Sensors to Earth Sciences, Kluwer Academic Publishers, 2003, Netherland.

Levitus S., 1998: Word Ocean Atlas 1998 series (WOA98), CDROM set, Ocean Climate Laboratory, National Oceanographic Data Center.

Maltrub M.E., R.D. Smith, A.J. Semtner, R.C. Malone, 1998: Global eddy-resolving ocean simulations driven by 1985-1995 atmospheric winds, J. Geophys. Res., 103, C13, 30825-30853.

MEDAR/MEDATLAS Group, MEDAR/MEDATLAS 2002 database. Cruise inventory, observed and analysed data of temperature and bio-chemical parameters (4 CD-Roms), 2002.

Meincke J., C. Le Provost and J. Willebrand, 2001: Dynamics of the North Atlantic circulation: simulation and assimilation with high resolution models (DYNAMO), Prog. in Oceanog., 48, N0 2-3, 121-335.

Paillet J. and H. Mercier, 1997: An inverse model of the eastern North Atlantic general circulation and thermohaline ventilation, Deep Sea Res., 44, 1293-1328.

Reynaud T., P. Legrand, H. Mercier and B. Barnier, 1998: A new analysis of hydrographic data in the Atlantic and its application to an inverse modeling study, International WOCE Newsletter, 32, 29-31.

Richardson AP.L., 1983: Eddy kinetic energy in the North Atlantic from surface drifters, J. Geophys. Res., 88, 4355-4367.

Richardson AP.L., 1985: Average velocity and transport of the Gulf Stream near 55°W, J. Marine Res., 43, 83-111.

Rio M.-H. and F. Hernandez, 2005: A Mean Dynamic Topography computed over the world ocean from altimetry, in-situ measurements and a geoid model, J. Geophys. Res., submitted.

Roemmich D. and C. Wunsch, 1985: Two transatlantic sections: meridional circulation and heat flux in the subtropical North Atlantic. Deep Sea Research, 33, 619-664.

Schmitz W.J., 1984: Abyssal eddy kinetic energy in the North Atlantic. J. Mar. Res., 42, 509-536.

Schmitz W.J., 1996: On the World Ocean Circulation: Volume 1, some global features of the North Atlantic Circulation. Woods Hole Oceanographic Institution Technical Report WHOI-96-03, 141 p.

Schott F., J. Fischer and L. Stramma, 1998: Transports and pathways of the upper layer circulation in the western tropical Atlantic, J. Phys. Oceanog, 28, 1904-1928.

Siedler G., Kuhl A., Zenk W., 1987: The Madeira Mode Water, J. Phys. Oceanog., 17, 1561-1570.

Smedstad O. M., H.E. Hurlburt, E.J. Metzger, R. C. Rhodes, J. F. Shriver, A. J. Wallcraft and A. B. Kara, 2003: An operational eddy resolving 1/16 degree global ocean nowcast/forecast system. J.Marine Systems, 40-41, 341-361.

Stramma L. and F. Schott, 1996: Western Equatorial circulation and interhemispheric exchange. In Krauss, The Warmwatersphere of the North Atlantic Ocean (Chapter 7) Berlin-Stuttgart: Gebrüder Bornträger.

Sy A., 1988: Investigation of large scale circulation patterns in the central North Atlantic: the North Atlantic Current, the Azores Current, and the Mediterranean Water Plume in the area of the Mid-Atlantic Ridge. Deep Sea Res., 35, 383-413.

Tokmakian R., and J.L. McClean, 2003: How realistic is the high-frequency signal of a 0.1° resolution ocean model?, J. Geophys. Res., 108(C4), 3115, doi:10.1029/2002JC001446.

Townsend T.L., H.E. Hurlburt and P.J. Hogan, 2000: Modeled Sverdrup flow in the North Atlantic from 11 different wind stress climatologies, Dyn. Atmosph. Oceans, 32, 373-417.

Willebrand J., B. Barnier, C. Boning, C. Dieterich, P.D. Killworth, C. Le Provost, Y. Lia, J.M. Molines and A. L. New, 2001: Circulation characteristics in three eddy-permitting models of the North Atlantic, Progress in Oceanography, 48 (2001), 123-161.

Worthington L.V., 1959: The 18ºC water in the Sargasso Sea. Deep Sea Research, 5, 297-305.

Chapter 21

OPERATIONAL OCEANOGRAPHY IN THE U.S. NAVY: A GODAE PERSPECTIVE

John Harding and James Rigney

U.S. Naval Oceanographic Office, Stennis Space Center, Mississippi, USA

Abstract: Operational oceanography, throughout its long history, has continually evolved in order to remain relevant to the changing operational needs of the U.S. Navy. This transformation continues today as operational oceanography plays a key role in the rapidly shrinking tactical decision timelines of the naval warfighter. Understanding the warfighter's tactical questions and translating the information available from the ocean sciences into tactically useful answers is key to this transformation. Appreciating and applying advances in oceanography, such as those provided by the international Global Ocean Data Assimilation Experiment (GODAE), assures the highest technical quality of these answers. We address these transformation and translation processes and the nature of U.S. Navy operational oceanography today with specific emphasis on its interaction with GODAE.

Keywords: Operational oceanography, GODAE, U.S. Navy.

1. Introduction

Transformation, in response to the U.S. naval warfighter's changing environmental needs, captures the essence of the evolution of U.S. Naval operational oceanography. From its 19th century genesis under Matthew Fontaine Maury and his immediate predecessors to the 21st century needs of today, operational oceanography has continued to adapt. U.S. naval oceanography first grew from Navy concerns for safety of navigation, more specifically the creation and archiving of navigational charts.

By the middle of the 20th century, World War II required increasing oceanographic support for mine warfare (MIW), amphibious warfare (AW), and antisubmarine warfare (ASW) (Pinsel, 1982). With the threat of undersea launched nuclear missiles, ASW became the dominant concern of naval operational oceanography in the last half of the 20th century through the end of the Cold War in the late 1980s.

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 467-482.

With the Cold War over, Navy leadership again reemphasized the importance of nearshore warfare requirements (Dalton et al., 1994; Kelso et al., 1992) such as MIW and AW and with a growing important role for naval special warfare (NSW). ASW, no longer the dominant warfare area, remained a significant requirement and again has arisen in importance in the 21st century (Kreisher, 2004). Technological change has greatly influenced the nature of these transformations. The longer time scales of navigational charting and ocean data base creation, while still important, are being superseded by more immediate response capabilities where near-real-time oceanographic knowledge can now provide relevant environmental information within the warfighter's tactical decision loop (Burnett et al., 2002).

Key to this faster response time is the operational oceanographer's ability to understand the language and needs of the warfighter. This ability allows the translation of oceanographic knowledge into operationally significant information, allowing the warfighter to more efficiently and safely perform his job.

Participation by the U.S. Navy in the Global Ocean Data Assimilation Experiment (GODAE) is a key element in providing these relevant, real-time products. While the regions within ~200 km of coastal regions are presently major concerns for the U.S. Navy, the GODAE emphasis, namely the creation of assimilative global and regional ocean prediction capability (Smith and Lefebvre, 1997), is also required. These larger domain prediction systems can supply important information to open-ocean ASW needs, provide boundary conditions to high resolution coastal models, and eventually supply global capability of sufficiently high horizontal resolution to be directly applicable to the warfighter in nearshore areas.

2. What is operational oceanography?

For the U.S. Navy, operational oceanography encompasses the fields of physical and optical oceanography, hydrography, bathymetry, acoustics, and marine geophysics. However, we will limit our discussion mainly to that sub-discipline of primary interest to GODAE, namely physical oceanography.

"Providing relevant oceanographic knowledge to the warfighter" succinctly summarizes the goal of U.S. Navy operational oceanography. To accomplish this requires smart data collection, focused analysis, and responsive delivery. The essential data must be collected either ahead of an operation, based on anticipated operational needs, or in near-real-time, directly associated with a specific operation. The specific analysis may focus on direct processing of these data via either statistical methods or more effectively through their assimilation into dynamic ocean models, thus resulting in an ocean prediction system. Finally, the results of this analysis

must be returned to the warfighter within his decision timelines and in a form that he can readily understand and use.

How is operational oceanography distinct from oceanography performed in a research and development (R&D) environment? The U.S. naval operational oceanographer has a defined customer, namely the fleet user, whereas the R&D oceanographer focuses more on the scientific question(s) he seeks to answer. The operational product supports actual naval operations and exercises. The warfighter expects a timely and useful product derived from specific product requests pertaining to his operation. The operational oceanographer typically must rely on "operational" vice "research" quality data when attempting to answer a warfighter question. That is, he must make use of the limited data available rather than have the advantage of a systematic, preplanned research effort targeted at a specific scientific question. Similarly, the operational oceanographer must typically rely on rapid analyses that, given the limited data available, may not approach the statistical rigor of analyses required of publication in scientific journals. Finally, the operational oceanographer requires the capability for rapid, near-real-time monitoring and assessment of ocean product performance in order to quickly gauge its value in answering a warfighter request.

Who, more specifically, is this warfighter requesting environmental information and what sort of questions might he ask? Naval Special Clearance Team swimmers engaged in neutralizing mines in a nearshore MIW operation may ask if their operation will be limited by currents too strong or water depths too shallow. This holds true for any swimmer or low speed vehicle in the water, such as an NSW swimmer approaching a harbor or a deployed autonomous vehicle measuring nearshore bathymetry. Navy surface ships evacuating noncombatants from hostile territory, amphibious ship commanders planning amphibious assaults, or the joint (Army/Navy) warfighter planning logistics support across the beach after a successful amphibious operation may all ask if their respective operations will likely be adversely affected by surface waves and surf too large or tides too low for safe operations. ASW or MIW sonar operators searching for an adversary's submarines or mines, respectively, may ask for the marine acoustic environment to best understand detection ranges or even what search patterns to conduct. Submarine operators or NSW swimmer delivery vehicle drivers may want to know the threat of broaching due to internal wave or soliton-induced density changes. An NSW swimmer might ask whether bioluminescence will compromise his surreptitious approach to a harbor by leaving a luminous wake trailing behind as he approaches shore. This limited sampling of warfighter questions, while only beginning to address the numerous possibilities of environmental impact on naval operations, provides a flavor of the potential impact of operational oceanography.

3. U.S. NAVY operational oceanography structure

Burnett et al. (2002) provides a comprehensive description of how the U.S. Navy conducts operational oceanography with two production centers and several geographically-distributed regional centers and detachments associated with the various U.S. fleets around the world. In keeping with the history of changing requirements and the necessity to provide ever more user-focused, ocean information within the warfighter's decision loop, this geographic-centered approach began changing in late 2004 (Rear Admiral Timothy McGee, Commander, Naval Meteorology and Oceanography Command (CNMOC), personal communication). From its former regional product delivery structure, U.S. naval operational oceanography is now transforming into a more centralized, efficient business model specifically focused on nine warfare (or warfare-supporting) disciplines. These include ASW, NSW, MIW, and Navigation, each with its own military program director. The Naval Oceanographic Office (NAVOCEANO), the primary production center for Navy oceanographic products, is already structured by technical discipline and matrixed by warfare area thus allowing it to readily adapt to this new approach. This current structure is shown in figure 1.

Both short-term requests from naval warfighters as well as longer term requirements, based on future anticipated warfighter needs, are coordinated through the Plans, Programs, and Requirements Department, currently the unifying node shown at the center of figure 1. NAVOCEANO warfare area program managers currently within this department work with the CNMOC warfare area directors to coordinate the specific warfighter requirements with the remainder of the NAVOCEANO, discipline-focused departments. The following description provides a brief synopsis of the various department activities with more departmental detail available at https://www.navo.navy.mil/pao/other/departments.htm.

From a GODAE-centric view, the Oceanography Department most closely overlaps with GODAE-related activities by collecting and processing real-time oceanographic data and assimilating these data into ocean prediction models. This, however, is only a limited subset of the department's activities that will be related in more detail later. The Hydrography Department primarily collects and processes both shallow water and deep water bathymetry to support safety of navigation but also provides the bathymetric data bases specifically required for ocean prediction models. The Geophysics/Acoustics Department collects and processes data on water column and bottom properties primarily related to ASW and MIW and is therefore a consumer of the GODAE-related oceanographic predictions from the Oceanography Department. The Warfighting Support Center primarily engages in remotely-sensed imagery analysis but also creates composite products, especially of use to NSW, incorporating Oceanography Department data bases and model output. The Ocean Projects Department collects and analyzes targeted environmental

data focused on short-term projects with special emphasis on exploring future technologies applicable to the other departments. One such example is the use of long-range autonomous underwater vehicles for potential applications in order to increase hydrographic survey efficiencies. The final three technical departments serve primarily in critical support roles to the others. The Survey Operations Center coordinates the use of NAVOCEANO's major collection assets including the seven-ship survey fleet as well as the high-bandwidth, near-real-time ingest of survey data from the ships. Survey Operations also coordinates both shipboard and airborne buoy deployments including the NAVOCEANO participation in the international ARGO drifting buoy program (Roemmich and Owens, 2000). The Engineering Department provides support to all the sensors used on the survey ships, coordinates the primary in-house information technology support, and also maintains the environmental software libraries designed for on-scene use.

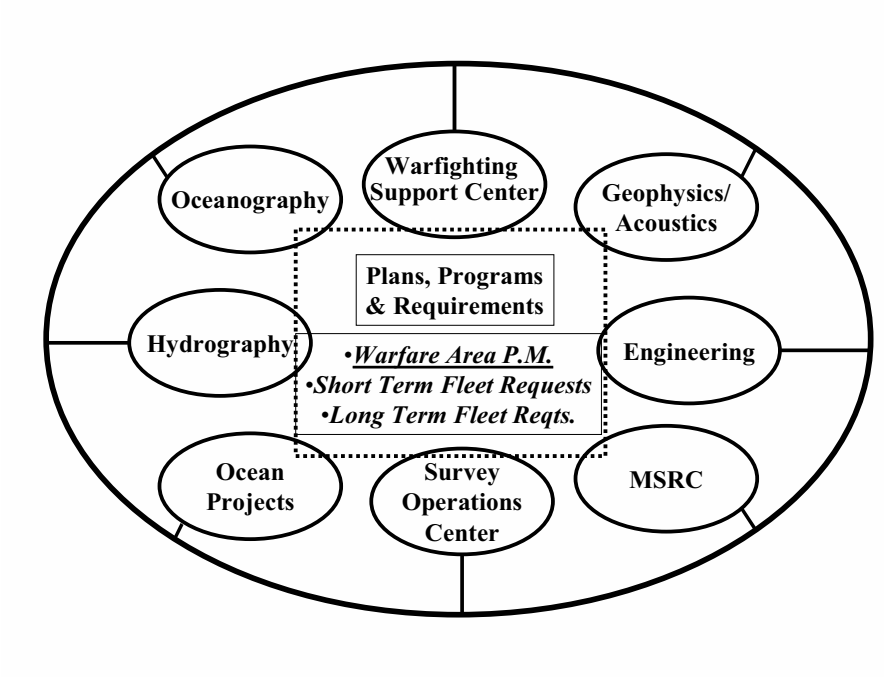

Figure 1. Current NAVOCEANO departmental structure designed to coordinate warfighter-related "business lines" to ocean science disciplines.

The final department deserves special attention as it provides the critical computing capability required to conduct GODAE-related, real-time ocean predictions within the Oceanography Department. The Major Shared Resource Center (MSRC) is a U.S. Department of Defense R&D high performance computing (HPC) activity housed at NAVOCEANO. NAVOCEANO manages this facility and, in return for a 15% Navy investment, has use of 15% of this terascale computing resource for its

operational oceanographic applications, primarily numerical ocean prediction. The MSRC is one of the most capable HPC environments in the world and, given its bi-annual hardware refresh rate, regularly ranks in the top 10 computer centers globally. This facility serves a nationwide user community of over 4,000 scientists and engineers and currently includes a cumulative 30 teraop computing resource combined with petabyte hierarchical storage and multi-gigabit local/wide area network.

4. NAVOCEANO oceanography and GODAE

The Oceanography Department's efforts in ocean prediction depend on numerous interrelated factors outlined in figure 2. The primary technology transfer of ocean prediction capability transitions via the research and development activities of the Naval Research Laboratory Oceanography Division. These activities in turn take advantage of the larger investment in academic science and technology funded via the Office of Naval Research and others (Harding et al., 1999).

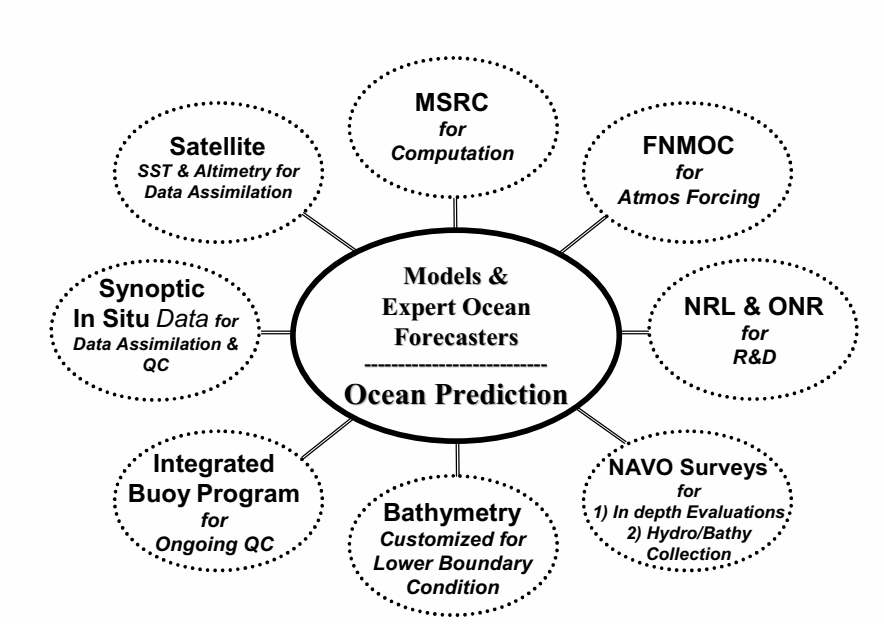

Figure 2. Operational ocean prediction relationships and dependencies.

Oceanographic information collected during NAVOCEANO oceanographic surveys provides data for in-depth evaluations of ocean prediction products as well as the creation of ocean climatologies used directly or indirectly by the forecast systems. Bathymetry comes from the data collection, processing, and data basing by the Hydrography Department as referenced earlier. The Integrated Drifting Buoy Program of the Survey

Operations Center provides data for ongoing quality control of prediction products. Synoptic data available from both the World Meteorological System Global Telecommunication System (GTS) as well as directly from Navy ships and aircraft provide measurements available for data assimilation and real-time quality control. Remote-sensing data, primarily multi-channel sea surface temperature (MCSST) and altimetry from both national and international satellites, yield wide-area, near-real-time information available for data assimilation. The MSRC has already been noted as the primary source for HPC computing for the large-scale prediction systems. Fleet Numerical Meteorology and Oceanography Center (FNMOC), the sister production center to NAVOCEANO, provides the global and regional atmospheric forcing to drive the ocean models (Rosmond et al., 2002; Hodur et al., 2002).

4.1 Ocean data collection

Oceanographic data serve multiple purposes as applied to the ocean prediction challenge: for building ocean climatologies, for creating synoptic analyses, for assimilating directly into dynamic models, and for providing quality control and evaluation information for ocean prediction systems. Conductivity/Temperature/Depth (CTD) probes, expendable bathythermographs (XBT), AC9 optical probes, Acoustic Doppler Current Profilers (ADCP), and conventional current meters used during a typical NAVOCEANO oceanographic survey mission provide primarily the data supporting the production of ocean climatologies and to a lesser extent, evaluation of ocean prediction products.

Real-time, remote *in situ* data collection supports not only the longer term input for climatologies but also, more importantly, the data required for synoptic statistical analyses and assimilation into real-time Navy prediction systems. This data collection effort includes the Integrated Drifting Buoy Program that makes use of MiniMet, WOCE, APEX, and Davis Drifters to relay real-time oceanographic data back to NAVOCEANO. This program also represents the NAVOCEANO participation in the international ARGO profiler program. Real-time XBT data are accessed directly from Navy aircraft and ships while XBT, CTD, ADCP, and mooring data available from other national and international sources are accessed via the GTS.

Finally, remotely-sensed SST, altimetry, and ocean color measurements from satellite supply data for analyses, assimilation, evaluation and quality control. Polar orbiting and geostationary satellites provide MCSST data (Robinson, 2005, this volume). Altimetry data, processed within the Oceanography Department Altimetry Data Fusion Center, provide sea surface height (SSH) data presently from the Navy GEOSAT Follow On (GFO), NASA/CNES JASON, and ESA ENVISAT satellites (Jacobs et al., 2002). Finally, while not yet assimilated, ocean color data yield useful

synoptic analyses as well as near-real-time evaluation capability for certain dynamic features in the ocean prediction models.

4.2 Data processing and relevance to NAVY applications

As noted in the introduction, the tendency for U.S. naval operational oceanography has been toward more synoptic and real-time capabilities with less emphasis on ocean climatologies that support longer-term requirements. In this light, there has been growing emphasis toward surge efforts directly supporting warfare areas such as ASW, NSW, and MIW. Today, emphasis is placed on rapid response capabilities for description of ocean parameters such as currents, temperature and salinity, waves, and optical properties. A key change is the emphasis on including forecasters in the process in order to interpret prediction model output and to generate targeted products of immediate use to the warfighter.

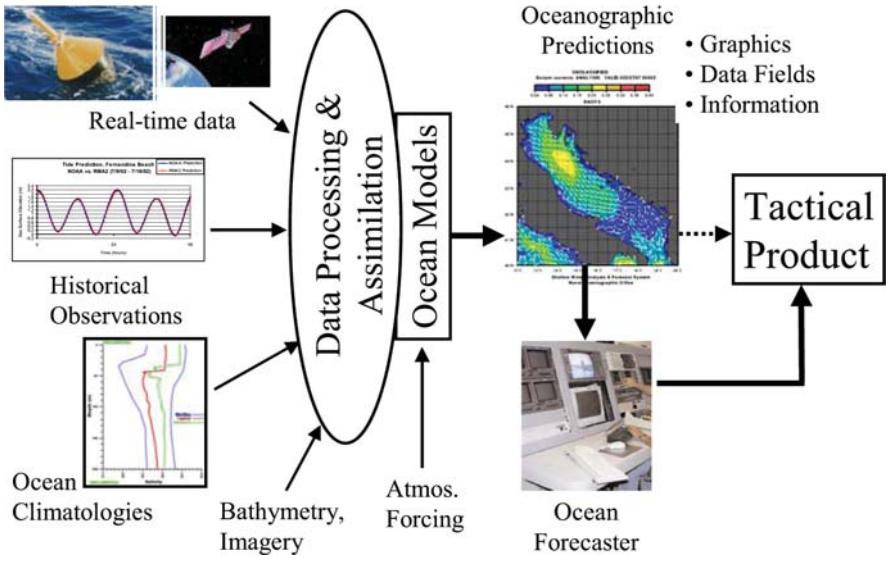

Figure 3. Components and flow of the operational ocean forecasting process emphasizing the role of the ocean forecaster.

The forecaster role within the larger ocean prediction process is more clearly delineated in figure 3, illustrating the flow from data through model/data assimilation to oceanographic products. The tactical product that provides a relevant answer to the warfighter question is best filtered through a knowledgeable operational ocean forecaster. Recalling the warfighter questions identified earlier, what are some of the specific naval applications of these ocean prediction models? A sampling includes currents and depths

for diver/swimmer mission windows; currents for underwater vehicle operations, mine drift, and scour prediction; waves and tides for NSW small boat and amphibious operations; currents and waves for oil/contaminant spill prediction; and sound speed for mine and submarine detection.

Drift issues were of particular concern during operations in the Persian Gulf both in 2003 as well as in the early 1990s. At that time, the NAVOCEANO Shallow Water Analysis and Forecast System (Horton et al., 1994), based on the sigma coordinate Princeton Ocean Model design of Blumberg and Mellor (1987), provided outputs allowing both forecasts and hindcasts of objects or substances in the water. Applied to an oil drift problem in the Persian Gulf in the early 1990s, the SWAFS combined with the Navy regional atmospheric prediction capability (Hodur et al., 2002) and the NOAA oil fates code GNOME, provided guidance in protecting Saudi desalinization facilities.

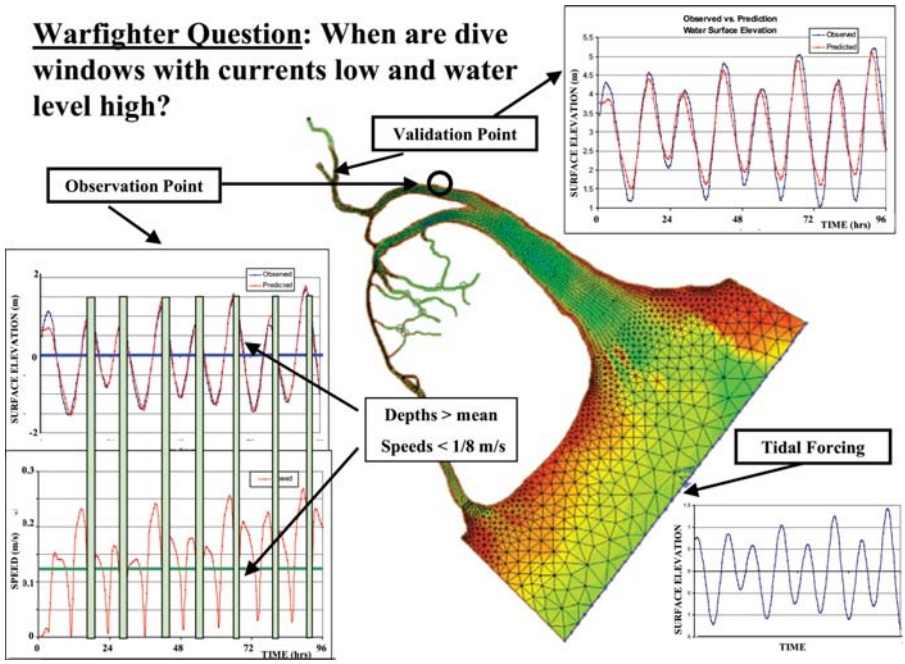

Figure 4. Answering warfighter questions - an example from counter-mine warfare. When are my safe dive windows?

During Operation Iraqi Freedom, a tidally-driven, finite-element, riverine/coastal model in use at NAVOCEANO, developed by the U.S. Army Corps of Engineers, provided guidance for appropriate dive windows for mine clearance divers to remove tethered mines that were restricting relief supplies from reaching civilians in Um Qsar. The upper right-hand side of figure 4 illustrates the good agreement of modeled tide heights relative to those from a historical tide gauge station. The lower left-hand side of figure 4 illustrates, with the vertical shaded bands, those time periods

when speeds would be low and water level high in order to provide the safest environment for the divers responsible for eliminating the mine threat. As noted earlier, R&D-quality evaluation data are rarely available in this sort of operational situation; however, anecdotal feedback from the MIW squadron responsible suggested this environmental information reduced the mine countermeasure operations by 2 to 3 months (Steve Haeger, NAVOCEANO, personal communication).

4.3 Link to GODAE

While U.S. naval oceanography has a major focus on nearshore, high-resolution capabilities, a key issue connects it to the GODAE effort. The U.S. Navy has global responsibilities with different geographic areas having potentially different dominant ocean dynamics affecting the warfighter. For instance, in a limited domain such as the northern Persian Gulf, tides and wind-forced currents may be the dominant oceanographic impact on NSW diver operations. In a more open coastal area like the U.S. west coast, front and eddy features associated with mesoscale dynamics or coastally trapped waves generated by meteorological events hundreds or even thousands of kilometers distant may add significant time-dependent variability that can affect a coastal mission. Until computer resources become sufficient to run a global model at resolutions to accurately describe the required dynamics, a nesting approach (figure 5) is required. Burnett et al., (2002) provides a table of current operational or near-operational prediction systems, from ocean currents to thermal structure to surface waves, and how they are used. The present global, dynamic ocean prediction capability is based on a combination of the Navy Layered Ocean Model (NLOM) and the global application of the Navy Coastal Ocean Model (NCOM) as described by Rhodes et al., (2002). By 2007, NAVOCEANO plans to operationally run a system based on the Hybrid Coordinate Ocean Model (HYCOM) as described by Chassignet et al.,(2005, this volume). The present operational, regional capability, SWAFS, was described earlier with several regional applications of NCOM planned for transition. For high-resolution coastal domains, NAVOCEANO presently uses several models with the expected future direction focusing increasingly on the finite element approach used in the ADvanced CIRCulation model (ADCIRC) (Blain et al., 2002). In general, the Navy must be prepared to take a global view to get a coastal environmental description.

An obvious example is the propagation of swell energy generated from southern Pacific storms that can affect surf forecasts on the U.S. West Coast. Westward propagating mesoscale eddies can bring open-ocean mesoscale energy into the South China Sea through the Luzon Strait from the Pacific (Metzger and Hurlburt, 2001) or affect the southeast coast of the Arabian Peninsula after propagating westward across the Arabian Sea. Global prediction capabilities are also important to operational forecasters in areas

where a rapid answer may be required and a higher resolution coastal model is not available.

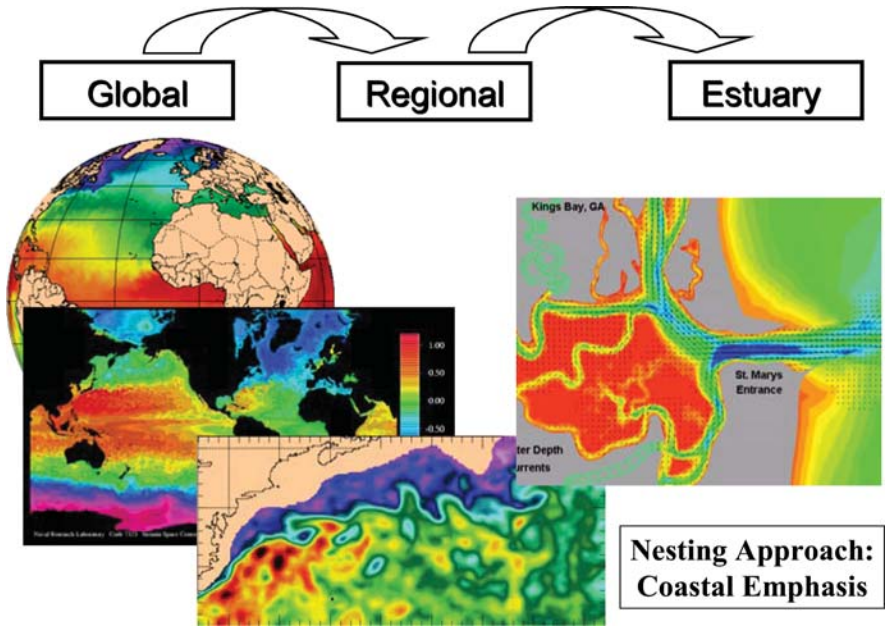

Figure 5. The U.S. Navy nested model approach to ocean prediction.

Until computing resources allow operational global models to attain sufficiently high resolution, large-area regional models, also related to the GODAE effort, are equally important. A real-time Gulf of Mexico/Caribbean NCOM, currently run in R&D context by NRL and planned for transition to NAVOCEANO, supplies an example. Dr. Dong Shan Ko of NRL provides skill assessments of computed storm surge by comparing model-predicted coastal ocean heights to independent tide gauge measurements along the northern Gulf. Ko also demonstrates predictability of the mesoscale eddy structure of the Gulf of Mexico/Caribbean by comparing it to independent ocean color data. For more detail see www.oceans.nrlssc.navy.mil/iasfcst.

Fox et al. (2002) provides examples of the impact of such mesoscale oceanography on both acoustic ray paths and propagation loss, demonstrating the difference when one uses a climatological data base compared to the current real-time operational statistical interpolation system. When crossing a Gulf Stream-like frontal feature, the real-time result provides increased probability of detection due to bottom bounce ray paths while at the same time potentially missing a detection of a near-surface adversary as the surface acoustic duct disappears.

Internal waves/solitons can also have a significant impact, altering surface layer depth both temporally as they propagate through an area and

directionally depending if acoustic propagation is along or across crests. Sperry et al. (2003) describe "interesting time-and-range-variable population of the acoustic normal modes" due to tidally generated soliton packets in the shelf-slope region off the eastern United States. Ramp et al. (2002) present the interesting result that some of the largest soliton signals in the northeastern South China Sea are generated, not locally in the South China Sea but in the Luzon Strait, another example of the importance of remote forcing. NAVOCEANO, in partnership with ONR and Dave Fratantoni of Woods Hole Oceanographic Institution, deployed Slocum Gliders in the Philippine Sea in 2004. They found internal waves to be a notable temporal signal in the upper ocean thermal structure in the open western Pacific with significant impact on the sound speed structure of the upper ocean (Lorens, 2004, personal communication). Figure 6 compares the primarily tidal temporal variability associated with one such glider with a regional East Asian Seas NCOM including tidal forcing and global NCOM without tides. The overall qualitative agreement of the global model with the measured thermal structure is reasonable, although the thermostad between 50-100 m in the data is not well represented. The regional model provides a better comparison, albeit again without the thermostad, where the tidal signal begins to appear but without the expected amplitude.

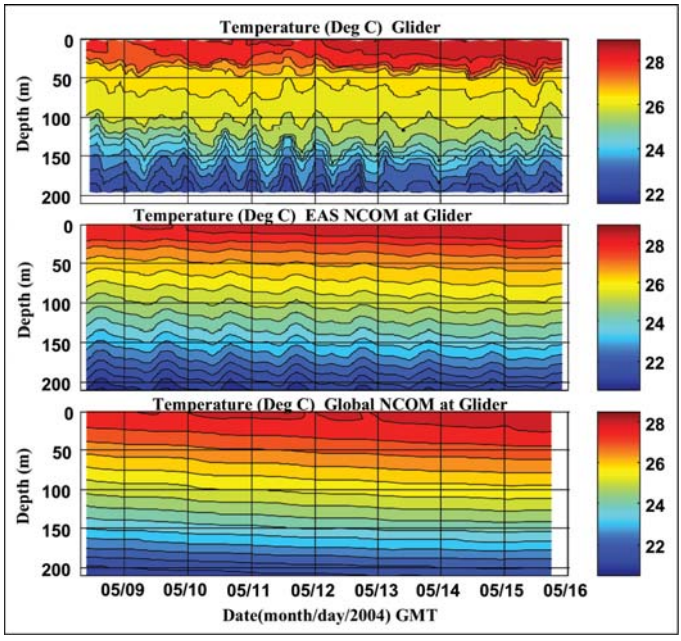

Figure 6. Measured western Pacific temperature profiles vs. time compared to global NCOM, without tidal input, and a regional NCOM, including tides, illustrating the need for additional ocean prediction R&D (Courtesy Dr. Rob Lorens, Oceanography Department, NAVOCEANO).

Higher horizontal resolution and especially improved data assimilation may be required to better resolve what appears to be the result of a tidal/topographic interaction (figure 7). A non-assimilative, regional, western Pacific NCOM tidal simulation with ~5 km resolution suggests that topographic interaction is the source for these open-ocean internal wave signals (Paul Martin, NRL, personal communication). Initialized with representative uniform stratification and forced solely by barotropic tides, the simulation yields reasonable amplitudes and complex interference patterns in the 300 m temperature with the generation zones appearing at the locations of significant bathymetric features. Panel (a) represents the initiation of the tide forcing with panels (b-d) representing each subsequent 48 hours. Note especially the strong variability associated with the Luzon Strait consistent with the strong South China Sea soliton generation region proposed by Ramp et al. (2002). The strong signals in the open waters of the Philippine Sea are similarly consistent with the temporal thermocline variability in figure 6.

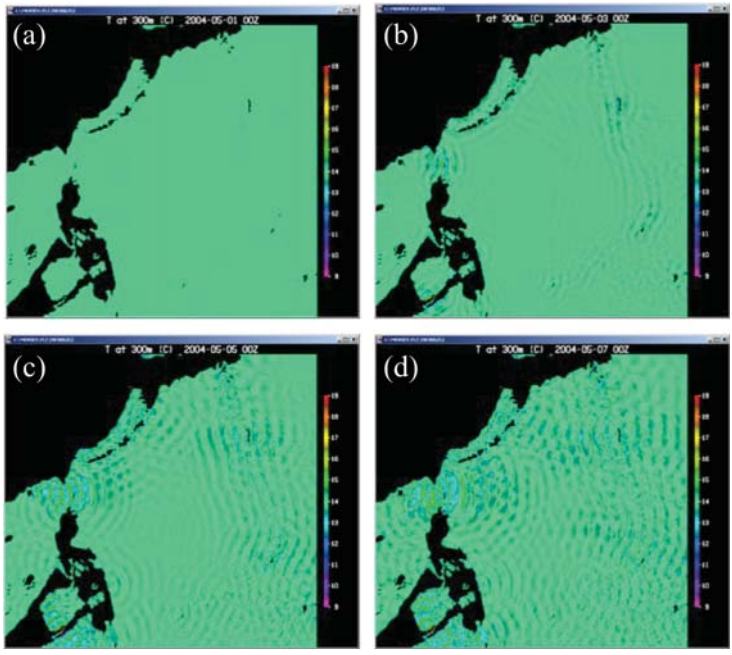

Figure 7. Western Pacific high resolution NCOM initialized with a single representative temperature/salinity profile and forced solely by tides. Panel (a) represents the initiation of the tide forcing with panels (b-d) representing each subsequent 48 hours. (Panels are extracted from an animation courtesy of Paul Martin, Oceanography Division, NRL).

5. Research & development needs

These last examples are specifically provided to illustrate that, while useful to an operational oceanographer, the present operational and near-operational capabilities still require significant R&D effort to realize their full potential. These efforts can be roughly divided into four areas.

The first area includes data collection, fusion, and exploitation, focused primarily on coastal domains around the globe but including deep water areas as well. These efforts include the development of new sensors, sensor technologies, platforms (e.g., autonomous underwater vehicles), and adaptive sampling techniques to name a few. Improving techniques to extract remotely sensed data near coasts would serve to improve the limited volume of data now available for assimilation into coastal models. Similarly, any development that improves coastal ocean data sharing and communications could increase the quantity of available data on the GTS.

Second, high-resolution coastal analysis and prediction efforts are needed to specifically address improved coastal and estuarine models and assimilation techniques. One key area is the need for improved quality control techniques for coastal data to minimize the amount discarded. This would partially alleviate the lack of adequate coastal information available for assimilation and validation in many regions.

Third, and most closely related to GODAE, are the global/regional analysis and prediction R&D efforts targeting improvements to global and regional models and assimilation techniques. The need to account for dynamic processes with high temporal and spatial variability, as highlighted at the end of the previous section, provides a notable example.

The fourth and final area covers real-time and near-real-time evaluation, visualization, and applications. The operational oceanographer needs tools to rapidly evaluate prediction system output in order to assign confidence measures to any product sent to the warfighter. Direct warfighter applications are also required, such as the drift tools for search and rescue.

6. Summary

U.S. naval operational oceanography has a long history of evolving as different ocean warfare challenges arise, fade, and rise again. The role of the operational oceanographer is to translate the available oceanographic knowledge into accurate information relevant to the warfighter, namely within his tactical decision loop and specifically focused on his particular warfare challenge. Finally, GODAE plays a key role providing state-of-the-science, global capability supporting open-ocean warfighter issues including ASW, as well as providing boundary conditions to high resolution, coastal prediction systems applicable to those warfare areas with greater shallow water emphasis such as NSW, AW, and MIW.

Acknowledgements

The authors greatly thank Dr. Rob Lorens of the NAVOCEANO Oceanography Department for providing the basis for figure 6 and Paul Martin of the NRL Oceanography Division for the animation used to create figure 7. Special thanks are due to Veronica Nichols of the NAVOCEANO technical publications group for her extensive editing of the original manuscript.

References

Blain, C.A., Preller, R.H., and Rivera, A.P., 2002, Tidal prediction using the Advanced Circulation Model (ADCIRC) and a relocateable PC-based system, *Oceanography*, 15: 77-87.

Blumberg, A.F. and Mellor, G.L., 1987, A description of a three-dimensional coastal circulation model, In: *Three-Dimensional Coastal Circulation Models*, ed. N. Heaps, pp. 208, Washington, American Geophysical Union.

Burnett, W., Harding, J., and Heburn, G., 2002, Overview of operational ocean forecasting in the U.S. Navy: Past, present & future, *Oceanography*, 15: 4-12.

Dalton, J.H., Boorda, J.M., and Mundy, C.E., 1994, Forward… From the sea, *Proceedings of the United States Naval Institute*, December: 46-49.

Fox, D.N., Barron, C.N., Carnes, M.R., Booda, M., Peggion, G, and Van Gurley, J, 2002, The Modular Ocean Data Assimilation System, *Oceanography*, 15: 22-28.

Harding J.M., Carnes, M.C., Preller, R.H. and Rhodes, R., 1999, The Naval Research Laboratory role in naval ocean prediction. *MTS Jour*., 33, 67-79.

Hodur, R.M., Pullen, J., Cummings, J., Hong, X., Doyle, J.D., Martin, P., Rennick, M.A., 2002, The Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS), *Oceanography*, 15: 88-97.

Horton, C., Clifford, M., Schmitz, W., Hester, B., 1994, SWAFS: Shallow Water Analysis and Forecast System overview and status report, Technical Report, Naval Oceanographic Office, Stennis Space Center, MS 39522, pp. 53.

Jacobs, G.A., Barron, C.N., Fox. D.N., Whitmer, K.R., Klingenberger, S., May, D., and Blaha, J.P., 2002, Operational altimeter sea level products, *Oceanography*, 15: 13-21.

Kelso, F.B., Mundy, C.E. and O'Keefe, S., 1992, … From the sea. *Proceedings of the United States Naval Institute*, November, 93-96.

Kreisher, O., 2004, An underwater threat re-emerges, Navy renews emphasis on ASW, *Sea Power*, October, http://www.military.com/NewContent/0,13190,NL_ASW_100404-P1,00.html.

Metzger, E.J. and Hurlburt, H.E., 2001, The importance of high horizontal resolution and accurate coastline geometry in modeling South China Sea inflow, *Geophys. Res, Let.*, 28:1059-1062.

Pinsel, M.I., 1982, *150 Years of Service on the Seas, A Pictorial History of the U.S. Naval Oceanographic Office*, Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, pp. 56-69.

Ramp, S.R., Chiu, C.S., Bahr, F.L., Lynch, J., Duda, T., Tang, D., Liu, A.K., 2002, Tracking the generation sites and packet variability of internal solitons in the south China Sea, *J. Acoust. Soc. Am.*, 12: 2449.

Rhodes, R.C., Hurlburt, H.E., Wallcraft, A.J., Barron, C.N., Martin, P.J., Metzger, E.J., Shriver, J.F., Ko, D.S., O.M., Cross, S.L., and Kara, A.B., 2002, Navy real-time global modeling systems, *Oceanography*, 15: 29-43.

Roemmich, D. and W.B. Owens, 2000, The ARGO Project: Global ocean observations for understanding and prediction of climate variability. *Oceanography*, 13, 45-56.

Rosmond, T.E., Teixeira, J., Peng, M., Hogan, T.F., and Pauley, R., 2002, Navy Operational Global Atmospheric Prediction System (NOGAPS), *Oceanography*, 15: 99-108.

Smith, N, and Lefebvre, M., 1997, The Global Ocean Data Assimilation Experiment (GODAE). Monitoring the oceans in the 2000s: an integrated approach. *International Symposium*, Biarritz, October 15-17.

Sperry, B.J., Lynch, J.F., Gawarkiewicz, G., Chiu, C.S., Newhall, A., 2003, Characteristics of acoustic propagation to the eastern vertical line array receiver during the summer 1996 New England shelfbreak PRIMER experiment, *IEEE/JOE*, 28: 729-749.

Chapter 22

APPLICATIONS OF OCEAN FORECAST INFORMATION FOR ECONOMIC ADVANCEMENT IN DEVELOPED AND DEVELOPING SOCIETIES

Mary Altalo

Enterprise Solutions, Science Application International Corporation, U.S.A,/UK, and The London School of Economics, UK

Abstract: It is critical that the new environmental forecast information resulting from the full deployment of the Global Observing System be fully assimilated and exploited to serve the needs of societies and economies. Both developed and developing nations require the information to be "mainstreamed" into the business practices of their major industries and development programs. Providers as well as the users of the information must partner to create a "voice" to policy makers to support the global vision. Understanding the various industry drivers for the use of the information can aid in defining the optimal requirements for the design and operation of a system configured to serve societies. All segments of the economies require this information for optimal management of their operations and for informed strategic planning. Energy and water utilities, construction companies, tourism industries, the financial services sector, healthcare operations, transport, food security as well as defense all use environmental forecast information for daily operations from revenue forecasting and load management to infrastructure siting and supply chain management The "value" of information can be determined by "industry trials" in developed as well as developing nations which demonstrate the impacts on the performance of the business, whether by increases in safety margins, enhancements in revenue generation or improvements in reliability. Cost estimates of savings are presented. The production of decision support systems for the industry from environmental information is critical to activate the information to serve societies needs. Dual sets of performance metrics, which measure the progress of both the technology deployment as well as the socioeconomic use in meeting their development goals is critical.

Keywords: Societal and business applications, weather, climate and ocean observation and information systems, decision support systems, economic impacts, industry trials, GEO, GEOSS.

E. P. Chassignet and J. Verron (eds.), Ocean Weather Forecasting, 483-505.

1. Introduction

The purpose of this chapter is to introduce the concepts behind the application of modeled observing system information to the needs of the various economic sectors of the global business community. We will use examples from developed and developing nations to demonstrate how the information is used and the process by which it is "mainstreamed" into the business practices of industry. We will introduce concepts of marketplace dynamics and their vulnerability to environmental variability.

There are a number of ideas that will be explored in this chapter. The first is related to the "provider push" of the information –that is the "build it and they will use it" approach, versus the "user pull" whereby businesses and economies truly recognize the importance of the information and create the demand for the data and create a "voice" to policy makers to support the provision of that data. The second concept that we wish to explore is that of understanding the various industry drivers for the use of the information—that is what makes them sit up and take notice that they need the information—this may be everything from regulation to good citizenship. The third concept is to introduce the various industry segments that are in need of this information. Since there has been much attention paid to natural resource management, marine transport and oil and gas operations, this chapter will concentrate on the less frequently examined areas of the economy such as regional utilities management, construction, and tourism. The fourth concept is an introduction to the approach that we use to ascertain and "value" how information is used and how it impacts the performance of business. Here the method of an "industry trial" or "beta testing" and audit analysis are introduced. In order to illustrate the method, a number of case studies will be explored with particular emphasis on the power industry, and the tourism or leisure industry. In the case of the power industry, examples are used from developed as well as developing nations. The next concept explored is the production of decision aids for the industry and methods for linking multiple decision support tools to optimize operational performance. Finally, the need to create a sound family of business plans including financial, marketing, governance, science etc in order to "execute" the transition is demonstrated.

2. Provider "push" vs. user "pull" for observing system information

The application of scientific information gleaned from the deployment of new technologies is often difficult. Part of the problem stems from the need for "translation" between the provider and user community, but even within the academic community, the disciplines required in this transition phase

may reside in different schools within the same university. In essence, the principles used to guide the transition are in the information technology discipline which includes the field of Decision Science. These principles are taught to a large degree within the schools of management, business, and to a lesser degree within the schools of government and economics. Concepts dealt with here are those of operations management and strategic planning so critical to the overall success of any organization. Also, if environmental information is to be mainstreamed within the company "process", organizational change may have to occur in which the structure of organizations changes to follow the information "flow". This is the subdiscipline of Business Process Reengineering. The last issue is one of market understanding. Virtually every business is impacted by weather. Obtaining enhanced atmospheric and ocean information to improve the meteorological forecast accuracy is of highest priority. This engages inland users as well as coastal users and forces the land bound nations to invest in global earth observing as well.

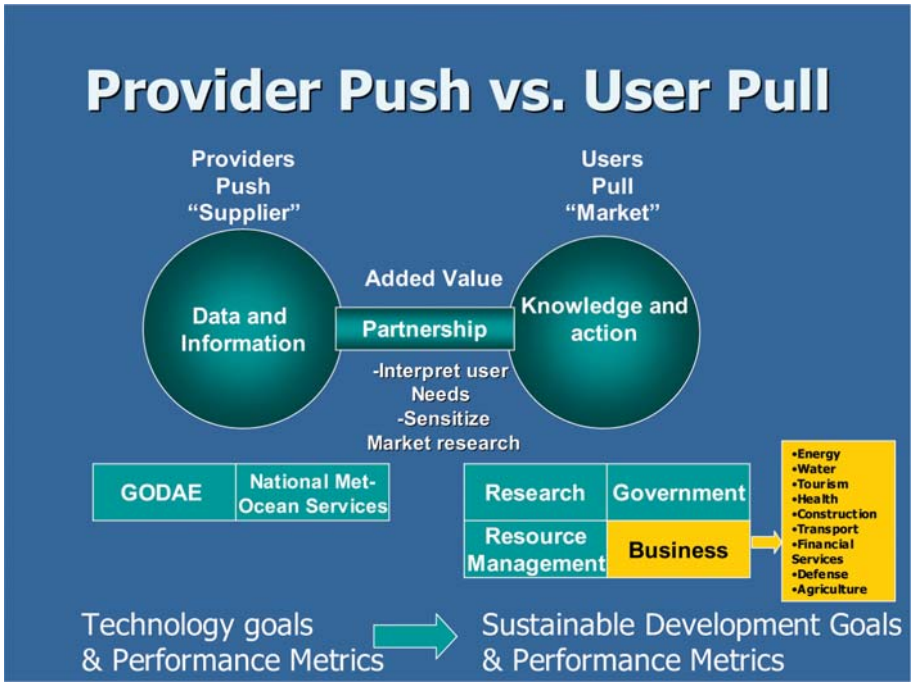

Figure 1. The polar relationship between the "provider" and the "user" of the Environmental Information. The business sectors requiring the information are shown at bottom right. Both technology as well as societal performance metrics must be applied criteria for success.

Figure 1 demonstrates the polar relationship between the data provider and the data user. Ideally, what we would like to create is a partnership between these two communities where users' needs are interpreted and market research is tapped. GODAE, other earth observing programs within

GEOSS (Global Earth Observing System of Systems) and The National Meteorological and Hydrological Services (NMHS) are on the provider side. Normally, governments and the research community are most often though of as the provider community. However, many of the private sector businesses also have observing systems which are increasingly being incorporated into the framework of the global observing system. The user community is a diverse one with the research, government, resource management and Business community all being considered users. The translation of the data and information into sector specific knowledge and action is the goal. Note that either "the push" or "the pull" can make this happen. The key is in the resourcing. As governments decide to endorse this effort, making the information from technology advancement relevant to their National development goals is needed. Producing and not distributing the information to societies is a serious omission resulting in widespread loss of human life and property (i.e., tsunami of December 2004).

The sustainability of each nation is tied to the development and best practices management of economies, the fair governance and protection of societies and the cultivation of a habitable environment. Only when all aspects of national well-being are addressed will Nations thrive in the Global context. Many nations of the world have set individual development goals to measure their progress. In addition, the United Nations has set global development goals, the Millennium Goals, for the alleviation of poverty, reducing spread of disease, reduction of childhood mortality, etc. The progress towards these goals needs to be continuously monitored, and in many cases, earth observations are needed to demonstrate this. Identifying and designating the earth observing system as part of a "Critical National and International Infrastructure" to safeguard and benchmark social, economic and environmental progress is a necessary step to the continuous operation of the "system of systems". Allowing the observing system products to inform national policy, to guide market decisions, and safeguard the environment will engender the financial support of governments and businesses. This will change the focus from designing a "sustained" observing system to an observing system for sustainable development. What does this refocusing mean? It means that there needs to be new performance metrics, new awareness, new teams, new products. This is particularly true in setting the criteria for success of the observing system. At present the success is measured in how well the technology performs—that is the accuracy of the sensors, the skill of the models, the technical performance of the integrated system as a whole. These have been set by the research agenda and were very appropriate for the design, deployment and testing stage of the system. However, while these criteria need still to be in place in order to continue the refinement, upgrading, improvement and quality assurance of the technology and models, additional criteria need to be added at this stage as the information products of the system are applied to National

Development goals. Sustainability indicators such as the GRI, Moody, Standard and Poor Indices may guide in the development of these new criteria. Thus in essence we will have **dual sets of performance metrics** - one for the performance of the technology and another for the performance of the information products stemming from the research and technology.

3. The market structure for global observing system information: Who uses it, why do they use it, and what do they use it for?

The market drivers for using environmental information are varied, with safety of life and property being the overarching driver. Market economics is a major driver (i.e., profit and loss, balancing supply and demand, efficiency and cost effectiveness) where there is need to internalize the externalities of the environment to reduce the impact on operations is essential. Risk reduction at all levels of the organization is the goal and is reflected in the shareholder value and reported in annual reports of company performance. Recently there has been another driver emerging- that of sustainability, where corporate social responsibility is evaluated. The final and most compelling driver is regulation—less responsible corporations move only to this motivator.

Figure 1 also illustrates the various markets for the information from the observing systems. The user community is complex and includes the research community, governments, the business community as well as the natural resource management community (environment). The information needs for the research community have been developed extensively and, in fact, it is upon these needs that the configuration of the present observing system and modeling requirements are determined. The information requirements for the economic and government sectors are less well defined. In the government arena, much of the information produced by one agency is used by other government agencies for policy making, for the protection and regulation of national economic and social needs. Thus the major "customer" of information produced by a "research agency" is often an "applied agency". The information needs of the business community are diverse and must be examined in detail. It is to the definition of their requirements and the understanding of how these industries use the data that much of this chapter is devoted.

There are diverse business applications for information from the global observing systems. The information used includes real time, historical and forecast information. Major applications are in the following sectors of the economy.

Energy: The energy sector has one of the greatest needs for the information in terms of both their operations and their strategic planning.

They are a sophisticated industry and thus have the capability to uptake and manipulate the environmental information to manage their risk. The needs of the oil and gas community have been examined extensively and include requirements for exploration, platform operation, product transport, pipeline laying, ship routing, refining and distribution. The power utility industry has been less examined but has critical requirements for balancing electricity loads and optimally dispatching power into the Grid. Determining fuel choice and guiding its wise use in the next decades is critical to the concept of energy security, and knowledge of regional climate information is a priority to planning of wind farms for example. Building energy control systems use "smart sensing" feedback technology to monitor the external temperature as well as pollutants to regulate heating and cooling systems.

Healthcare: In the healthcare industry, connections have been made between disease and weather and climate patterns such as drought and flood. In the ocean, red tide prediction can aid in the management of shellfish bed closures and swimming beach safety. New applications are now being made to managing the inventory supply chain as well as the scheduling of hospital and clinic emergency room staff and beds using weather event forecasting. Military bases use potential injury to personnel due to storms, icing, and earthquakes to determine base closures.

Transport: In the transport industry, ship, truck and aviation routing have long relied on the current state as well as the forecast information particularly around critical threshold temperatures (snow, ice, rain), and storm tracking. With the trend today toward the seamless integration of all forms of transport or "intermodal" transport, strategies are being developed to optimize the transport of goods and personnel in a safe, efficient and cost effective way. Using new environment forecast information is a way to mitigate unwanted delays due to "congestion" and route disruption.

Finance: The financial services sector underpins all of the other industries of the nations and has diverse needs. Today, banks are considering factors such as sea level rise, beach erosion rates, storm surge penetration and other coastal factors before approving lending for capital investment projects. Risk ratings for "environmental friendliness" are being calculated for companies, and weather derivatives are major instruments for trading and hedging.

Tourism: In the tourism industry, which is one of the largest economic sectors of the global economy, atmospheric and ocean forecasts are essential to safety and risk management in the cruise ship industry and coastal hotel industry, long term resort construction development, and facility staff and supplies scheduling. Room rate setting, revenue projections, inventory management and food service planning all rely on this information.

Water Management: Another important industry operation "at risk" from weather and ocean variables is water management where drinking

water safety, sanitation management, hydropower and irrigation all require environmental information for optimal reservoir management. New decision support for wastewater and storm water management requires the knowledge of precipitation, storm surge, winds, for optimal hedging of risk. Also, the salt water intrusion into the water table due to climate change associated variables such as sea level rise and droughts is of major concern to the coastal communities of several countries.

Food Security: The agriculture industry has long been an important user of environmental information for food and silage security and crop supply chain management. The seasonal to inter-annual information on temperature, winds, dew point and precipitation are of considerable interest to this community. Information is used for selecting seed stocks, determination of schedules for planting, irrigation, fertilization, pesticide application, harvesting, inter-modal transport logistics, crop pricing and futures trading.

Construction: The construction industry is business which, like the financial services sector, underpins all of the infrastructure needs of the other industries. Infrastructure design and siting relies heavily on metocean data for structural design codes, location and materials selection and compliance.

Defense: This industry is a diverse one including national governments as well as the private sector which serve as the civilian support arm. Force projection as well as homeland security head up the major functions. All branches of the armed services use information from the observing system for operations to gain tactical advantage to strategic planning for force readiness.

4. Organizational use of environmental information: Information needs within a company's value chain

There is an increasing tendency for businesses to be organized around their value chain or supply chain. By this we mean the functional aspects or product/service line of their company. Even national agencies and ministries are reexamining their structure and reorganizing in this fashion. The driver for this type of reorganization is that it allows accountability for each of the units in terms of how well they fulfill their function to the organization as a whole. It is important to look at this structure to determine where the environmental information being operationally produced by the met-ocean community will have the most impact in the company- that is which units are most at risk from the environment.

An example of how the value chain is reflected in the organizational structure of a company is demonstrated in Figure 2, a diagram of the organization of a major international construction company. BOVIS Lend Lease (BLL), like other leading global construction companies are

expanding their operations to include more than the construction phase of the project. This company's value chain now spans from the financing of a project through the land development and facility design phase, through the construction phase and into new territory, the operations and maintenance of the property. These added responsibilities are increasing their need for environmental information so as to decrease their risk and vulnerability. BOVIS can minimize the financial risk to capital infrastructure investment by incorporating new environmental forecast information such as regional temperature, air quality, winds, precipitation, seismic, frost levels, etc estimates out to 30 years. This enables BLL's investment portfolios to be compiled which have in them scientifically-based environmental predictions about the future conditions of the building site (before , during and after the construction phase), often used for insurance purposes as well. BOVIS can incorporate environmental forecast information into all phases of the construction siting, design, materials, scheduling and compliance reporting of a project. BOVIS can factor in the real time, and seasonal to inter-annual weather and climate predictions into the operations and management phase of the Business such as optimal building energy management using real-time environmental information of temperature, humidity, winds, cloud cover, pollutants for heating and cooling cost savings. Real time and short term weather forecast information can be used as well in the scheduling of staff resources as well as to "trigger" the required emergency management actions in case of a disaster.

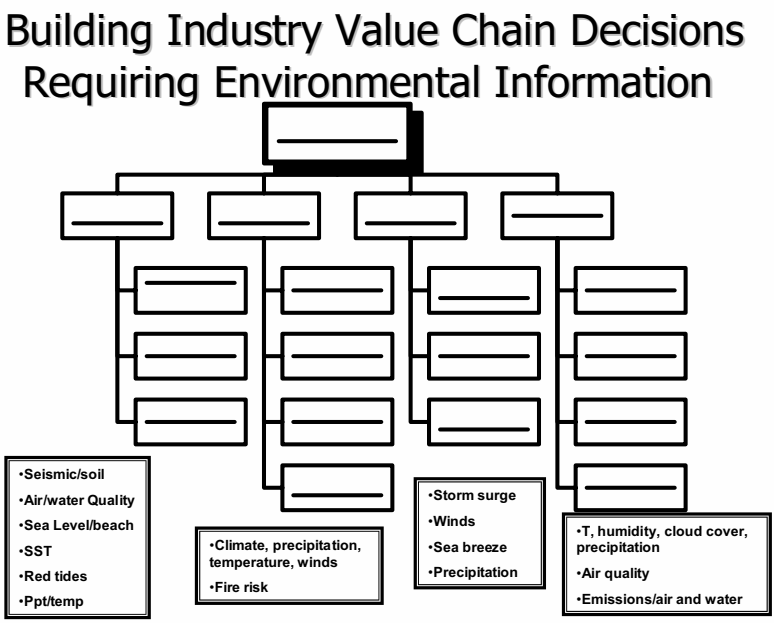

Figure 2. An example of how environmental information is assimilated into the operational decisions of the construction Value Chain of Bovis Lend Lease.

The 3-6 month, local forecasts of temperature, icing, precipitation, for the advanced procurement of bulk items such as salt for freezing/snow conditions; plywood for use in hurricane conditions; fertilizer applications carried out only under specific weather conditions.

Factoring the 1 month to 1 year regional climate forecasts particularly in coastal areas allows building maintenance regimes to factor in projected impacts such as salt erosion, high winds, sand battering, and toxic mold production leading to rapid deterioration of structures, helping to set maintenance, renovation, retrofit and materials replacement schedules. This may also help mitigate potential adverse human health impacts.

5. Vulnerability of business operations to metocean processes

Optimal operations of the industry are of chief concern to the CEO and environmental impacts are noticed when they interrupt the smooth operations of the company. Thus to examine the vulnerabilities of the industries to metocean events, their impact on all management processes need to be examined. These include supply chain, inventory, fiscal, and emergency management plans and procedures. If the impacts are seen to be substantial, the normal operating procedures and possibly even the organizational structure will need to be changed. This is known as business process reengineering.

The insurer AON estimates that weather and climate have a substantial impact on over 70% of the businesses worldwide. For example in the case of weather risk, it is well documented that businesses are vulnerable and that the costs are substantial. In an analysis of annual reports of major energy corporations, much of the poor earnings performance is blamed on weather. Risk is in the form of cash flow and earnings and much of the risk comes from imperfect forecasting. In coastal regions, forecast inaccuracies are often due to an under sampling in the adjacent ocean environment.

In addition, there are critical cross cutting industry needs such as demand forecasting, emergency management and policy formulation or company governance that require the types of environmental information provided by observing systems.

6. Defining industry's information requirements

A series of research projects funded by NOAA has examined the environmental information requirements for weather climate and ocean data for major sectors of the economy. The studies spanned all time scales of operations and planning from hours to climate scale. The impacts on all

aspects of operations, planning, markets and policy making involved in the respective industry were defined. Selected results from several of those investigations are shown in Figures 3 and 4 whereby the affected operations are plotted on the timescale of forecast information needed.

In the case of the energy industry, weather, climate and ocean forecast information at all lead times is used routinely throughout the industry. The operations currently using the information most rigorously are load balancing, both for single utility and grid management, generation commitment, fuel mix choice (fossil fuel, hydro, wind), dispatch scheduling, power marketing and trading, power pricing, fuel pricing, procurement strategies, tariff scheduling, natural gas storage management, revenue projections, infrastructure siting and construction management. The most critical forecasting needs are for sub-day, 2-14 day and seasonal forecasting.

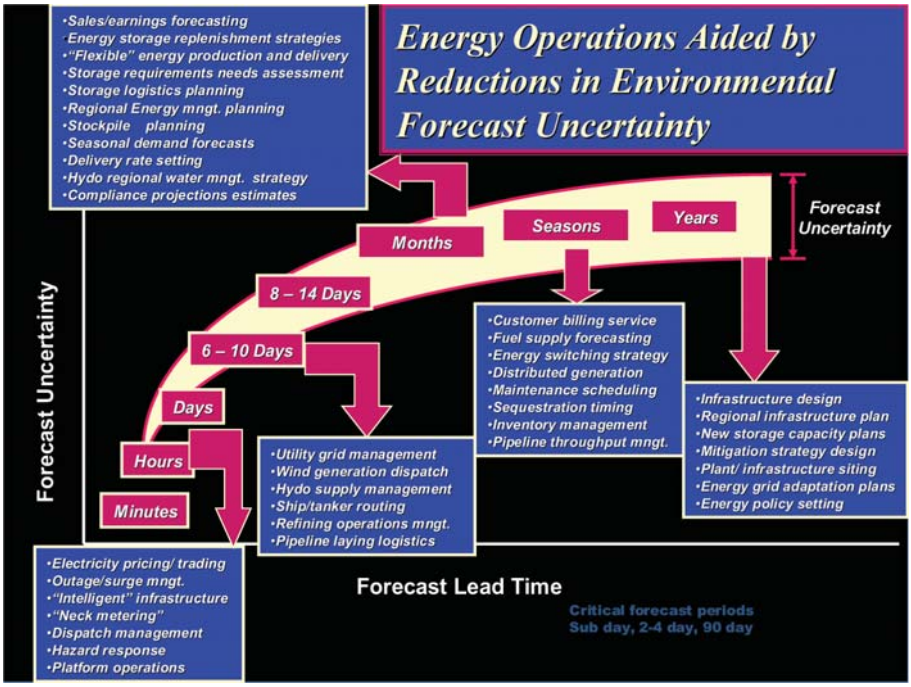

Figure 3. Requirements of the energy industry operations for improved environmental forecast data.

If we examine a similar needs diagram for the water industry, several key operations that are in need of forecast data are identified. Figure 4 illustrates the needs for hydropower, drinking water, sanitation and storm water management and recreational uses. Forecast information is used at time scales from "real time" in emergency management procedures such as issuing "boil water" orders, through seasonal forecasts for planning for

irrigation of crops, to interannual for the siting and construction of dams and reservoirs.

Interestingly enough the forecast lead time needs hold true across all sectors of the economy with transport, health, recreation and tourism, 'financial services and construction industries' operations and planning all impacted by weather and climate uncertainties.

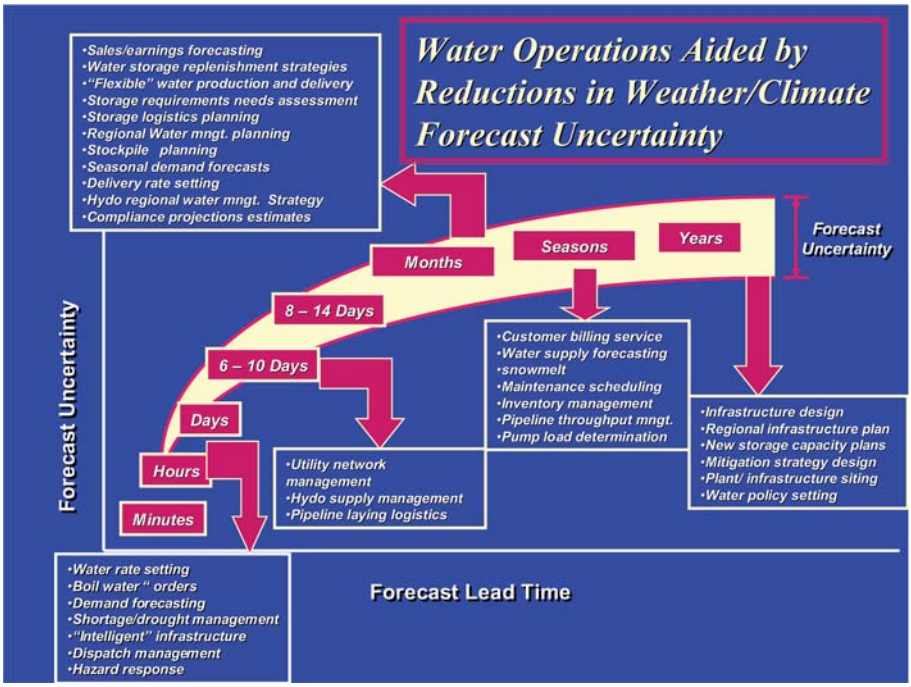

Figure 4. Environmental information forecast needs for the water industry.

7. Mapping the flow of environmental information in business decisions and outcomes

Figure 5 illustrates how the environmental **data** product is transformed into **information** through the **knowledge** chain to become business information products to inform **actions** needed to effect the desired **outcome** so as to mitigate any risk imposed by the environment.

This diagram defines the operational decision process and bridges the information provider and the business user. Starting on the left, met-ocean data and/or modeled forecast products are obtained by the business user (represented by the red box) either directly through the operational agency such as the NMHS or as provided through a service provider, which may

take the information and adjust it for the operational area or other needs of the user. Many companies, recognizing the importance of getting the weather forecast "right" may actually buy "tailored" forecasts from several vendors. In some instances they average them, in others they do a weighted forecast (i.e., terrain dependent) to which they apply their own regional knowledge as to the accuracy of the data. Other analytical services will provide statistical analysis of the data in order to reduce the data error for the operational area. The "refined" forecast is now inputted into the industry "demand forecast" model, in this case load forecasting model for power demand forecasting which takes a "day ahead" temperature forecast product and uses a rolling average out to 7 days. Other inputs to this model are market dynamics, present demand, yesterday's demand, prices and regulatory constraints, etc. The output is a demand forecast which is converted into megawatts of generation to put on the electricity grid to meet the demand. The generation scheduler then "buys" or commits the generation from the various generators to meet the need and the "dispatch" scheduler puts the power over the transmission lines to the end user. Other operational decisions which are based on the demand forecast include financial projections, power pricing, trading, etc. By effectively balancing the load (supply=demand), the desired outcome of no fines, good profit margin, increased reliability, decreased liability, etc. will follow. However, when the weather error is large and the demand forecast is substantially off leading to stresses on the system, fiscal loss, and possible threat of blackout conditions. Thus getting the weather input right is of highest concern.

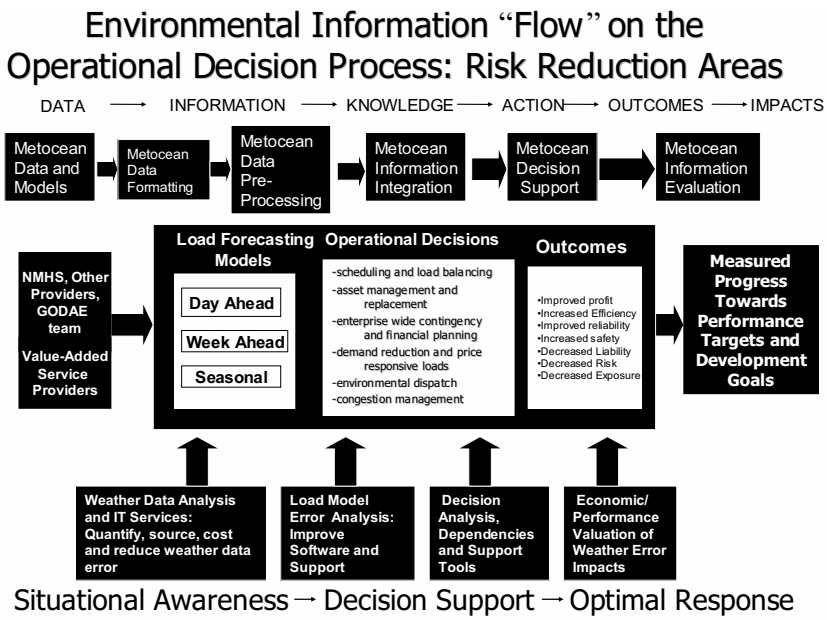

Figure 5. Environmental information "flow" in business operations.

8. Case studies from the energy industry

Weather can be a significant "tipping point" between: cost effective and non-cost effective energy infrastructure operations; reliable and unreliable operations; and, periods where emphasis is placed on either supply options or demand options. With greater uncertainty being introduced in many facets of energy supply chain delivery, load optionality and the use of weather related derivatives are becoming increasingly important.

Weather information is critical to different participants in the electric energy value chain. One example concentrates on demonstrating the value of improved forecast information to major energy operators in the Northeast U.S. and in California. To regional transmission organizations such as the New England ISO and the California ISO, both long range and short-range forecasts are critical to evaluating the ability of the system to deliver the power needed for weather sensitive loads. Each of the ISOs is responsible for 1) operating the regions built power system—a power grid of generators and high voltage transmission lines, and 2) administrating the regions wholesale electricity marketplace. The mission is to ensure a reliable power system for customers from the Northeast and California guaranteeing the equal access to electricity and transmission systems. They oversee the operation of a fair, efficient wholesale electricity market.

For local distribution utilities, such as Consolidated Edison of New York or San Diego Gas and Electric, it is critical to know about wind and ambient weather conditions so that an evaluation of the adequacy of its back up power and local distribution system substation capability during peak periods can be assessed. This information is also crucial for the major energy users, like the State University of New York system, which has the responsibility of purchasing gas and electricity for over 30 campuses.

9. Method of assigning a value to the information product: The industry trial or beta test

There are many approaches to estimating value of information from global observing systems. The first is through a cost benefit analysis approach, examples of which have been done on a number of observing systems in various countries. However, this method often underestimates the value particularly if the product is new or not recognized as useful by the customer. The second, rather different, but complementary approach is through an "Industry trial" or "Beta Test". It is a proactive approach where a Business End Use Partner and an Information Product Provider team to run an information product "performance assessment" in real life operations. Here, the environmental information product from the observing system is inserted into an operational business model and is transformed into a business or policy decision tool (i.e., demand forecast). The business models

are run with and without the environmental information, scenarios are generated, consequences analyzed, cost factors applied and performance assessed. As the skill of the environmental forecast product increases so too does the skill of the business and policy forecast. Thus the "Demand" grows for the information product. This approach may require integrated information from ALL the observing systems elements-ocean, terrestrial, weather and climate. The "trial" approach actively engages the end user and the organizational unit in which s/he functions. The outcomes are used to guide marketplace decisions, to inform national policy and to prioritize the research and development strategies of nations.

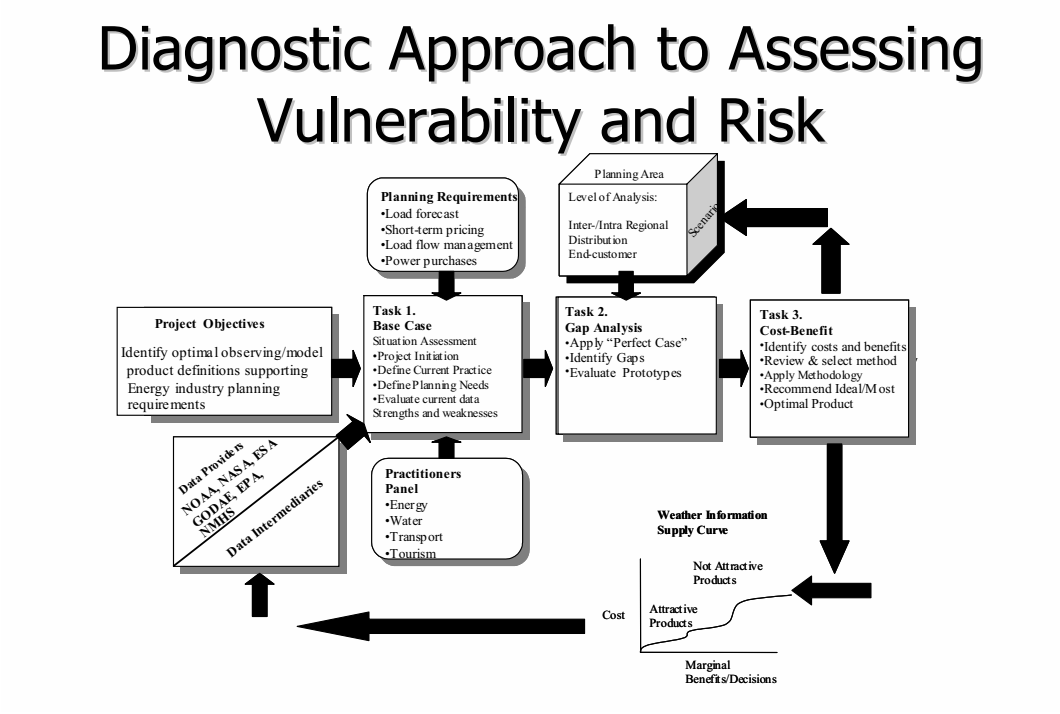

Figure 6. Diagrammatic approach to estimating the value of forecast information from the global observing system in industry operations.

10. Sourcing the environmental error

Once it is established where the weather error impacts the operation, it is important to diagnose the environmental cause of the error. Figure 7 illustrates the concept of how a weather induced demand error is traced back to uncaptured sea breeze events, afternoon thunderstorms, back door fronts or fog. Only by identifying the ultimate cause of the weather error can the appropriate observing systems be enhanced to capture the information needed for the industry operations.

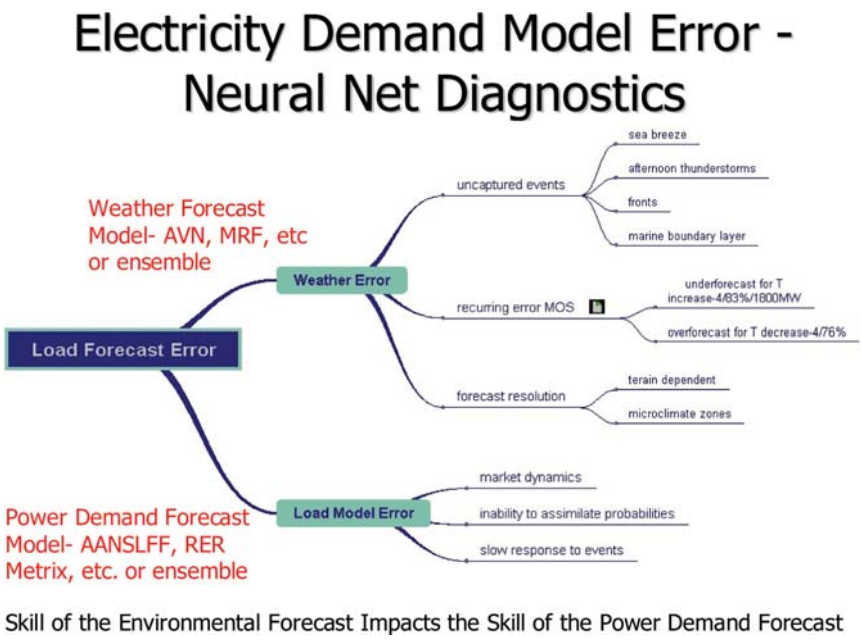

Figure 7. A diagnostic approach to sourcing the electricity load forecast error to weather error.

11. Cost findings

While it is beyond the scope of this paper to examine all of the cost implications of weather induced business error, a few examples will illustrate the magnitude of the impact. In general, errors in weather and climate forecasts significantly affected the operations and planning of the power industry.

Over the past years, several projects have been undertaken to evaluate the risk (Altalo et al, 2001, 2002, 2003, 2004). In a detailed study in the Northeast U.S. the demand forecasting process was shown (Figure 8) to be particularly vulnerable to unpredicted temperature fluctuations caused by seabreeze, afternoon thunderstorms, back door fronts. The Northeast Energy project 2002 estimated that the benefits of improving day ahead weather forecast accuracy by one degree F or by reducing forecasting error by 50% for days 2-7 is:

- ~$20-25 million per year for a regional transmission authority
- ~$1-2 million/year for a major distribution utility.

If these savings were generalized to other regional transmission organizations, and regional transmission authorities the total savings would be for the entire Northeast Region:

- $100-140 million/year for ISO's. Scaling up the impact to the entire nation, the cost savings would be in the billions of dollars to the U.S.

- $30-60 million for electric distribution companies in the Northeast. Scaled up to a National basis, the estimates would be in the 100's of millions.

Furthermore, capturing the "events" such as seabreeze, fronts, "pop-up" afternoon thunderstorms etc., on top of this will yield significantly higher savings (10's millions/day regionally).

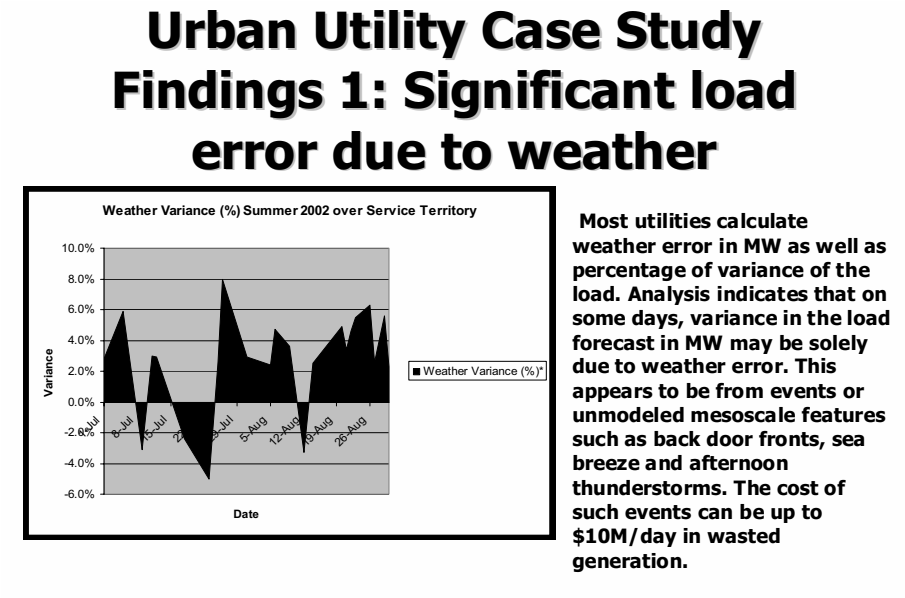

Figure 8. Impact of weather error on electricity load error.

In another study on the West Coast of the U.S. significant error in the day ahead temperature forecasts can lead to a significant error in the electricity demand forecast resulting in high financial consequences, in this case several million dollar loss per day, as well as grid instability. Figure 9 illustrates the impact of one such temperature error on the load error of the grid operator, nearly bringing the system to instability

In an additional example, a west coast sea breeze phenomenon known as the Delta Breeze contributes strongly to the load forecast error for the same grid operator as shown in Figure 10. To mitigate these effects, new information from operational observing system from buoys, radar, satellite, needs to be brought in and captured in the numerical prediction models for that area. In many cases, the weather prediction models need to be refined through the use enhanced modeling techniques such as ensemble based forecasting which may yield significant improvement in forecast accuracy and have been shown to reduce overall summer peak electricity error (Smith et al 2005, Altalo and Smith 2004). In addition, the utility load forecast models must be often be reconfigured to accept the new data or the probabilistic information from ensembles. This is where a partnership comes in.

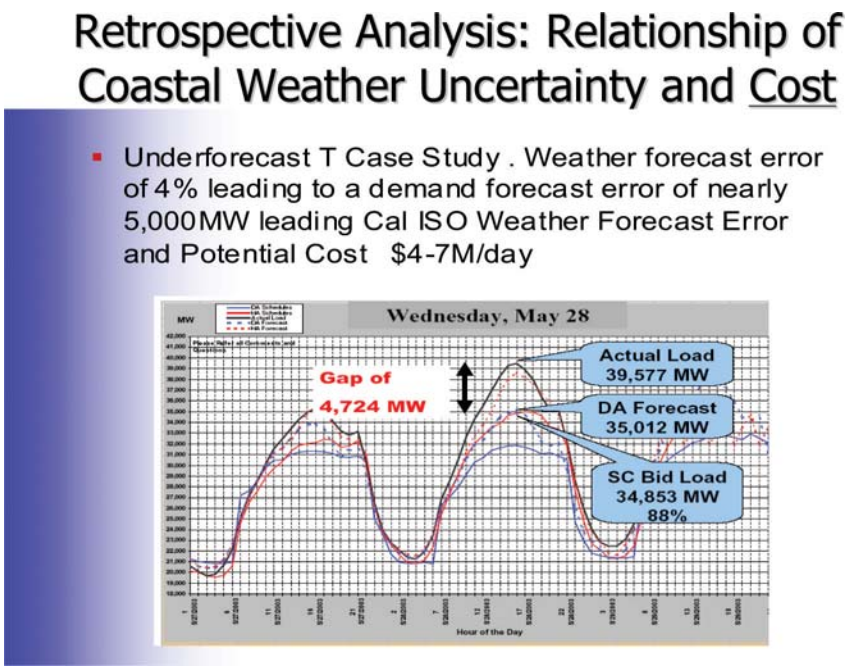

Figure 9. Temperature underforecast and impact on bid load. The shortfall of nearly 5000 MW of generation caused the ISO to buy costly "on spot" power.

Cal ISO Mean Daily Forecast Error

-Delta breeze and weather/load forecast errors contribute to major errors in prediction of Delta Breeze effects.
-Delta breeze is defined as the conditions when the wind speed is > 12 knots and the direction is between 190 degrees and 280 degrees.
-Delta Breeze can change load by 500MW
-Direct Costs: 250k per breeze day; 40 events per year
-An over forecast problem

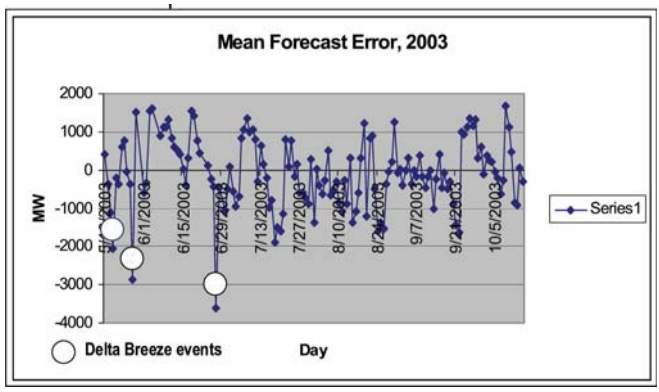

Figure 10. California sea breeze impact on load forecast error and resultant fiscal loss.

Requirements

Environmental Information in the Power Value Chain

Energy Operations / Environmental Dependency	Fuel choice	Generation	Pricing	Transmission	Transport (gas)	Distribution	Energy Procurement	Demand Side management
Sea breeze						x		x
Offshore wind prediction		x						
Fog						x		x
Back Door fronts						x		
Sea/lake Level/wave height	x	x			x			
Air Quality		x						
Microclimates						x		
Precipitation Forecasts						x	x	x
Radiation						x		x
Water Temperature		x						
Air Temperature			x	x	x	x	x	x

Table 1. An example of the requirements of the power value chain for information from the Global Observing System

Studies such as these lead to the definition of a set of requirements by the industry for observing system information as shown in Table 1. Real time winds, fog, dew point, cloud cover as well as temperature forecasts are needed to help meet operational as well as strategic planning needs. Observation systems can and need to be configured with dual goal of meeting research objectives and satisfying national needs

12. The enhanced value of the environmental forecast product to developing nations

Developing nations, as well as developed nations are in critical need of the weather, climate and ocean forecast information that observing systems can provide for the optimal management and planning of their business interests. This applies across all sectors but is particularly important in the infrastructure planning of roads, water, electricity, mining, agriculture and

health. In regions where hydropower makes up significant fraction of electrical generation capacity, incorporating climate-induced weather information into the daily operations of the facility could head off disasters.

An example is taken from a study conducted in Ethiopia with EEPCo (Ethiopian Electric Power Company) which is the public agency responsible for generation, transmission, and distribution. In this country, 97% of electrical generation is from hydropower produced by Koka Dam, with major use by heavy industry such as cement, metal, and sugar. Energy generation averages growth 10%/year. There is large scale inter-annual variability of rainfall in this country leading to periods of droughts and floods. Seasonal to inter-annual predictions offer great benefits if hydrological indicators are used for planning by the industry. The value of the advanced environmental information can be used as the basis of a risk analysis to mitigate flash flood hazards and identify periods of water scarcity. There is a need as well to incorporate surface variables (precipitation and temperature) into hydrological forecasts. Information on the timing, location and duration of flash floods is used for water releasing schemes from dams by Ministry of Water Resources; and, as dam capacity is impacted by the erosion in a basin, such information is of value to the construction engineers.

During drought, power rationing leads to major revenue loss. In a study conducted under the IRI program, the linkage effect of power production and customer revenue loss was examined and determined to be approximately $8M, enough to destabilize local economies. From these types of studies, it was recommended that EEPCo must include seasonal forecasts into its long term plan. It was also recommended that they produce power demand scenarios based on seasonal rainfall outlooks. For the "provider" community, it was recommended that the development of future models include hydrological parameters in addition to meteorological variables, and that they combine variability of rainfall, complex topography and behavior of rainfall on sub grid level. The skill score of climate forecasts needs to be improved as well as performance metrics set for the improvement in business operations and planning.

In another IRI benchmarking study with Kenya Power and Lighting Company (KPLC) and KenGen, it was determined that drought-induced rationing decreased production to 40% and they had to secure an, emergency power credit of $72M to purchase fuel since there were no sources of fossil fuel internal to the country. In addition, to help mitigate the crisis, the World Bank granted $47M to import and operate generators. The economic losses from rationing and power failure were estimated at $2M/day and KPLC lost $20M/6 months with expenditure of $141M for fuel. The effect was 12 hr. power rationing which translated to massive layoffs. Seventy percent of the manufacturing firms not willing to increase investment in Kenya if power reliability issues were not addressed. In this case, the energy crisis loss to

economy was $100M/month. Drought was one of the contributing factors to poor reservoir management with deforestation, environmental degradation, poor agriculture practices in catchment areas contributing as well. The result was low water levels and under performance of turbines, cascade operations sustained thru low level outlet releases. The recommendations to KPLG and the government from this study included incorporating seasonal forecasts into its long term plan for hydropower management. For the prediction community there is a need to increase rainfall prediction tools, and to incorporate soil and evaporation for calculation of water losses.

13. Entraining environmental information in decision support tools

One of the biggest challenges to both the provider community as well as the business community is to turn an environmental forecast into a business decision support tool. From the user point of view, the high level goal of the entire process is to "morph" the environmental information into something that the industry can use – or "Turning a metocean forecast into a business forecast" to enhance decision making. The steps are straightforward and consist of

1. Improving the metocean product to deliver the information on an operational basis and in a format usable by the business community;
2. Establishing the "relationships" or dependencies between the environmental event and the business effect (i.e., the power curve of wind speed and electricity generation)
3. Turning that relational information into a predictive business tool
4. Using the predictive tool to formulate an action plan

Figure 11 illustrates how appropriately designed decision support tools (this one designed for MEMA, the Maryland Emergency Management Agency) can "pull in" a hurricane track forecast, link it with a critical infrastructure list to feed into a consequence risk assessment prediction toolset (CATS) that is capable of informing decisions on which transmission lines to restore (Expert Grid) in order to ensure the hospitals and shelters in the disaster area have power. This method creates the situational awareness to map the "at-risk" critical infrastructure with reliable environmental forecasts to activate disaster management action plans. The work helps emergency managers to identify, prioritize and coordinate the protection of critical infrastructure, and can be integrated into comprehensive emergency management systems. Ensuring that the Global Observing system infrastructure is also on the critical infrastructure lists of nations is essential to national security.

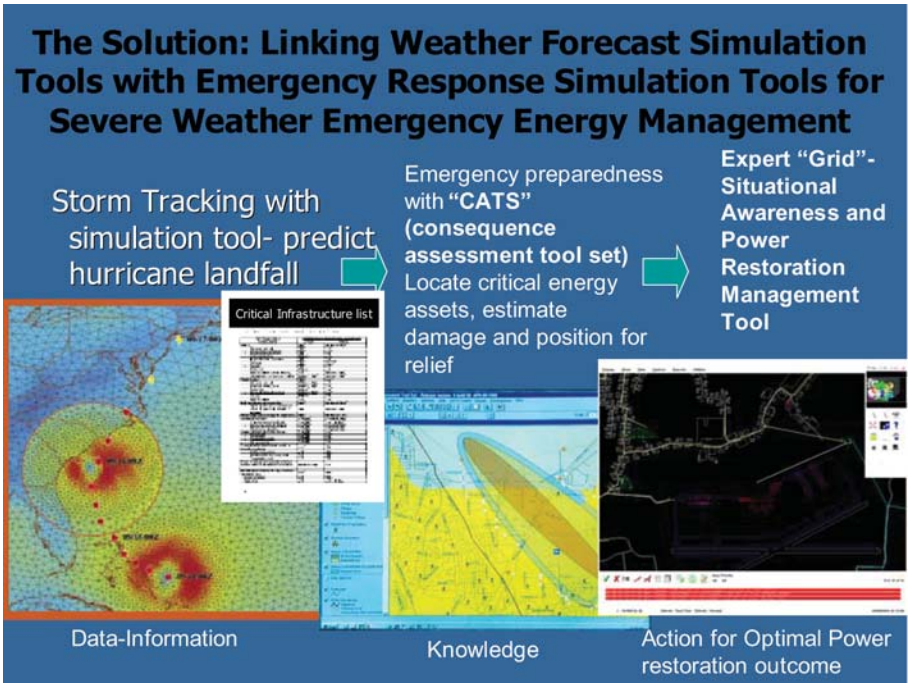

Figure 11. Linking environmental forecast information with decision support tool sets and critical infrastructure needs lists contributes to the successful deployment of disaster relief emergency operations, in this case for the restoration of essential services (power) to hospitals.

14. Business case considerations

While it is important to stress the scientific design, observational strategy, data assimilation and distribution scheme and delivery mechanisms to users, it is of equal importance to consider the Business plans for the overall operation and management of such a system. Businesses will not incorporate the environmental information into their decision making if they think that the information may not be sustained. Thus there is a pressing need to demonstrate that the information will be available long term or operationally. A credible business plan for the long term financial strategy to sustain the observing system must be developed and may include traditional funding sources such as science and technology agencies as well as less conventional sources such as more applied agencies and ministries, business councils, the private sector and even venture capital. New studies showing the changing nature of R&D funding (i.e., Mullin), show that many projects are focusing on this topic. Legacy planning must start now and all of the potential investment partners must be invited to the table in the beginning. There are many models in existence to determine the appropriate course of action. Technology transition and commercialization strategies may shed

some light on the issue. Also the changing roles of investment strategies with a many financial backers can be taken under consideration.

15. Summary and conclusions

The major conclusions of this chapter are that: GODAE provides situational awareness to enable optimal operational management, strategic planning and policy support to the business community thus enhancing the development of national economies to meet their individual development goals. Retrospective analysis of the dependencies of metocean events with business performance can help set priorities for research with potentially high payoff to society. "Beta testing" of new environmental information products through an industry trials approach can lead to the rigorous estimation of the costs mitigated with the incorporation of the new environmental forecast information. It must be realized that the impact of new forecast information may be highest in developing nations. Institutionalizing the new information in a decision support tool is critical, and, developing a long term business strategy to financially sustain the information product will ensure the success of the program.

References and suggested further reading

Altalo, M., 2000. Defining the Requirements of the U.S. Energy Industry for Climate, Weather, and Ocean Information. 140 pp. DOC, NOAA technical report http://www.saic.com/weather/decision-support

Altalo, M., 2001. The Critical Role of New Environmental Information and Technology in National Energy Needs. Conference Proceedings. DOC/DOE Energy Roundtable, 17 July 2001. 157pp. http://www.saic.com/weather/decision-support

Altalo, M., K. Cook, O. Anastasia, and M. Marshall, 2002. Public-Private Partnerships for New Government Initiatives and Organizational Structures. DOC NOAA, Technical Report Series

Altalo, M., 2002. Critical use of environmental information in industry operational decision aids and scenario building, Proceedings of Oceanology International 2002, London.

Altalo, M., M. Hale, O. Anastasia, and H. Alverson, 2002. Requirements of the U.S. Recreation and Tourism Industry for Climate, Weather, and Ocean Information. 237pp. DOC NOAA Report. http://www.saic.com/weather/decision-support

Altalo, M., C. Summerhayes, N. Flemming, and P. Bernal, 2003. Demand side "pull" for Eurogoos products: six case studies of market and policy decisions impacted by new environmental information. In Summerhayes. EuroGOOS Proceedings.

Altalo, M., P. Bogden, C. Colgan, H. Dantzler, M. Davidson and P. Mundy, 2004. The Business Case for the Global Observing System. Oceanography. vol. 16:68-76.

Altalo, M., and M. Hale, 2004. Turning Weather Forecasts into Business Forecasts, Environmental Finance May, pp. 20-21.

Altalo, M., 2004. Contributions to the California Energy Security Program. NOAA Technical Report.

Altalo, M., and L. Smith, 2004. Using Ensemble Weather Forecast to manage utilities risk. Environmental Finance. October, 2004.

Archer, E.R.M., 2003. Identifying underserved end user groups in the provision of climate information. BAMS:84 (11) 1525-1532.

Bernal, P., and M. Altalo, 2001. Observing Systems for Sustainable development: Towards Improvement in Economies and Societies through New Environmental Information, RIO Plus Ten, UNESCO Paris

Burroughs, W.J., 1997. Does the Weather Really Matter?: The Societal Implications of Climate Change. Cambridge University Press.

Chagnon, S., J. Chagnon, and D. Chagnon, 1995. Uses and Applications of Climate Forecasts for Power Utilities. Bulletin of the American Meteorological Society: 76(5)

Callahan, B., et al., 1999. Policy Implications of Climate Forecasts for Water Resources Management in the Pacific Northwest. Policy Sciences: vol. 32:3

Hale, M., 2003. Environment, Weather and Climate Information in the Financial Services Sector. In "Practice". Bulletin of the Institute of Ecology and Environmental Management (IEEM), No. 42 p.9-12, Dec. 2003

Mullin, 2001. International Social Science Journal, 168.

National Research Council, 1999. Our Common Journey: A Transition Towards Sustainability. National Academy Press. 363 pp.

Scavia, D. et al., 2003. NOAA's Ocean Observing Programs. Oceanography, 16:4, 61-67.

Smith, L., M. Altalo and Ziehmann, 2005. Predictive Distributions from an Ensemble of Forecasts: Extracting Electricity Demand from Imperfect Weather. Physica D, in press.

Stern, P., and W. Easterling. Eds. 1999. Making Climate Forecasts Matter. National Academy Press, 175pp.

Stewart, T.R., R. Pielke, Jr., and R. Nath, 2004. Understanding user decision making and the value of improved precipitation forecasts: Lessons from a case study. BAMS 85(20, 223-235.

Thompson, R.D. and A. Perry, Eds. 1997. Applied Climatology. Routledge, London. 352pp.

Chapter 23

FORECASTING THE DRIFT OF OBJECTS AND SUBSTANCES IN THE OCEAN

Bruce Hackett, Øyvind Breivik, and Cecilie Wettre
Norwegian Meteorological Institute, Oslo, Norway

Abstract Forecasting the drift of floating objects, ships and oil spills is an important ocean application. Most nations support services for ship safety, oil spill combatment and search-and-rescue, all of which may benefit from drift forecasts. Examples from Norwegian services are discussed. The models for drifting things themselves are founded on hydrodynamic principles (ship drift), empirical parameterizations (floating objects) and oil-water chemistry. An overview of these models is given. All the drift models share a crucial reliance on geophysical forcing data. In operational services, these data are obtained from weather, wave and ocean forecast models. Currently, ocean forecasts are the component with greatest scope for improvement. Effective interfacing of drift forecasting services to the users - the emergency response services - is vital for obtaining optimal benefit from the forecasts.

Keywords: Oil spill, search and rescue, ship drift, ocean forecasting.

1. Introduction

Ocean forecasting is founded on the operational prediction of the prognostic variables in hydrodynamic models: water level, temperature, salinity and currents. While these variables are essential for describing the state of the ocean, they are in themselves of limited interest outside of the scientific community. Most people are not concerned with, for example, tomorrow's forecast for surface current 2 km out to sea - even those who reside on the coast. Of more interest to public is how these physical variables affect other things in the ocean. In the case of currents, there is considerable interest in how various things in the ocean

E.P. Chassignet, Jacques Verron (eds.), Ocean Weather Forecasting, 507-523.

are transported from place to place and how they are altered along the way. The transport of objects and substances by currents is commonly referred to as *drift*, and is the subject of this paper.

There are a number of drifting objects and substances in the ocean that are of concern or interest to us, most important of which are the nutrients and phytoplankton that form the basis for life in the ocean. However, in this work we will focus on a special class of drifting things: those that derive from human activities and have potentially serious consequences for human activities and the marine ecosystem. Specifically, we will look at (large) floating objects and oil spills. These are quite different things, but in the present context of ocean forecasting, they share a dependence on knowledge of the physical conditions in the ocean, chiefly currents. Furthermore, they share a potential for negative impact sufficient to warrant the development of services, including the prediction of their drift and fate, to mitigate those impacts.

In the case of drifting objects, we are concerned with objects of great value in themselves, such as a human body, and those that (also) pose a threat to maritime safety, such as a floating container. The task at hand is either to find and recover a lost object (search-and-rescue, or SAR) or to track a known object until remedial action can be taken. Many countries have established national SAR services for handling emergency situations, and some also run trajectory models to aid in the search. The objects that are of interest in this type of service range from smaller than a human (debris from a wreck) to large ships, although the scope and administrative organization of the services vary from country to country. In this work, we will focus on examples for SAR and ship drift.

Oil spills are quite different inasmuch as the object is a fluid that spreads and can change properties quite dramatically once it is spilled into the ocean. The source of the oil may be on the surface (a ship) or subsurface (an offshore wellhead blowout). Also, oil on the surface can be mixed down into the water column by wave action and be transported by subsurface currents. Oil spills fall generally into two classes: small spills from ships, typically from illicit flushing of ballast tanks, and large catastrophic spills, either from tanker accidents or from offshore installations. Small spills from shipping are so numerous that they account for the bulk of the worldwide input of oil to the sea, but it is the large accidents that are responsible for the most dramatic damage to the environment. In most countries, detection and combatment of oil spill incidents is a national service that includes a drift forecast component.

In the following, we will look at current practises in forecasting the drift of floating objects and oil spill fate, using examples from operational services in Norway. Two aspects of these services will be addressed: the

scientific basis for the forecast numerical models used and the information infrastructure required for operational service. Since the theme of the Summer School is ocean forecasting, we will only briefly describe the modeled processes and concentrate more on the practical implementation and operational aspects. Indeed, models for drifting objects, ship drift and (especially) oil fate are quite complex and a fuller understanding of the modeled processes is beyond the scope of this summer school. However, some insight into the workings of these models is necessary to understand the operational data input and dissemination requirements. The sections 2 and 3 will introduce modeling of floating objects and oil spills, including typical methods for implementing them. It will become apparent that the geophysical data needed to drive these models in an operational service are quite similar. Therefore, section 4 will deal with the structure of operational services for both floating objects and oil.

2. Drift of floating objects

The motion of a drifting object on the sea surface is the net result of several forces acting upon its surface (water currents, atmospheric wind, wave motion, and wave induced currents), and its center of mass (the gravitational force and the buoyancy force). It is possible to estimate the trajectory given information on the local wind, surface current, and the shape and buoyancy of the object.

The position of a floating object is computed by numerically integrating the total drift velocity V_{drift} of the object, given by

$$\mathbf{V}_{drift} = \mathbf{V}_{curr} + \mathbf{V}_{rel}, \tag{1}$$

where $\mathbf{V}_{curr}$ denotes the ocean current velocity relative to the earth, and $\mathbf{V}_{rel}$ denotes the object drift velocity relative to the ambient water. The ocean current is made up of two components: the surface current, which includes the Ekman drift, baroclinic motion, tidal and inertial currents, and the Stokes drift induced by waves. $\mathbf{V}_{curr}$ is assumed to influence all floating objects in the same manner. It is typically equated with the (near-)surface current obtained from a numerical ocean model, a parameterization on the wind velocity and/or local observations. $\mathbf{V}_{rel}$ results from the wind and wave forces acting on the object, and is strongly dependent on the characteristics of the object.

The basic model in eq. 1 may be separated into two modules, based on the forces that determine $\mathbf{V}_{rel}$. A well-known result from hydrodynamics is that wave effects are small when the length scale of the object is smaller than the wave length and increase dramatically when the lengths are about the same (Grue and Biberg, 1993, Hodgins and Hodgins, 1998). Thus, one module is for relatively small objects (in practise, less than

some 10's of meters), where wave forces may be ignored and wind forces are of variable importance, depending on the overwater structure of the object. Objects in this class include wreckage, bodies, rafts, small craft, etc. Drift due to wind forces is commonly referred to as *leeway* drift, also called the windage. The other module is for larger objects - conveniently lumped together under the heading of "ships" - where both wind and wave forces on the object must be taken into account.

2.1 Leeway drift

The maritime term *leeway* refers to an object's motion relative to the wind. It is well known that, due to the asymmetry of almost any floating object, there will be a net side force causing the object to drift at a certain angle to the wind. Thus, we can decompose the leeway drift velocity vector into two components: a downwind and a crosswind leeway component, as shown in Figure 1. The concept of leeway drift

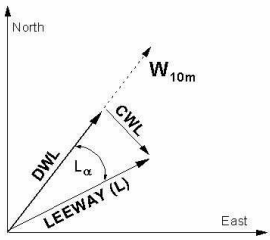

Figure 1. Relationship between leeway drift velocity vector (**L**) and wind velocity vector $\mathbf{W}_{10m}$. DWL = *downwind leeway* component, CWL = *crosswind leeway* component, L_α = *leeway angle* (measured positive for leeway to the right of the wind direction).

is an empirical approach to the very difficult problem of determining the net force on a drifting object. Compounding the difficulty is the wide range of objects (size, form) that we may want to track. Thus, empirical studies of actual objects have so far been the most fruitful method. Allen, 1999 reports field experiments carried out to determine how different classes of objects respond to the wind. The DWL and CWL components for each class of object are recovered from linear regressions on the windspeed. The standard deviations about the DWL and CWL coefficients are identified as "error bars" on the drift properties, and must be interpreted as the total error associated with the wind and current measurements as well as the inherent variation in leeway properties of two ideally identical objects.

Allen and Plourde, 1999 have assembled tables of the leeway parameters for 63 leeway categories of floating object. These tables consist of the linear regression coefficients, and their standard deviations, for each of the 63 categories. They are mainly based on observations and field experiments, although some of the values have been extracted by converting values derived from other ways of calculating the off-wind drift of a floating object.

In implementing the leeway drift in the **met.no** operational service, the DWL and CWL components are calculated in a straightforward manner from the linear regression formulae as functions of the wind speed, once the object type is specified. The standard deviations are used in estimating the uncertainty in the drift speed and direction. An interesting result of the field experiments is that both positive and negative values of the CWL component can occur for a given category of objects. Apparently, small differences in the initial orientation of the object relative to the wind can result in the object drifting either to the left or to the right of the wind direction, with about the same likelihood. The initial orientation is normally unknown, so the prediction is obliged to account for both possibilities.

2.2 Ship drift

The drift of ships has been approached in a different, more analytical manner, based on knowledge from marine architecture and hydrodynamics. The ship drift model at **met.no** is based on work at Det norske Veritas (DnV), reported by Sørgård and Vada, 1998 in which estimates of the wind and wave-generated forces acting on the vessel, rather than empirical regressions, are used to calculate the velocity relative to the water ($\mathbf{V}_{rel}$). The advantage of this approach is, of course, that knowledge of the object may be reduced to a few key parameters.

Case studies by Sørgård and Vada, 1998 show that the relative drift speed of the ship will increase rapidly (in a matter of 2-10 minutes) towards a stationary solution. Therefore it is not necessary to integrate the acceleration over time when the relative drift speed is calculated for simulations over several hours or more; the stationary solution may be applied to good approximation. This means among other things that it is not necessary to know the mass of the ship.

The balance of forces acting on the ship may be written

$$\mathbf{F}_{wind} + \mathbf{F}_{wave} + \mathbf{f}_{form} + \mathbf{f}_{wave} = 0. \tag{2}$$

$\mathbf{F}_{wind}$ is the wind drift force acting on the vessel; it depends on the vessel length, the keel-to-deck height, the momentary draft and the lateral area

of the superstructure. It may be formulated in the well-known form

$$\mathbf{F}_{wind} = \frac{1}{2}\rho_a(A_h + A_s)C_d\|\mathbf{U}_w\|\mathbf{U}_w, \tag{3}$$

where ρ_a is the density of air, A_s is the superstructure area, A_h is the wind-exposed hull area, C_d is a drag coefficient and $\mathbf{U}_w$ is the wind velocity. $\mathbf{f}_{form}$ in eq. 2 is the form drag or damping force exerted by the water on the hull due to the relative motion; it depends on the wet lateral area of the hull, i.e., the length and draft. $\mathbf{F}_{wave}$ is the wave drift force acting on the hull, while $\mathbf{f}_{wave}$ is the wave damping a counterforce that occurs as the moving hull generates its own wave field. Much research has been done to determine the wave drift and damping. Numerical simulations of individual tanker hulls and idealized objects at DnV have shown that representing hulls with a simple rectangular box of similar dimensions is a fair approximation. Thus, a ship may be parameterized using just the length, beam and draft. The wave forces on a given hull are calculated as functions of the wave spectrum. DnV have tabulated transfer functions in wave frequency space for both the wave drift and wave damping forces, for a range of box hull dimensions. The forces for a particular ship may then be estimated by interpolation in the database.

Figure 2 sketches the forces acting on a drifting ship. Usually, the wind and wave forces will act in about the same direction, but for the sake of generality they are given different directions. Observations (Sørgård and Vada, 1998) have revealed a difference in drift direction depending on whether the wave (and wind) direction is towards the port or starboard side of the ship. Due to limited data it was not possible from those observations to separate the wind and wave effects. However, the correlation with wave direction was observed to be the more significant. In the **met.no**/DnV model, this modification is included by altering the wave drag force direction by an empirically determined angle.

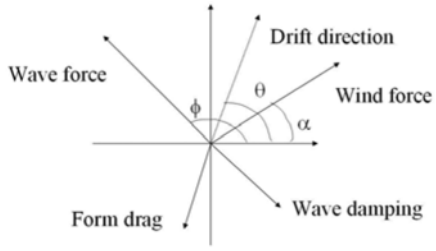

Figure 2. Forces acting on a drifting ship.

2.3 Stochastic approach to drift prediction

In predicting the drift of objects on the sea surface, we are faced with the challenge of accounting for uncertainties in almost all aspects of the task. We have already seen that the models for drift of objects and ships utilize empirical parameterizations (or empirically calibrated formulae) and imperfect approximations to the hydrodynamical laws. In addition, we often lack information about the object itself and where it is (or was at some time). And even if we did have this information, there is still the unavoidable uncertainty in the wind, wave and current data we use to drive the drift models. Accounting for uncertainty is most readily approached in a probabilistic framework. By assigning probabilities to the relevant parameters, an ensemble of numerical integrations can be performed where the various parameters are perturbed in a stochastic fashion. The perturbations are dictated by the pertinent probability distributions. Thus, we get a cloud of "candidate" positions for the drifting object. This cloud is itself a measure of search object's most probable location. Such a technique is known as a Monte Carlo integration and has been extensively used across many scientific disciplines (see e.g., Press et al., 1993).

The last known position. The first task in any real SAR or tracking operation is to determine the last known position (LKP). For SAR, this is a critical step, since the accuracy of this information is decisive for the outcome of the search. In the case of ship drift, the position at any time is often known very accurately. In the stochastic approach, an uncertainty is then assigned to the LKP, both in space and time. If the LKP is assumed to be very precise (e.g., a distress call is received from a ship with a GPS unit), a small radius may be assigned to the datum and all candidate objects (ensemble members) can be released at the same point in time. In the other extreme - a situation where little is known about the time and location of the accident - then a wide radius and a long period of time must be used. This will result in a cloud of possible initial positions scattered over a large portion of the sea surface released over an extended period of time. Thus, the various members of the ensemble will meet very different fates under the influence of differing current, wind and wave conditions. Obviously, the choice of initial distribution of ensemble members will affect the future search area seriously. It is the task of the Rescue Coordinator to estimate the LKP and its distribution.

Uncertainties in forcing fields and drift properties. In addition to the uncertainty assigned to the LKP, we also need to address the uncertainties present in the forcing data that are used and the drift properties of the object. The spread of the ensemble is thus a function of pertur-

bations to the time-invariant leeway coefficients (accounting for experimental variance) and time-varying perturbations to the wind field. The latter represent a random walk perturbation. A discussion of more advanced stochastic methods can be found in Griffa, 1996 and Berloff and McWilliams, 2002. Particularly the random flight model is useful when studying surface drift. However, for SAR purposes, it suffices to observe that the random walk will represent an upper bound on the inflation of search areas as it represents the maximum dispersion of particles in a given flow. Using random walk means erring on the side of caution in the sense that the size of the search areas is not underestimated. As search areas should be conservative (rather too large than too small), this makes sense with an operational SAR model.

The orientation of a drifting object. A final random factor is the orientation of the object with regard to the local wind direction. As the leeway drift for most objects contains a substantial cross wind component, there will be a significant discrepancy between the downwind direction and the direction of propagation of the object. Whether the object drifts to the left or to the right of the wind cannot be known in advance and unless more is known about the object we must assign equal probability to the two outcomes. Search areas are thus naturally bimodal, meaning that there will be two disjoint areas of high probability.

Furthermore, one could even "perturb" the object class if nothing is known about the object. In practice, however, this is done by running several integrations from the same initial conditions, once for, say, a person in water (PIW), then for a life raft, then for a swamped boat, etc. Overlaying the different trajectories will give a total search area.

Implementation of stochastic initial positions. The ensemble is O(500) and all members are positioned using a 2D normal distribution with a standard deviation equal to half the radius input by the user. The mean position of the particle cloud is a great circle arc and the release time varies linearly throughout the ensemble. This approach is flexible: it allows on one hand for point release in space and time and, in the other extreme, a contiuous release in space and time in a "trumpet" shaped area with one radius in one end and another radius in the other end of the seeding area (see figure).

3. Oil spill fate modeling

Oil spill fate models tend to be considerably more complicated than the models for surface objects, due to the range of oil types and the complex chemical processes that oil undergoes in the ocean (weathering). We will in the following outline the types of models in use and the

processes that they include. For a fuller review of the state of oil spill modeling, see Reed et al., 1999.

3.1 Oil and weathering processes

Oil spilled into the ocean ranges from unrefined crude oil to heavily refined products, with a corresponding wide range of chemical composition. In accidents, there can be several types of oil product spilled, for example crude oil cargo and diesel fuel. The chemical composition of the particular oil type and how it is spilled into the ocean determine to a large degree how weathering processes will transform the spill (Daling et al., 2003) and, consequently, what kind of remedial action can be considered.

The main weathering processes are evaporation, emulsification and natural dispersion:

Evaporation. For some oil types high in volatile fractions (e.g., many crude oils), evaporation from the surface slick removes a significant portion of the total mass within a short time, while other types (e.g., heavy fuel oils) lose relatively little to evaporation. Evaporation algorithms depend on the boiling point of the oil components, the ambient temperature, wind speed, film thickness and exposure time, although there is debate as to which parameters are important. The algorithms in use vary widely in computational expense and interpolation into empirical databases is a common practical approach.

Emulsification and natural dispersion. The uptake of water into the oil forms an emulsion, which may differ dramatically in drift behavior from the pure oil. Natural dispersion is the uptake of oil droplets of diminishing size into the water until they are no longer part of the oil slick in any practical sense. These two processes, illustrated in Figure 3, are competitors inasmuch as each reduces the rate of the other. In particular, the rate of natural dispersion is reduced with emulsion formation and, consequently, the lifetime of the slick is extended. Changes in the slick lifetime, in turn, affect the choice of response and possibility for environmental impact. All algorithms currently used for modeling emulsification and natural dispersion are curve fits to empirical data. Oil spill models must also track changes in the basic physical properties of the oil, such as density and viscosity, under the weathering process. These calculations are typically based on empirical tables. Inaccuracies can have important ramifications for the drift prediction. For example, the density of oil is close to that of water, and slight increases can result in increased downward mixing under the action of waves.

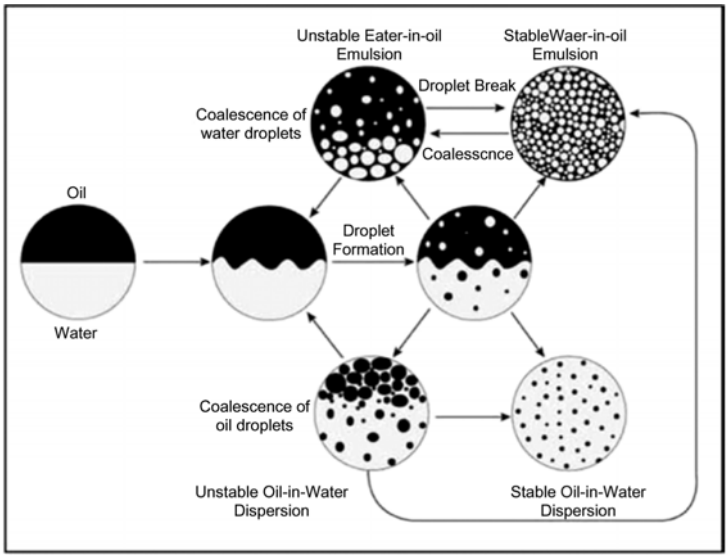

Figure 3. Illustration of emulsification and natural dispersion processes active when oil is spilled onto seawater. The action of waves initiates a complex mixing of oil droplets into water and water droplets into oil, resulting in stable emulsion and dispersion. From Daling et al., 2003.

At present, setting up algorithms to determine the properties of weathered oil starting from various oil products still relies heavily on field or laboratory experiments. These are time-consuming and expensive, and a methodology for parameterizing these properties on, say, basic petroleum assay data would be a valuable aid.

3.2 Transport processes: spreading and advection

By spreading we mean the spread of the oil from its source as a light fluid on top of a more dense fluid (water). This process may be described by fundamental gravity-viscous equations. Spreading affects the weathering processes since it influences film thickness. Reciprocally, evaporation and emulsification change the viscosity and density of the oil, thereby altering the spreading. At some point, the spreading process ceases. For most crude oils, this occurs quite early in the spill. Thereafter, the movement of the oil is dominated by the geophysical advective forces: currents, wind, waves and associated turbulence. During this initial phase, advective processes at scales of 10's to 100's of m are important, and it is believed that Langmuir cells play a central role. These

scales are smaller than those typically covered by the hydrodynamic forecast models used to drive oil spill models; these sub-gridscale effects must be parameterized from the available hydrodynamic data. Advection at scales larger than 1 km can be estimated from the hydrodynamic model data.

Spreading and advection of oil are also strongly influenced by the release conditions. So far, we have implicitly assumed a spill on the surface, or near enough that the oil rises to the surface unaltered. Deep, underwater releases may result in quite different initial surface slick conditions. The oil may be transformed during its rise to the surface and it will be advected by subsurface currents. In the case of a wellhead blowout, the oil may be accompanied by gas, which forms gas hydrates in contact with the water. Some oil spill model systems include special initialization modules for deep sources (Johansen, 1998, Wettre et al., 2001).

3.3 Model implementation

Representing an oil slick on the surface of the ocean and accounting for its motion, deformation and tendency to break up into smaller slicks, not to mention the changes due to weathering described above, is a formidable task. Over the years, several concepts have been introduced, ranging from simple center-of-mass trajectories to complex polygon representations. Perhaps the most popular type of model today treats the oil as a collection of discrete particles, each representing a certain mass of oil. The model currently in use at **met.no** is of this type, and we will explore it as an example. The main advantages of the particle representation are that it is inherently Lagrangian, it is amenable to a probabalistic approach and it reduces to a series of independent particle integrations. The disadvantage is that the results don't look much like an oil slick.

The **met.no** model OD3D has been developed together with SINTEF Applied Chemistry, who develops the weathering algorithms. An oil spill is described by a position, time of start, duration, spill rate, oil type, plus other optional parameters. Given this information, a prescribed number of particles is added (seeded) every time step (typically 1 hour) for the spill duration. The mass of each particle is determined by the flow rate, and remains constant throughout the simulation; mass loss due to evaporation and dispersion is effectuated by removing particles. Advective and weathering processes are applied on a particle-by-particle basis. The model is forced by atmosphere, ocean and wave data from a selection of sources, ranging from manually entered values to hindcast

and forecast model data. Particle information is output to a history file at hourly intervals for analysis and graphical rendering.

The particle based model described here is very similar to the probabalistic approach used in the SAR model described above. In both cases, a cloud of particles is spread and advected by geophysical forces. Aside from some differences in the seeding strategy, the real difference lies in the peculiarities of the particles and how we interpret the results; floating object particles are robust but have very special drift characteristics, while oil particles are fairly simple drifters but with complicated lifelines. The particle approach can be (and is) utilized in models of other things in the ocean, such as fish eggs and larvae.

4. Operational services

We have seen that, despite the fact that life rafts, tankers and oil are quite different things in the ocean, the models that are used to predict their drift have much in common. Primarily, they share a reliance on the same kinds of geophysical forcing data. When these models are to be implemented in an emergency service for real-time response to incidents, immediate access to prognostic forcing data is essential. The need is especially acute for SAR services, where minutes saved can mean lives saved. Consequently, forecast services for SAR, ship and oil drift are closely allied to operational centers for weather and ocean forecasting. Forcing data not only need to be available quickly, but also in the form of products suitable for the drift models. This requires preparation and testing of the full data production chain. The various drift models also share a need for efficient interfacing with the users - the crisis response teams in the field. Attaining optimal performance of the services is dependent on end-to-end testing and validation of the systems through regular exercises.

In order to facilitate rapid and reliable national response services for emergency drift episodes, Norway has implemented drift models for SAR, ship and oil drift at **met.no**. These models are directly interfaced to the operational forecast models for the atmosphere, ocean and waves. Many countries have similar arrangements. The responsibility for action in an emergency lies with the Joint Rescue Coordination Centres of Norway (JRCC) for SAR and ship drift, and with the Norwegian Coastal Authority for oil spills; they will request drift forecasts based on information at hand. A cardinal rule for these services is that a forecast should always be returned to the requesting party, even if the best available data basis is uncertain. Thus, backup alternatives to opera-

tional forecast model data are required, and uncertainty assessment is an essential part of the forecast information returned.

4.1 Geophysical forcing data access

An important task for the operational implementation is accessing the best possible forcing data at a given time and location. This can be a complicated task. In a SAR case, for example, the LKP may be many hours, even days old. At **met.no**, forcing data sets covering the last 7 days are maintained for rapid retrieval to meet such an eventuality. Furthermore, there may be several candidate forecast models, with different horizontal extent and resolution, capable of supplying the same type of forcing data (Figure 4). The choice of model data set to use for a drift forecast will in principle depend on the location and the presumed forecast accuracy of the data. However, in practical implementation, the choice is limited to models that are considered "officially operational" in the sense of established quality and robustness (e.g., supported by automatic backup systems, computer redundancy, archiving, etc.). In a typical national service, there will be a small number (1-2) of *operational* models for weather, ocean and wave forecasting, together with several *pre-operational* models being tested in the daily routine with the aim of replacing or supplementing the existing operational models. At **met.no**, the drift services currently obtain their atmosphere and wave forcing data from a selection of operational models, including **met.no**, ECMWF and UK Met Office, while ocean data are obtained from one operational model at **met.no**. The default is the **met.no** *operational* models (cf. Figure 4). In the event of total failure to obtain model forecast data, an operator may enter uniform values of wind, wave and current manually.

Recent developments in global ocean modeling and, not least, data exchange capability (e.g., the European Mersea project) are making it feasible to access adequate ocean forcing data from other operational forecasting centers. Thus, there is potentially a wide range of alternative data sets available. The **met.no** drift forecast service is being extended to allow selective access to a fuller range of forcing data sets, from local, high resolution in-house models to global data sets obtained from external sources.

The challenge of this approach is devising methods to determine which forcing data sets are best for a given emergency situation. For external data sets, one must ensure that they are reliably available and archived (e.g., for post mortem reruns), as well as make the necessary agreements on formats, data product requirements and delivery

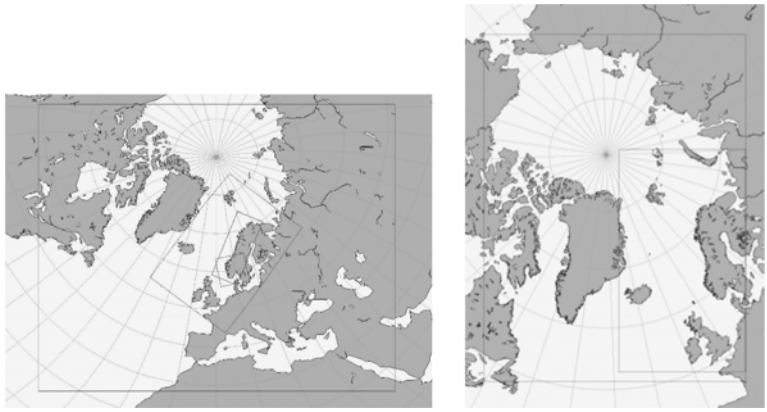

Figure 4. Geographical extent of operational models at **met.no**. Left: numerical weather prediction models. The largest rectangle is operational HIRLAM at 20 km resolution. The smaller domains are nested pre-operational models at resolutions of 10, 5 and 4 km. Right: ocean models. The largest rectangle is a pre-operational coupled ocean-ice model at 20 km resolution. The inset covering the Nordic Seas is the operational model at 4 km resolution.

schedules. These issues have been solved for atmospheric and wave data, through WMO (World Meteorological Organization) data exchange conventions. The situation for ocean forecast data is less mature, but is being vigorously addressed in several international initiatives (e.g., GODAE, GOOS, Mersea). Furthermore, the drift forecasting services need to find the optimal method of utilizing external data sets. Two options are: applying the external data directly to the drift model, and nesting local in-house models. Nesting may be done on a routine basis, as is typically done in weather forecasting (e.g., European national weather services nest limited area models in ECMWF global model data), or on a case-by-case basis using so-called "relocatable" models. Each method has advantages and difficulties, and the local drift forecasting service must judge what is best. Given the increasing number of forcing data options, it is imperative that the drift forecast services offer the right balance of forecast alternatives and simple, easily understood drift information to the field teams. This can only be done through comparative testing and validation of the alternatives. At present, skill assessment of drift forecasts is not well-established.

4.2 User interface

In the Norwegian emergency drift response services, a request is typically made by the duty officer to a meteorologist on watch at **met.no**, who, in turn, starts a forecast run of the relevant model; the drift forecast information is then sent back to the requesting officer in an agreed form. Since services for SAR, ship and oil drift have developed more or less independently over the years, the interface between **met.no** and the user has been somewhat different. However, the current development is moving away from manual operation and towards an automated production via similar web-page request forms for user input. The returned forecast information, on the other hand, is tailored to the needs of the particular emergency agency. Typically, the user will require some graphical products for quick assessment, but also forecast data to feed into their own crisis management tools, such as GIS. The Leeway user interface may serve as an example.

Figure 5 shows the Leeway request form that is filled out in a web browser by the duty officer at JRCC in the event of a SAR emergency or exercise. The request results in an automatic run of the Leeway model using the default operational atmosphere and ocean model forcing data (cf. Figure 4). Forecast data are returned as a compressed data file via email. The file is formatted so as to be readable by JRCC's SAR management tool. This tool has features tailored specifically to the JRCC's operations, such as overlaying on digital sea charts and calculating polygonal search areas. **met.no** maintains an in-house capability for graphical rendition of the forecast results; this serves both as a backup for the JRCC tool and as a development tool.

5. Outlook

Forecasting the drift of oil, ships and other floating objects have become standard ocean applications that address a clear demand from society. Modeling techniques have advanced considerably over the past 30 years, from rule-of-thumb models ("3% of wind speed and 15^o to the right") to complex numerical and empirical models. The models governing the fate of the drifting things - ship hydrodynamics, small object taxonomies and oil chemistry - are capable of giving increasingly detailed information on their specific behavior in the sea. There are still, however, significant deficiencies in these models; for example, object taxonomies need to be expanded to cover more object classes.

Improvements in drift forecast skill are currently being sought in the geophysical forecast data used to drive the drift models. Wind and wave forecasts are generally considered to be of good quality in the drift

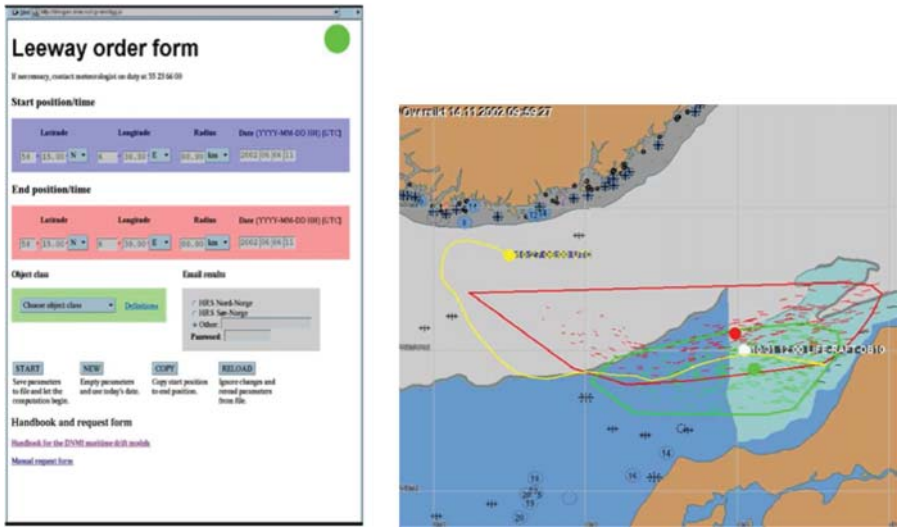

Figure 5. Example of a user interface to a SAR forecast service: **met.no** Leeway interfaced to Joint Rescue Coordination Centres of Norway (JRCC). Left: Snapshot of web browser request form. Sending the request starts a model forecast run. Results are returned JRCC by email. Right: Snapshot of drift forecast data presented in JRCC's management tool (SARA). Short line segments show particle paths over 1 hour: red = leeway to the left, green = leeway to the right (see text). Large red and green spots indicate centroid of corresponding particle clouds; white spot is centroid for all particles. Red and green polygons enclose corresponding particle clouds, indicating possible search areas. Yellow line is quick estimate of path of centroid for all particles. Data are overlaid on digital sea chart.

forecasting context, at least out to a day or two. The situation is less satisfactory for ocean currents and hydrography, which reflects the fact that ocean models exhibit variable forecast skill at the small scales that often are important in drift emergencies. However, the skill of ocean models is steadily increasing with improvements in computing capacity, observations and assimilation methods. An important aspect is the emerging capacity for global ocean forecasting, which is expected to give two benefits to drift forecasting services. One is an improvement in regional and local ocean forecasts via nesting of hydrodynamic models. The other is a capability for drift forecasting anywhere in the global ocean with improved skill. At the other end of the spatial scale, local operational ocean models are moving to higher resolution, giving increasingly improved definition of coastlines and topography, and consequently small scale dynamics. Since most SAR operations occur within 5 km of the coast, this is an important development.

Finally, the interaction of forecast providers with the people responsible for taking emergency action in the field needs to be maintained and enhanced. The task for drift forecast services is helping the response teams to use the forecasts and use them intelligently. This means making forecast products that are quickly understandable in a crisis situation; it also means attacking the difficult problem of estimating forecast accuracy. Education of response teams needs to be complemented by feedback from regular field exercises and post-crisis assessments.

References

Allen, A. A. (1999). Leeway divergence. Technical Report CG-D-XX-99, US Coast Guard Research and Development Center, Groton, CT, USA.

Allen, A. A. and Plourde, J. V. (1999). Review of leeway: Field experiments and implementation. Technical Report CG-D-08-99, US Coast Guard Research and Development Center, Groton, CT, USA.

Berloff, P. S. and McWilliams, J. C. (2002). Material Transport in Oceanic Gyres. Part II: Hierarchy of Stochastic Models. *J. Phys. Oceanogr.*, 32(March):797–830.

Daling, P. S., Moldestad, M. Ø., Johansen, Ø., Lewis, A., and Rødal, J. (2003). Norwegian testing of emulsion properties at sea - the importance of oil type and release conditions. *Spill Science & Technology Bulletin*, 8(2):123–136.

Griffa, A. (1996). Applications of stochastic particle models to oceanographic problems. In Adler, R, Muller, P, and Rozovskii, B, editors, *Stochastic Modelling in Physical Oceanography*, pages 113–128. Birkhauser, Boston.

Grue, J. and Biberg, D. (1993). Wave forces on marine structures with small speed in water of restricted depth. *Applied Ocean Research*, 15:121–135.

Hodgins, D. O. and Hodgins, S. L. M. (1998). Phase II leeway dynamics program: development and verification of a mathematical drift model for liferafts and small boats. Technical Report Project 5741, Canadian Coast Guard, Nova Scotia, Canada.

Johansen, Ø. (1998). Subsea blowout model for deep waters. SINTEF Report STF66 F98105, SINTEF Applied Chemistry, Trondheim, Norway.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1993). *Numerical Recipes in C*. Cambridge University Press, Cambridge.

Reed, M., Johansen, Ø., Brandvik, P. J., Daling, P. S., Lewis, A., Fiocco, R., Mackay, D., and Prentki, R. (1999). Oil spill modeling towards the close of the 20th century: overview of the state of the art. *Spill Science & Technology Bulletin*, 5(1):3–16.

Sørgård, E. and Vada, T. (1998). Observations and modelling of drifting ships. DNV Technical Report 96-2011, Det norske Veritas, Høvik, Norway.

Wettre, C., Johansen, Ø., and Skognes, K. (2001). Development of a 3-dimensional oild drift model at dnmi. Research Report 133, Norwegian Meteorological Institute, Oslo, Norway. 50 pp.

Chapter 24

ON THE USE OF DATA ASSIMILATION IN BIOGEOCHEMICAL MODELLING

Andreas Oschlies
School of Ocean and Earth Science, Southampton Oceanography Centre, Southampton, UK

Abstract A main objective of applying data assimilation methods to marine ecosystem models is the optimisation of often poorly known model parameters or even of the model's functional form. Recent efforts in this direction are reviewed. Results obtained so far indicate that presently available data sets can constrain not more 10 to 15 different ecological parameters. This raises questions about the use of more complex models. On the other hand, none of the optimised models yields a satisfactory fit to the observations, suggesting that present ecosystem models are overly simplistic. Implications of these apparently contradictory findings are discussed and a data assimilative strategy for future improvement of marine ecosystem models is suggested.

Keywords: Marine ecosystem models, data assimilation, parameter optimisation.

1. Introduction

Interest in prognostic models of marine biogeochemical cycles arises to a large extent from our need to better understand, quantify, and eventually predict the ocean's role in the global carbon cycle. This includes cycles of related elements, such as nitrogen, phosphorus or iron, that can act as limiting nutrients for phytoplankton growth. Other aspects addressed by biogeochemical and ecological modelling include the prediction of harmful algal blooms [*Schofield et al.*, 1999], and a quantitative understanding of oceanic food webs up to fish [*Loukos et al.*, 2003], birds, and humans, as well as the possible impact of marine sulfur emissions on the formation of cloud condensation nuclei [*Gabric et al.*, 2004]. In this chapter, I will focus on the carbon issue.

E.P. Chassignet Jacques Verron (eds.), Ocean Weather Forecasting, 525-547.

Carbon fluxes in the ocean are often described in terms of solubility pump and biological pump. The abiotic solubility pump is caused by increasing solubility of CO_2 (as of other gases) with decreasing temperature. For present climate conditions, deep water forms at high latitudes and average ocean temperatures are colder than average sea surface temperatures. The solubility pump then ensures that the volume averaged carbon concentration is larger than the surface averaged one, yielding a vertical CO_2 (or, more precisely, dissolved inorganic carbon, DIC) gradient with higher concentrations at depth than at the surface (Figure 1).

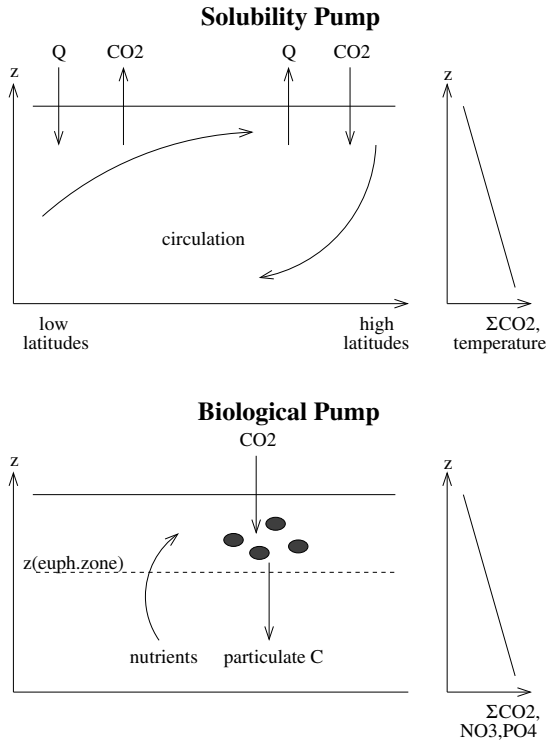

Figure 1. Schematic representation of the solubility pump (top) and the biological pump (bottom), both acting to maintain the vertical gradient in total dissolved inorganic carbon (ΣCO_2) in the ocean. Q is the surface heat flux, with oceanic heat uptake corresponding to outgassing and cooling to CO_2 uptake, z(euph.zone) is the depth of the euphotic zone which describes the surface region with light levels sufficient to allow for photosynthesis (typically about 100 m).

The term "pump" reflects that carbon is transported against the mean vertical gradient. Closer analysis reveals that of the presently observed

vertical DIC gradient, only about a quarter can be explained by the solubility pump [*Sarmiento et al.,* 1995] with the remaining three quarters being attributed to the "biological" carbon pump [*Volk and Hoffert,* 1985]. The driving agent of the biological pump is photosynthesis that generates organic carbon and thereby reduces DIC concentrations, and accordingly the partial pressure of CO_2, in the surface ocean. Respiration of organic carbon by metabolic processes in bacteria, higher trophic levels, and in the photosynthetically active phytoplankton itself reverses this process. As a result of mixing and advection along the vertical light gradient and because of the formation of biogenic particles that sink through the water instead of moving with it, respiration occurs generally deeper in the water column than photosynthesis. This decoupling of photosynthesis and respiration generates vertical gradients of DIC. To make things more complicated, some organisms form calcium carbonate "hard parts" which, on formation, sinking, and dissolution also affect the carbonate chemistry of sea water and result in an alkalinity pump. Because the formation of calcium carbonate in surface water increases surface pCO_2, this constitutes a counter pump in terms of CO_2 which partly compensates the pCO_2 effect of the organic carbon pump. A robust mechanistic understanding of the formation and biotically aided dissolution of calcium carbonate shells is not yet available, and many models so far assume that a fixed fraction of all biogenic particulate carbon sinking out of the light-lit euphotic zone (roughly the upper 100 m) is associated with calcium carbonate formation.

A close interaction of biology and physics arises not only from the interplay of physical and biological transport mechanisms on the vertical DIC gradient, but also from the fact that phytoplankton growth requires the presence of both light and nutrients, which usually have opposite vertical gradients. Accordingly, light and nutrient levels experienced by a phytoplankton cell are very sensitive to physical transport processes that may upwell or entrain deeper and nutrient-rich waters, or may mix or advect phytoplankton cells down into the dark ocean interior. This physical control on biological production has to be taken into account when attempting to simulate the marine carbon cycle. A standard strategy is to couple marine ecosystem models into circulation models. Validation of such coupled models is not straightforward. For example, the strong sensitivity of the marine biota to physical transport processes makes it difficult to separately evaluate the individual model components. For many applications one can at least safely neglect the biological impact on the physics via changes in the absorption profile of solar radiation [*Oschlies,* 2004]. While this allows to evaluate the physical model component individually, the reverse is not true for the

impact of the physics on the marine biogeochemistry. Here, a potential mapping of errors of the physical model onto the predicted ecosystem fields makes the separate evaluation of the ecosystem model component difficult. This is not necessarily a disadvantage: Because of the marine biology's strong sensitivity and fast response time of the order of days to changes in the light or nutrient supply, coupling ecosystem and circulation models may actually help to identify deficiencies of physical transport processes, particularly in the upper ocean [e.g., *Oschlies,* 1999].

The following section will give a brief overview over the field of biogeochemical models and presently used marine ecosystem models. Section 3 discusses some aspects of observations that are relevant for data assimilation, and section 4 addresses the potential use of combining data and biogeochemical models. Data assimilation methods are presented in section 5, and this article ends with a discussion of some achievements and perspectives of data assimilation in the field of biogeochemical modelling.

2. Biogeochemical modelling

Compared to numerical modelling of the ocean circulation, the field of biogeochemical modelling is much less mature. In particular, there is no known equivalent to the Navier-Stokes equations. In principle, these describe the motion of sea water exactly, but an exact solution does not (yet?) exist. The rules are thus clear for physical models, and different numerical models basically attempt to find different approximations to the unknown exact solution.

The situation is different in the field of biogeochemical modelling. Although there are some reliable, albeit mainly empirical, laws that describe transformations among various inorganic compounds dissolved in sea water, things become relatively shaky once life, and thereby transformations among organic and inorganic chemical compounds, comes into play. In practice, biogeochemical models are generally composed of an inorganic chemistry component and an ecosystem component, of which the latter is the by far more complex, expensive, and problematic part. In the following I will focus on this ecosystem model component and often use the term ecosystem model as synonym for the entire biogeochemical model.

Marine ecosystem models usually attempt to describe life's action on marine biogeochemical tracers by partitioning the marine ecosystem into a handful of boxes, often called compartments, such as phytoplankton (plants), zooplankton (animals), or detritus (dead organic matter). Sometimes, a further distinction is made between particular and

dissolved dead organic matter which does not sink but moves passively with the water. Besides the different transport properties, the distinction among dissolved and particulate organic matter is also useful in terms of elemental ratios: while the elemental composition of particulate organic matter is, on average, found to be close to the Redfield ratio [*Redfield,* 1934; *Redfield et al.,* 1963], dissolved organic matter often contains several times more carbon than the Redfield ratio would predict [*Williams,* 1995; *Kähler and Koeve,* 2001].

Using mass conservation as underlying concept, the compartments simulate stocks of atoms of the relevant element, and fluxes such as primary production, grazing, or mortality all describe the transfer of atoms among the different compartments. Often only a single element (usually one associated with a potentially limiting nutrient, e.g., nitrogen for nitrate, phosphorus for phophate) is modelled explictly, and its concentration in each of the compartments becomes a prognostic variable. Concentrations and fluxes of other elements (in particular, carbon) are usually diagnosed via the Redfield ratio. While this seems to be consistent with the analysis of average inorganic remineralisation products [e.g., *Anderson and Sarmiento,* 1995], more detailed investigations reveal local and temporal systematic deviations [*Körtzinger et al.,* 2001; *Sterner and Elser,* 2002; *Klausmeier et al.,* 2004]. A few recent models have therefore begun to explicitly resolve the cycling of different elements [*Moore et al.,* 2002].

For each marine ecosystem model, the particular choice of its compartments and of the parameterisation of fluxes between the compartments contains subjective elements, which may for example be influenced by operational measurement protocols, historical paradigms, or taxonomic nomenclature. Such an approach is, of course, valid and probably necessary in a field in which a strong theoretical framework is not yet available (and in which key species may not even be discovered yet). Progress will be made by trying to construct models that can explain the observations and at the same time tell a plausible story, and by more or less steadily changing the story as new observations add new information. In this process it is important to keep in mind that the underlying rules that make up a particular ecosystem model are generally assumed rather than demonstrated and hence are subject to change.

After these cautionary remarks about the theoretical foundations of marine ecoystem models, it is time to point out that these very models may greatly help to improve our understanding of marine ecosystems by allowing us to test the assumed hypothetical rules against observations in a quantitative way. In the following I will try to present my subjective view of how this can be achieved in practice.

2.1 Ecosystem model types

Today, a large variety of marine ecosystem models exist, probably similar in number to the number of researchers in the field. Although strict categories do not exist, present models roughly fall into three main groups (Figure 2):

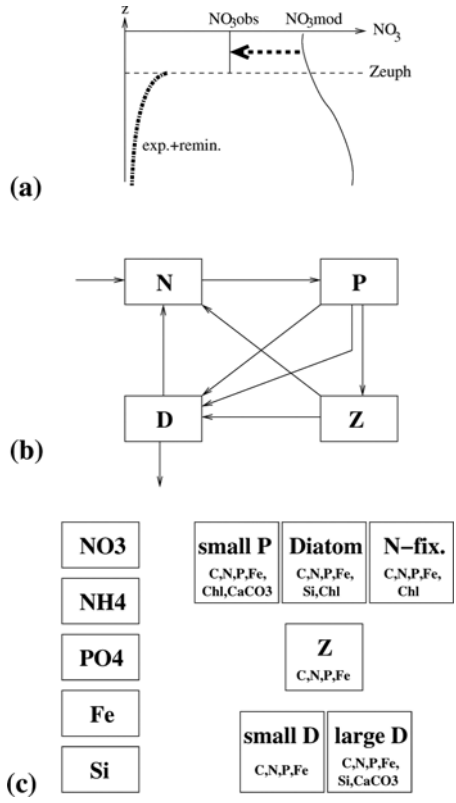

Figure 2. Schematic representation of various ecosystem model concepts: (a) Restoring of nutrients to observed or zero surface concentrations and immediate export and remineralisation according to a prescribed vertical remineralisation profile (e.g., *Bacastow and Maier-Reimer* [1990]). (b) Nutrient-Phytoplankton-Zooplankton-Detritus (NPZD) model (e.g., *Oschlies and Garçon* [1999]). (c) Multi-element multi-functional group model (after *Moore et al.*, [2002]). Each biological compartment is composed of sub-compartments for each of the prognostic elements and chemical compounds indicated. For clarity of the illustration, the $O(100)$ fluxes among the various (sub)-compartments are not shown.

- Nutrient-restoring models. These models do not explicitly resolve ecosystem components other than a (usually) single nutrient.

Biological production is simulated by restoring to either zero or observed nutrient concentrations in the light-lit surface layer, and instant sinking and remineralisation are accounted for by a prescribed remineralisation profile. Examples are the models of *Bacastow and Maier-Reimer* [1990], *Najjar et al.,* [1992] and the models used during phases 1 and 2 of the Ocean Carbon Model Intercomparison Project (OCMIP, [*Orr,* 1999]). Depending on whether or not dissolved organic matter is explicitly resolved, these biogeochemical/ecosystem models typically have 2 to 4 parameters. They are most widely used in models that address time scales much longer than a year, and applications to seasonal or shorter time scales will be problematic because of the absence of any particulate organic-matter storage pools.

- NPZD-type models. Although NPZD stands for Nutrient, Phytoplankton, Zooplankton, and Detritus, such models may contain a few more prognostic variables like bacteria or dissolved organic matter. Most of these models are descendants of a configuration proposed by *Fasham et al.,* [1990] and they explicitly simulate the cycling of either nitrogen or phosphorus. They have been applied to general ocean circulation models ranging from coarse resolution [*Sarmiento et al.,* 1993; *Fasham et al.,* 1993; *Chai et al.,* 1996; *McCreary et al.,* 1996; *Six and Maier-Reimer,* 1996] to eddy-permitting [*Oschlies and Garçon,* 1998, 1999] and eddy-resolving resolution [*Oschlies,* 2002]. Typically, these ecosystem models have 10 to 30 parameters.

- Functional-group type models. Going beyond the NPZD-type structure, these recently emerging models attempt to resolve different species or groups of phytoplankton and zooplankton. According to their special ecological function (e.g., nitrogen fixation, calcification) these are often called functional groups. These different groups require (and allow) to explicitly resolve the cycling of different biogeochemical elements. Examples are the models described by *Moore et al.,* [2002] and *Aumont et al.,* [2004], as well as the European Regional Seas Ecosystem Model (ERSEM, *Ebenhöh et al.,* [1997]) and the evolving Dynamic Green Ocean Model [*Le Quéré et al.*, Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, submitted to *Global Change*]. They typically have far more than hundred parameters.

3. Observations

A trivial statement is that the ocean is severly undersampled and that we need more data to better understand what is going on out there. We also have data of very different quality. There is a large number of data which are difficult to interpret in terms of ecological variables or processes resolved by dynamical models. Examples include wet zooplankton weight, satellite ocean colour data (which contain information on water-leaving radiance at a few wavelengths, but not immediately on surface chlorophyll or even primary production), or uptake of carbon isotopes into particulate matter (which is related but not identical to primary production, e.g., *Dring and Jewson* [1982]). Particular care has to be taken when different measurement methods that attempt – and often claim – to measure the same quantity (e.g., chlorophyll concentration, primary production) in fact measure different things . In contrast to more straightforward measurements of concentrations of standing stocks of organic or inorganic matter, direct observations of processes or rates (e.g., growth, grazing, sinking, exudation, mortality) are usually difficult to carry out without perturbing the system under investigation and, accordingly, are very limited in number and often have large random and systematic errors.

Available observations are also often biased towards the spring and summer season, with generally very few ship-based winter or autumn observations, particularly in mid and high latitudes. The same sampling bias holds for measurements of physical variables, but may be more critical for ecological properties for which the amplitude of the local seasonal cycle can be as large as the global range of the respective annual mean property.

Valuable observational information can also be taken from laboratory studies. Investigations using cultured species may, for example, help to reveal physiological information on the impact of environmental conditions like nutrient concentrations, light intensity, turbulence levels, or temperature on growth rates. A caveat to be kept in mind is that those species that have been cultured so far are not necessarily representative of the open-ocean plankton community. Considering that the number of generations separating our domestic plants and animals from their wild ancestors is reached by phytoplankton in only a few years, culture species may also be affected by selection and mutation.

It appears that information on the loss side (e.g., grazing, mortality) is more difficult to obtain than on the production side (growth). There is a (perhaps related?) tendency of marine ecosystem modellers to increase model complexity preferentially on the production side rather than on

the loss side. The net impact of marine biology on biogeochemical cycles is, however, controlled by the balance of production and loss processes. Because marine phytoplankton seems to invest relatively more into defence structures (mineral cell walls, spines, chains and colonies) than do land plants, which seem to compete more for fastest growth, one might even argue that marine ecosystems are more sensitive to loss processes than are terrestrial ecosystems [*Smetacek,* 2001].

4. Motivation for data assimilation

In a situation in which our understanding of marine ecosystem dynamics is still relatively poor and in which observations and data types are distributed unevenly, data assimilation may be seen as promising tool to interpolate in time and space as well as among different data types. Dynamical, albeit hypothetical, rules, e.g., in form of model equations, help to go beyond purely statistical interpolation and to link the observational information according to these rules. As such models have various poorly known parameters and functional relationships, data assimilation can at the same time provide a vehicle to estimate these parameters and parameterisations. This is conceptually different from state estimation that attempts to find a model state that agrees best with the observations and possibly a previous model forecast.

State estimation is used frequently in the field of meteorology to initialise new forecast simulations. For marine biogeochemistry, this aspect is generally less relevant, although it has already been applied for operational planning of research cruises [*Popova et al.,* 2002]. Forecast times are typically limited to a few weeks. The dissipative character of the dynamics that we believe to hold for marine ecosystems and that we use in our models [*Popova et al.,* 1997] and the strong seasonal and intraseasonal forcing in form of light, temperature, and mixing regimes lead to a quick memory loss of the initial conditions in typically much less than a year.

With respect to improving longer term forecasts, e.g., for climate prediction purposes, it seems to be more promising to rely on parameter estimation (and "parameterisation estimation") to improve our quantitative understanding of marine ecosystem dynamics. Data assimilation then provides a tool to test various hypothetical model dynamics against the available observations in an organised and quantitative way. The following sections attempt to give some overview over recent activities in this area.

5. Data assimilation methods used in marine biogeochemical modelling

5.1 Sequential methods

Sequential assimilation methods are constructed to accumulate information gathered from both observations and model predictions in time with the aim to generate an optimal state estimate. This approach is widely used in operational systems for which fast and robust delivery of information is a crucial aspect, ranging from instrument-guided aircraft landing, and related less peaceful applications, to atmospheric and oceanic weather forecasting. Most methods are approximations to, or descendants of, the Kalman Filter [*Kalman*, 1960]. Its basic principle is that of optimal interpolation between an observation and its simulated counterpart.

To illustrate its concept, we start by considering a model with a single state variable, x, for which observational counterparts are available. For Gaussian errors of the observation, x_{obs}, and model forecast, x_f, the best linear unbiased estimate (BLUE) of the true state vector is then given by:

$$x_a = \frac{\frac{x_f}{\sigma_f} + \frac{x_{obs}}{\sigma_{obs}}}{\frac{1}{\sigma_f} + \frac{1}{\sigma_{obs}}} \tag{24.1}$$

where σ_f and σ_{obs} are expected rms errors of model forecast and observation, respectively. The expected rms error of the analysed state, x_a, is

$$\sigma_a = \frac{1}{\frac{1}{\sigma_f} + \frac{1}{\sigma_{obs}}} \quad . \tag{24.2}$$

The fact that σ_a is smaller than both σ_f and σ_{obs} is consistent with x_a containing more information about the true state vector than any of the model forecast and observation alone. The analysed state, x_a, can then be used as initial condition to integrate the model until the next observation becomes available and the above process is repeated. Note, however, that the model forecast error will, in general, not be constant in time. For example, it may be large during particulate phases of the annual cycle (e.g., during the spring bloom), or it may be smaller when a lot of observations have been assimilated in the recent past. Computing the evolution of the model forecast error is the main contribution of the Kalman Filter. For a linear model, this is achieved in the following way:

$$\sigma^2_{n+1} = \mathbf{A}\sigma^2_n\mathbf{A}^t + \mathbf{e}^2_n \tag{24.3}$$

where $\mathbf{A}$ is the matrix that computes the evolution of the state vector $\mathbf{x}$, composed of the individual prognostic model variables, from time step

n to $n+1$ via $\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n$, $\mathbf{A}^t$ is its transpose, and $\mathbf{e}$ is some intrinsic uncertainty of the model that at each time step increases the forecast error. This equation is written in vector form to point out the main computational burden of the Kalman Filter: for an N-dimensional state vector $\mathbf{x}$, σ^2 is a $N \times N$ matrix. This means that the computation of σ^2_{n+1} is as expensive as stepping forward the state vector $2N$ times. In many applications, updating the covariance matrix will dominate the computational effort and often render it impractical.

Note, that, in addition to computational constraints, the above approach will provide optimal solutions only for linear systems and for Gaussian error distributions. Both conditions are not generally met for marine biogeochemical systems, and a number of adaptations to the original Kalman Filter approach have been developed. To cope with nonlinear systems, the so-called Extended Kalman Filter (EKF) steps forward the error covariance equation by the tangent linear operator (Jacobian) of the full model operator [*Evensen,* 1992]. While this works well for weakly nonlinear systems, the Ensemble Kalman Filter (EnKF) approach [*Evensen,* 2003] can also be applied to strongly nonlinear systems. It uses Monte-Carlo generated state vectors to initialise an ensemble of model forecast runs from which the error distribution of the model forecast is estimated. More recent developments include the Singular Evolutive Extended Kalman (SEEK) filter that has been applied to assimilate ocean colour data by *Carmillet et al.,* [2001]. A sequential method that is not directly related to the Kalman Filter is the Sequential Importance Resampling (SIR) filter [*Bertino et al.,* 2003].

A common feature of all sequential methods is the generation of a model trajectory that is "only" piecewise self-consistent. Whenever observations become available, merging the respective model forecast and observation into a new analysed state generates unsteady "jumps" in the state vector trajectory. Special care has to be taken if one wants, for example, to ensure mass conservation across these jumps. Similarly, any analysis of output from an assimilation experiment has to account for fluxes or perturbations associated with the assimilation steps.

5.2 Variational methods

While sequential assimilation methods attempt to estimate a "best" state vector at each instant observations become available, variational methods attempt to find a "best" model trajectory. The strong-constraint variational method ensures that over the entire time interval considered, the "best" model trajectory exactly obeys the model dynamics. Minimising a model-data misfit thereby becomes a constrained optimisation

problem, with the constraint being the model dynamics. Weak-constraint variational methods, on the other hand, allow for some uncertainty in the model dynamics and do not require that the "best" trajectory is an exact solution of the model equations [*Losa et al.*, 2004]. In principle, the uncertainty in the model dynamics can also be accounted for by the strong-constraint method: by explicitly introducing error terms that can be turned on or off by adjustable parameters to be optimised during the optimisation process, a strong-constraint method can also account for (and quantify) model errors. For this reason, only strong-constraint variational methods will be discussed here.

The clue to solving constrained optimisation problems is to identify the so-called control parameters on which the solution, or trajectory, of the dynamical model depends. Such parameters may be initial conditions, boundary conditions (e.g., nutrient supply from outside the model domain), or internal model parameters (e.g., maximum growth rates, mortality rates). Together, they form a control parameter vector $\mathbf{p}$, and any particular choice of $\mathbf{p}$ will, for the model under consideration, uniquely determine the temporal evolution of the model trajectory. The total model data misfit over the considered time interval is then a function of $\mathbf{p}$ only. This function is often called cost function. Assuming a total of M observations d_j, $j = 1, \ldots, M$, and model counterparts m_j, one simple choice for the cost function is:

$$J(\mathbf{p}) = \sum_{j=1}^{M} [d_j - m_j(\mathbf{p})]^2 \tag{24.4}$$

Any prior information about the parameter values (e.g., physiological constraints, positiveness) or the model trajectory (e.g., possible deviations from a stationary seasonal cycle, smoothness) should enter the cost function in form of additional terms. Constructing appropriate terms should always be possible, and any information we cannot measure in this way is probably useless anyway.

In principle, things are easy now: finding the "best" model trajectory is equivalent to finding the parameter vector $\mathbf{p}^{opt}$ that minimises the cost function $J(\mathbf{p})$. At closer inspection, however, things are a little more complicated. In general, we will have different kinds of measurements, i.e., for different j the corresponding observations d_j (and model counterparts m_j) may have different units as well as different error statistics. This is usually dealt with by introducing a scale factor S_j and replacing $[d_j - m_j]$ by $[d_j - m_j]/S_j$. Various choices of S_j have been used so far, e.g., $S_j = \sigma(d_j)$, $S_j = d_j$, $S_j = m_j$, $S_j = d_{ave}$, $S_j = d_{max}$, and little emphasis is usually put on investigating the implications of the actual choise made [*Evans*, 2003].

Another issue is the functional form of the terms entering the cost function. The most common approach of using sums of squares in the cost function gives the same weight to positive and negative misfits, and minimising least squares corresponds to a maximum-likelihood estimate only as long as the error distributions are Gaussian. In most cases this will not be correct. For example, there is strictly zero probability that nutrient concentrations are negative. This could, for example, be reflected in the cost functions by terms that go to infinity as simulated nutrient concentrations approach zero (e.g., by a high negative power of the nutrient concentration). A further aspect to consider is a possible correlation of different observations in space and in time. Sometimes it is attempted to take these into account by weighting different observations by the number of measurements taken. More frequent observations are then assumed to be correlated and accordingly downweighted in the cost function, whereas rare observations get relatively more weight. If prior information on the parameter values is available, the respective cost-function terms have to be weighted against the model data misfit terms as well.

When all information is accounted for by appropriately weighted cost function terms, the resulting cost function defines the metrics that measures the qualitiy of any parameter set. Construction of the cost function will always involve some subjective elements regarding the functional forms or weights of the individual terms. This basically reflect that, as usual in life, different people have different views on what is "best". The optimisation results will always depend on the quality of the cost function which therefore should be crafted as carefully as possible.

5.3 Minimisation methods

In principle, one could just explore the cost function "landscape" in parameter space by explicitly evaluating $J(\mathbf{p})$ for a large number of different choices of the parameter vector $\mathbf{p}$. In practice, however, this will usually not be possible. Even for a simple NPZD-type ecosystem model with, say, 15 parameters, a very coarse sampling of only 10 possible values per parameter would require 10^{15} evaluations of the cost function. As soon as models have more than a handful parameters, more efficient minimisation methods are needed.

A large variety of such minimisation methods exist, most of which have been developed outside oceanography. They can be devided into methods that make use of the cost function's gradient, i.e., information about the downhill direction in the cost-function landscape, and into methods that do not use this information and therefore do not require

the often expensive computation of the gradient. In the following, a brief overview will be given of methods that have so far been applied to marine ecosystem models.

Gradient descent methods. A standard conjugate gradient method, that alters search directions in consecutive iterations, has been applied by *Fasham and Evans* [1995] to optimise a model at the site of the North Atlantic Bloom Experiment (47°N, 20°W). The cost function's gradient was approximated by varying the individual parameters by a finite amount and computing the corresponding difference quotient. A finite-difference approximation of the cost function gradient was also used by *Prunet et al.,* [1996a,b]. It is generally and without detailed investigation assumed that the cost function is smooth enough so that the estimated gradient (times -1) points at least downward in the cost function topography and that errors in its exact direction and size will only slow, but not hinder, convergence of the descent algorithm. *Prunet et al.,* [1996a,b] in addition made the probably more critical assumption of a locally parabolic shape of the cost function at each iteration of a gradient descent method. Their sensitivity analysis indicated that this method did not generally yield robust parameter estimates, and that posterior error estimates were too small compared to the results of their sensitivity experiments.

In order to improve the quality of the gradient computation, the adjoint method has received considerable attention. It was first used in the context of marine ecosystem models by *Lawson et al.,* [1995]. The adjoint method computes the exact gradient of the cost function $J(\mathbf{p})$ by resorting to the method of Langrangian multipliers. This method has been widely used in statistical mechanics to derive the Euler-Lagrange equations. In essence, a Lagrange function L is defined as the cost function augmented by a additional terms that contain the model equations E_j multiplied by a corresponding (and a priori unknown!) Lagrangian parameter λ_j:

$$L(\mathbf{p}, \lambda, \mathbf{x}) = J(\mathbf{p}) + \sum_{j}^{j_{max}} \lambda_j E_j(\mathbf{x}) \quad . \tag{24.5}$$

At first sight, things now look much more complicated than for the minimisation of the cost function $J(\mathbf{p})$ only: The Lagrange function depends not only on the parameter vector $\mathbf{p}$ but also on a vector of Lagrangian multipliers λ and on the model state vector $\mathbf{x}$. However, because the function contains all the information we have, i.e., the cost function and the model dynamics, there are no further constraints to be considered, and the minimisation of the Lagrange function becomes an uncon-

strained problem. Accordingly, the minimum of $L(\mathbf{p}, \lambda, \mathbf{x})$ can "simply" be found by setting all its partial derivatives to zero. It turns out that the derivatives with respect to the components of the Lagrangian multipliers return the model dynamics, whereas the derivatives with respect to the components of the state vector will, after repeated application of the chain rule, return what is called the "adjoint model". The adjoint model can be understood as a model that runs the dynamics of the original "forward" model backward in time while being forced by the model-data misfits. A single backward run of the adjoint model, which uses the final state of a run of the forward model as initial condition, returns the full gradient of the cost function, $\nabla_{\mathbf{p}} J(\mathbf{p})$, at the position of the actual parameter vector $\mathbf{p}$. A gradient descent algorithm will then be needed to find a new parameter vector to start the next iteration of forward model run and adjoint model run.

The main advantage of the adjoint technique is its very efficient computation of a complete gradient in N-dimensional parameter space: Only a single forward and a single backward model integration are needed, whereas the other methods mentioned above require order N model integrations to compute an approximate version of the gradient. There is, of course, nothing like a free lunch: construction of the backward, or so-called adjoint model is a major effort. (Semi-)automatic compilers exist than can help turning the computer code of a forward model into its adjoint counterpart (e.g., described by *Marotzke et al.,* [1999]).

Applications of the adjoint method to parameter estimation for marine ecosystem models have shown some success [*Spitz et al.,* 1998, 2001; *Friedrichs*, 2002; *Gunson et al.,* 1999]. However, *Schartau et al.,* [2001] reported that the likely existence of local minima in the cost function may require restarting the gradient search from a large variety of initial estimates of the parameter vector. This is a problem any gradient-descent minimisation method will have. So far, however, it has not been conclusively demonstrated that local minima of the cost function do indeed exist. They are very difficult to identify in N-dimensional parameter space. Visualisations of two-dimensional sections through the N-dimensional cost function topography often show local minima [*Athias et al.,* 2000; *Vallino,* 2000; *Schartau et al.,* 2001], but it is by no means clear whether these are also minima in the other $N-2$ directions.

Stochastic minimisation methods. In order to avoid the expensive and often cumbersome computation of the cost function's gradient and also to cope with possibly existing local minima without having to fully scan the complete parameter space in a "brute force" mode, search algorithms have been developed that contain stochastic elements. These

can explore large regions of the parameter space and reduce the chance of getting trapped in a local minimum early on. An example is the Markov Chain Monte Carlo method [*Harmon and Challenor*, 1997], that also addresses estimating the posterior error distribution of the optimal parameter values. Construction of other stochastic minimisation methods was guided by attempts to understand the emergence of apparently optimal structures in nature. One of these is the concept of simulated annealing that was used by *Matear* [1995] to optimise parameters of a suite of ecosystem models to observations at Station P in the subpolar North Pacific. The simulated annealing technique is analogous to the thermodynamics that describe cooling and cristallisation of liquids. It consists of an iterative random selection of state vectors within a slowly narrowing probability distribution around the "best" parameter vector of the previous iteration. The width of this probability distribution decreases with increasing iteration number as does the probability of new parameter vectors being accepted with a higher cost function value than the "best" one of the previous iteration. A finite probability of uphill steps in the cost-function landscape is required to escape local minima. The probability is formulated in terms of the Boltzman factor that describes energy fluctuations in statistical mechanics, where the probability of transitions to more energetic states increases with temperature. If the temperature is decreased slowly enough to avoid getting trapped in local minima, this can eventually lead to a liquid cristallising into a perfect lattice that represents the state with lowest possible potential energy. Replacing potential energy by the cost function value, this behaviour is mimicked by the minimisation algorithm with a "temperature" parameter in the Boltzman factor decreasing with increasing iteration number. For typical ecosystem-model applications, several tenthousand evaluations of the cost function are needed to obtain a robust result. Still, convergence to the global minimum cannot generally be proven.

Hurtt and Armstrong [1996] employed simulated annealing to minimise the model-data misfit for a new, implicitly size-structured ecosystem model at Bermuda and later [*Hurtt and Armstrong,* 1999] extended this approach to test simultaneous optimisation of a similar model at the distinct sites of the Bermuda Atlantic Time series Study (BATS, 32°N, 64°W) and Ocean Weather Ship India (59°N, 19°W). While they found that different ecological processes had to be considered at the two locations to achieve a reasonable fit, *Schartau and Oschlies* [2003a,b] reported that simultaneous optimisation of a NPZD model at three sites (BATS, OWS India, and the site of the North Atlantic Bloom Experiment at 47°N, 20°W) worked almost as well (or as badly!) as separate optimisations at the individual locations. Their optimisation method

of choice was a (micro-)genetic algorithm. Although this method also requires several tenthousand evaluations of the cost function, a genetic algorithm was found to be slightly more efficient than simulated annealing in an idealised model study by *Athias et al.*, [2000]. The genetic algorithm basically looks after a population of parameter vectors $\mathbf{p}_i$ that are allowed to reproduce according to a fitness measured by their cost-function value $J(\mathbf{p}_i)$. Gene crossover in the reproduction step can be accounted for by exchanging various components of the parameter vector, and mutation can be included as a random perturbation of the individual parameter values in the reproduction step (Figure 3). Both

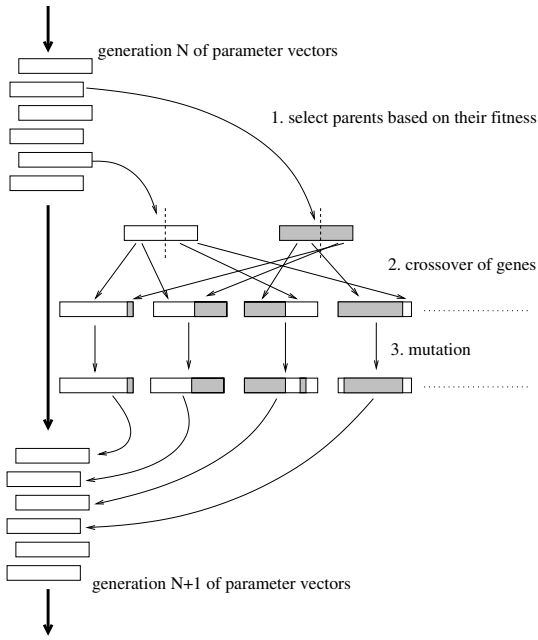

Figure 3. Schematic representation of the genetic algorithm. Out of an initial generation of parameter vectors, the fittest ones are selected according to their cost function value. In a second step, the selected parameter vectors can recombine (often the parameter vector is written as a binary string for this step), and in a third step there is low probability random mutation for individual elements of the parameter vector. For a micro-genetic algorithm, the fittest parameter vector of the parent generation is inherited to the next generation without any alterations. This elitism principle ensures that the lowest cost function value of each generation is at least as low as that one of the previous generation.

processes imply that even remote areas of the parameter space can be explored, whereas the gene-based reproduction concept ensures that information about good (i.e., "fit") parameter vectors (or their sub-sets)

is remembered by the algorithm. In many applications, the parameter "gene" is defined as the sequence of the components of the parameter vector written in binary notation, though a binary notation is not a necessary element of the method. As is the case with simulated annealing, convergence cannot generally be proven and will in practice depend on tunable parameters of the algorithm (mutation rate, population size).

6. Achievements and perspectives

The above assimilation and minimisation methods have so far been applied to ecosystem models that essentially all belong to the NPZD-type category. At present, there is no clear consensus on which method might be most efficient for this type of models. Sequential assimilation methods seem to have largest prospects in operational or near real-time applications that require good state estimates and that do not care that much about occasional jumps in the model trajectory. Variational methods, on the other hand, seem to be better suited for research issues that can take advantage of the dynamically self-consistent model trajectory. In principle, it is possible to obain a smooth model trajectory from sequential methods as well by applying a so-called smoother, which essentially consists of an integration backward in time. For linear systems it can be shown that sequential and variational algorithms can indeed produce identical results [*Bennett,* 1989].

A main advantage of stochastic optimisation techniques is the easy access to information about the posterior error of the parameter estimate. The large number of cost-function evaluations gives a reasonable picture of the cost function's sensitivity to changes in the individual parameters [e.g., *Schartau and Oschlies,* 2003a]. Particularly for strongly non-linear systems this may be more informative than local evaluations of the cost function's curvature via the Hessian matrix [*Fennel et al.,* 2001].

What is, to my knowledge, common to all of the assimilation studies performed so far is that at most 10 to 15 ecological model parameters could be constrained by the available observations. Although all these studies have so far employed relatively simple NPZD-type models, there were always a few parameters (or linear combinations of parameters) that could not be constrained. This indicates that even NPZD-type models have too many degrees of freedom and that models with fewer parameters should be able to reproduce the observations similarly well. Looking closer at how "well" the models can actually reproduce the data, one finds that even the optimised models fit the data very poorly. Usually, model-data misfits still amount to several standard deviations of the estimated prior errors. Such poor fits indicate that models need

more degrees of freedom to get closer to the observations. There are thus two contradictory statements about the required model complexity. This may to some extent result from errors in the physics used to drive the ecosystem models, but the problem is persistent even when physical observations are used to provide a physical environment as realistic as possible in a one-dimensional framework. Another possible explanation for the apparently contradictory statements about ecosystem model complexity is that the NPZD-type models employed so far may not have the right structures and hence are inherently inconsistent with the yet unknown ecological rules of real marine ecosystems. It is not clear whether more complex function-group type models would, in this data assimilative respect, perform any better. Although their order-of-magnitude larger number in adjustable parameters should allow for a much better fit to the data, such a fit would be of little value (and correspond to overfitting few data points by a high-order polynomial) for applying such models to other climate conditions unless all of the model parameters can be constrained by observations.

A possible strategy to clarify these issues is to undertake a systematic search for a model of minimum complexity that fits the available observations. Such an effort should start from a very simple model, perhaps similar to a nutrient-restoring one. Complexity should then be added only after careful analysis of the residual model-data misfits and some educated guess about the direction of complexity enhancement.

This approach for future model improvements should consider not only applications to the open ocean, where biogeochemical measurements are sparse and difficult to take, but also to controllable and nipulable field experiments, e.g., in artificial enclosures, and to physiological studies on cultures in the lab. A promising example is the application of ecosystem models and parameter optimisation methods to mesocosm experiments [*Vallino,* 2000]. Mesocosms are enclosed and generally gently mixed water volumes (typically several cubic meters in size), either in sea-water filled tanks on land, or in large plastic bags in the sea. They can be regarded as essentially homogeneous zero-dimensional systems that, in contrast to typical liter-sized incubation bottles, are large enough to keep boundary and enclosure effects small for a few generation times, i.e., days. By allowing for manipulation of environmental conditions and virtually unlimited access to observations, combining such experiments with modelling studies via data assimilative approaches may greatly help to better constrain our ecosystem models. On longer time scales, the same models will also have to be validated against open-ocean data collected within time-series programs and process studies. Physical models run in data assimilation mode may pro-

vide an optimal description of the physical environment for the marine ecosystem models to be run at these sites. Covering much longer time scales and more extreme climate states, even paleo information can help to constrain marine ecosystem models. By combining these very different time and space scales and the information provided by the different observational data sets with numerical models in an data assimilative approach, we will hopefully gain a better mechanistic understanding of marine ecosystem dynamics and their effects on biogeochemical cycles as well as their sensitivity to a changing climate.

Acknowledgements

I thank Eric Chassignet and Jacques Verron for allowing me to participate in this very enjoyable Summer School. This work is a contribution to the MERSEA Project. Partial support of the European Commission under contract SIP3-CT-2003-502885 is gratefully acknowledged.

References

Anderson, L. A., and J. L. Sarmiento, 1995: Global ocean phosphate and oxygen simulations. *Global Biogeochem. Cycles, 9*, 621–636.

Athias, V., P. Mazzega, and C. Jeandel, 2000: Selecting a global optimization method to estimate the oceanic particle cycling rate constants. *J. Mar. Res., 58*, 675–707.

Aumont, O., E. Maier-Reimer, S. Blain, and P. Monfray, An ecosystem model of the global ocean including Fe, Si, P co-limitations. *Global Biogeochem. Cycles, 17*(2), doi:10.1029/2001GB001745, 2003.

Bacastow, R., and E. Maier–Reimer, 1990: Ocean–circulation model of the carbon cycle. *Climate Dynamics, 4*, 95–125.

Bennett, A. F., 1989: The Kalman smoother for a linear quasi-geostrophic model of ocean circulation. *Dyn. Atmos. Oceans, 13*, 219–267.

Bertino, L., G. Evensen, and H. Wackernagel, 2003: Sequential data assimilation techniques in oceanography. *Int. Stat. Rev., 71,* 223–242.

Carmillet, V., J.-M. Brankart, P. Brasseur, H. Drange, G. Evensen, and J. Verron, 2001: A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical-biochemical model of the North Atlantic ocean. *Ocean Modelling*, 3, 167–192.

Chai, F., R. T. Barber, and S. T. Lindley, 1996: Origin and maintenance of high nutrient condition in the equatorial Pacific. *Deep-Sea Res. II, 43*, 1031–1064.

Dring, M. J., and D. H. Jewson, 1982: What does ^{14}C uptake by phytoplankton really measure? A theoretical modelling approach. *Proc. R. Soc. Lond., B 214*, 351–368.

Ebenhöh, W., J. G. Baretta, and J. W. Baretta, 1997: The primary production module in the marine ecosystem model ERSEM II with emphasis on the light forcing. *J. Sea Res., 38,* 173–193.

Evans, G. T., 2003: Defining misfit between biogeochemical models and data sets. *J. Mar. Syst., 40–41,* 49–54.

Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res., 97*(C11), 17,905–17,924.

Evensen, G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics, 53*, 343–367.

Fasham, M. J. R., H. W. Ducklow, and S. M. McKelvie, 1990: A nitrogen–based model of plankton dynamics in the oceanic mixed layer. *J. Mar. Res., 48*, 591–639

Fasham, M. J. R., J. L. Sarmiento, R. D. Slater, H. W. Ducklow, and R. Williams, 1993: Ecosystem behavior at Bermuda Station "S" and Ocean Weather Station "India": a general circulation model and observational analysis. *Global Biogeochem. Cycles, 7*, 379–415.

Fasham, M. J. R., and G. T. Evans, 1995: The use of optimisation techniques to model marine ecosystem dynamics at the JGOFS station at 47N 20W. *Phil. Trans. Roy. Soc. Lond. B, 348*, 203–209.

Fennel, K., M. Losch, J. Schröter, and M. Wenzel, 2001: Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *J. Mar. Syst., 28*, 45–63.

Friedrichs, M. A. M., 2002: Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean. *Deep-Sea Res. II, 49*, 289–319.

Gabric, A. J., R. Simo, R. A. Cropp, A. C. Hirst, and J. Dachs (2004), Modeling estimates of the global emission of dimethylsulfide under enhanced greenhouse conditions, *Global Biogeochem. Cycles, 18*, GB2014, doi:10.1029/2003GB002183.

Gunson, J., A. Oschlies, and V. Garçon, 1999: Sensitivity of ecosystem parameters to simulated satellite ocean colour data using a coupled physical-biological model of the North Atlantic. *J. Mar. Res., 57*, 613–639.

Harmon, R., and P. Challenor, 1997: A Markov chain Monte Carlo method for estimation and assimilation into models. Ecol. Model., 101, 41–59.

Hemmings, J. C. P., M. A. Srokocz, P. Challenor, and M. J. R. Fasham, 2003: Assimilating satellite ocean-colour observations into oceanic ecosystem models. *Phil. Trans. R. Soc. Lond., A 361,* 33–39.

Hurtt, G. C., and R. A. Armstrong, 1996: A pelagic ecosystem model calibrated with BATS data. *Deep-Sea Res. II, 43*, 653–683.

Hurtt, G. C., and R. A. Armstrong, 1999: A pelagic ecosystem model calibrated with BATS and OWSI data. *Deep-Sea Res. I, 46*, 27–61.

Kähler, P., and W. Koeve, 2001: Marine dissolved organic matter: can its C:N ratio explain carbon overconsumption? *Deep–Sea Res. I, 48*, 49–62.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic. Eng., 82D,* 35–45.

Klausmeier, C. A., E. Litchman, T. Daufresne, and S. A. Levin, 2004: Optimal nitrogen-to-phosphorus stoichiometry of phytoplankton. *Nature*, 429, 171–174.

Körtzinger, A., W. Koeve, P. Kähler, and L. Mintrop, 2001: C:N ratios in the mixed layer during the productive season in the Northeast Atlantic Ocean. *Deep-Sea Res. I, 48*, 661–688.

Lawson, L. M., Y. H. Spitz, E. E. Hofmann, and R. B. Long, 1995: A data assimilation technique applied to a predator-prey model. *Bull. Math. Biology*, 57, 625–651.

Loukos, H., P. Monfray, L. Bopp, and P. Lehody, 2003: Potential changes in skipjack tuna (*Katsuwonus pelamis*) habitat from a global warming scenario: modelling approach and preliminary results. *Fisheries Oceanography, 12,* 474–482.

Losa, S. N., G. A. Kivman, and V. A. Ryabchenko, 2004: Weak constraint parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data? *J. Mar. Syst., 45*, 1–20.

Marotzke, J., R. Giering, K. Q. Zhang, D. Stammer, C. Hill, and T. Lee, 1999: Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity. *J. Geophys. Res., 104*, 29,529–29,547.

Matear, R. J., 1995: Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. *J. Mar. Res., 53*, 571–607.

McCreary, J. P., K. E. Kohler, R. R. Hood, and D. B. Olson, 1996: A four-component ecosystem of biological activity in the Arabian Sea. *Prog. Oceanog., 37*, 193–240.

Moore, J. K., S. C. Doney, J. A. Kleypas, D. M. Glover, and I. Y. Fung, 2002: An intermediate complexity marine ecosystem model for the global domain. *Deep–Sea Res. II, 49*, 403–462.

Najjar, R. G., J. L. Sarmiento, and J. R. Toggweiler, 1992: Downward transport and fate of organic matter in the ocean: simulations with a general circulation model. *Global Biogeochem. Cycles, 6*, 45–76.

Orr, J. C., 1999: Ocean Carbon-Cycle Model Intercomparison Project (OCMIP), *Research GAIM, 2*, 7.

Oschlies, A., 1999: On spurious interactions among a mixed layer model, convective adjustment, and isopycnal mixing in ocean circulation models. *Mon. Weather Rev., 127*, 1920–1927.

Oschlies, A., 2002: Can eddies make ocean deserts bloom? *Global Biogeochem. Cycles, 16*, 1106, doi:10.1029/2001GB001830.

Oschlies, A., 2004: Feedbacks of biotically induced radiative heating on upper-ocean heat budget, circulation, and biological production in a coupled ecosystem-circulation model. *J. Geophys. Res.,* 110, doi:10.1029/2004JC002430.

Oschlies, A., and V. Garçon, 1998: Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean. *Nature, 394*, 266–269.

Oschlies, A., and V. Garçon, 1999: An eddy-permitting coupled physical-biological model of the North Atlantic. 1. Sensitivity to advection numerics and mixed layer physics. *Global Biogeochem. Cycles, 13*, 135–160.

Popova, E. E., M. J. R. Fasham, A. V. Osipov, and V. A. Ryabchenko, 1997: Chaotic behaviour of an ocean ecosystem model under seasonal external forcing. *J. Plank. Res., 19*, 1495–1515.

Popova, E. E., M. A. Srokosz, and D. A. Smeed, 2002: Real-time forecasting of biological and physical dynamics at the Island-Faeroes Front in June 2001. *Geophys. Res. Lett., 29*, 10.1029/2001GL013706.

Prunet, P., J. F. Minster, D. Ruiz–Pino, and I. Dadou, 1996: Assimilation of surface data in a one–dimensional physical–biogeochemical model of the surface ocean. 1. Method and preliminary results. *Global Biogeochem. Cycles, 10*, 111–138.

Prunet, P., J. F. Minster, V. Echevin, and I. Dadou, 1996: Assimilation of surface data in a one–dimensional physical–biogeochemical model of the surface ocean. 2. Adjusting a simple trophic model to chlorophyll, temperature, nitrate and P_{CO_2} data. *Global Biogeochem. Cycles, 10*, 139–158.

Redfield, A. C., 1934: On the proportions of organic derivatives in sea water and their relation to the composition of plankton. James Johnstone Memorial Volume, Liverpool, 176–192.

Redfield, A.C., B. H. Ketchum, F. A. Richards, 1963: The influence of organisms on the composition of sea water. In *The Sea*, Vol. 2, M. N. Hill (Ed.), Interscience, New York, 26–77.

Sarmiento, J. L., R. D. Slater, M. J. R. Fasham, H. W. Ducklow, J. R. Toggweiler, and G. T. Evans, 1993: A seasonal three–dimensional ecosystem model of nitrogen

cycling in the North Atlantic euphotic zone. *Global Biogeochem. Cycles, 7*, 417–450.

Sarmiento, J. L., R. Murnane, and C. Le Quéré, 1995: Air-Sea CO_2 transfer and the carbon budget of the North Atlantic. *Philos. Trans. R. Soc. London Ser. B, 348*, 211–219.

Schartau, M., A. Oschlies, and J. Willebrand, 2001: Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method. *Deep-Sea Res. II, 48*, 1769-1800.

Schartau, M., and A. Oschlies, 2003: Simulataneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I–Method and parameter estimates. *J. Mar. Res., 61*, 765–793.

Schartau, M., and A. Oschlies, 2003: Simulataneous data based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part II–Standing stocks and nitrogen fluxes. *J. Mar. Res., 61*, 795–821.

Schofield, O., J. Grzymski, W. P. Bissett, G. J, Kirkpatrick, D. F. Millie, M. Moline, and C. S. Roesler, 1999: Optical monitoring and forecasting systems for harmful algal blooms: possibility of pipe dream? *J. Phycol., 35*, 1477–1496.

Six, K. D., and E. Maier-Reimer, 1996: Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochem. Cycles, 10*, 559–583.

Smetacek, V., 2001: A watery arms race. *Nature, 411*, 745.

Spitz, Y. H., J. R. Moisan, M. R. Abbott, and J. G. Richman, 1998: Data assimilation and a pelagic ecosystem model: parameterization using time series observations. *J. Mar. Syst., 16*, 51–68.

Spitz, Y. H., J. R. Moisan, and M. R. Abbott, 2001: Configuring an ecosystem model using data from the Bermuda Atlantic Time Series (BATS). *Deep–Sea Res. II, 48*, 1733–1768.

Sterner, R. W., and J. J. Elser, 2002: *Ecological Stoichiometry: the biology of elements from molecules to the biosphere*, Princeton Univ. Press, Princeton, 439pp.

Vallino, J. J., 2000: Improving marine ecosystem models: Use of data assimilation and mesocosm experiments. *J. Mar. Res., 58*, 117–164.

Volk, T., and M. I. Hoffert, 1985: Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes, in *The Carbon Cycle and Atmospheric CO_2: Natural Variations, Archean to Present*, edited by E. Sundquist and W. Broecker, pp. 99–110, AGU Geophysical Monograph, 32, Washington D.C.

Williams, P. L. le B., 1995: Evidence for the seasonal accumulation of carbon-rich dissolved organic material, its scale in comparison with changes in particulate material and the consequential effect on net C/N assimilation rates. *Mar. Chem., 51*, 17–29.

Chapter 25

OCEAN FORECAST AND ANALYSIS MODELS FOR COASTAL OBSERVATORIES

John L. Wilkin and Lyon Lanerolle[1]

Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey, USA
[1]Now at: Coast Survey Development Laboratory, National Ocean Service, NOAA, Silver Spring, Maryland, USA

Abstract: Physical circulation processes in the coastal ocean affect air-sea interaction, sediment transport, the dispersal of nutrients and pollutants from terrestrial sources, and shelf-wide ecosystem dynamics and carbon cycling. A burgeoning network of coastal ocean observatories is expanding our ability to study these processes by simultaneously observing coastal ocean physics, meteorology, geochemistry and ecology at resolutions suited to quantitative interdisciplinary analysis. Complementary developments in ocean modeling have introduced more accurate numerical algorithms, improved parameterizations of unresolved sub-grid-scale mixing and boundary layer processes, and a transition to higher resolution on parallel computing platforms. The formulation and capabilities of modern coastal models are illustrated here with two examples from applications in the Mid-Atlantic Bight region of the northeast U.S. continental shelf. These are the Coupled Boundary Layers and Air-Sea Transfer (CBLAST) program centered on the Martha's Vineyard Coastal Observatory, and the Lagrangian Transport and Transformation Experiment (LATTE) centered on the Hudson River plume. The studies utilize the Regional Ocean Modeling System (ROMS) as a forecast tool to assist in the deployment of moveable instrumentation, and as a synthesis tool to aid the interpretation of observations. It is shown that regional models have the resolution and accuracy to capture the dominant features of the coastal ocean heat and salinity budget on diurnal to weekly time scales in regions with strong tides, vertical stratification, and highly variable bathymetry.

Keywords: Coastal oceanography, ocean modeling, coastal observatories.

E.P. Chassignet, Jacques Verron (eds.), Ocean Weather Forecasting, 549-572.

1. Introduction

The continental shelf seas represent 8% of the surface area of the World Ocean, yet are regions of far greater proportionate importance to human *activities. In addition to hosting up to 30% of the total global ocean primary* productivity, and more then 90% of the world's fish catch (Longhurst 1995), the coastal ocean represents a major chemical filter that transforms and accumulates nutrients and sediments from river and atmospheric sources, with the majority of the terrestrial inputs and biological productivity being re-mineralized on broad shelves with little export to the continental slope (Walsh et al. 1988, Biscaye et al. 1994). How nutrients, carbon and pollutants are introduced into eastern U.S. coastal waters from terrestrial sources, how they are transformed and transported while resident on the shelf, and how these physical, chemical, and ecological factors interact to regulate variability in primary productivity and higher trophic levels, is knowledge that is critical for assessment of climate change and human-induced effects on coastal ecosystems.

This paper describes, by example, how our knowledge of processes affecting coastal ocean transport and biogeochemistry is being expanded through the use of new observing technologies and improved coastal ocean models. The emphasis here is principally on results from a model of the Martha's Vineyard Coastal Observatory (MVCO) – this being the more mature coastal forecast system currently operated by the Rutgers Ocean Modeling Group. A description of the model design, implementation and validation is presented. This includes results from real-time forecasts and reanalyses of intensive observing periods at MVCO during the Coupled Boundary Layers and Air-Sea Transfer experiment (CBLAST). Additional preliminary results from related projects are presented to illustrate further aspects of the capabilities of a modern integrated coastal observation and forecast system.

On the U.S. East Coast, the discharge of many urbanized rivers is modified in estuaries, to a greater or lesser extent depending on residence time and other factors. For example, most of the nitrogen discharged into the Chesapeake Bay is assimilated by phytoplankton within the estuary (Malone 1996) and exported to the shelf as organic nitrogen, whereas 90% of the human source nitrogen load entering the New York lower estuary is exported unassimilated to the coastal ocean (Garside et al. 1976). The fate and transport of this material is controlled not only by biological and chemical processes, but also by the transport dynamics of river plumes as they enter the coastal ocean and interactions between the physical structure of a plume and the rates at which biogeochemical processes act. As human populations continue to grow along coastal margins, near-shore waters are subjected to increasing impacts from nutrients and pollutants, and outbreaks

of introduced species and harmful or nuisance organisms are increasing (Smayda 1990, Hallegraeff 1993, Anderson 1995).

On shelf-wide scales, it is recognized that the coastal northwest Atlantic is highly productive and plays an active role in the regional and global cycling of carbon and other elements (e.g., O'Reilly et al. 1987, Walsh et al. 1987). The limiting nutrient for phytoplankton production in the continental shelf ecosystems of the North Atlantic is nitrogen, in part due to significant benthic denitrification that results in a major loss from this system to N_2 gas (Seitzinger 1988, Seitzinger and Giblin 1996, Devol and Christensen 1993). In the Mid-Atlantic Bight (MAB), this loss is estimated to exceed the input of nitrogen from land and atmospheric sources. The high primary productivity is sustained by added nitrogen inputs from "onwelling" – various shelf-sea/deep-ocean exchange circulation processes that produce net onshore transport of new nutrients derived, ultimately, from regeneration in the deep ocean (Seitzinger 1988). Calculating a nitrogen budget for the MAB shelf therefore demands consideration of both biogeochemical processes, such as denitrification and primary production, but equally physical circulation and mixing.

Observational efforts to quantify shelf-sea/open-ocean exchange of nutrients and carbon include the Shelf Edge Exchange Processes (SEEP) experiments I and II (from 1983 to 1989) and the Ocean Margins Program (OMP) experiment (from 1994 to 1996). Results from SEEP showed that in the northern MAB only a small proportion of the particulate organic carbon (POC) produced on the shelf is exported while most of it is recycled by consumption or oxidized on the shelf (Anderson et al. 1994, Biscaye et al. 1994, Falkowski et al. 1988, Walsh 1994). However, the SEEP study assumed that carbon flux within the coastal ocean, and export off the shelf to deeper waters, is dominated by POC (typically measuring the larger sinking fraction of POC). The OMP experiment explicitly included contributions from dissolved organic and inorganic carbon (DOC, DIC) and suspended POC (Verity et al. 2002). We now know that the DOC pool in the MAB is one to three orders of magnitude larger than the POC pool (Bauer et al. 2001) and, being dissolved, is therefore readily transported by ocean circulation.

The salinity of shelf waters in the MAB and further north is significantly lower (3 to 7 psu) than adjacent open ocean waters due to the trapping of the circulation on the shelf by the shelf-slope front (Chapman and Beardsley, 1989). Significant inter-annual variations in this shelf water salinity, and to a lesser extent temperature, have been documented in the MAB by Mountain (2003). These anomalies have their origin in circulation processes acting outside the MAB, entering the region via the Gulf of Maine but being ultimately driven further upstream on the Scotian shelf or beyond in the Labrador Sea. As occurs for salinity, strong gradients in dissolved organic matter (DOM) concentration exist between coastal waters and the open ocean (Hopkinson et al. 2002). Mixing of shelf and slope waters by physical

circulation is therefore an important term when calculating a carbon budget for the MAB continental shelf. By analogy to passive tracers such as salinity, we also expect there to be inter-annual variability in nutrient and carbon reservoirs and fluxes in the MAB. Variability on these inter-annual time scales, and predicted longer term climate change, has important consequences for ecosystems (Boesch et al. 2000). For localized regions such as Narragansett Bay, time series studies suggest that climate change may affect plankton community composition, seasonal succession, and trophic interactions (Li and Smayda 1998, Smayda 1998), and for dominant species such as the copepod *Calanus finmarchicus*, basin-scale variability has been related to climate indicators observed or implied in both the Northeast Atlantic (Ottersen et al. 2001) and the Northwest Atlantic (Greene et al. 2003).

Our understanding of these processes in open shelf waters remains rudimentary, due in part to our historically limited ability to make multidisciplinary and multi-scale observations in an environment that is highly variable in space and time. But this is changing. We are poised to pursue studies of the coastal ocean through the widespread application of new observational technologies that allow long-term monitoring and adaptive sampling of physical, biological and chemical ocean conditions. A network of Coastal Ocean Observatories is evolving, globally, pioneered in many respects by the Long-term Ecosystem Observatory (LEO) established by Rutgers University on the New Jersey coast in the mid 1990s. A 10-km long electro-optic cable that powers and returns data from two sub-sea nodes in 15 m of water enables the deployment of vertically profiling temperature, salinity and bio-optical instruments returning long-term time series in real time. A concerted effort to acquire data from multiple satellites and the installation of a surface current radar system completed the backbone of what has become the Coastal Ocean Observation Laboratory, or the COOL Room (Glenn and Schofield 2003).

A series of Coastal Predictive Skill Experiments (CPSE) during the summers of 1999 to 2001 incorporated the LEO data in a coastal ocean model to deliver an ensemble of 3-day ocean forecasts that could be factored into the decision-making process for directing ship-based observations so as to adapt subsurface sampling to the evolving circulation. A re-analysis of the 2001 CPSE forecasts by Wilkin et al. (2004) explored the skill of the modeling system with respect to a set of subsurface validation mooring data (temperature and currents) recovered after the real-time experiment completed. It was found that the model had significant intrinsic predictive skill, and that this could be improved with the assimilation of sub-surface observations from ship-based towed-body observations and surface current data from radar.

Drawing on experience with the CPSE program, and wishing to employ models as a complement to the burgeoning network of coastal ocean observatories, we have formulated a number of limited area coastal ocean

models that share the objective of achieving reasonably faithful simulations of regional momentum, heat and freshwater transports on space scales of a few kilometers and time scales from tidal through diurnal cycle to several weeks. In this paper we describe two similar on-going efforts to coordinate ocean observing systems and predictive modeling in the Mid-Atlantic Bight region of the Northeast North American shelf. These are the Coupled Boundary Layers and Air-Sea Transfer (CBLAST) program centered on the Martha's Vineyard Coastal Observatory (MVCO) on the south coast of Massachusetts, and the Lagrangian Transport and Transformation Experiment (LATTE) centered on the Hudson River plume that emanates from New York Harbor. Though the two projects differ in the circulation and ecosystem processes on which they focus, and in some features of the observing technologies being applied, both adopt a similar approach to formulating a companion ocean modeling capability. In both projects, the ocean model is used both as a forecast tool to assist in the deployment of moveable observational assets, and as a synthesis tool to aid the interpretation of observations and their integration into conceptual frameworks that allow rigorous testing of hypotheses regarding coastal dynamics.

2. Coastal observatories

Coastal ocean observational capabilities have advanced greatly in the past decade. New observational technologies include CODAR (Coastal Ocean Dynamics Application Radar) systems for mapping surface currents over broad swaths of the coastal zone, autonomous underwater vehicles (AUV) with physical and optical sensors that can profile the water column out to the shelf break, and the advent of cabled observatories that allow the long-term deployment of sub-sea time series instrumentation such as profiling conductivity-temperature-depth (CTD) sensors, acoustic Doppler current profilers (ADCP) and biological recorders. The availability of data in real-time from these sources opens opportunities for operational oceanographic applications in support of environmental quality monitoring, fisheries management, maritime operations, and fundamental research.

In the following subsections we describe, briefly, two coastal observing systems in operation in the Mid-Atlantic Bight. These are two among the several regional observatories already established from Cape Hatteras to Nova Scotia that collectively form the Northeast Observing System partnership.

2.1 Martha's Vineyard coastal observatory and CBLAST-Low

The inner continental shelf of southeastern New England south of Cape Cod, Massachusetts, encompasses a variety of circulation regimes delineated by the geography of the region. The waters of Nantucket Sound are shallow and relatively sheltered by the islands of Martha's Vineyard and Nantucket. Tidal mixing on the Nantucket Shoals vertically mixes the water column throughout the year, while the waters south of Martha's Vineyard undergo a seasonal cycle of stratification and mixing. The Martha's Vineyard Coastal Observatory (MVCO) is a permanently cabled site, like LEO, that sits 3 km from the southern shore of Martha's Vineyard. The MVCO includes a meteorological mast on land, an undersea node at a depth of 12 m, and, in waters 15 m deep, an offshore air-sea interaction tower (ASIT) that spans the water column and extends 20 m into the atmosphere. Routinely obtained MVCO measurements now include wind velocity, air temperature, and solar radiation at the meteorological mast, near-bottom temperature and salinity at the undersea node, velocities throughout the water column from an ADCP at the undersea node, and estimates of Reynolds stress from an array of acoustic Doppler velocimeters (ADVs) near the bottom at the ASIT.

The CBLAST-Low program is a study of air-sea interaction at low wind speeds – a regime for which atmosphere and ocean boundary layer processes are modulated significantly by thermal forcing. In addition to the fixed instrumentation at MVCO and ASIT, additional instruments were deployed in the environs of MVCO during CBLAST Intensive Observing Periods in the summers of 2001, 2002 and 2003 (Hutto et al. 2003). In the most comprehensive field season, 2003, the first set of atmosphere and ocean flux observations from the ASIT became available, an array of 6 'heavy' moorings south of ASIT measured meteorological boundary layer properties (winds, temperature, humidity and radiation) at the surface buoy plus ocean temperature and velocity profiles to the seafloor, a further 9 'light' moorings observed sub-surface temperature profiles, surface currents were measured from a CODAR site on Nantucket, research aircraft profiled the marine boundary layer, and vertical thermistor strings were towed by ship through thermal features evident in satellite and aircraft imagery. The U.S. Navy operated the COAMPS (Coupled Ocean Atmosphere Mesoscale Prediction System) model in a multiply-nested configuration; the 27-km resolution West Atlantic operational product (Hodur et al. 2002) was refined through a 9-km resolution intermediate mesh to a very high (3-km) resolution experimental forecast specifically for the CBLAST-Low study (Wang 2004).

Observations from the fixed ASIT tower are being used to quantify the vertical fluxes of turbulent kinetic energy, momentum, mass, and heat in the oceanic mixed-layer and atmospheric boundary layer. With coincident local measurements of radiative fluxes, the independent air-side and water-side flux estimates can be compared to those derived from the bulk formulae

(Fairall et al. 1996) in widespread use by ocean and atmosphere modelers alike. The validated fluxes and in-water profile observations are well suited to evaluating the suite of closure options used to parameterize vertical turbulence in ocean models. However, this comparison is possible only if the model captures the essential features of the ocean heat budget on diurnal to several day time-scales, and spatial scales of order 1 km. This is required because, as noted above, the environs of Martha's Vineyard and Nantucket are characterized by complex bathymetry, significant stratification, and tidal eddy heat transport and mixing. As a result, 3-dimensional ocean circulation leads to lateral stirring and advection that constitute a significant term in any heat budget calculation at the MVCO site. Ocean modeling for CBLAST, which we describe in section 3, therefore has the dual objectives of critically evaluating numerical parameterizations within the model and complementing the interpretation of the field observations by quantifying unobserved lateral divergence of heat.

2.2 New Jersey shelf observing system and LATTE

Buoyant coastal currents extend along much of the U.S. East Coast and consist of a series of estuarine plumes that are fed by rivers with typical maximum discharge rates on the order of 1000 m^3/s. Among these the Hudson River is typical, yet it may dominate the transport of nutrients and chemical contaminants to the coastal ocean. For well over 100 years it has been the most urbanized estuary in the U.S. For example, only recently has Los Angeles's population exceeded that of New York in 1900; today over 20 million people live in its watershed. New York and New Jersey Harbor account for some 4% of sediment loadings in the Virginian Province (Cape Code to Chesapeake Bay), yet are responsible for 90% of sediments in this Province that exceed the EPA standard for total PCBs and 69% of those that exceed the mercury standard (Adams et al. 1998). Levels of nutrients and metal complexes in the dissolved phase can be an order of magnitude higher than in ambient shelf waters. It is arguably the most contaminated estuary on the East Coast.

Now termed the New Jersey Shelf Observing System (NJSOS), Rutgers' coastal observatory has been refocused from the LEO region to the apex of the New York Bight by an expansion to the network of CODAR instruments, the installation of a local X-band satellite receiver, and operation of a fleet of long-duration glider-type AUVs for subsurface physical and bio-optical observations. The NJSOS concept is designed, in part, to focus attention on the fate and transport of dissolved and particulate organic matter, inorganic nutrients and metals discharged onto the shelf in the Hudson River plume.

Using the NJSOS as a foundation, the centerpiece of the LATTE program is a series of dye tracer experiments (over 3 years) during the spring

peak in river flow. By tracking the dye with continuous underway sampling using towed vehicles, biological and chemical transformations can be observed in a Lagrangian perspective. Tracking the dye makes it possible to distinguish biogeochemical processes from physical processes that transport material in the buoyant plume, or mixing that merely dilutes it. Furthermore, the physical structure of river plumes differs in upwelling and downwelling wind conditions. Downwelling leads to a narrow near-shore buoyant plume that is thick, typically bottom attached, and rapidly transports material alongshore. Under these conditions enhanced turbulence retains particulate matter throughout the water column causing low light levels that constrain phytoplankton growth and colored DOM photo-bleaching rates. During upwelling, plumes can detach from the coast, spread, thin, and transport water directly offshore (Munchow and Garvine 1993, Fong and Geyer 2001, Hallock and Marmorino 2002). The enhanced stratification reduces turbulence and particulate matter settles out of the plume, elevating light levels and fostering phytoplankton growth, photo-bleaching, and the potential for bio-accumulation of metals.

The motivations for making ocean modeling an integral component of LATTE are rather different to those for CBLAST. The principal observing systems used in CBLAST (moorings and MVCO/ASIT) were fixed platforms whose locations had to be chosen in advance of the experiment. In LATTE, the use of relocatable sampling methods (dye, AUVs, towed bodies) enables the observing system to be adapted to the flow, and ocean forecasting becomes of rather greater utility.

The potential value of forecasting the plume trajectory under varying wind conditions became evident in the LATTE pilot experiment in April 2004. The first dye release coincided with the onset of upwelling winds that drove the patch onto the south shore of Long Island, prematurely curtailing the Lagrangian experiment. A second dye injection was made shortly thereafter, and a switch to downwelling winds favored the formation of a more classical coastally trapped plume that carried the dye south along the Jersey Shore.

The LATTE program is in its infancy having passed only its first milestone, namely the 2004 pilot program dye release. Accordingly, only preliminary modeling results for LATTE are available. These will be presented in section 3 as an illustration of some of the issues to be addressed for coastal ocean forecasting in this situation. The data synthesis aspects of the modeling, not yet realized, are quite different from CBLAST. Firstly, variational data assimilation (Moore et al. 2004, Weaver et al. 2003) applied to the measured dye distribution will be used to provide high a resolution hindcast of the evolving plume trajectory and physical structure. Secondly, a coupled physical/biological model that explicitly computes the bio-optical features of suspended and dissolved matter (Bissett et al. 1999a, 1999b) will be used to hindcast the depth-dependent distribution of phytoplankton and DOM and predict the inherent optical properties of the water column. This

will contribute to testing the hypotheses regarding the interaction of plume structure with photo-chemistry, productivity rates and bio-accumulation.

3. Coastal ocean modeling

Developments in observing systems are matched by progress in recent years in the capabilities of coastal ocean models, due in part to increases in computer technologies but more significantly through improved methods in computational fluid dynamics, attention to the physical veracity of subgridscale parameterizations, the application of advanced data assimilation methods, and widespread experimentation in the formulation of ecosystem and biogeochemical models.

3.1 The Regional Ocean Modeling System – ROMS

The model we have adopted for the LEO, CBLAST, LATTE and other similar high to medium resolution coastal modeling applications is the Regional Ocean Modeling System (ROMS) (http://marine.rutgers.edu/po/index.php?model=roms). ROMS is a versatile, free-surface, hydrostatic primitive equation ocean circulation model developed by collaborators at several institutions, but led by specialists at Rutgers and UCLA. The model is being used for applications from the basin to coastal and estuarine scales (e.g. Haidvogel et al. 2000, Marchesiello et al. 2003, Lutjeharms et al. 2003, Peliz et al. 2003, Dinniman et al. 2003, MacCready and Geyer 2001).

ROMS is formulated in a vertical terrain-following 's-coordinate' similar to classic sigma-coordinate models, but with generalizations that allow selective weighting of the vertical distribution of points toward the free surface or seafloor, or both. The terrain-following coordinate is attractive for coastal applications because provides an accurate representation of the vortex stretching term that dominates coastal-trapped wave dynamics and the bathymetric steering of coastal currents. The optional weighted stretching of the vertical coordinate allows for enhanced resolution in the upper ocean mixed layer and turbulent bottom boundary layer. The horizontal discretization is by an orthogonal curvilinear Arakawa C-grid.

Shchepetkin and McWilliams (1998, 2003, 2004) describe in detail the algorithms that comprise the ROMS computational kernel, and these have been summarized recently by Haidvogel (2004). They include careful formulation of the time-stepping algorithm to allow both exact conservation and constancy preservation for tracers, while achieving enhanced stability and accuracy in coastal applications where the free surface displacement is a significant fraction of the total water depth. A redefinition of the barotropic mode reduces the mode splitting error associated with solving the vertically-integrated momentum equation on a much smaller time-step than the tracer

equations, i.e. the 'split-explicit' formulation popular in free surface ocean models. Conservative parabolic-spline discretization in the vertical significantly reduces the pressure-gradient truncation error that has previously plagued terrain-following coordinate models.

Tangent linear and adjoint versions of ROMS have been developed and are being turned to applications including sensitivity analysis, stability analysis, ensemble prediction and variational data assimilation (Moore et al. 2004). The ROMS code has been structured for efficient performance on parallel-computing platforms (using MPI or OpenMP).

The parameterization of vertical turbulence in coastal models can have ramifications for details or even some qualitative features of the flow, especially the transport of suspended matter and sediments (Durski et al. 2004, Wijesekera et al. 2003, Warner et al. 2005). Among the options for the parameterization of vertical mixing in ROMS are the k-profile parameterization (KPP) of Large et al. (1994), the level 2½ Mellor and Yamada (1982) scheme, and the suite of two-equation closures (one equation for turbulence kinetic energy and a second equation for a generic turbulence length scale quantity) proposed by Umlauf and Burchard (2003) that implement the widely used k-epsilon and k-omega closures, and a revised form of k-kl (Mellor-Yamada level 2½). The two-equation closures are completed by stability functions based on the parameterizations of Galperin et al. (1988), Kantha and Clayson (1994), or Canuto et al. (2001). As an adjunct to vertical turbulence closure, the parameterization of bottom boundary stress includes, optionally, quadratic drag or the effects of wave-current interaction, moveable beds, and ripples (Soulsby 1995, Harris and Wiberg 2001, Li and Amos 2001). This implementation of an extensive suite of turbulence closures in a single 3-dimensional ocean model allows the systematic comparison of the schemes in the context of the CBLAST observations.

3.2 ROMS modeling for CBLAST

In order to make a meaningful comparison of the CBLAST-Low observations with the modeled regional heat budget, ROMS must capture the essential features of the 3-dimensional heat transport on diurnal to several day time-scales, and spatial scales of order a few kilometers. To achieve this objective, we have employed a high degree of realism in the configuration of model bathymetry and forcing. The model has fine grid spacing (1 km) and realistic bathymetry from a 3-arc-second Coastal Relief Model (NGDC, 2004). Open boundary conditions are specified following the method of Marchesiello et al. (2001): Orlanski-type radiation is applied to tracers and baroclinic velocity in conjunction with relaxation (with timescale of 0.5 days on inflow and 10 days on outflow) to a regional bi-monthly climatology of shelf circulation from the semi-diagnostic model of Naimie et al. (1994).

The free surface and depth-integrated velocity boundary conditions use the method of Flather (1976) with the external values specified by the climatology plus tidal variability (harmonics M_2, N_2, S_2, K_1, O_1, M_4 and M_6) from an ADCIRC model simulation of the western Atlantic (Luettich et al. 1992). Air-sea fluxes of momentum and heat were computed using bulk formulae (Fairall et al. 1996) applied to the modeled sea surface temperature and atmospheric marine boundary values (10-m wind, 2-m air temperature, sea level pressure and relative humidity) and downward shortwave and long-wave radiation from the 3-km resolution COAMPS forecast. Mellor-Yamada (1982) mixing and quadratic bottom drag complete the model configuration.

During August-September of 2003, ocean forecasts were produced each day as an aid to daily operations of the field program. Data from 72-hour COAMPS forecasts that started each day at 0000 UTC and 1200 UTC were delivered to Rutgers after the forecast completed, typically about 12 hours after the start of the forecast cycle. The subsequent ROMS forecast, initialized from the previous ocean model run, was therefore at best a 60-hour forecast depending on the timeliness of the COAMPS data transfer, which was sometimes delayed. Summary 3-hour interval forecast results were posted on the web each day showing near surface currents and sea temperature with the coincident COAMPS winds, plus a summary for each forecast cycle of simulated Lagrangian float paths as a visualization the pathways of lateral heat advection.

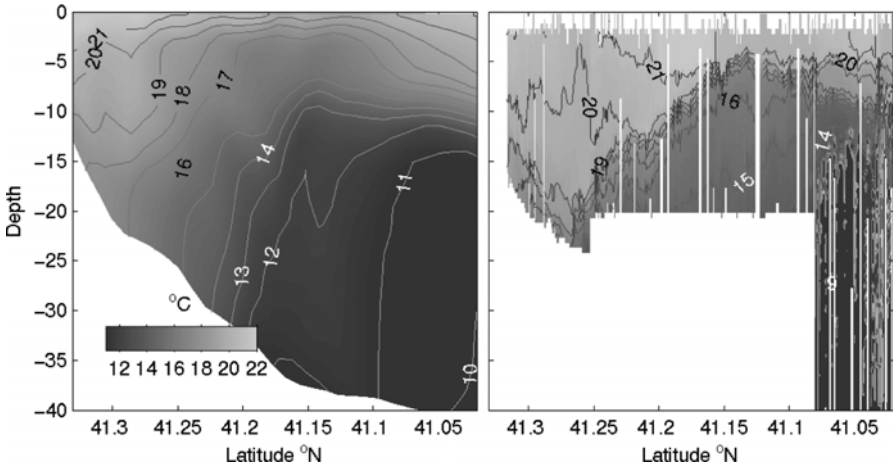

Figure 1. Vertical cross-section of temperatures along a line from Martha's Vineyard to mooring A of the 2003 CBLAST array. Left: ROMS model on 19-Aug-2003. Right: Observed by a glider-type autonomous underwater vehicle (AUV).

A limited validation of the 2003 forecast system that operated in real time is offered in Figure 1 showing a vertical cross-section of the

temperature forecast for August 19 in comparison to in situ data gathered over the 3 days August 18-21 by a Webb Research coastal glider AUV that traveled from the coast of Martha's Vineyard due south 30 km to the site of CBLAST mooring 'A' (Hutto et al. 2003). Adjacent to the south coast of Martha's Vineyard, at the left of the section, a 'bowl' of warm water is portrayed in both forecast and observations. To the south, a sharp thermocline at 10 to 15 m depth with a temperature difference of 8 to 10 ^{o}C is also forecast well. The intervening zone is characterized by a shallow mixed layer overlaying a more diffusely stratified water column. A more comprehensive validation, plus an investigation of the physical origins of the observed spatial patterns, has been made for hind-casts of the 2002 field season.

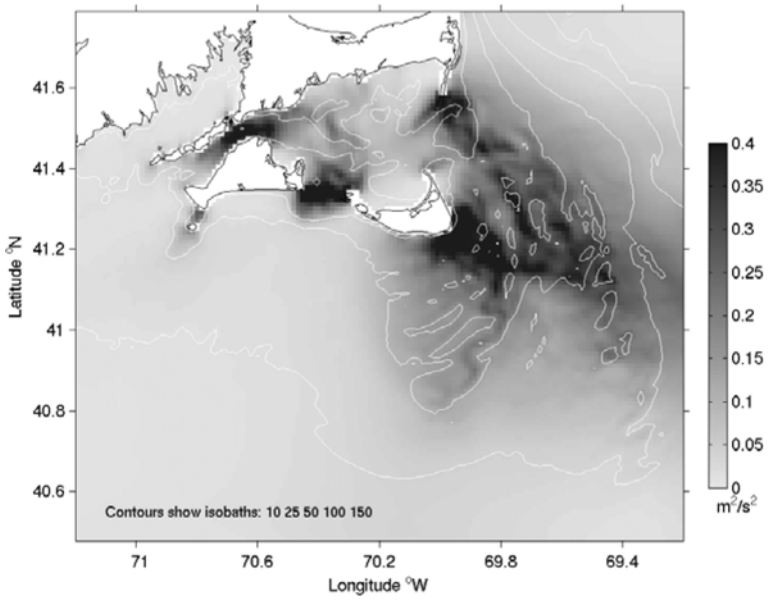

Figure 2. July 2002 mean eddy kinetic energy. Tidal mixing generates a region of perpetually cold SST on the eastern flank of the Nantucket Shoals.

For the 2002 simulations the downward long-wave and shortwave radiation data from COAMPS were not archived completely so these were supplanted by radiometer observations at MVCO (Hutto et al. 2003) in the ROMS forcing conditions, but we do not expect this loss of spatially resolved radiative heating to be a limitation because radiation observations at five surface meteorological moorings show little spatial variability. In all other respects the model configuration is the same as for the 2003 forecasts.

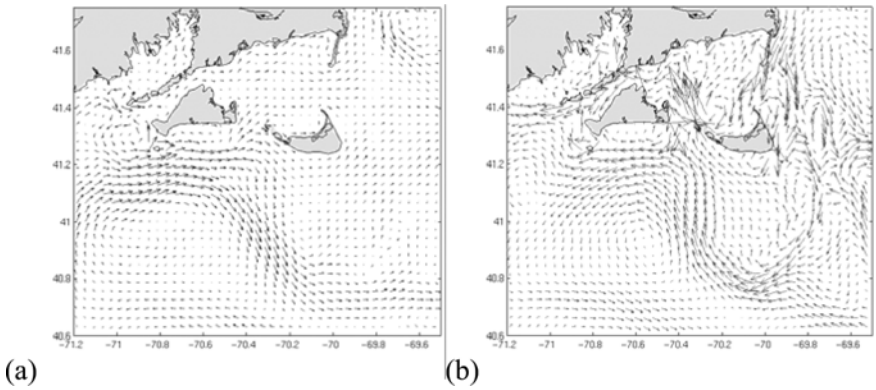

Figure 3. Depth-averaged mean currents for idealized simulations of the CBLAST region with (a) winds only forcing, and (b) tides only forcing.

The circulation throughout much of the model domain is influenced strongly by the tides. In a broad region of high eddy kinetic energy (Figure 2) on the flank of the Nantucket Shoals the tides vertically mix the water column maintaining cold sea surface temperatures throughout the summer. A second region of elevated tidal energy is the Muskeget Channel between the islands of Martha's Vineyard and Nantucket. On the ebb tide, water from Massachusetts Bay and the Gulf of Maine is swept westward into Nantucket Sound through Pollock Rip, while others waters that have warmed while within the Sound are ejected through Muskeget Channel and the Vineyard Sound passage between Martha's Vineyard and Cape Cod. The eddies exiting Muskeget Channel transport warm water toward MVCO, causing a substantial mean lateral transport of heat and producing the persistent 'bowl' of warm water near the south coast of the Vineyard that was observed by the glider section in 2003.

We have used the model to separate the competing influences of winds and tides by conducting two idealized simulations that omit, separately, the tide and wind forcing (Figure 3). In the absence of tides, Figure 3a shows that south of Martha's Vineyard winds drive upwelling favorable eastward circulation. This is opposed by a westward mean current that branches from a strong tidal rectified anti-cyclonic flow encircling the Nantucket Shoals (Figure 3b). The mean circulation with all forcing terms included (Figure 4) shows that the tidally driven mean flow prevails south of the MVCO region. The mean depth-averaged heat budget (Figure 5) shows net air-sea flux (Q_{net}) is greatest east of Nantucket Sound in the region of consistently cool SST, and is largely balanced by horizontal divergence associated with tidal mixing. Ocean temperature increase (storage) during July is largest south of The Islands, where surface heating is still warming the water column whereas in shallower water the temperature has reached its summertime equilibrium. Horizontal divergence is low south of Martha's Vineyard, indicating an approximate 1-D vertical heat balance.

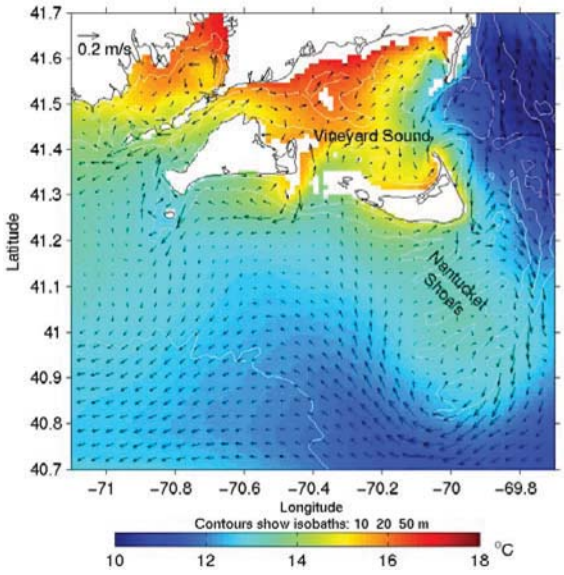

Figure 4. July 2002 mean temperature and currents at 5 m depth. A tongue of warm water issuing from Vineyard Sound through Muskeget Channel encroaches on MVCO but does not warm the region due to opposing mean flow.

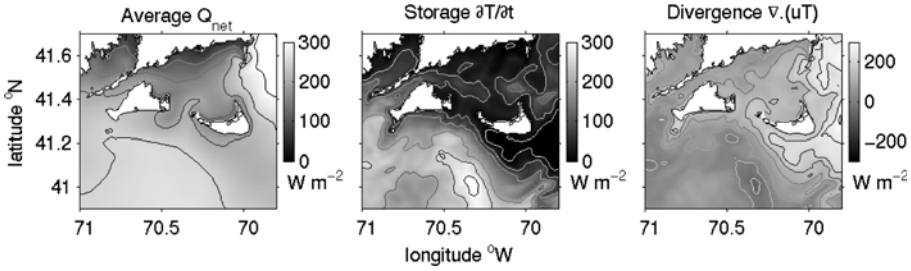

Figure 5. Modeled mean July 2002 depth-integrated heat budget terms.

The contributions of air-sea flux and advection, over time, to the heat budget for a box enclosing MVCO is shown in Figure 6 in terms of the equivalent heating in °C. In contrast to the area further south, lateral heat transport is significant near MVCO, with only half the air-sea flux going to warming the water column while half is removed by lateral divergence. Of this, time mean advection cools the MVCO box at, on average, 200 W/m^2, while the temporal eddy divergence <u'T'> warms the region at 50 W/m^2. Figure 6 also shows that strong episodic positive divergence (cooling) events briefly arrest the warming trend.

This analysis of the terms contributing to the heat budget shows that lateral advection is significant in the environs of the MVCO site. This will need to be quantified further in order to interpret the air-sea flux and vertical mixing observations made during 2003 at the ASIT tower and the various moorings.

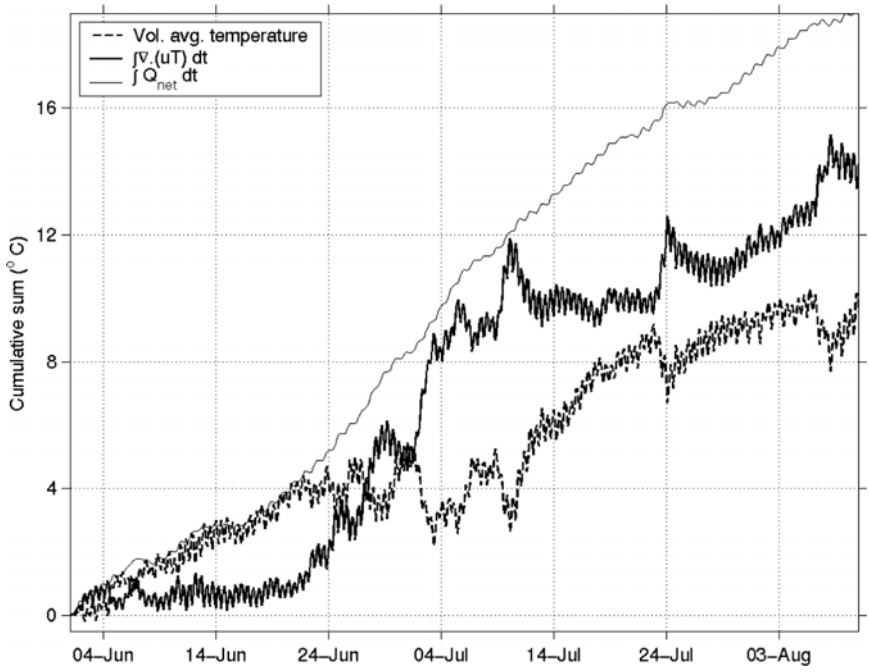

Figure 6. Time series of terms in the heat budget for a box enclosing MCVO. Light solid line: air-sea heating. Heavy solid line: depth-integrated advection. Dashed line: net heating.

Figure 7 shows time series of subsurface temperature at Mooring-F, the closest to MVCO of the five moorings deployed in 2002 (Hutto et al. 2003). For comparison, we show the corresponding time series modeled by ROMS for three different vertical turbulence closures: Mellor-Yamada, KPP, and the k-epsilon parameterization within the generic length scale scheme. In agreement with the experience of other recent turbulence closure comparisons (Wijesekera et al. 2003, Warner et al. 2005), the qualitative features of solution using these three schemes is similar. However, on inspection, there is a sense that the KPP scheme, at least at this location, performs rather better than the other two. This comparison will be pursued further in future analyses of the more extensive 2003 CBLAST data set. Results of a more comprehensive, quantitative validation of the model by comparison to the full set of 2002 mooring time series, including velocities, will be presented elsewhere (J. Wilkin, Modeling the summertime heat budget of southeast New England Shelf waters, in prep.), but it is clear from the simple comparisons presented above for both 2002 and 2003 that the model is able to capture the essential features of the circulation in the region.

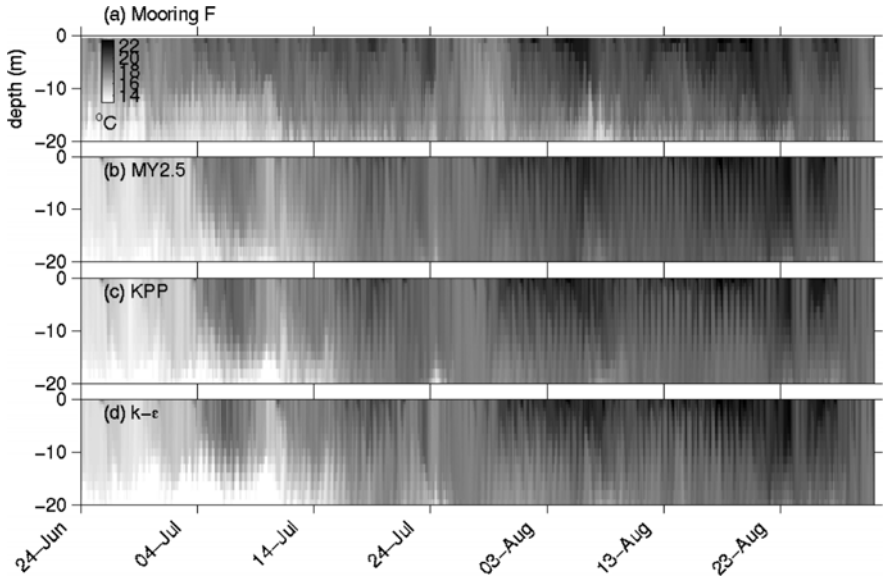

Figure 7. (a) Time series of the vertical temperature profile at mooring F deployed 10 km south of MVCO during the 2002 CBLAST Intensive observing period. ROMS hind-cast for different vertical turbulence closure options: (b) Mellor-Yamada (c), KPP, (d) k-epsilon.

A demonstrated capability to simulate the characteristics of the regional circulation has opened up other potential applications for the modeling system that were not envisioned at the outset of this project. Among these is a study of the processes controlling seasonal variability of phytoplankton biomass over the inner shelf. Using the cabled observatory, bio-optical sensors are being deployed to measure time series at MVCO of in situ optical properties using fluorometers, spectral absorption and scattering sensors, an experimental optical nitrate sensor, a submersible flow cytometer (Olson et al. 2003, Sosik et al. 2003) and an image-in-flow submersible microscope. These instruments provide time series information about phytoplankton abundance, community structure, and physiological growth rate of some phytoplankton groups (Sosik et al. 2003). These observations can be used to infer the time rate of change of phytoplankton biomass at MVCO, but closure of a 'phytoplankton budget' time series requires an estimate of the role of lateral transport of phytoplankton and nutrients past the MVCO site. From satellite chlorophyll imagery we know that there is considerable short length scale patchiness in phytoplankton distributions in the area. Much of this heterogeneity is intrinsic to phytoplankton distributions in general, but it is potentially amplified in the MVCO region by ocean circulation.

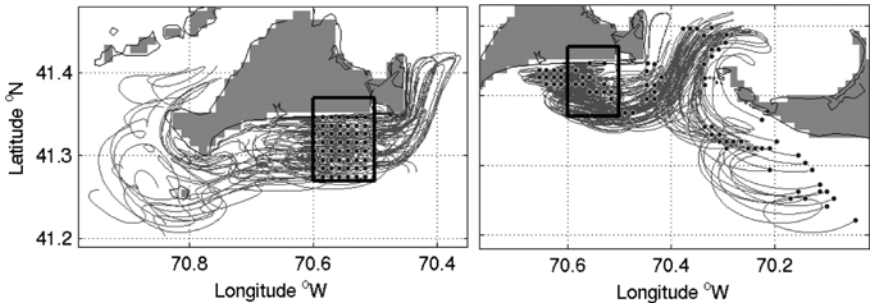

Figure 8. 48-hour simulated drifter tracks. Dots show the release locations. Left: Particles that are released in a box near MVCO. Right: Particles who end their 2-day journey in the MVCO box.

This is demonstrated in Figure 8 which shows two sets of 48-hour simulated drifter tracks. The top panel shows that within 2 days, particles that start near MVCO will be repeatedly swept past the observatory by the tide, but can ultimately be transported some 20 km away. In the lower panel, the dots show the starting locations, 2 days prior, of a set of particles that end up near MVCO. Their journey could commence as far as 50 km away if they are caught in the tidally-driven mean flow circulating along the southwest side of the Nantucket Shoals. Most enter and exit Nantucket Sound several times through the Muskeget Channel. These trajectories will carry particles through a variety of average water temperatures, stratifications, and light regimes. Regionally varying vertical mixing rates could affect the availability of nutrients sourced from deeper, cooler, waters circumnavigating the Nantucket Shoals.

3.3 ROMS modeling for LATTE

Modeling studies conducted to date for LATTE are limited to preliminary forecast simulations for the 2003 Pilot Program dye release. Differences in the model configuration, compared to CBLAST, are few. The major distinction is the need, obviously, to include the time dependent inflow of the Hudson River. Eight rivers are included in the model domain, of which the Hudson is by far the dominant. Hudson River daily flow data are available on the internet from USGS automated stream-flow gauges. We adopt a rule-of-thumb that the discharge at the first model grid point, near the southern end of Manhattan Island, is 1.3 times the sum of gauges on the Mohawk River and the Hudson River at Fort Edward. We have not yet found a source of prognostic hydrological data, so in order to provide a ROMS forecast we use observed daily averaged river flow up until the initialization

of each forecast. From that time onward climatological mean daily river flow is specified.

The second difference is the source of meteorological forcing. CBLAST used a high-resolution COAMPS forecast specially run for the intensive observing periods. Instead, we used the 72-hour ETA-12 forecast from the U.S. National Centers for Environmental Prediction (NCEP). This choice of a standard forecast product in widespread use was also partly motivated by the wish to evaluate these data for forcing operational coastal ocean models for any North American coastal observatory. An attractive feature of the ETA-12 forecast is the availability of hourly data, as soon as each day of the forecast completes, through an OPeNDAP (http://opendap.org) server at NCEP. This means we are able to start the ocean forecast promptly and are relieved of many data management tasks. The ETA-12 product includes forecast downward shortwave and long-wave data, so that we need not locate alternate sources of these as was necessary for the CBLAST 2002 hind-casts.

The remaining differences in configuration are minor. The initial conditions on April 16, 2004, were zero velocity, and temperature and salinity from an along-shelf average of all historical hydrographic station data for April within 100 km of Hudson Canyon. The focus on a short-term forecast for the duration of the dye experiment meant we neglected any detailed treatment of the open boundary climatology.

The first dye release of the LATTE Pilot Program was on May 3, 2003, just off Sandy Hook, New Jersey. Rather than proceed in a classic coastally trapped buoyant plume, the dye promptly headed due east toward Long Island and 36 hours later was dispersed on the coast. A second dye injection on May 4 in a weak plume proceeded down the coast but reversed and dispersed on May 5 with the onset of southerly winds. ROMS qualitatively captured all of these flow changes (Figure 9).

This example is presented to illustrate that a coastal ocean modeling system with useful forecast skill can be relatively easily constructed with ROMS to support a coastal observational program. While the science objectives of the LATTE program focus on fundamental studies of buoyant plume transport and biogeochemical processes, an aspect of the program is to demonstrate the capabilities of a coastal observatory comprised principally of re-locatable observational systems; namely, gliders, CODAR and ships. A readily configured model that does not require tuning or in situ observations for initialization or open boundary conditions, and for which forcing data (both river and atmosphere) are easily accessible by internet, contributes to this objective.

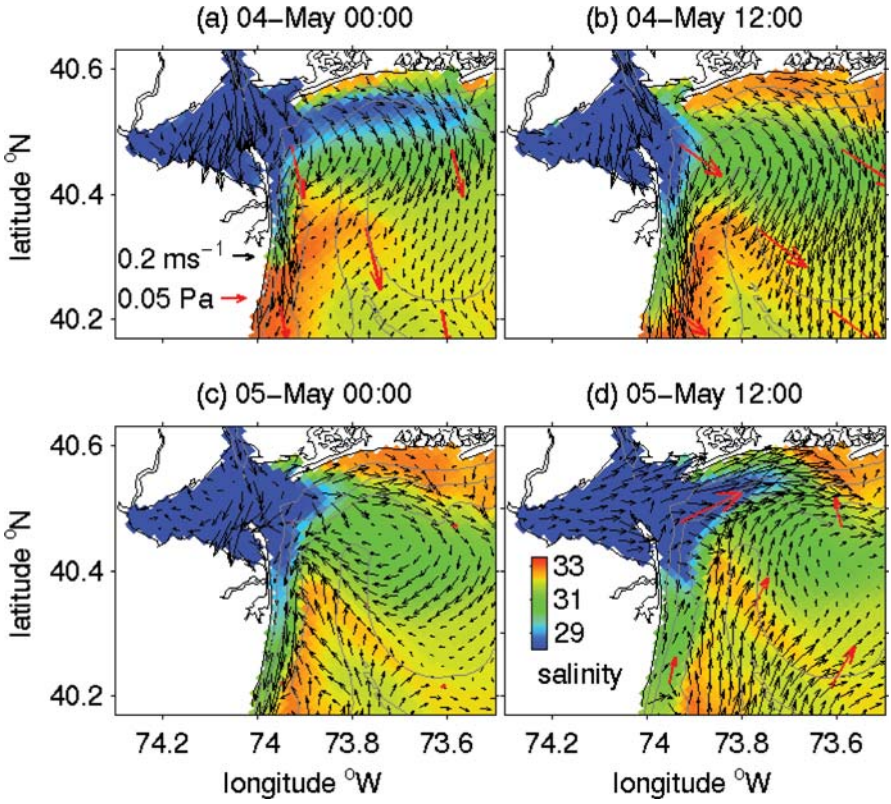

Figure 9. 48-hour ROMS forecast salinity during LATTE 2004 dye release experiments. Dates and times are GMT.

4. Summary

Circulation processes in the coastal ocean affect human activities on a broad range of space and time scales. On the shortest time scales, sea temperatures affect marine weather which in concert with tides and other rapidly varying currents significantly affects sediment transport, shipping and other maritime operations. On longer time scales, shelf circulation affects marine ecosystem dynamics and fisheries, and the cycling of carbon with implications for global climate regulation.

New observing technologies are rapidly expanding our ability to study these processes by simultaneously observing coastal ocean physics, geochemistry and ecology at resolutions suited to quantitative interdisciplinary analysis. These developments are matched by advances in ocean modeling through the adoption of more accurate numerical algorithms, re-structuring to use parallel supercomputers, and attention to realistic parameterizations of unresolved vertical subgridscale mixing and

boundary layer processes. The capabilities and skill of modern coastal ocean models, and the formulation of these models for hind-cast and forecast systems were presented by example.

As the CBLAST example shows, regional ocean models can now have the spatial resolution and numerical accuracy to capture the dominant features of the coastal ocean heat budget on short time scales in a region with strong tides, vertical stratification and significantly heterogeneous 3-dimensional mesoscale circulation in a complicated spatial domain of islands, channels and rough bathymetry.

The availability of high resolution surface meteorological forcing data, in real-time, opens opportunities for implementing ocean models as an integrated component of the burgeoning network of coastal observatories. Preliminary experience with the LATTE Pilot Program in 2004 showed that a straightforward configuration of ROMS for a limited area of the New York Bight had the ability to forecast qualitative features of the variability of Hudson River plume as is entered the shelf ocean.

In these examples an ocean model was used over a short time period as a forecast tool to complement the operation of a coastal observatory. The models clearly have the capability to become operational oceanographic forecast systems. More significantly, analysis of the CBLAST model was able to quantify the contributions of different terms in the local heat budget, and separate the roles of tides and winds in driving the regional mean circulation. Thus within the framework of a model with realistic bathymetry and forcing, the selective removal of individual forcing terms represent an application more akin to an idealized process-oriented study. By modeling Lagrangian pathways of simulated particles, or the dispersion of river runoff, in situ data can be placed in a broader geographic context, aiding the interpretation of data from coastal observing systems. In these ways, coastal ocean models contribute to exploring conceptual ideas and hypotheses regarding coastal ocean processes.

Acknowledgements

We thank M. Adani, F. Castruccio, M. Goes, R. Helber and D. Prosperi for helpful comments on the manuscript. This research is funded by U.S. Office of Naval Research grant N00014-04-1-0383 and National Science Foundation CoOP grant OCE-02-38957. The ROMS model development is funded by ONR grant N00014-1-0227. Computational resources were provided by the DOD High Performance Computing Modernization Program.

References

Adams, D.A., J.S. O'Connor, and S.B. Weisberg. 1998, Final Report: Sediment Quality of the NY/NJ Harbor System – An Investigation under the Regional Environmental Monitoring and Assessment Program (REMAP). EPA/902-R-98-001.

Anderson, D. M., 1995, Toxic red tides and harmful algal blooms: A practical challenge in coastal oceanography. *Reviews in Geophysics, Supplement*, 1189-1200.

Anderson, R.F., G.T. Rowe, P. Kemp, S. Trumbore, and P.E. Biscaye, 1994, Carbon budget for the mid slope depocenter of the Middle Atlantic Bight, *Deep-Sea Res. II*, 41, 669-703.

Bauer, J.E., E.R.M. Druffel, D.M. Wolgast, and S. Griffin, 2001, Cycling of dissolved and particulate organic radiocarbon in the northwest Atlantic continental margin, *Global Biogeochemical Cycles*, 15, 615-636.

Biscaye, P.E., C.N. Flagg, and P.G. Falkowski, 1994, The Shelf Edge Exchange Processes experiment, SEEP-II: an introduction to hypotheses, results and conclusions, *Deep-Sea Res. II*, 41, 231-252.

Bissett, W.P., K.L. Carder, J.J. Walsh, and D.A. Dieterle, 1999a, Carbon cycling in the upper waters of the Sargasso Sea: II. Numerical simulation of apparent and inherent optical properties. *Deep-Sea Research I*, 46, 271-317.

Bissett, W.P., J.J. Walsh, D.A. Dieterle, and K.L. Carder, 1999b, Carbon cycling in the upper waters of the Sargasso Sea: I. Numerical simulation of differential carbon and nitrogen fluxes. *Deep-Sea Research I*, 46, 205-269.

Boesch, D. F., J. C. Field, and D. Scavia (Eds.), 2000, The potential consequences of climate variability and change on coastal areas and Marine resources sector team, U.S. national assessment of the potential consequences of climate variability and change, U.S. Global Change Research Program. NOAA Coastal Ocean Program Decision Analysis Series No. #21. NOAA Coastal Ocean Program, Silver Spring, MD.

Canuto, V.M., A. Howard, Y. Cheng, and M.S. Dubovikov, 2001, Ocean turbulence I: one-point closure model. Momentum and heat vertical diffusivities. *J. Phys. Oceanogr*. 31, 1413–1426.

Chapman, D.C. and R.C. Beardsley, 1989, On the origin of shelf water in the Middle Atlantic Bight. *J. Phys. Oceanogr*., 19, 384-391.

Devol A.H., and J.P. Christensen, 1993, Benthic fluxes and nitrogen cycling in sediments of the eastern North Pacific, *J. Mar. Res.*, 51, 345-372.

Dinniman, M.S., J.M. Klinck and W.O. Smith, Jr., 2003, Cross shelf exchange in a model of the Ross Sea circulation and biogeochemistry, *Deep Sea Research II,* 50, 3103-3120.

Durski, S.M., S.M. Glenn, D.B. Haidvogel, 2004, Vertical mixing schemes in the coastal ocean: Comparison of the Level 2.5 Mellor-Yamada scheme with an enhanced version of the K-profile parameterization, *Journal of Geophysical Research*, 109, C01015, doi:10.1029/2002JC001702.

Fairall, C., E. Bradley, D. Rogers, J. Edson, and G. Young, 1996, Bulk parameterization of air-sea fluxes for TOGA COARE, *Journal of Geophysical Research*, 101, 3747-3764.

Flather, R. A., 1976, A tidal model of the northwest European continental shelf. *Mem. Soc. Roy. Sci. Liege, Ser. 6*, 10, 141–164.

Falkowski, P.G., C.N. Flagg, G.T. Rowe, S.L. Smith, T.E. Withledge, and C.D. Wirick, 1988, The fate of a spring phytoplankton bloom: export or oxidation?, *Cont. Shelf Res.*, 8, 457-484

Fong, D.A.and W.R. Geyer, 2001, Response of a river plume during an upwelling favorable wind event. *J. Geophys. Res*., 106, 1067-1084.

Galperin, B., L. H. Kantha, S. Hassid, and A. Rosati, 1988, A quasi-equilibrium turbulent energy model for geophysical flows. *J. Atmos. Sci.* 45: 55-62.

Garside, C., T.C. Malone, O.A. Roels and B.A. Sharfstein, 1976, An evaluation of sewage derived nutrients and their influence on the Hudson River estuary and the New York Bight. *Estuar. Coastal Mar. Sci.* 4, 281-289.

Greene, C. H. and others 2003, Trans-Atlantic responses of Calanus finmarchicus populations to basin-scale forcing associated with the North Atlantic Oscillation. *Prog. Oceanogr.* 58, 301-312.

Haidvogel, D.B., A. Shchepetkin, and H. Arango, 2004, The Regional Ocean Modeling System: New Time-stepping Algorithms to Reduce Mode-splitting Error and to Ensure Constancy Preservation.

Haidvogel, D.B., H.G. Arango, K. Hedstrom, A. Beckmann, P. Malanotte-Rizzoli, and A.F. Shchepetkin, 2000, Model Evaluation Experiments in the North Atlantic Basin: Simulations in Nonlinear Terrain-Following Coordinates, *Dynamics of Atmospheres and Oceans*, 32, 239-281.

Hallegraeff, G. M. 1993, A review of harmful algal blooms and their apparent global increase. *Phycologia* 32: 79-99.

Hallock, Z. R., and G. O. Marmorino, 2002, Observations of the response of a buoyant estuarine plume to upwelling favorable winds, *J. Geophys. Res*., 107(C7), 3066, doi:10.1029/2000JC000698.

Harris, C.K. and P.L. Wiberg. 2001, A two-dimensional, time-dependent model of suspended sediment transport and bed reworking for continental shelves. *Computers and Geosciences*, 27: 675-690.

Hodur, R. M., J. Pullen, J. Cummings, X. Hong, J.D. Doyle, P. J. Martin, and M.A. Rennick, 2002, The Coupled Ocean/Atmospheric Mesoscale Prediction System (COAMPS), *Oceanography*, 15, 88-98.

Hopkinson, C.S., J.J. Vallino, and A. Nolin, 2002, Decomposition of dissolved organic matter from the continental margin, *Deep-Sea Res. II*, 49, 4461-4478.

Hutto, L., J. Lord, P. Bouchard, R. Weller, and M. Pritchard, 2003, SecNav/CBLAST 2002 field experiment deployment/recovery cruises and data report, F/V Nobska, September 4 and 9, 2002, mooring data June 19-September 9, 2002. Woods Hole Oceanographic Institution Technical Report WHOI-2003-07; UOP Technical Report 2003-03, 114pp.

Kantha, L.H., and C.A. Clayson, 1994, An improved mixed layer model for geophysical applications. J. *Geophys. Res.*, 99, 25235–25266.

Large, W.G., J.C. McWilliams and S.C. Doney, 1994, Oceanic vertical mixing: A review and a model with a nonlocal k-profile boundary layer parameterization, *Reviews of Geophysics*, 32, 363-403.

Li, Y., and T. J. Smayda. 1998, Temporal variability of chlorophyll in Narragansett Bay, 1973-1990. ICES *J. Mar. Sci.* 55: 661-667.

Li. M. Z. and C. L. Amos, 2001, SEDTRANDS96: the upgraded and better calibrated sediment-transport model for continental shelves. *Computers & Geosciences*, 27, 619-645.

Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill, 1995, An estimate of global primary production in the ocean from satellite radiometer data, *J. Plank. Res.*, 17, 1245-1271.

Luettich, R. A., J.J. Westerink, and N.W. Scheffner, 1992, ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries, Tech. Report DRP-92-6, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Lutjeharms, J.R.E., P. Penven and C. Roy, 2003, Modelling the shear edge eddies of the southern Agulhas Current, *Continental Shelf Research*, 23, 1099-1115.

MacCready, P. and W.R. Geyer, 2001, Estuarine salt flux through an isohaline surface, *Journal of Geophysical Research*, 106, 11629-11637.

Malone, T., D.J. Conley, T.R. Fisher, P.M. Glibert, L.W. Harding, and K. Sellner. 1996, Scales of nutrient limited phytoplankton productivity in Chesapeake Bay. *Estuaries*, 19(2B): 371-385.

Marchesiello, P., J.C. McWilliams, and A. Shchepetkin, 2001, Open boundary conditions for long-term integration of regional ocean models, *Ocean Modelling*, 3, 1-20.

Marchesiello, P., J. C. McWilliams, and A. Shchepetkin, 2003, Equilibrium Structure and Dynamics of the California Current System, *Journal of Physical Oceanography*, 33, 753–783.

Mellor, G.L. and T. Yamada, 1982, Development of a Turbulence Closure Model for Geophysical Fluid Problems, *Reviews of Geophysics and Space Physics*, 20, 851-875.

Moore, A.M., H.G. Arango, A.J. Miller, B.D. Cornuelle, E. Di Lorenzo and D.J. Neilson, 2004, A Comprehensive Ocean Prediction and Analysis System Based on the Tangent Linear and Adjoint Components of a Regional Ocean Model, *Ocean Modelling*, 7, 227-258.

Mountain, D. G., 2003, Variability in the properties of Shelf Water in the Middle Atlantic Bight, 1977–1999, *J. Geophys. Res*., 108(C1), 3014, doi:10.1029/2001JC001044.

Munchow, A and R. Garvine, 1993, Buoyancy and wind forcing of a coastal current current. *J. Mar. Res*., 51 293-322.

NGDC (National Geophysical Data Center), 2004, Coastal Relief Model, Data Announcement 04-MGG-01, http://www.ngdc.noaa.gov/mgg/coastal/coastal.html.

Olson, R. J., A. A. Shalapyonok, and H. M. Sosik. 2003, An automated submersible flow cytometer for pico- and nanophytoplankton: FlowCytobot. *Deep-Sea Research I,* 50: 301-315.

O'Reilly, J. E., C. Evans-Zetlin, and D. A. Busch. 1987, Primary production, p. 220-233. In D. W. Bourne [ed.], Georges Bank. The MIT Press.

Ottersen, G., B. Planque, A. Belgrano, E. Post, P. C. Reid, and N. C. Stenseth. 2001, Ecological effects of the North Atlantic Oscillation. *Oceanologia* 128: 1-14.

Peliz, A., J. Dubert, D. B. Haidvogel, and B. Le Cann, 2003, Generation and unstable evolution of a density-driven eastern poleward current: The Iberian Poleward Current, *Journal of Geophysical Research*, 108, 3268, doi:10.1029/2002JC001443.

Shchepetkin, A.F. and J.C. McWilliams, 1998, Quasi-monotone advection schemes based on explicit locally adaptive dissipation, *Monthly Weather Review*, 126, 1541-1580.

Shchepetkin, A.F. and J.C. McWilliams, 2003, A Method for Computing Horizontal Pressure-Gradient Force in an Oceanic Model with a Non-Aligned Vertical Coordinate, *Journal of Geophysical Research*, 108, 3090, doi:10.1029/2001JC001047.

Shchepetkin, A.F. and J.C. McWilliams, 2005, The Regional Ocean Modeling System: A split-explicit, free-surface, topography following coordinates ocean model, *Ocean Modelling*, submitted.

Smayda, T. J. 1990, Novel and nuisance phytoplankton blooms in the sea: Evidence for a global epidemic. In E. Graneli, B. Sundstrom, L. Edler and D. M. Anderson (eds.), Toxic marine phytoplankton. Elsevier.

Smayda, T. J., 1998, Patterns of variability characterizing marine phytoplankton, with examples from Narragansett Bay. ICES *J. Mar. Sci,*. 55.

Seitzinger S.P., and Giblin A.E., 1996, Estimating denitrification in North Atlantic continental shelf sediments, *Biogeochemistry*, 35, 235-260.

Sosik, H. M., R. J. Olson, M. G. Neubert, A. Shalapyonok and A. Solow, 2003. Growth rates of coastal phytoplankton from time-series measurements with a submersible flow cytometer. *Limnology and Oceanography*, 48, 1756-1765.

Soulsby, R. L., 1995, Bed shear-stresses due to combined waves and currents, In: M. J. F. Stive et al., *Advances in Coastal Morphodynamics*, 4-20 – 4-23.

Umlauf, L. and H. Burchard, 2003, A generic length-scale equation for geophysical turbulence models, *Journal of Marine Research,* 61, 235-265.

Verity, P.G., J.E. Bauer, C.N. Flagg, D.J. DeMaster, and D.J. Repeta, 2002, The Ocean Margins Program: an interdisciplinary study of carbon sources, transformations, and sinks in a temperate continental margin system, *Deep-Sea Res. II*, 49, 4273-4295.

Walsh, J.J., 1994, Particle export at Cape Hatteras, *Deep-Sea Res. II,* 41, 603-628.

Walsh, J.J., D. A. Dieterle, and M.B. Meyers, 1988, A simulation analysis of the fate of phytoplankton within the Mid-Atlantic Bight, *Cont. Shelf Res*., 8, 757-787.

Walsh, J. J., T. E. Whitledge, J. E. O'Reilly, W. C. Phoel, and A. F. Draxler. 1987, Nitrogen cycling on Georges Bank and the New York Shelf: A comparison between well-mixed and seasonally stratified waters, p. 234-246. In R. H. Backus and D. W. Bourne [eds.], Georges Bank. The MIT Press.

Wang, S., Q. Wang, Z. Gao, J. Edson, R. Weller and C. Helmis, 2004, Evaluation of COAMPS Real Time Forecast for CBLAST-Low Summer Experiments 2002/2003, *Eos Trans. AGU Suppl.*

Warner, J.C., C.R. Sherwood, H.G. Arango and R. P. Signell, 2005, Performance of four turbulence closure methods implemented with a generic length scale method, *Ocean Modelling,* 8(1-2), 81-113, doi:10.1016/j.ocemod.2003.12.003

Weaver, A. T., J. Vialard, and D. L. T. Anderson, 2003, Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific Ocean. Part 1: Formulation, internal diagnostics and consistency checks. *Mon. Wea. Rev*., 131, 1360-1378.

Wijesekera, H.W., J.S. Allen, and P.A. Newberger, 2003, Modeling study of turbulent mixing over the continental shelf: Comparison of turbulent closure schemes. *J. Geophys. Res*. 108 (C3), 3103, doi:10.1029/2001JC001234.

Wilkin, J.L., H. G. Arango, D. B. Haidvogel, C. S. Lichtenwalner, S. M. Glenn, and K. S. Hedström, 2004, A Regional Ocean Modeling System for the Long-term Ecosystem Observatory, Journal of Geophysical Research, in press.

APPENDIX

List of attendees

ADANI Mario, *INGV, Bologna, Italy*
ALTALO Mary, *Science Applications International Corp, Potomac, USA*
ANDERSON David, *ECMWF, Reading, UK*
BAHUREL Pierre, *MERCATOR Océan, Ramonville St Agne, France*
BAQUERO-BERNAL Astrid, *Max Planck Institut fur Meteorology, Hamburg, Germany*
BARTON Andrew, *NOAA, Silver Spring, USA*
BEG PAKLAR Gordana, *Institute of Oceanography & Fisheries, Split, Croatia*
BELL Mike, MET Office, Exeter, UK
BERLINE Léo, *LEGI, Grenoble, France*
BLAYO Eric, *LMC-IMAG, Grenoble, France*
BONAZZI Alessandro, INGV, Bologna, Italy
BRANKART Jean-Michel, *LEGI, Grenoble, France*
BRASSEUR Pierre, *LEGI, Grenoble, France*
BROQUET Grégoire, *LEGI, Grenoble, France*
BUNGE Lucia, *LODYC, Paris, France*
BUONGIORNO NARDELLI Bruno, *CNR/ISAC, Roma, Italy*
BURUD Ingunn, *Norwegian Meteorological Institute, Oslo, Norway*
CAHILL Bronwyn, *University of Rhode Island, Narragansett, USA*
CALIL Paulo, *University of Hawaii, Honolulu, USA*
CALZADA Amilcar, *Institute of Meteorology, Havana, Cuba*
CASAGRANDE Gaëlle, *SHOM, Toulouse, France*
CASTRUCCIO Frédéric, *LEGI, Grenoble, France*
CHANGXIANG Yan, *Chinese Academy of Sciences, Beijing, China*
CHASSIGNET Eric, *University of Miami, RSMAS, Miami, USA*
CHHAK Kettyah, *University of Colorado, Boulder, USA*
CROSNIER Laurence, *MERCATOR Océan, Ramonville St Agne, France*
DE CAMARGO Ricardo, *University of Sao Paulo, Sao Paulo, Brazil*
DE MUIZON Marc, *SHOM, Toulouse, France*
DOUGLASS Elisabeth, *Scripps Institution of Oceanography, La Jolla, USA*
DRANSFELD Steffen, *CLS, Toulouse, France*
FERRY Nicolas, *MERCATOR Océan, Ramonville st Agne, France*
FUENZALIDA Rosalino, *Universidad Arturo Prat, Iquique, Chile*
FUJII Yosuke, *Japan Meteorological Agency, Tsukuba, Japan*
FUKUMORI Ichiro, *Jet Propulsion Laboratory, Pasadena, USA*
GARCES José, *Universidad de Concepcion, Concepcion, Chile*
GASPAR Philippe, *CLS, Toulouse, France*
GNANASEELAN Chellappan, *India Institute of Tropical Meteorology, Pune, India*
GOES Marlos, *University of Sao Paulo, Sao Paulo, Brazil*
GRIFFIES Stephen, *NOAA, Princeton, USA*
GUINEHUT Stéphanie, *CLS, Toulouse, France*
GUNDUZ Murat, *METU, Erdemli-Mersin, Turkey*
HACKETT Bruce, *Norwegian Meteorological Institute, Oslo, Norway*
HARDING John, *Naval Oceanographic Office, Stennis Space Center, USA*

HELBER Robert, *University of South Florida, St Petersburg, USA*
HERDIES Dirceu, *INPE, Cachoeira Paulista, Brazil*
JOHNSTON Shaun, *Scripps Institution of Oceanography, San Diego, USA*
KIM Young Ho, *Seoul National University, Seoul, Korea*
KIM Sung-Dae, *Korea Ocean Research & development Institute, Seoul, Korea*
KOURAEV Alexei, *State Oceanography Institute, St Petersburg, Russia*
LACHKAR Zouhair, *LSCE, Gif-sur-Yvette, France*
LANGLAIS Clothilde, *LSEET, Toulon, France*
LARGE William, *NCAR, Bouler, USA*
LEWIS Katy, *Plymouth Marine Laboratory, Plymouth, UK*
LOGOUTOV Oleg, *Harvard University, Cambridge, USA*
MAGALDI Marcello, *University of Miami, Miami, USA*
MANZANO SARABIA Marlenne, *CIBNOR, La Paz, Mexico*
MATTHEWS Dax, *University of Colorado, Broomfield, USA*
MÉNARD Yves, *CNES, Toulouse, France*
MESSIE Monique, *LEGOS, Toulouse, France*
MEZA Eustorgio, *IPN, Altamira, Mexico*
MICHEL Sylvain, *IFREMER, Plouzané, France*
MIN HongSik, *Korea Ocean Research & development Institute, Seoul, Korea*
MINSTER Jean-François, IFREMER, *Issy Les Moulineaux, France*
MURUKESH Nuncio, *National Institute of Oceanography, Panjim, India*
NERGER Lars, *AWI, Bremerhaven, Germany*
NODET Maëlle, *Université de Nice, Nice, France*
NUNEZ-RIBONI Ismael, *AWI, Bremerhaven, Germany*
OH Kyung-Hee, *Korea Ocean Research & Development Institute, Seoul, Korea*
OSCHLIES Andreas, *SOC, Southampton, UK*
PARK Jong Jin, *Seoul National University, Seoul, Korea*
PASCUAL Ananda, *CLS, Toulouse, France*
PEZZI Luciano, *INPE, Cachoeira Paulista, Brazil*
PINARDI Nadia, *SINCEM, Ravenna, Italy*
POULIQUEN Sylvie, *IFREMER, Plouzané, France*
PROSPERI Davide, *Weizmann Institute of Science, Rehovot, Israel*
RABIER Florence, *CNRM, Toulouse, France*
RAMIREZ Isabel, *CICESE, Ensenada, Mexico*
RIO Marie-Hélène, *CNR/ISAC, Roma, Italy*
ROBINSON Ian, *SOC, Southampton, UK*
ROZIER David, *LEGI, Grenoble, France*
RUSAKOV Alexander, *Russian Academy of Sciences, Moscow, Russia*
SANTORO Francesca, *University of Venice, Venice, Italy*
SARASON Christian, *University of Washington, Seattle, USA*
SCHILLER Andrea, *CSIRO, Hobart, Australia*
SCHROETER Jens,, *AWI, Bremerhaven, Germany*
SEMANE Noureddine, *Centre National de Recherches Météorologiques, Casablanca, Morocco*
SEND Uwe, *Scripps Institution of Oceanography, La Jolla, USA*
SIMON Ehouarn, *LMC-IMAG, Grenoble, France*
SKACHKO Sergey, *LEGI, Grenoble, France*
SMITH Neville, *BMRC, Melbourne, Australia*
STANEVA Joanna, *National Institute of Meteorology & Hydrology, Sofia, Bulgaria*
TJIPUTRA Jerry, *University of Wisconsin Madison, Madison, USA*
TORRES-NAVARRETE Carlos, *Universidad Autonoma de Baja California, Ensenada, Mexico*
TREGUIER Anne-Marie, *LPO, Plouzané, France*
VERRON Jacques, *LEGI, Grenoble, France*
WILKIN John, *Rutgers University, New Brunswick, US*

INDEX