Subduction in an Eddy-Resolving State Estimate of the Northeast Atlantic Ocean

by

Geoffrey Alexander Gebbie
B.S., University of California, Los Angeles, 1997

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
and the
WOODS HOLE OCEANOGRAPHIC INSTITUTION
June 2004

Author .....
Joint Program in Oceanography – Massachusetts Institute of Technology/
Woods Hole Oceanographic Institution
June 2004

Certified by .....
Carl Wunsch
Cecil and Ida Green Professor of Physical Oceanography
Massachusetts Institute of Technology
Thesis Supervisor

Certified by ....
Patrick Heimbach
Research Scientist – Massachusetts Institute of Technology
Thesis Supervisor

Accepted by .....
Joseph Pedlosky
Chair – Joint Committee for Physical Oceanography
Massachusetts Institute of Technology/Woods Hole Oceanographic Institution

Subduction in an Eddy-Resolving State Estimate of the Northeast Atlantic Ocean

by

Geoffrey Alexander Gebbie

Submitted to the Joint Program in Oceanography – Massachusetts Institute of Technology/ Woods Hole Oceanographic Institution
on June 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Relatively little is known about the role of eddies in controlling subduction in the eastern half of the subtropical gyre. Here, a new tool to study the eastern North Atlantic Ocean is created by combining a regional, eddy-resolving numerical model with observations to produce a state estimate of the ocean circulation. The estimate is a synthesis of a variety of *in-situ* observations from the Subduction Experiment, TOPEX/POSEIDON altimetry, and the MIT General Circulation Model. A novel aspect of this work is the search for an initial eddy field and eddy-scale open boundary conditions by the use of an adjoint model. The adjoint model for this region of the ocean is stable and yields useful information despite concerns about the chaotic nature of eddy-resolving models. The method is successful because the dynamics are only weakly nonlinear in the eastern region of the subtropical gyre. Therefore, no fundamental obstacle exists to constraining the model to both the large scale circulation and the eddy scale in this region of the ocean. Individual eddy trajectories can also be determined.

The state estimate is consistent with observations, self-consistent with the equations of motion, and it explicitly resolves eddy-scale motions with a $1/6°$ grid. Therefore, subduction rates, volume budgets, and buoyancy budgets are readily diagnosed in a physically interpretable context. Estimates of eddy subduction for the eastern subtropical gyre of the North Atlantic are larger than previously calculated from parameterizations in coarse-resolution models. Eddies contribute up to 40 m/yr of subduction locally. Furthermore, eddy subduction rates have typical magnitudes of 15% of the total subduction rate. To evaluate the net effect of eddies on an individual density class, volume budgets are diagnosed. Eddies contribute as much as 1 Sv to diapycnal flux, and hence subduction, in the density range $25.5 < \sigma < 26.5$. Eddies have a integrated impact which is sizable relative to the 2.5 Sv of diapycnal flux by the mean circulation. A combination of Eulerian and isopycnal maps suggest that the North Equatorial Current and the Azores Current are the geographical centers of eddy subduction. The findings of this thesis imply that the inability to resolve or accurately parameterize eddy subduction

in climate models would lead to an accumulation of error in the structure of the main thermocline, even in the eastern subtropical gyre, which is a region of comparatively weak eddy motions.

Thesis Supervisor: Carl Wunsch
Title: Cecil and Ida Green Professor of Physical Oceanography
Massachusetts Institute of Technology

Thesis Supervisor: Patrick Heinbach
Title: Research Scientist – Massachusetts Institute of Technology

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. Carl Wunsch, who had confidence in me even when I did not. His overarching sense of perspective often guided me through the inevitable intricacies and technical details. The substance of this work probably stems from earlier work of Carl and his collaborators, Detlef Stammer and Jochem Marotzke. I'm fortunate that they entrusted me with this line of research.

Patrick Heimbach has served as my co-advisor, and in many ways, was the most indispensable member of my thesis committee. He has been the man behind the scenes, unassuming as always. Its been a pleasure to be a friend of a true Adjoint Wizard.

My thesis committee was invaluable in providing feedback. Jim Price was especially helpful with Chapter 1, and John Marshall had many opinions on Chapter 4. Breck Owens was a questioning voice, although one with much interest in the subject. Last but not least, the Chair of the Defense, Paola Rizzoli, worked with incredible speed throughout the review process, and the defense itself!

Many outside collaborators should be acknowledged here. Dan Lea, Tony Lee, Tom Haine, Ichiro Fukumori, Raffaele Ferrari, Michael Spall, Bob Weller, and Amit Tandon have all made useful contributions to this thesis. Chris Hill has also been a major player in adjoint studies.

This thesis has received support through grants, fellowships, and computer allocations. A NASA Earth System Science Fellowship has been my primary source of funding. In addition, computer time at the University of Texas has been provided by a NPACI PRAC grant, "State Estimates of the Time-Evolving Three-dimensional Ocean Circulation with Eddy Resolution." Grant #6857100 through CalTech and the Jet Propulsion Laboratory, as well as Grant #6892952 through NASA-Goddard Flight Center for the synthesis of the World Ocean Circulation Experiment, have supported me, too.

This thesis has relied upon, and is a contribution of, the ECCO (Estimating the Climate and Circulation of the Ocean) Consortium. Detlef Stammer has also provided large amounts of computer time at the San Diego Supercomputer Center (NPACI) and the Naval Oceanographic Office MSRC at Stennis Space Center. Annette May, the

program administrator at NAVO, was instrumental in garnering extra computer hours.

The Academic Programs Office at WHOI, and the 9th floor at MIT have been invaluable. I would especially like to thank Julia Westwater, Marsha Gomes, Ronni Schwartz, and Carol Sprague. They are a strong resource of the program, and a good counterbalance to the faculty! On the 14th floor, Lisa Mcfarren has always been there when time is running out and deadlines are approaching.

In grad school, its necessary to have solid support group. My entering class, Kerim, Zan, Alison, Heather, Donnan, Bill, Xingwen, and Baylor, were just that. Greg and Peter seemed to be there from start to finish, too. Payal went out of her way for me many times. Thanks to Yu-Han for the thesis-writing camaraderie. And Véronique Bugnion was my original role model with adjoint models (and skiing/snowboarding). Many other faces came and went; I once mentioned that I had 14 different office mates at MIT. That's not quite true, but I would like to thank Chris, Jubao, Natalia, Richard, Rong, Alison, Taka, Pere, Ariane, Dave, and Asher for keeping the spirit and the fish of 54-1419 alive.

To my parents, thanks for making me an independent researcher at an early age, when you told me that you wouldn't and couldn't help me with my 7th grade math class. And to Tricia, who showed me that graduating in 2004 isn't so bad.

Contents

Chapter 1

Introduction

1.1 Subduction and the general circulation

Throughout the subtropical regions of the world ocean, the atmosphere has a window to influence the structure of the main thermocline and upper ocean; the window is opened in the process of subduction. Subduction is the transfer of fluid from the mixed layer into the interior thermocline by combined vertical and horizontal flow, or by thermodynamic forcing. The process is typically quantified by the volume flux of subducted fluid per unit horizontal area, known as an entrainment velocity. In general, subduction carries surface properties of the ocean downward and out of direct atmospheric contact. Therefore, the water-mass characteristics of the mid-latitude upper ocean directly reflect the process of subduction. The mid-latitude upper ocean has an enormous heat capacity and plays an obvious role in climate studies (Broecker 1991; Hartmann 1994). In addition, subduction primarily determines the pathways of influence and information flow. For example, tropical-subtropical exchanges primarily take place through subducted water and through pathways made available by subduction (McCreary and Lu 1994; Deser and Blackmon 1995; Malanotte-Rizzoli et al. 2000; Lazar et al. 2002). The sensitivity of the composition of the subtropical ocean to atmospheric forcing raises concern because of global climate change; however, the historical record of subduction rates is exceedingly sparse. The impact of the atmosphere on a large class of water masses is not quantifiable

without understanding the process of subduction.

Subduction influences more than the water-mass properties of the upper ocean; due to the strong coupling of the density field and the velocity field by geostrophy, subduction helps set the inherent timescales of oceanic motions. The depth and slope of the main thermocline reflect how fast the ocean is moving. Thermocline depth fundamentally determines baroclinic wave properties (Pedlosky 1987), and the thermocline slope is related to velocity through thermal wind balance (Pond and Pickard 1983). Studies have hypothesized that the timescales of subduction may also set the frequency of climate oscillations, such as the North Atlantic Oscillation (Czaja and Frankignoul 2002) or the El Niño-Southern Oscillation (ENSO) (Gu and Philander 1997). As shown here, subduction is an important process that influences the "clock" of both the internal ocean circulation and atmosphere-ocean coupling.

1.1.1 Review of subduction

The original theories describing subduction were based upon gross large-scale observations of the ocean (also see Price (2001) for a detailed review of subduction theory). From North Atlantic atlases of temperature and salinity (Wüst 1935; Defant 1936), Montgomery (1938) suggested that Ekman convergence in the near-surface ocean drove fluid into the deeper ocean. A volume budget calculation in a "stream-tube" confirmed that the rate of fluid transfer has the same order of magnitude as the Ekman pumping rate. Montgomery's idea of subduction by vertical velocity at the base of the mixed layer is the precursor to today's concept of subduction. In fact, almost all of the later work in ocean theory is based upon the idea that the Ekman layer can force the deeper geostrophic circulation. The region of negative wind stress curl, and hence Ekman convergence, generally defines the "subtropical gyre" (Pedlosky 1996). Iselin (1939) showed the striking similarity between a meridional section of late-winter mixed-layer properties and a vertical profile of temperature and salinity in the North Atlantic. He suggested that surface layer properties slide down density surfaces to set the properties of the interior ocean. As an aside, Iselin did not call upon mass lateral movement to explain

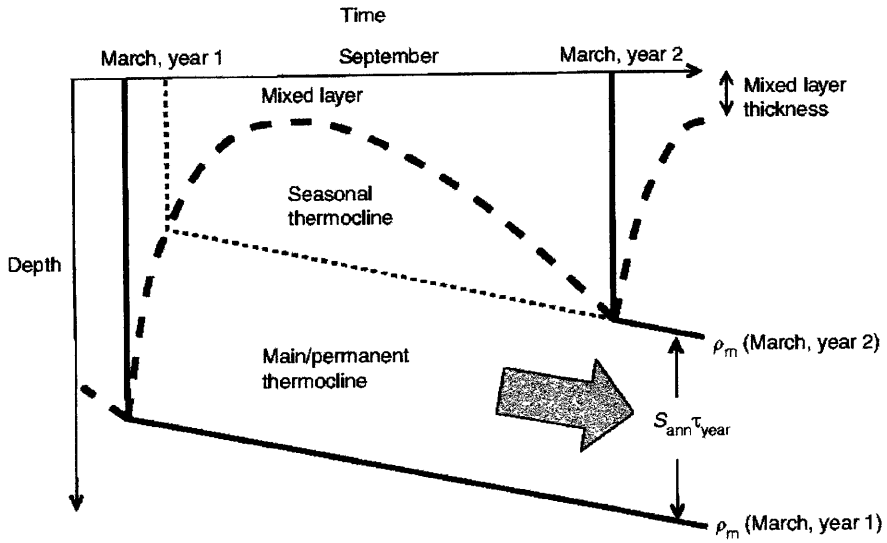

Figure 1-1: A depth-time schematic of Stommel's mixed-layer demon. An upper ocean water column with seasonally-varying mixed-layer depth (*thick, dashed line*) and downward Ekman pumping leads to net transfer of fluid from the seasonal to main thermocline. Effective subduction only occurs in a short time period because subducted water is re-entrained into the mixed layer. The last permanently-subducted water of year 1 (*thin, dashed line*) leaves the mixed layer in March. From Williams et al. (1995).

the connection between surface and depth, but instead remarked that "lateral turbulence" could be responsible. Forty years passed before Stommel (1979) explained why late-winter surface properties reflect those at depth. He showed that the typical seasonal excursion of the mixed layer is larger than the vertical displacement of water, and hence, only late-winter subducted water avoids entrainment back into the mixed layer (see Fig. 1-1). Later, a primitive equation model showed that the so-called "mixed-layer demon" did indeed allow only a short window for subduction to affect the main thermocline (Williams et al. 1995). All of these previous studies showed the great extent to which the ocean's large-scale hydrographic structure is explained by subduction.

The relationship between the mixed layer, the main thermocline, and wind forcing was made explicit in the steady thermocline model of Luyten et al. (1983). Earlier

mathematical models, such as the work of Robinson and Stommel (1959) and Welander (1961), sought similarity solutions to a steady thermocline externally driven by an Ekman layer, but did not provide much physical insight. More than twenty years later, the theory of the ventilated thermocline (Luyten et al. 1983) introduced a layered model that explained the "bowl-like" shape of the subtropical main thermocline. The ventilated thermocline circulation was steady, inviscid and geostrophically balanced below the mixed layer, and driven by Ekman pumping at the surface. The division of the ocean into separate vertical layers, particularly the separation of the mixed layer and underlying stratum, advanced our physical understanding. The direct influence of the atmosphere on oceanic properties in the surface layer was termed "ventilation", which conjures the image of exposure to air. Below the surface layer, the "subducted" layers conserved potential vorticity and were adiabatic. In the limit of a many-layered or continuous model, subduction and ventilation are identical (Cushman-Roisin 1987; Huang 1990). The ventilated thermocline theory predicted ocean domains with distinct dynamics due to differing pathways of subducted water. As foreseen by Montgomery's stream-tube model, a large portion of the gyre subducts water southward and downward. Nevertheless, subducted water does not pass through the unmoving eastern boundary region, termed the shadow zone. Conversely, the western boundary has an unventilated region with homogenized potential vorticity (Rhines and Young 1982). These theoretical studies used potential vorticity as a framework to view the ocean circulation. The theory of Luyten et al. (1983) provides the basic concepts that many later studies of subduction rely upon.

One key omission in ventilated thermocline theory was a realistic mixed layer with variable thickness and thermodynamics. When the mixed layer has spatially-varying thickness, horizontal velocity causes subduction. The lateral flow of fluid across a sloping mixed-layer base is called *lateral induction* (Huang 1990). Near strong currents like the Gulf Stream, lateral induction typically produces subduction rates of 100 m/yr or more, even though the average Ekman pumping rate is only 30 m/yr (Woods 1985; Marshall and Nurser 1991). Another shortcoming of the ventilated thermocline model was the lack

of mixed-layer thermodynamics. Subduction undoubtedly affects the water masses of the interior ocean from a kinematic point of view, but mixed-layer thermodynamic forcing is the primary way that new water masses are formed (Walin 1982; Speer and Tziperman 1992; Garrett et al. 1995). The kinematics of subduction and the thermodynamics of the mixed layer were reconciled in the work of Marshall et al. (1999), where accurate diagnosis of mixing and entrainment in a general circulation model (GCM) showed that the two processes are intimately related. In summary, the addition of a more realistic mixed layer is necessary to quantify accurately the many processes which affect subduction.

According to a recent textbook (Wunsch 1996), "the central distinguishing feature of oceanography as a branch of fluid dynamics is the extreme difficulty of obtaining observations." This is still true. However, with the advent of satellite measurements and the continuation of intensive field programs, oceanographers now have greater capability to observe the ocean than ever before. The unprecedented supply of new data shows clearly that the ocean moves on all space and time scales and must be studied as such. With subduction, recent work has begun to consider the net impact of small-scale motions. The role of "eddies", small-scale motions with a characteristic lengthscale of $100 - 400km$, is especially murky. Eddies act to diffuse tracers as well as providing an effective advection by a "bolus velocity". Marshall (1997) showed that the bolus velocity (Gent et al. 1995) is responsible for eddy-induced subduction (Figure 1-2). Hence, regions with large bolus velocities have large subduction rates due to eddies. The numerical model study of Hazeleger and Drijfhout (2000) showed intense eddy subduction near the Gulf Stream, a region with large bolus velocities. Furthermore, baroclinic instability associated with oceanic fronts provided a mechanism for subduction (Spall 1995; Follows and Marshall 1996). In the face of high-resolution observations, large-scale, steady theories may be irrelevant. Will these theories stand up to quantitative analysis? The inherently turbulent character of the observed ocean forces the revision of recent theories of subduction.

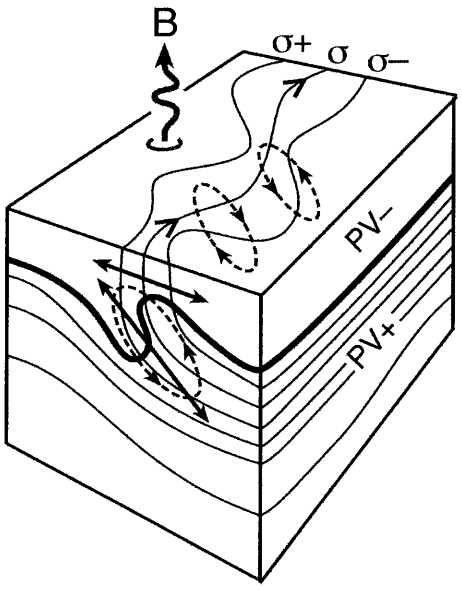

Figure 1-2: Schematic of eddy-driven subduction. Time-variable motions near a density front, marked by tightly packed isopycnals *thin lines*, can transport fluid below the mixed-layer base, marked by the boundary between high and low potential vorticity (*bold line*). Following Walin (1982). Figure from J. Marshall (pers. comm.).

1.1.2 The Subduction Experiment

Goals of the field study

The overall goal of the Subduction Experiment was to understand the sequence of events leading to subduction, and the subsequent movement and transformation of subducted water. Subduction is primarily due to large-scale forcing by the atmosphere, especially by the wind. The accurate measurement of the large-scale atmospheric forcing was therefore a necessary goal of the experiment. From a purely kinematic point of view, the Subduction Experiment also sought the large scale mean surface flow and its convergence, because this forces water downward. Connections between the kinematic and thermodynamic viewpoints were specifically sought by the experiment; in other words, the basic dynamic balances in the ocean were unknown. Finally, the degree of nonlocality in the process of subduction was to be determined as well. Furthermore, the Subduction Experiment was part of the much larger World Ocean Circulation Experiment (WOCE), and the goals stated here are but a subset of the overall WOCE goals (Siedler et al. 2001).

To achieve these goals, the eastern subtropical North Atlantic Ocean was chosen as the site of the Subduction Experiment. The region has a large-scale pattern of negative wind stress curl (Stommel 1979; Moyer and Weller 1995) and observations have shown that subduction occurs there (Jenkins 1987). Also, the eddy kinetic energy is low in relation to western boundary currents or the tropics (Stammer 1997). The experiment comprised three separate field deployments between June, 1991, and June, 1993. An array of five moorings observed both oceanic and meteorological fields (Brink et al. 1995). They were spaced in a "X" pattern over with typical separation of 1000 kilometers in order to quantify largescale changes in atmospheric variables. Mooring locations are marked in Figure 1-3. The meteorological component of moorings measured short and longwave radiation, humidity, wind speed, temperature, and rainfall. The large scale Bermuda-Azores high dominated the atmospheric variability in the region (Moyer and Weller 1995). Below the surface, the moorings measured subsurface velocities (typically

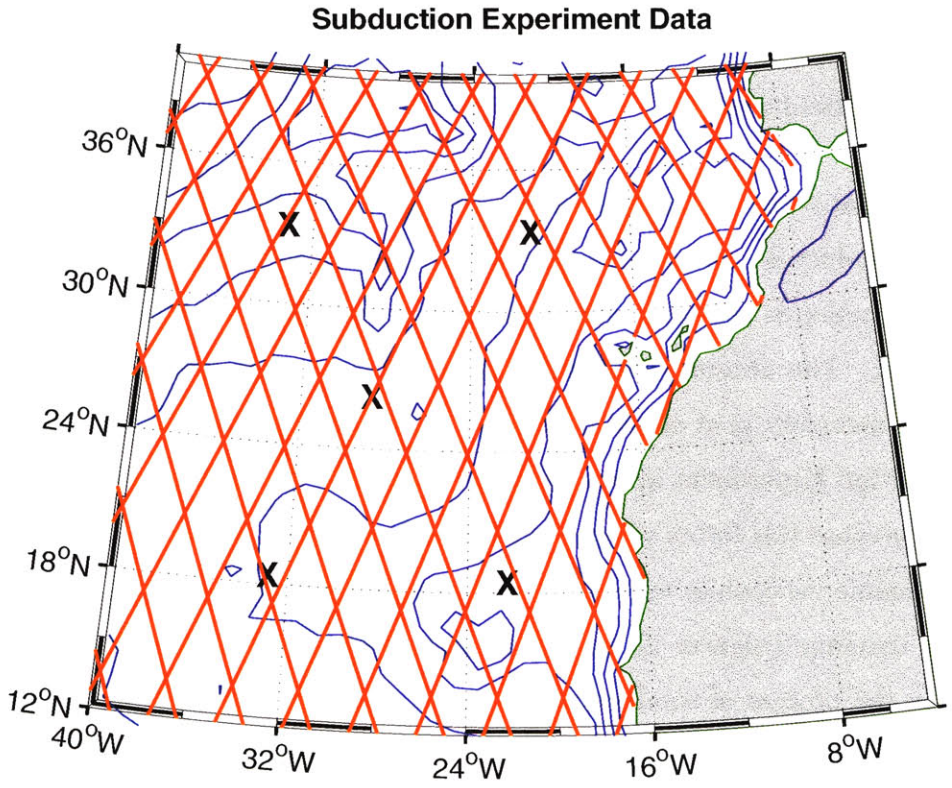

Figure 1-3: The Subduction Experiment was an intensive field experiment designed to study the subduction of fluid from the mixed layer into the main thermocline. This study uses 5 moorings (marked by "X") with temperature, velocity, and meteorological observations. TOPEX/Poseidon altimetry (marked by bold, solid tracks) is also used here. The thin solid lines are depth contours with an interval of 1000 m.

with Vector Measuring Current Meters and with Acoustic Doppler Current Profilers) and temperatures to a depth of 1500 meters. All of the moorings were deployed south and west of the Azores Current in order to remain in a low eddy kinetic energy region of the ocean, presumably because the original experiment planners believed that high values of eddy energy would obscure their findings. This thesis takes the viewpoint that the eddy energy is an intrinsic part of the process, and that it can not be ignored without careful analysis. As can be seen above, the deployment of the five moorings had specific science objectives in mind, and this study reviews whether the specific objectives were met.

State Estimate Obs.	Withheld Obs.	Previously used Obs.
Mooring Temperature	WOCE hydrography	Bobber, SOFAR, and ALACE floats
Mooring Velocity	Mooring heat fluxes	Sea-Soar profiles
TOPEX/POSEIDON altimetry		NATRE, Tritium-Helium

Table 1.1: Summary of the observations. The state estimate observations were used explicitly to constrain the model. The center column indicates observations that were later used as an independent check on the state estimate. Previous studies have used the assortment of observations in the third column, but they were not directly used here.

In addition to the mooring data, other quantities were measured. The moorings were refurbished every 8 months, so there were many hydrographic transects during transit. Over 800 standard CTD stations and thirteen surveys with a SeaSoar towed profiler were taken (Pallant et al. 1995; Joyce et al. 1998). The near-surface flow field was measured with the drifters of P. Niiler and J. Paduan, and deeper measurements by twenty-eight SOFAR (Sound Fixing and Ranging) and bobber floats characterized the flow in the region (Sundermeyer and Price 1998). Bobber floats rested at a preprogrammed density level, and profiled in a pre-specified density band every other day. Eleven ALACE (Autonomous LAgrangian Circulation Explorer) floats of R. Davis were also in the region. Approximately eighty other floats of A. Bower, P. Richardson, and W. Zenk specifically studied the Mediterranean Outflow. Dye and dye-like studies were also carried out simultaneously. The North Atlantic Tracer Release Experiment (NATRE) occurred very near the central Subduction Experiment mooring during the same time period (Ledwell et al. 1993). Tritium-Helium observations of W. Jenkins also characterize rates of subduction and dispersion of water masses. Last but not least, the TOPEX/POSEIDON sea surface height observations began in October, 1992, and overlap half of the Subduction Experiment. As a whole, the Subduction Experiment was an intensive field study with a wide variety of instrumentation.

Results from the Subduction Experiment

Close comparison of the Subduction Experiment and ocean theory gave rise to startling differences. Theories of ventilation and subduction have focused on the large-scale and steady ocean (Luyten et al. 1983; Woods 1985). In contrast, mesoscale eddy energy was a ubiquitous feature of all observations and it was not obvious that it can be ignored. Joyce et al. (1998) showed that SeaSoar profiles of subducted water have mesoscale variability that is not damped by the process of subduction. Mixing after initial formation was crucial to the evolving water mass properties of subducted fluid. From these observations, Joyce et al. (1998) made objective maps of the mesoscale eddy field on a scale of 100 kilometers. Sundermeyer et al. (1998) used the ALACE floats of the Subduction Experiment to calculate particle dispersion rates and strain rates of mesoscale eddy field. The results of the Tracer Release Experiment (Ledwell et al. 1993) confirmed the similar diffusive effect of the small scale ocean circulation. Other differences to ocean theory came from geographic complications. Helium-tritium observations (Robbins et al. 2000) showed that the Azores Current acted as a barrier to the net mass flux of subduction (Figure 1-4). According to ventilated thermocline theory, this would create a pool of homogenized potential vorticity (PV) behind the barrier (Rhines and Young 1982), but such a PV distribution is not observed. Robbins et al. (2000) appealed to the diffusive nature of subduction in this case, which is reminiscent of the net effect of the mesoscale eddy field. Weller (2003) and Weller et al. (2004) further remark that "mean advection [alone] cannot explain how water is carried into the mixed layer ... and eddy transport processes should be considered." Perhaps these differences to theory should not be so surprising; the observational view of the ocean as fundamentally turbulent sometimes opposes theoretical tradition.

Moyer and Weller (1995) focused on the impact of the moored meteorological measurements. They showed the inability of climatological datasets of atmospheric forcing to adequately represent the forcing at the mooring sites. Large errors in heat flux and oversmoothing were deficiencies in the climatologies. Systematic biases reach 50% of the total signal. Moyer and Weller (1995) warned that mean subduction rates or mean

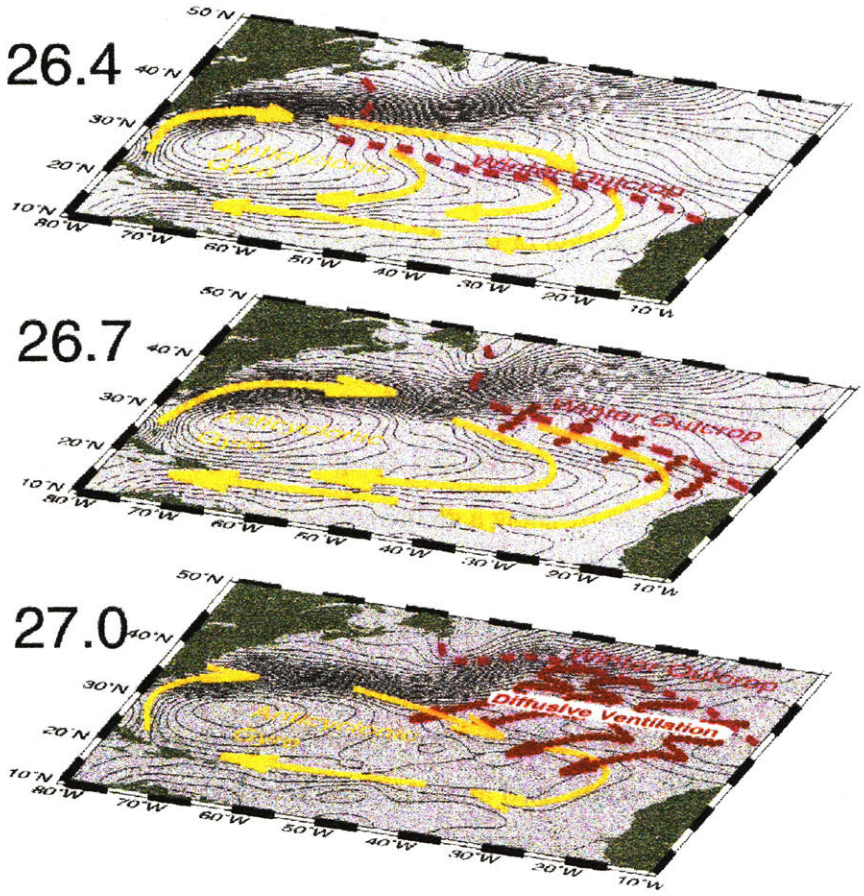

Figure 1-4: Schematic of the pathways of ventilation on three isopycnal surfaces. Each surface is defined by its σ_θ value. Montgomery streamfunction (*thin, black lines*), the mean circulation (*yellow arrows*), and the winter outcrop line (*magenta dashed line*) are plotted for each surface. Different mechanisms must explain the variety of observed subduction paths. From Robbins et al. (1998).

Ekman pumping rates calculated from these climatological datasets (i.e., Woods 1985; Marshall et al. 1993) may not be representative. For example, Marshall et al. (1993) calculate a mean subduction rate of $80 - 100\ m/yr$ in the eastern subtropical gyre, although measurements for the Subduction Experiment time period were much lower. In summary, ocean modelers either need improved forcing fields, or they should consider the model output to be very uncertain.

A hierarchy of models has been used to simulate the dynamics of the Subduction Experiment region. This was (and remains) a necessary avenue of research because the spatial and temporal resolution of the observations was not high enough to diagnose accurate dynamical balances. The hierarchy of models ranged between the "pipe" model of Robbins et al. (2000), the two-layer quasi-geostrophic model of Sundermeyer and Price (1998), and the primitive equation models of Spall (1990) and Spall et al. (2000). In particular, Spall et al. (2000) attempted to quantify subduction rates, dynamical balances, and the role of eddies by using a global coarse resolution Climate System Model (CSM) of the National Center for Atmospheric Research (NCAR). Typical subduction rates were over $100\ m/yr$ in the wall of the North Atlantic Current and $40\ m/yr$ away from that region, with a $5 - 10\%$ contribution from eddy motions (Figure 1-5). This was the first attempt to make a region-wide quantitative analysis of the Subduction Experiment dynamics. The study used an eddy-parameterization scheme to describe the role of eddies in subduction. A qualitative comparison of the model with observations was also made. This thesis aims to extend and improve the line of research started by Spall et al. (2000).

1.1.3 Unresolved questions

The original goals of the WOCE experiment have not been fully achieved by the Subduction Experiment. According to the WOCE AIMS document (1997), a major goal was the quantification of transport estimates, water-mass formation rates, and a description of variability. Although air-sea fluxes are known very well at the moorings, the uncertainty of climatologies away from those sites makes the atmospheric forcing very poorly

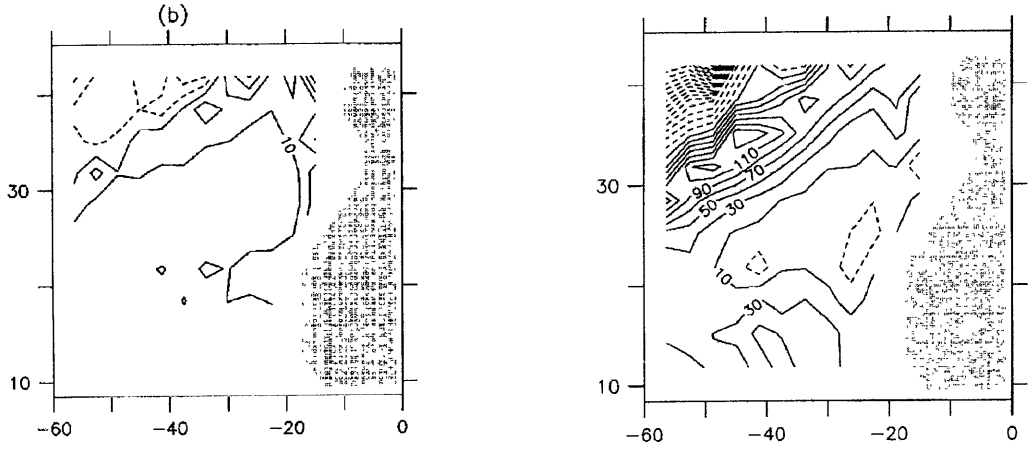

Figure 1-5: *Left panel:* Thermodynamic estimate of eddy-driven subduction [m/yr] by the diapycnal bolus transport of heat in the mixed layer. *Right panel:* Estimate of subduction by the mean flow [m/yr]. Both calculations were from a coarse-resolution numerical model with an eddy-parameterization scheme. From Spall et al. (2000).

known in general. Even meteorological re-analyses are highly uncertain over the open ocean. An improved estimate of the true atmospheric forcing everywhere is a prerequisite to progress. Mean subduction rates and subduction rates from coarse resolution models have been calculated, but these values do not have strong observational support of the Subduction Experiment. A description of variability does exist, although its relation to the large-scale circulation is unknown. The need for basic quantification of the Subduction Experiment parameters and variables still exists.

Another major goal of WOCE is the understanding of dynamical balances of the ocean. Although the impact of the small-scale variability of the ocean has been noted in all observations, the role of these motions in dynamical balances has only been quantified recently. No consensus exists regarding the impact of eddies on the net product of subduction, for example. With respect to dynamical balances and water mass transformation, are eddies relevant in the eastern subtropical gyre?

Although previous research in the Subduction Experiment has achieved much with individual data forms and models, only the recent, independent study of Weller et al. (2004) has attempted to compare and collate the large collection of the available

information. A more trustworthy and self-consistent picture of the ocean physics arises from an integration of the many forms of observations and a model. In contrast to Weller et al. (2004), this thesis aims to be a quantitative *synthesis* and an extension of the previous research through rigorous mathematical methods. The quantification of the ocean dynamics over the entire domain of the Subduction Experiment is the overarching goal. This thesis has already introduced the observations available, but to properly carry out a synthesis, a numerical model is also essential.

1.2 Novel aspects of the thesis

1.2.1 Approach: synthesis of observations

To create a model-observation synthesis, a realistic model of the Subduction Experiment region is necessary. As carried out in this thesis, this endeavor has side benefits, although many are technical. The formulation of open boundary conditions is crucial for any regional ocean model. No standard method for open boundaries has yet been adopted by oceanographers. Ocean models also have many systematic errors such as improper mixed-layer parameterizations. Deficiencies in ocean models, or discrepancies between models and observations, lead to improvement in ocean models themselves. In short, the attempt to realistically simulate the ocean is an important one in itself, and has been the subject of entire books (e.g., O'Brien 1986; Haidvogel and Beckmann 1999).

The methodology of combining observations with models has fundamental importance in its own right. These methods are important for a more general science and engineering audience, such as the fields of computer science, economics, biology, and any other field with mathematical models. Some of the first methods to combine models and observations in geophysics were forms of objective mapping used in meteorology (e.g., Gilcrest and Cressman 1954; Sasaki 1970). In oceanography, large datasets are now available, and the synthesis of large and disparate forms of information is logically handled by combining all the observations with a model. This leads to a state estimate of the ocean (to be defined in more detail in Section 2.1) which is our best estimate of

what the ocean actually does. Relatively recently, oceanographers have used the Kalman filter (e.g., Fukumori et al. 1993; Miller et al. 1994) and the method of Lagrange multipliers (e.g., Thacker and Long 1988; Tziperman and Thacker 1989; Sheinbaum and Anderson 1990; Marotzke and Wunsch 1993) to combine models and data. This thesis presents novel research with the latter technique, the method of Lagrange multipliers, otherwise known as the adjoint method (see Section 3.2). The effects of nonlinearity in an extremely large dimensional space are explored here. In the future, the methods of this thesis and related methods are expected to be in widespread use in oceanography and the wider scientific community.

1.2.2 Eddy-resolving model with open boundaries

The model used in the present study is the Massachusetts Institute of Technology Ocean General Circulation Model (MIT GCM) with the complementary state estimation codes of the ECCO (Estimating the Circulation and Climate of the Ocean) Consortium. It is a z-coordinate model which employs the incompressible Navier-Stokes equations under the Boussinesq approximation and hydrostatic balance (Marshall et al. 1997a; Marshall et al. 1997b). The dynamical core of the model is discussed in more detail in Appendix A. The intent is to realistically simulate the Subduction Experiment region for one year: June, 1992, to June, 1993. Also, the model is designed to explicitly simulate the mesoscale eddy field. The Rossby radius of deformation is between $25 - 45$ km in this region, and the resolution we have chosen for the model is approximately 15 km, or $1/6°$. To completely resolve the eddy field, much higher resolution, e.g. $1/12°$ or even $1/20°$, is probably needed. At such high resolution, it is impractical computationally to run a global model, or even a complete North Atlantic model. Consequently, the model domain contains most of the eastern subtropical gyre of the North Atlantic (see Figure 1-6). At $1/6°$, eddy kinetic energy of the model is typically $50 - 75\%$ of TOPEX/POSEIDON observations. Although the domain is small, it was chosen such that all of the Subduction Experiment is within the interior and well away from the boundaries. Because this is a regional model, open boundaries have been implemented. The north, south and west

Horizontal Resolution	$1/6°$x$1/6° \approx (14.2 - 18.2)\ km$ x $18.5\ km$
Vertical Resolution	10-500 m
Grid Points	192 x 168 x 23 vertical levels
Time Step	900 s = 15 min.
Wind Stress Period	43200 s = 0.5 days
Heat/Freshwater Flux Period	86400 s = 1.0 day
Horizontal Viscosity/Diffusivity (biharmonic)	$\nu_h^2 = \kappa_h^2 = 0\ m^2/s$ $\nu_h^4 = \kappa_h^4 = 2\text{x}10^{11}\ m^4/s$
Vertical Viscosity	$\nu_z = 1\text{x}10^{-3}\ m^2/s$
Vertical Diffusivity	$\kappa_z = 1\text{x}10^{-5}\ m^2/s$

Table 1.2: Model parameters

boundaries are open, but the Mediterranean Sea is only opened in special experiments (see subsection "open boundaries" below).

This regional model is nested in the global, 2° state estimate of the ECCO Consortium (Stammer et al. 2002). This is a great advantage because all the time-dependent boundary values of the regional model are taken from the global estimate. For example, the initial temperature and salinity here are taken from the global state estimate. Preliminary model runs use the National Center for Environmental Prediction (NCEP) Reanalysis daily sensible and latent heat fluxes and twice-daily surface windstresses. Some modelers claim that the European Centre for Medium-Range Weather Forecasts (ECMWF) surface forcing is superior in this region, however (L. Yu, personal communication). The atmospheric forcing fields are improved and estimated here, so a reasonable first guess suffices for the first model runs. In conclusion, the MIT GCM is a state-of-the-art numerical model which makes it possible to simulate realistically the Subduction Experiment region.

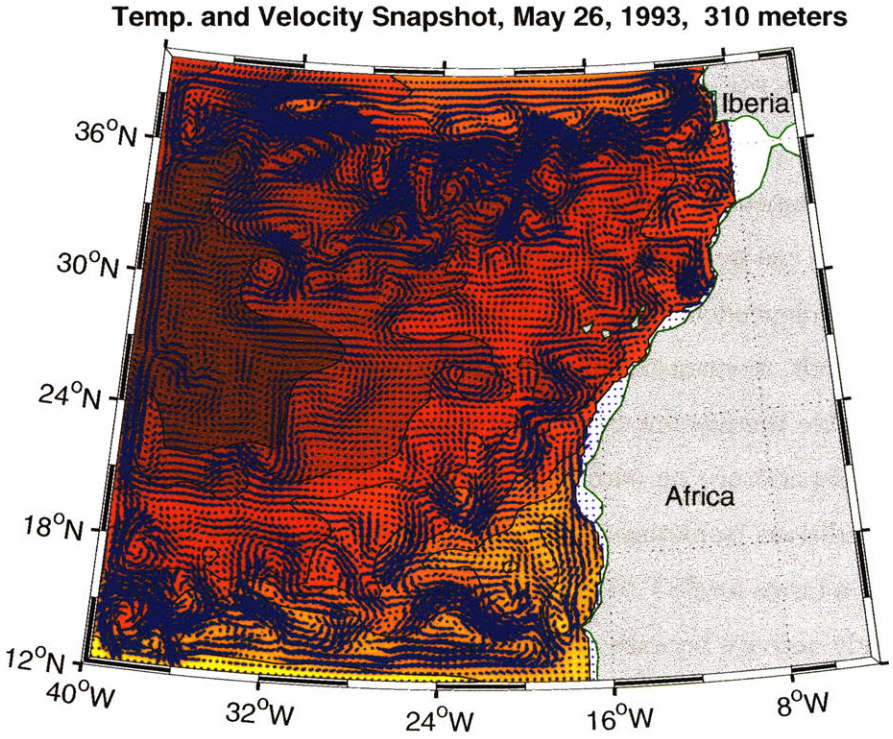

Figure 1-6: Snapshot of the 1/6° model temperature and velocity fields at 310 meters depth. Temperature has 1° contour intervals from 15°C to 21°C. The full model domain and three open boundaries are shown. This snapshot represents our first guess at the true ocean state on June 1, 1993. The model was started one year earlier, June 1, 1992.

K-profile parameterization

Previous simulations with GCM's have shown serious systematic errors in the mixed layer. Summertime mixed layers were too shallow and the sea surface temperature (SST) became unrealistically warm in the seasonal cycle. For this reason, the K-profile-parameterization (KPP) scheme is used here (Large et al. 1994). The scheme improves the model by parameterizing wind deepening of the boundary layer, by enhancing shear instability in the upper ocean, and by reducing the dependence on surface restoring conditions. The KPP model calculates increased diffusivities for underrepresented and unresolved ocean processes through the similarity theory of turbulence (Tennekes 1973). Another improvement of this boundary-layer model is its nonlocal behavior; heat, salt, and momentum can be fluxed through vertically homogeneous regions. Turbulent fluxes are therefore independent of local gradients, which is frequently the case in the mixed layer. As a result, momentum input at the surface can cause the boundary layer to penetrate the stable thermocline by wind-stirring. The improved model physics with KPP reduces the dependence on surface restoring conditions (Sausen et al. 1988). Surface restoring conditions (sometimes called flux corrections, especially with coupled models) are relaxation terms for SST to prevent systematic bias. These terms force the model to suppress eddy activity because of the constraint to a large scale SST field. The overall model performance is much improved in comparison with observations when the KPP model is added (see Chapter 3).

The KPP model has several weaknesses. In general, mixed-layer depths are still shallower than observed. The wind-stirring parameterization in KPP reduces the discrepancy but does not completely eliminate it. In coastal regions, the mixed-layer model has numerical problems when the mixed layer reaches the sea floor. There, the model behavior is noisy and nondifferentiable (see Section 3.3.3 for a definition and discussion), and nonphysical bottom fluxes are present. Continental shelves were removed in this model to eliminate the problem as they are not the focus of the research. A major practical problem with KPP is that the scheme analyzes vertical columns independently. Computational noise frequently develops in the horizontal direction. An ad-hoc solution,

used here, is the introduction of a horizontal smoothing function. The model results are not highly dependent on this smoothing in the subtropical gyre. In conclusion, the KPP model represents the best boundary-layer model at present, but improvement is possible.

Open boundaries

The implementation of open boundaries has not traditionally been a standard feature of GCMs. Here, density and velocities from the global state estimate (Stammer et al. 2002) are used to constrain the boundary through a sponge layer. The boundary conditions are treated as adjustable parameters, so an estimate of improved boundary velocities emerges in the synthesis (see Chapter 3). The Mediterranean Sea outflow is closed in the early experiments of this thesis, and is open later. The open boundary conditions vary in time on a monthly basis. Also, they have been calculated to exactly balance the volume flux into the domain on a monthly basis. With our present level of knowledge, exact volume conservation is a reasonable null hypothesis over these timescales. This assumption is checked later in the thesis (see Section 2.4.2). The design and implementation of numerical code for control and estimation (inverse aspects) of open boundary conditions is potentially a major contribution of this thesis, and is discussed later (see Section 2.4). The formulation of the open boundaries of the forward model alone is discussed in the next paragraph.

Open boundaries which require the prescription of the full oceanic state for forward integration are overdetermined and formally ill-posed (Orlanski 1976; Oliger and Sundström 1978). The prescribed open boundary state usually contradicts the dynamical equations that describe the interior circulation. At every timestep, two pieces of information exist for the new open boundary state: the update from the equations of motion and the prescribed state for the next timestep. This problem is formally overdetermined because too many boundary conditions are supplied (Bennett 2002). The correct number of boundary conditions for a primitive equation model depends on the interior flow characteristics and the vertical structure of waves (Oliger and Sundström 1978). This

is much more complicated than the case for a quasigeostrophic or shallow-water model where the correct number of open boundary conditions is more easily calculated. In summary, the addition of open boundaries to a primitive equation model is ill-posed because the solution of an overdetermined problem usually does not exist.

There are two ways to resolve the ill-posedness of the open boundary problem with a general circulation forward model: impose the correct number of boundary conditions in the first place or discard extra information. Radiation boundary conditions, like those of Orlanski (1976) and Marchesiello et al. (2001), identify passive and active boundaries, then modify the passive open boundary values. In this process, they attempt to apply the correct number of boundary conditions. On the other hand, a sponge layer, as used in this thesis, keeps the transition between the boundary and interior smooth by adding a relaxation term to the dynamics. In the forward numerical model, the right hand side of the temperature equation (Equation A.4) includes advection and diffusion terms, symbolically written $\mathbf{G}_\theta$, and an extra term due to the sponge layer:

$$\frac{D}{Dt}\theta(x,z,t) = \mathbf{G}_\theta(x,z,t) + \frac{1}{\tau(x-x_{ob})}[\theta(x,z,t) - \theta(x_{ob},z,t)] \tag{1.1}$$

where τ is a relaxation timescale that depends on distance from the boundary, $(x - x_{ob})$. At the boundary, the timescale is formally zero; there $\theta(x_{ob}, z)$ is prescribed. The sponge layer width is 1°, in which the boundary solution smoothly transitions to the interior. Salinity and horizontal momentum are also relaxed to prescribed values in the 1° layer. The sponge layer is an ad-hoc and nonphysical solution; therefore, a state estimate which is highly sensitive to the sponge layer formulation should be rejected. The model-observation synthesis of Chapter 3 seeks adjusted open boundary conditions which are dynamically consistent with the interior solution. Bennett (2002) postulated that the treatment of the open boundaries as an inverse problem renders the problem well-posed. Nevertheless, finding well-behaved boundary conditions has not previously been done for an eddy-resolving, primitive equation model.

Quantity	Dimension/Size/Length
State Vector	3.14 x 10^6 elements
Control Vector	5.49 x 10^6 elements
Observations	1.28 x 10^7 observations
Model Input	7.98 x 10^7 forcing elements
Model Output	1.09 x 10^{11} estimated elements
Parallel Processors	24-48 processors
Computational Time	400 cpu hours/iteration with IBM 1.3 GHz Power4 processors
Search Iterations	$\approx$ 120 iterations
Total Computer Time	$\approx$ 50,000 hours (5.7 years)
Numerical Code	569 subroutines 322,895 lines of forward code 22,507 lines of adjoint code

Table 1.3: Dimension of the problem

1.2.3 Size of the problem

The integration of a realistic eddy-resolving model is expensive and has many uncertain parameters. The sheer size of the problem presents a challenge. First, the high resolution of the model gives a very large number of grid points and a great computational cost. In fact, there are over three million prognostic variables for the model (identified as the *state vector* in Table 1.3). Second, the search for a model solution which fits the observations leads one to vary the uncertain boundary conditions[1]. The important, uncertain boundary conditions are chosen to be *control parameters*, and are further defined in Section 2.1. Here, there are over five million control parameters and consequently the search occurs in a five-million-dimensional space. The thesis tests the assumption that the high-dimensionality of the problem does not alter its fundamental character. Of course, the computational cost is high and present-day limits of computing power are approached.

[1]Boundary conditions include initial conditions, open boundary conditions, and surface forcing.

Computational tools

A computationally-intensive model needs massively parallel supercomputers to run the code. The MIT GCM has a "WRAPPER" environment which easily allows this to be implemented. In this thesis, the code is parallelized using a domain decomposition approach where subdomains (also called "tiles") of the model are sent to separate processors (see Foster (1995) for an excellent introduction to parallel computing). The limiting factor of the computational scalability of the GCM is communication between processors. When the subdomain size shrinks below thirty by thirty grid points, increased communication time offsets the increased computer processing power. With the number of grid points, twenty-four processors are the optimal number here. At various times during the thesis, the model was run on the eighth[2] and eleventh[3] largest supercomputers in the world. The practical implementation of the numerical model would not be possible without the parallelized code and the access to massively parallel supercomputers.

Another technical aside is that the MIT GCM numerical code has been automatically differentiated with the TAF (Tranformations of Algorithms in Fortran) tool of Giering and Kaminski (1998). An automatic differentiation tool allows for the adjoint model code to be regenerated whenever there are necessary changes in the forward code. The adjoint model provides vital information for fitting the model to observations, and is fully introduced in Section 3.2. The forward model contains over 500,000 lines of code, so hand-writing and rewriting the adjoint code would take approximately one to two years of dedicated work (Yu and Malanotte-Rizzoli (1996) took two years to hand-code the adjoint of the MOM ocean model). Therefore, the compatibility of this particular model with the adjoint generator makes the entire thesis feasible.

[2]The IBM SP3 "blue horizon" of the San Diego Supercomputer Center has 1,152 375 MHz processors, the 8th largest unclassified supercomputer in the world upon its release in 2000. Inevitably, it no longer ranks in the top 50 after a mere two years. Source: www.top500.org.

[3]The IBM SP4 "marcellus" of the Naval Oceanographic Office Major Shared Research Center, Stennis Space Center, MS, is the 11th largest supercomputer in the world (2003) with a peak performance of 7.5 Teraflops.

1.3 Synopsis of the thesis

To review, Chapter 1 shows the widespread impact of subduction on the water mass properties of the main thermocline. Past studies of subduction have focused on the large scale and steady or seasonally-varying ocean circulation. Recent papers have begun to consider the impact of eddy-driven subduction and have shown that eddies are important in certain regions of the ocean. In the subtropical gyre, mesoscale eddy energy is a ubiquitous feature of all observations and it is not obvious that it can be ignored. The observations of the Subduction Experiment do not adequately resolve the eddy-scale motions of interest. A numerical model, the MIT GCM, statistically combined with the observations, produces a estimate of the ocean circulation at 1/6°. Using this state estimate, this thesis aims to understand subduction in a realistic, turbulent ocean.

Chapter 2 shows that the synthesis of a model and observations can be formulated as a giant least-squares problem. To advance the scientific agenda, a best estimate of the ocean circulation is sought from the combination of the Subduction Experiment observations and an eddy-resolving, regional general circulation model for June, 1992, to June, 1993. Measurements of temperature and velocity at five moorings, TOPEX/Poseidon satellite altimetry, Levitus climatologies and Reynolds sea surface temperatures are used as constraints on the model. The model trajectory is controlled by adjusting the initial conditions, boundary conditions, wind stresses, heat and freshwater flux. The goal is to vary the control parameters to find a model trajectory that fits the observations within their uncertainty.

Chapter 3 finds a model solution which fits both the large-scale and small-scale observational signal. The method of Lagrange multipliers [otherwise known as the adjoint method (Wunsch 1996)] is a logical way to combine oceanic datasets into one dynamically-consistent estimate. For field campaigns where all the data has been compiled and collected, the adjoint method uses all the data at once and the method enables estimation from data collected in future time. The method is also computationally feasible because it does not require a extraordinarily large number of perturbed model simulations, nor does it need to compute large error covariance matrices. Practical

implementation and solution of the minimization problem is detailed in this chapter. In particular, the nonlinearity of the model constraint is shown to be a fundamental factor in the optimization problem. Despite concerns of the published literature (Lea et al. 2000; Köhl and Willebrand 2002), trajectories of the eddy-resolving Subduction Experiment model diverge quasi-linearly in time and the adjoint model is stable. Consequently, the adjoint-computed gradients give adjusted initial conditions which do lead to an improved model trajectory. After fifty iterations of the forward-adjoint model, the method decreases the data-model misfit nearly to the level of the expected error in the observations. For this study, there appears to be no fundamental obstacle to adjusting the model trajectory into complete consistency with the observations and their prior estimated error. The adjoint method is successful because the forward model itself is only weakly nonlinear in the region. The model is not extremely sensitive to the initial conditions, and the problems associated with chaotic dynamics do not interfere. The result is a dynamically-consistent, three-dimensional, time-varying, nested, high resolution estimate of the ocean circulation. The Subduction Experiment model suggests a wide potential for the adjoint method in oceanography, and this is a major result in itself.

Chapter 4 illuminates the role of eddies in subduction. This chapter uses the state estimate to diagnose quantities of interest which can not be measured directly. A preliminary step is to compare subduction in the state estimate to classical theory. As expected, the seasonal cycle and the mixed-layer demon influence the properties of subducted water, but the pathways of subduction do not resemble those of an idealized ocean model. The pattern of annual subduction rates has a small-scale signature and suggests a significant contribution of eddies to subduction. The goal of this thesis is to quantify the relative importance of eddy-driven subduction to the total subduction. In the state estimate, eddy-induced volume fluxes across the base of the mixed layer are 15% of the total subduction, and consequently are locally important. When subduction is calculated in density coordinates, eddy-subduction is seen to be important in the density range of $25.5 < \sigma < 26.5$, which encompasses both the Azores Current and the

North Equatorial Current. From these findings, the eddy scale motions are an additional and sizable source of subducted water near fronts in the eastern North Atlantic Ocean.

Chapter 5 summarizes the findings of the thesis. The novel scientific results of this thesis, as well as advances in the methodology, are reviewed. Finally, the limitations of the thesis are discussed, with speculation for future research.

Chapter 2

The Model-Observation Least-Squares Problem

2.1 Overview of the concept

To shed new light on subduction, this thesis creates a new estimate of the ocean circulation during the Subduction Experiment. The goal is to estimate the ocean circulation as realistically as possible. In a world of imperfect models and sparse, noisy observations, how can one determine the "goodness" of an estimate? A set of criteria, sometimes called the *performance* in control theory (i.e., Dahleh (1999)), are determined by the observations and characteristics of the ocean. Mathematically, the performance criteria are written as a giant least-squares minimization problem. This chapter defines the specific least-squares problem at hand: the search for an eddy-resolving regional model trajectory that fits the Subduction Experiment observations within their uncertainty.

Definitions

Before proceeding, it is instructive to be more specific about our stated goals. We wish to estimate the circulation of the ocean as described by the three-dimensional, time-varying density, velocity, and surface elevation fields. From the temperature, salinity, and horizontal velocity fields, all physical quantities of interest are computable (see

Appendix A). Hence, temperature, salinity, and horizontal velocity completely describe the previous history of the ocean circulation and are called the *state*[1]. The useful combination of model and observations is called a *state estimate* as we are explicitly interested in the circulation (i.e., the state) as it evolves. State estimation problems are frequently solved by methods that have been developed in the field of *control theory*, the study and search for forces or controls that drive an observed system in a desired way. An ocean model is driven by forces which can be considered controls, like the relatively unknown atmospheric fields over the open ocean. Of course, the actual ocean is not controllable due to engineering limits, but instead one wishes to control an ocean model to behave in a way which is consistent with observations. Much like control theory, the controls themselves are considered important quantities to be estimated[2]. Hence, observations contain knowledge of the true boundary conditions, not just the interior ocean where the observations were taken.

The methodology used here does not solely come from control theory. Many of the methods are also classified as *inverse methods*, which are methods used to solve problems that are not posed in the usual mathematical way (Tarantola 1987; Wunsch 1996). Inverse methods are unique in that they consider uncertainty to be an essential part of the solution. This problem is also classified as a part of *optimization theory*, which has a large set of available tools, although many were developed for small-dimensional systems (Luenberger 1984; Gill et al. 1986). Optimization includes both maximization and minimization problems, such as the least-squares problem here.

When dealing with combinations of models and observations, many atmospheric scientists and oceanographers prefer to use the term *data assimilation*. Some researchers denote both state estimation and forecasting as parts of the wider field of data assimilation. Unfortunately, data assimilation now has the connotation of the particular methods developed for the atmosphere and has little meaning to the entire scientific community. Therefore in an effort to use a terminology that is meaningful to those outside

[1]For the numerical model, the number of variables needed for a restart at any time is larger than the state described here.

[2]Estimated controls contain both an estimate of the true boundary conditions as well as model error. Separating these two contributions is not usually trivial.

of atmospheric and oceanic physics, the methodology here is termed state estimation.

Performance criteria

Observations undoubtedly provide information on the state of the circulation. Nevertheless, measurements are imperfect. They contain some error due to the instrument which should be accounted for. Also, observations are irregularly distributed in space and time and typically miss some features of interest. This is true in the Subduction Experiment, where five moorings can not be expected to give much spatial coverage despite their decent temporal coverage. The diagnosis of budgets, such as subduction rates, are especially difficult with mooring observations. Although observations are sometimes seen as the only source of "sea-truth", they alone are not adequate to make an estimate which fulfills our criteria.

Like observations, the laws of physics themselves provide meaningful information which can be used to improve a state estimate. However, the laws of physics, embodied here as a general circulation model (GCM), are uncertain as well. Model trajectories are uncertain because of both poorly-known oceanic forcing fields and inaccurate dynamics. On the positive side, a model provides information with high-resolution, only limited by computer power. The well-distributed coverage of model output makes possible the computation of sensible budgets. A complete state estimate must use the laws of physics because of the useful information they provide.

At this point, an estimate that best uses all available information necessarily contains both observations and a model. A further criterion is that the estimate provides a statistical blend of both sources that depends on their relative uncertainty. In cases where the error in both sources is assumed to be jointly normal, the proper statistical blend can be proved to be the maximum likelihood solution, the best estimate of truth (Van Trees 1968). A statistically-rigorous combination will also allow for the careful assessment of the uncertainty of the final solution, a desirable quantity. The result of our combination of data and model could be called *dynamic interpolation*; a dynamic model interpolates and fills the missing information between given observational points.

The estimate need not go through the exact observational values, however. Because of the observations' uncertainty, this would not be the best solution anyway. Consequently, a model can be used to distinguish between signal and noise in observations. As can be seen, there are many reasons to form an estimate from both model and observations.

When does a model serve as an adequate dynamic interpolator? *Cross-validation*[3] is the comparison of an estimate to withheld data, and it evaluates the model's ability to predict the ocean circulation in the absence of observations. The MIT GCM shows promise as a dynamic interpolator for two reasons. One, the first-guess model trajectory is reasonably close to the observations. This model trajectory uses none of the observations in the cost function; it withholds all the data points. Two, the model compares well with observations that were not included[4] in the estimation process. A WOCE hydrographic section is used for this purpose later (see Section 3.5.4). Cross-validation is one way to give the investigator more confidence in the state estimate.

To be explicit, our performance criteria can be listed:

- Follow what was observed within its uncertainty
- Adhere to the laws of physics within their uncertainty at all times
- Combine all information in a statistically rigorous way

The performance criteria are objectified into one number, the *cost function*: a sum of squared elements of the model-data misfit. A small cost function represents a solution which follows all of the performance wishes. Of course, "small" is a relative term which must be defined later. Second, we identify uncertain parameters in the model which can be adjusted. These parameters are known as the *control variables*, because they are the parameters that allow control of the model. The goal of combining the model and observations can now be restated: adjust the control variables such that the cost function has an appropriately small value (see Figure 2-1). More specifically, the cost function and its individual elements must satisfy the prior error statistics, which include

[3]Cross-validation is perhaps a misleading term because true model validation is not possible; only falsification is possible.

[4]A best state estimate, however, would use all available information in the cost function.

specification of the overall error as well as the distribution of individual errors (see Section 3.4.2). One added criterion is:

- Do not allow unrealistically large controls

Together, these criteria become the mathematical statements which allows us to unambiguously define the problem of combining observations with a model.

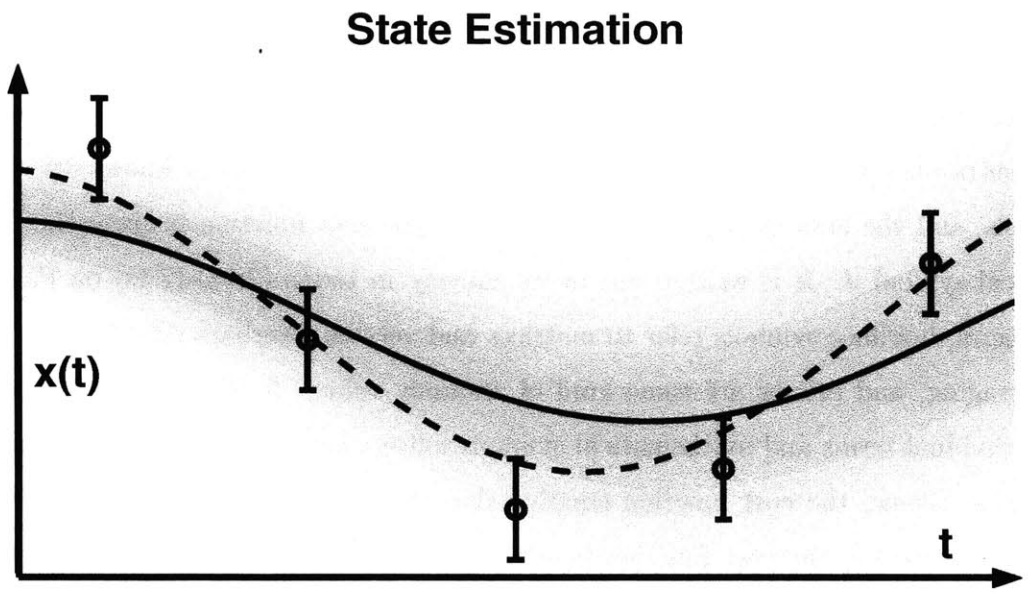

Figure 2-1: Schematic of state estimation. The goal is to find a model trajectory that is within observational uncertainty (*O's with error bars*). The model trajectory is also subject to uncertainty due to model error and uncertain model parameters (*shown as a gray probability distribution cloud*). Here, the first-guess model simulation (*solid black line*) is not within the observational uncertainty at all times. However, there is a model trajectory (*dashed line*) that is consistent with both the observational and model uncertainty. This improved model trajectory is the state estimate.

As a reminder, it is not necessarily true that all the performance criteria can be met. In practice, these criteria actually form a very stringent test. In case of failure of one or more items, all is not lost. Such a result gives the investigator information about the inconsistencies between various observations or could possibly force the investigator to rethink the accuracy of measurements. Another possibility is the rejection of the model

as inadequately realistic. This serves as a call for model improvement. In any case, the performance criteria serve the useful purpose of quantifying the problem at hand, whether they can be satisfied or not.

2.2 Cost function

The form of the cost function is a squared misfit between the estimate and all *a priori* information. The problem of combining a model and observations is reduced to a least-squares problem, albeit a giant one. In this section, the thesis systematically introduces the contributions to the cost function. The observations, the prior knowledge of the controls, and the laws of physics all play a role. The cost function is given the mathematical symbol J. It is written out in its entirety in terms (2.1a)-(2.1s) on Page 43. In general, boldface symbols refer to matrices and vectors, overbars refer to some kind of averaging, and primes are some kind of anomaly value. A more detailed guide to the individual terms and mathematical symbols follows in the next sections. To repeat an earlier theme, the cost function simply takes the form of a sum of squared differences. Minimizing the cost function is equivalent to solving a least squares problem, although many contributions must be considered. The first five terms (2.1a)-(2.1d) are the observational misfit terms, the goodness of fit to the observations. The next three terms (2.1e)-(2.1g) are the climatological misfits; they constrain the estimate to ocean climatologies with considerable leeway. The next fourteen terms (2.1h)-(2.1s) are *control penalty terms*; they constrain the control parameters to lie within a certain range of their initial guess or to adhere to dynamical rules. The control penalty terms take the place of an explicit model error term in our cost function. The next sections explain the cost function in a term-by-term manner.

2.2.1 Role of weights

The generic form of the cost function (Equation (2.1)) has a weighting matrix, $\mathbf{W}$, with each term. Critics of inverse problems claim that the weighting matrices determine the

$$
\begin{aligned}
J &= \sum_{t}^{12} \sum^{5} (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{moor})^T \, \mathbf{W}_T \, (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{moor}) && \text{(2.1a)} \\
&+ \sum_{t}^{12} \sum^{5} (\overline{\mathbf{U}} - \overline{\mathbf{U}}_{moor})^T \, \mathbf{W}_{VEL} \, (\overline{\mathbf{U}} - \overline{\mathbf{U}}_{moor}) && \text{(2.1b)} \\
&+ \sum_{t}^{12} \sum^{5} (\overline{\mathbf{V}} - \overline{\mathbf{V}}_{moor})^T \, \mathbf{W}_{VEL} \, (\overline{\mathbf{V}} - \overline{\mathbf{V}}_{moor}) && \text{(2.1c)} \\
&+ \sum_{t}^{360} (\boldsymbol{\eta}' - \boldsymbol{\eta}'_{tp})^T \, \mathbf{W}_{tp} \, (\boldsymbol{\eta}' - \boldsymbol{\eta}'_{tp}) \; + \; (\overline{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}_{tp})^T \, \mathbf{W}_{geoid} \, (\overline{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}_{tp}) && \text{(2.1d)} \\
&+ \sum_{t}^{12} (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{Lev})^T \, \mathbf{W}_{Lev_T} \, (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{Lev}) && \text{(2.1e)} \\
&+ \sum_{t}^{12} (\overline{\mathbf{S}} - \overline{\mathbf{S}}_{Lev})^T \, \mathbf{W}_{Lev_S} \, (\overline{\mathbf{S}} - \overline{\mathbf{S}}_{Lev}) && \text{(2.1f)} \\
&+ \sum_{t}^{12} (\overline{\mathbf{T}}_{sfc} - \overline{\mathbf{T}}_{Rey})^T \, \mathbf{W}_{SST} \, (\overline{\mathbf{T}}_{sfc} - \overline{\mathbf{T}}_{Rey}) && \text{(2.1g)} \\
&+ (\mathbf{T}_0 - \mathbf{T}_{ecco})^T \mathbf{W}^*_{T_0} (\mathbf{T}_0 - \mathbf{T}_{ecco}) + (\mathbf{S}_0 - \mathbf{S}_{ecco})^T \mathbf{W}^*_{S_0} (\mathbf{S}_0 - \mathbf{S}_{ecco}) && \text{(2.1h)} \\
&+ \sum_{t}^{37} (\overline{\boldsymbol{\tau}}_x - \overline{\boldsymbol{\tau}}_{x_{ncep}})^T \mathbf{W}^*_{\tau_x} (\overline{\boldsymbol{\tau}}_x - \overline{\boldsymbol{\tau}}_{x_{ncp}}) + (\overline{\boldsymbol{\tau}}_y - \overline{\boldsymbol{\tau}}_{y_{ncp}})^T \mathbf{W}^*_{\tau_y} (\overline{\boldsymbol{\tau}}_y - \overline{\boldsymbol{\tau}}_{y_{ncp}}) && \text{(2.1i)} \\
&+ \sum_{t}^{37} (\overline{\mathbf{H}}_Q - \overline{\mathbf{H}}_{Q_{ncep}})^T \, \mathbf{W}^*_{H_Q} \, (\overline{\mathbf{H}}_Q - \overline{\mathbf{H}}_{Q_{ncep}}) && \text{(2.1j)} \\
&+ \sum_{t}^{37} (\overline{\mathbf{H}}_F - \overline{\mathbf{H}}_{F_{ncep}})^T \, \mathbf{W}^*_{H_F} \, (\overline{\mathbf{H}}_F - \overline{\mathbf{H}}_{F_{ncep}}) && \text{(2.1k)} \\
&+ \sum_{t}^{12} (\mathbf{T}_{o.b.} - \mathbf{T}_{o.b\,ECCO})^T \, \mathbf{W}_{T_{o.b}} \, (\mathbf{T}_{o.b.} - \mathbf{T}_{o.b._{ECCO}}) && \text{(2.1l)} \\
&+ \sum_{t}^{12} (\mathbf{S}_{o.b.} - \mathbf{S}_{o.b._{ECCO}})^T \, \mathbf{W}_{S_{o.b.}} \, (\mathbf{S}_{o.b.} - \mathbf{S}_{o.b._{ECCO}}) && \text{(2.1m)} \\
&+ \sum_{t}^{12} (\overline{\mathbf{U}}^z_{o.b} - \overline{\mathbf{U}}^z_{o.b._{ECCO}})^T \, \mathbf{W}_{V_{bt}} \, (\overline{\mathbf{U}}^z_{o.b.} - \overline{\mathbf{U}}'^z_{o.b._{ECCO}}) && \text{(2.1n)} \\
&+ \sum_{t}^{12} (\mathbf{U}'_{o\,b.} - \mathbf{U}'_{o.b._{ECCO}})^T \, \mathbf{W}_{V_{bc}} \, (\mathbf{U}'_{o.b.} - \mathbf{U}'_{o.b._{ECCO}}) && \text{(2.1o)} \\
&+ \sum_{t}^{12} (\overline{\mathbf{V}}^z_{o.b} - \overline{\mathbf{V}}^z_{o.b._{ECCO}})^T \, \mathbf{W}_{V_{bt}} \, (\overline{\mathbf{V}}^z_{o.b.} - \overline{\mathbf{V}}^z_{o.b._{ECCO}}) && \text{(2.1p)} \\
&+ \sum_{t}^{12} (\mathbf{V}'_{o.b.} - \mathbf{V}'_{o.b._{ECCO}})^T \, \mathbf{W}_{V_{bc}} \, (\mathbf{V}'_{o.b.} - \mathbf{V}'_{o.b._{ECCO}}) && \text{(2.1q)} \\
&+ \sum_{t}^{12} (\frac{\partial \mathbf{V}'_{\perp}}{\partial z} + \frac{g}{\rho_0 f} \frac{\partial \boldsymbol{\rho}}{\partial l})^T \, \mathbf{W}_{ageos} \, (\frac{\partial \mathbf{V}'_{\perp}}{\partial z} + \frac{g}{\rho_0 f} \frac{\partial \boldsymbol{\rho}}{\partial l}) && \text{(2.1r)} \\
&+ \sum_{t}^{12} (\overline{\mathbf{V}}^z_{\perp}{}^T \mathbf{A}_{lz})^T \, \mathbf{W}_{volflux} \, (\overline{\mathbf{V}}^z_{\perp}{}^T \mathbf{A}_{lz}) && \text{(2.1s)}
\end{aligned}
$$

entire solution and can be manipulated by the investigator to serve any purpose. They are correct that the weights determine the solution to the problem. However, the choice of weights should be physically motivated, and the final solution must pass posterior error tests. Lorenc (1986) has shown that the weight matrix, $\mathbf{W}$, should be the inverse of the noise covariance matrix, $\mathbf{R}_{nn}$, to have the minimum variance solution. In other words, the weight used in the cost function is inversely related to the acceptable error or noise in the misfit; a smaller acceptable error leads to a larger weight in the cost function. Lorenc (1986) further showed that this judicious choice of weights leads also to the maximum likelihood solution if the error statistics are jointly normal (Van Trees 1968). For statistical rigor, only *a priori* knowledge should be used to determine the weights. Then, the final estimate must have errors that satisfy the original specifications: a difficult posterior test to pass. Other critics point out that there are many different ways to adjust the controls to achieve the same goal. For example, the ocean model can be made warmer by either warming the initial condition or by imposing a heat flux at the surface. The weights distinguish which process is more likely. In summary, the weights are a ubiquitous feature of the cost function, and they are not manipulated in a haphazard fashion; knowledge of physics and *a priori* error statistics drives the choices.

Although the theory behind the weights may be sophisticated and well-developed, the practical application of such ideas is typically far from straightforward. For example, the misfit between observations and model may be due to a number of reasons. First, the observations themselves contain noise due to measurement error. This is typically a small error, although with satellite altimetry such measurement error rivals the signal we wish to observe. Second, there may be representation error due to the model. Representation error results because the model grid and the observational locations do not coincide. In such a case, the model must be mapped onto the observation's location via an imperfect interpolation scheme. With a high-resolution model such as the one in this thesis, this form of representation error is small because there is very little separation between grid points. Another representation error is due to missing physics in the model. All unresolved processes must be considered as possible sources of error

in a model-observation comparison.

Zang-Wunsch model of low frequency variability

To compute the expected representation error, the energy in various wavenumber and frequency bands is computed via the spectrum of Zang and Wunsch (2001, hereafter ZW, Figure 2-2). The energy that must be considered noise varies with data type and model resolution. ZW use a simple dynamical model of a linear, continuously-stratified, time-varying ocean and a knowledge of a wide variety of oceanographic measurements. Their model can be used to infer a universal shape of the frequency-wavenumber spectra, and also can be used to infer spectra that are not observable. The main weakness of the Zang-Wunsch model is the potential energy structure in the mixed-layer. Quasi-geostrophic dynamics do not describe this region of the ocean, so other assumptions must be made to account for the seasonal cycle. Nevertheless, the Zang-Wunsch model provides a reasonable *a priori* guess of the mesoscale eddy energy everywhere in the domain.

The recipe for calculating eddy energy from the Zang-Wunsch model follows. First, SSH variability from TOPEX/POSEIDON is used to calibrate the horizontal distribution of potential energy ($I(\phi, \lambda)$, equation (22), ZW). The horizontal pattern of energy used here is very similar to the original pattern in ZW. The vertical structure of energy is partitioned in the first three modes with a ratio of $1 : 1 : 1/2$. From the surface potential energy and the vertical structure, temperature variance is calculated at every level (equation (41), ZW). To account for the seasonal cycle, the Reynolds SST seasonal variance is calculated, and added to the previous temperature variance profile with an exponential decay scale of 200 meters. 200 meters is chosen to coincide with the deepest wintertime mixed layers in the region. Next, eddy kinetic energy is estimated. Surface potential energy is related to kinetic energy through geostrophy, as also used by Stammer (1997). Again, the vertical normal modes are used to extrapolate and estimate the vertical structure of kinetic energy. The prior estimates of eddy energy compare well with the observations of the Subduction Experiment. In addition, estimates of needed

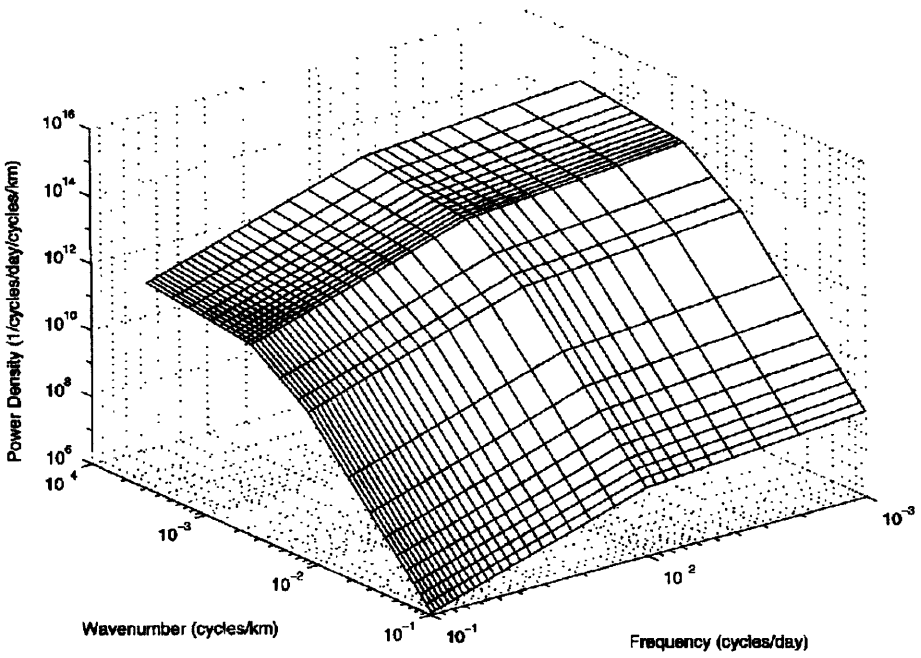

Figure 2-2: Universal frequency and wavenumber spectrum for the streamfunction of the Zang-Wunsch model of ocean variability. From Zang and Wunsch (2001).

but unobserved quantities can be made (Figure 2-3).

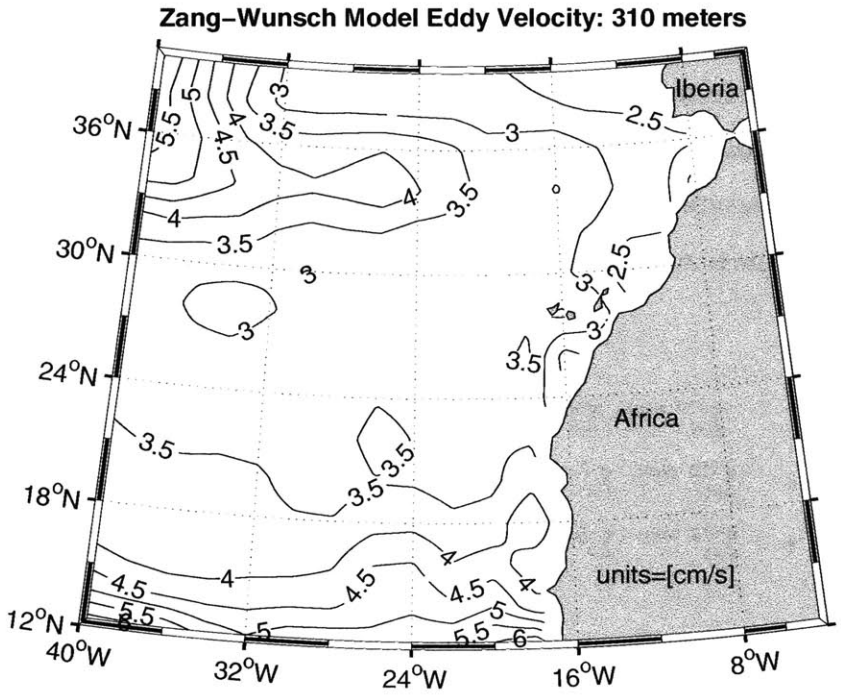

Figure 2-3: Standard deviation of time-variable zonal velocity at 310 meters in the Zang-Wunsch model. This thesis uses this eddy field as an *a priori* estimate for weights in the cost function. Notice two bands of higher eddy energy: the Azores Current and the North Equatorial Current.

2.2.2 Observational terms

Subduction Experiment moorings

The state estimate should accurately reflect the observations of temperature and velocity made at the five locations of the Subduction Experiment moorings. There is a greater density of temperature measurements, but there are also many velocity measurements by Vector Measuring Current Meters (VMCM's, Weller and Davis, 1980) in the upper 1000 meters. VMCM's provide both the u and v component of velocity which is directly

comparable[5] to model output. Unfortunately, there are no salinity measurements at the moorings. Also, several month-long failures are present in the data. In the vertical, measurements were concentrated in the upper 1000 meters. Measurements in the deep ocean were so sparse that they were ignored for this study; also, our primary objective is to understand the mooring data as it affects subduction and upper ocean processes, so deep ocean measurements have little influence in a short time period.

In the cost function, the misfit between the model and the Subduction Experiment moorings is:

$$\sum_t^{12\ mon} \sum^{5\ moor} (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{moor})^T\ \mathbf{W}_T\ (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{moor}) \tag{2.2}$$

$$+ \quad \sum_t^{12\ mon} \sum^{5\ moor} (\overline{\mathbf{U}} - \overline{\mathbf{U}}_{moor})^T\ \mathbf{W}_{VEL}\ (\overline{\mathbf{U}} - \overline{\mathbf{U}}_{moor}) \tag{2.3}$$

$$+ \quad \sum_t^{12\ mon} \sum^{5\ moor} (\overline{\mathbf{V}} - \overline{\mathbf{V}}_{moor})^T\ \mathbf{W}_{VEL}\ (\overline{\mathbf{V}} - \overline{\mathbf{V}}_{moor}) \tag{2.4}$$

where $\mathbf{T}$, $\mathbf{U}$, and $\mathbf{V}$ are the model temperature, zonal and meridional velocity, $\mathbf{T}_{moor}$, $\mathbf{U}_{moor}$, and $\mathbf{V}_{moor}$ are the observed temperature, zonal and meridional velocity, the overbar represents a monthly mean, and $\mathbf{W}_T$ and $\mathbf{W}_{VEL}$ are diagonal weighting matrices.

The weighting matrices take into account the instrumental error in the records as well as the representation error in the model. The temperature measurements are accurate within $0.01°C$ (Brink et al. 1995) and the current meters are assumed to measure within 0.005 m/s, although no error estimates were published. The numerical model does not accurately represent the physics below scales of 100 kilometers, a much bigger error. Those small scales are either completely unresolved, or mesoscale activity is underrepresented and overdamped by numerical friction. Wavenumber spectra of ocean properties drop off too quickly at scales smaller than 100 kilometers due to friction. This is an example of representation error in the model, and any energy in the observations at these scales will have to be considered noise in the observations. Using the model of Zang and Wunsch (2001, and Section 2.2.1), it is possible to calculate the ocean variability at scales less than 100 kilometers and at periods greater than a month (because

[5]If the current meters measured speed, this would be a nonlinear function of the model state, and could cause additional problems.

we are using monthly means for the comparison). Changes in model resolution require a revised estimate of representation error.

Mesoscale variability is a strong function of horizontal location and depth. For example, the eddy kinetic energy varies by a factor of five in the region, with a band of high energies in both the Azores Front and the North Equatorial Current. Also, the expected variability, and hence representation error, varies by a factor of ten in the vertical. In addition, the vertical structure itself changes throughout the region. In the northern, "mid-latitude" part of the basin which is observed by the central, northwest and northeast moorings, most of the eddy energy is equally partitioned between the barotropic and first baroclinic modes. For the southwest and southeast moorings, the second baroclinic mode contains much more energy. The vertical partition of horizontal kinetic energy is consistent between the mooring observations and the Zang-Wunsch model (also see (Wunsch 1997)). All of these subtleties are taken into account in our estimate of the expected errors. However, there are a few assumptions here that should be highlighted. The expected errors due to the misrepresented mesoscale eddy field are assumed to be isotropic, as evidenced by the identical $\mathbf{W}_{VEL}$ weighting matrices for both u and v. This assumption is actually quite good in this region without a strong western boundary current. Also, the Zang-Wunsch spectrum is not a function of time, and likewise our weighting matrices are not a function of time. Finally, no covariance is assumed between the model-observation misfit at different locations and times. This is clearly wrong, but is a first-order attempt to accurately guess the error statistics. As can be seen above, a knowledge of the physics has guided our choice for the mooring weights.

TOPEX/POSEIDON altimetry

Satellite altimetry offers a wealth of information that was not previously available. Although the satellite altimeter mission was not explicitly part of the Subduction Experiment, the sheer number of observations of sea surface height made by the TOPEX/POSEIDON satellite is staggering, and any estimate of the ocean circulation would be remiss to ig-

nore it. Here, direct comparisons are made between the satellite-measured sea surface height anomaly and the model on the satellite's ground tracks. The mean sea surface height is computed from 7 years of TOPEX/POSEIDON observations and put onto the model grid.

The misfit between the model and the TOPEX/POSEIDON satellite altimetry is:

$$\sum_{t}^{360\ days} (\boldsymbol{\eta}' - \boldsymbol{\eta}'_{tp})^T \mathbf{W}_{tp} (\boldsymbol{\eta}' - \boldsymbol{\eta}'_{tp}) \tag{2.5}$$

$$+ \quad (\overline{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}_{tp})^T \mathbf{W}_{geoid} (\overline{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}_{tp}) \tag{2.6}$$

where $\boldsymbol{\eta}_{tp}$ is the along-track sea surface height observed from the TOPEX/POSEIDON satellite, $\boldsymbol{\eta}$ is the model sea surface height on the same tracks, the overbar is a one year mean, primes represent the daily-averaged sea surface height anomaly, $\mathbf{W}_{tp}$ is the weight on sea surface height anomaly, and $\mathbf{W}_{geoid}$ is the weight on mean sea surface height field (primarily due to errors in the geoid).

Unlike many other observations used in this thesis, TOPEX/POSEIDON measurements have considerable instrumental noise. Sources of this noise include orbital tracking error and the E-M bias of ocean waves (Fu et al. 1994; Tai and Kuhn 1995). Therefore, $\mathbf{W}_{tp}$ takes into account a spatially-invariant and stationary background noise of 4.3 *cm*. For comparison, the signal we wish to track has magnitudes of $5 - 20$ *cm* in this region. Also, some percentage of the eddy energy will not be represented by the model. According to the Zang-Wunsch model, 6% of the sea surface height variance is at spatial scales less than 100 kilometers; this is also treated as acceptable noise. As with the mooring weights, $\mathbf{W}_{tp}$ accounts for spatial variations in the acceptable noise, but is not a function of time and is diagonal. The mean sea surface height field has errors of a different kind: errors in the absolute reference level or *geoid*. At scales less than 1000 *km*, geoid errors dominate the mean sea surface height signal. $\mathbf{W}_{geoid}$ is therefore taken from published error estimates of the EGM96 geoid (Lemoine et al. 1997; Wunsch and Stammer 1998). With the small domain of the Subduction Experiment, the mean sea surface field is only a marginal constraint.

2.2.3 Climatological terms

Although ocean climatologies are not usually considered "observations", they are a collection of observations and they do contain information. Here, the Levitus climatology of temperature and salinity and Reynolds sea surface temperature climatology (Levitus et al. 1994; Reynolds and Smith 1994) are monthly-averaged climatologies of the seasonal cycle. Their information content is used in the least-squares problem by adding terms to the cost function.

The misfit between the model and ocean climatologies is:

$$\sum_t^{12\ mon} (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{Lev})^T \mathbf{W}_{Lev_T} (\overline{\mathbf{T}} - \overline{\mathbf{T}}_{Lev}) \tag{2.7}$$

$$+ \quad \sum_t^{12\ mon} (\overline{\mathbf{S}} - \overline{\mathbf{S}}_{Lev})^T \mathbf{W}_{Lev_S} (\overline{\mathbf{S}} - \overline{\mathbf{S}}_{Lev}) \tag{2.8}$$

$$+ \quad \sum_t^{12\ mon} (\overline{\mathbf{T}}_{sfc} - \overline{\mathbf{T}}_{Rey})^T \mathbf{W}_{SST} (\overline{\mathbf{T}}_{sfc} - \overline{\mathbf{T}}_{Rey}) \tag{2.9}$$

where $\mathbf{T}$, $\mathbf{S}$, and $\mathbf{T}_{sfc}$ are the model temperature, salinity and sea surface temperature, $\mathbf{T}_{Lev}$, $\mathbf{S}_{Lev}$, and $\mathbf{T}_{Rey}$ are the Levitus temperature, Levitus salinity and Reynolds sea surface temperature, the overbar represents a monthly mean, and $\mathbf{W}_{Lev_T}$, $\mathbf{W}_{Lev_S}$ and $\mathbf{W}_{SST}$ are diagonal weighting matrices.

The Levitus climatology of temperature and salinity includes error estimates as a function of depth, and these are primarily used to compute the weights $\mathbf{W}_{Lev}$. The representativeness of a climatology for any particular year must be estimated. Interannual variability contributes to the misfit between the climatology and model fields. Upon further inspection, the published errors in Levitus's product are similar to the interannual variability as seen by Roemmich and Wunsch (1984) and Parrilla (1994). In addition, there are other forms of error in the climatologies. The uneven coverage of much of the ocean probably presents a large source of uncertainty, but because the actual distribution of data points has not been presented, one does not know how this would change the error estimates. On a different note, the Levitus compilation represents the large-scale density structure of the ocean and not the mesoscale eddy signature. Again, the Zang-Wunsch model is used to determine the energy of the mesoscale which

is not represented in the dataset. In this case, the climatology only represents energy at lengthscales larger than 400 kilometers because of its coarse gridding; 58% of mesoscale energy in temperature fluctuations is at smaller scales and is considered noise. For simplicity, the Reynolds weights $\mathbf{W}_{SST}$ are identical to the Levitus weights at the surface. After accounting for all of the above sources, the acceptable error in the model fit to the ocean climatologies is much larger than the acceptable error for an individual, *in-situ* observation. Because cost function weight is inversely proportional to acceptable error, terms (2.7)-(2.9) are downweighted relative to the other observational terms in the cost function. This does not automatically render the climatologies unimportant in the state estimation problem; the total number of independent pieces of information in a climatology determines its relative influence.

On the consistency of the multiple datasets

Although our ultimate goal is to combine a model with all forms of observations, one must first assure that the observations are consistent amongst themselves. A comparison between observations of differing data types, such as between the mooring temperature and satellite sea surface height, is difficult. Such a study would be a whole research project unto itself (Stammer 1997). This consistency check will be done automatically during the process of combining the model and observations, and can be determined by a final estimate statistics. Nevertheless, for the sake of bolstering confidence before more intensive endeavors, the mooring temperature dataset can easily be compared to the Levitus climatology for temperature. Figure 2-4 shows the squared difference between the two datasets as a function of depth. The two datasets are consistent within the prior error estimates. These error estimates consider the instrumental error in the dataset, as well as errors in representation. Consistency between datasets, as shown here, is a necessary condition to proceed.

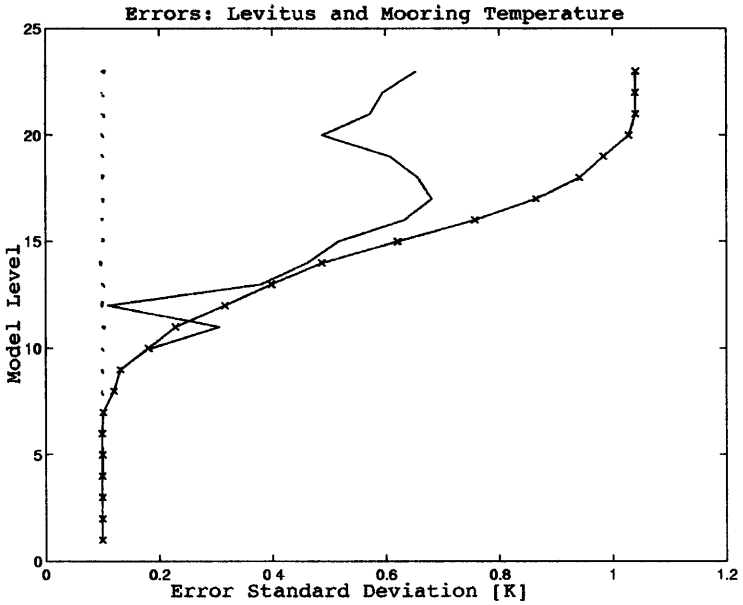

Figure 2-4: The consistency of the Subduction Experiment mooring temperatures with the Levitus climatology. The solid line with "X"'s is the prior error estimate in the Levitus temperature climatology as a function of model level (level 23 is the surface, and level 1 is the deepest level, 4900 meters). The solid line without "X"'s is the standard deviation of the difference between the Subduction Experiment moorings and the Levitus climatology. This line is generally to the left of the Levitus error estimate, which is a statement of the statistical consistency of the dataset and the climatology. Mooring data is only used in the upper ocean, levels 10-23.

2.3 Initial and surface controls

Control parameters are boundary conditions, forcing, or model parameters which are varied to control the trajectory of the solution. The term is borrowed from control theory, and is sometimes shortened to *controls*. The choice of control parameters is entirely up to the investigator. However, good controls share certain qualities. They are parameters which are somewhat unknown. Also, the controls should be identifiable as a major source of uncertainty in the model trajectory. A model is said to be *controllable* if changes in one or all of the control variables is capable of driving the model to any point in the permissible phase space (Dahleh and Diaz-Bobillo 1999). In an ocean model, there are many unknown parameters and forcing fields, and they are likely capable of controlling much of the model solution, although this has rarely been quantified (Fukumori et al. 1993). For the Subduction Experiment model, we have chosen the following control parameters:

- Initial Temperature and Salinity
- Surface Heat Flux and Freshwater Flux
- Meridional and Zonal Wind Stress
- Open Boundary Temperature and Salinity
- Open Boundary Normal and Tangential Velocity

There are 5,493,537 control variables.

2.3.1 Initial conditions

A properly-posed model integration requires the specification of the entire initial state. The initial state is relatively unknown and yet makes a huge impact on the model results over a one year time period. In our case, the initial density field, comprised of temperature and salinity fields, has a dominant effect on the early stages of the model integration and its elements will be chosen as control variables. The initial velocity

field is not explicitly controlled, but comes into equilibrium with the initial density in a few days through geostrophic adjustment. A second reason to adjust the initial density field is our relative lack of knowledge of it. The Levitus climatology could be used for the initial density field, but it does not include any effects of the mesoscale eddy field or the interannual variability. A better initial density field is the ECCO (Estimating the Climate and Circulation of the Ocean) 2° resolution state estimate (Stammer et al. 2002). We will use this improved global field, then improve the initial conditions once again with the regional model.

To keep the adjustments to the initial conditions within a physically reasonable range, we will add penalty terms to the cost function:

$$(\mathbf{T}_0 - \mathbf{T}_{0_{ECCO}})^T \ \mathbf{W}^*_{T_0} \ (\mathbf{T}_0 - \mathbf{T}_{0_{ECCO}}) \tag{2.10}$$

$$+ \quad (\mathbf{S}_0 - \mathbf{S}_{0_{ECCO}})^T \ \mathbf{W}^*_{S_0} \ (\mathbf{S}_0 - \mathbf{S}_{0_{ECCO}}) \tag{2.11}$$

where $\mathbf{T}_0$ and $\mathbf{S}_0$ are the initial model temperature and salinity, $\mathbf{T}_{0_{ECCO}}$ and $\mathbf{S}_{0_{ECCO}}$ are the ECCO 2° state estimate for temperature and salinity interpolated onto 1/6° for the same time, $\mathbf{W}^*_{T_0}$ and $\mathbf{W}^*_{S_0}$ are weighting matrices with nondiagonality marked by a star, "$*$".

The ECCO state estimate does not have a formal error estimate, but it is undoubtedly a better estimate of the initial conditions than the Levitus climatology. For this study, a conservative assumption is that the uncertainty is equal to that of the Levitus climatology. Therefore, the diagonal elements of $\mathbf{W}^*_{T_0}$ are identical to $\mathbf{W}_{Levt}$. The nondiagonal elements of this matrix are outlined below. A correlation length scale of 200 kilometers, used here, is a conservative choice relative to the peak of atmospheric energy in longer wavelengths ($\sim 1,000\ km$) (Peixoto and Oort 1992; Kalnay and coauthors 1996). However, recent scatterometer measurements (Chelton et al. 2001) show small-scale shifts in the winds over the Pacific cold tongue, so the correlation lengthscale may indeed be quite small in select regions over the open ocean. Further thought is necessary to provide more accurate atmospheric statistics. Isotropy is a good assumption in this

region away from boundary currents. The weights on the initial conditions therefore allow the addition of a mesoscale eddy field with the proper lengthscales.

Nondiagonal weighting matrices

Noisy control adjustments lead this study to implement nondiagonal weighting matrices. The controls have nondiagonal weight matrices here, because small-scale, unphysical features which represent model error should be repressed. Nondiagonal weights penalize noisy features because they require the fields to spatially covary. Small-scale structures in the control parameters are thereby eliminated.

Theoretically, the best nondiagonal matrix is the inverse of the error covariance matrix (Lorenc 1986). Unfortunately, the off-diagonal elements of the matrix are very poorly known *a priori*. Also, inversion of such a large matrix is not computationally feasible. Instead, we follow an approximate approach which follows the discussion in Lea (2001, Ph. D. thesis, p. 114) and Bennett (2002). For a vector $\mathbf{u}$ made of a two-dimensional scalar field, they showed

$$\mathbf{u}^T\ \mathbf{W}_0\ \mathbf{u} + (\nabla^2\mathbf{u})^T\ \mathbf{W}_1\ (\nabla^2\mathbf{u}) \approx \mathbf{u}^T\ \mathbf{B}^{-1}\ \mathbf{u} \tag{2.12}$$

where $\mathbf{W}_0$ and $\mathbf{W}_1$ are diagonal matrices, but $\mathbf{B}^{-1}$ is a nondiagonal matrix. For properly chosen diagonals in $\mathbf{W}_0$ and $\mathbf{W}_1$, $\mathbf{B}^{-1}$ can be made such that $\mathbf{B}$ is nearly a Gaussian covariance matrix,

$$\mathbf{B}(\mathbf{r}_1, \mathbf{r}_2) \approx Var(x_1, y_1)\ exp(-\frac{1}{2}\frac{|\mathbf{r}_1 - \mathbf{r}_2|^2}{l^2}), \tag{2.13}$$

which represents the covariance between points $\mathbf{r}_1 = (x_1, y_1)$ and $\mathbf{r}_2 = (x_2, y_2)$. The correlation lengthscale for the Gaussian covariance is 200 km for all the control parameters because of the large characteristic scales of the atmosphere. In summary, the addition of a smoothness constraint of the form of Equation (2.12) mimics a nondiagonal weighting matrix with a chosen Gaussian correlation lengthscale.

2.3.2 Surface forcing fields

Wind stress, heat flux, and freshwater flux are the driving forces of the ocean circulation. The first guess for the controls is the daily and twice-daily NCEP Reanalysis fields (Kalnay and coauthors 1996). The individual control adjustments are perturbations applied to the NCEP Reanalysis over a 10-day period.

The penalty for adjusting the surface forcing controls is added to the cost function Equation (2.1h)-(2.1k):

$$\sum_t^{37} (\overline{\tau}_x - \overline{\tau}_{x_{ncep}})^T \mathbf{W}^*_{\tau_x} (\overline{\tau}_x - \overline{\tau}_{x_{ncep}}) \tag{2.14}$$

$$+ \quad \sum_t^{37} (\overline{\tau}_y - \overline{\tau}_{y_{ncep}})^T \mathbf{W}^*_{\tau_y} (\overline{\tau}_y - \overline{\tau}_{y_{ncep}}) \tag{2.15}$$

$$+ \quad \sum_t^{37} (\overline{\mathbf{H}}_Q - \overline{\mathbf{H}}_{Q_{ncep}})^T \mathbf{W}^*_{H_Q} (\overline{\mathbf{H}}_Q - \overline{\mathbf{H}}_{Q_{ncep}}) \tag{2.16}$$

$$+ \quad \sum_t^{37} (\overline{\mathbf{H}}_F - \overline{\mathbf{H}}_{F_{ncep}})^T \mathbf{W}^*_{H_F} (\overline{\mathbf{H}}_F - \overline{\mathbf{H}}_{F_{ncep}}) \tag{2.17}$$

where τ_x and τ_y are the zonal and meridional windstresses, $\mathbf{H}_Q$ and $\mathbf{H}_F$ are heat fluxes and freshwater fluxes, $\tau_{x_{ncep}}$, $\tau_{y_{ncep}}$, $\mathbf{H}_{Q_{ncep}}$ and $\mathbf{H}_{F_{ncep}}$ are the respective NCEP Reanalysis fields, and $\mathbf{W}^*$ represents nondiagonal weighting matrices for each variable type.

There is a lack of information about the daily wind stress, heat flux, and freshwater flux over the open ocean. A simple comparison of different wind products reveals strong biases and systematic errors of 35–50% in the Subduction Experiment region (Moyer and Weller 1995). Therefore, the controls are allowed to change by the variance of the NCEP fields. The weighting matrices reflect this choice and vary spatially. The nondiagonal elements of the weighting matrices are handled as discussed in Section 2.3.1.

2.4 Open boundary control and estimation

A regional ocean simulation can only be completed with an additional source of information: the open boundaries. The open boundary conditions fundamentally influence the interior solution of the model. Simple changes in boundary conditions cause large differ-

ences in the interior circulation; for example, slip and no-slip conditions completely affect the circulation of an ideal gyre [(Pedlosky 1996), p.76; Adcroft and Marshall (1998); G. Ierley and W. Young, personal communication]. The open boundary state can control the circulation to a greater extent. In addition, the proper open boundary conditions are very uncertain. Unlike temperature and salinity, no climatology of open-ocean velocities exists. Open boundary conditions are ideal control variables; they influence the model profoundly yet they are relatively unknown.

Open boundaries control the solution of a regional model to a great extent, as will be further shown in Section 3.5.3. Because the open boundaries affect the interior of an ocean model, observations in the interior conversely convey some information about the correct open boundaries. In principle, this allows an investigator to estimate open boundary conditions which are realistic, not just boundary conditions which yield a realistic interior. In this thesis, the goal will be both control of the interior through the open boundaries, and estimation of realistic open boundary conditions.

Review of open boundary estimation

A review of the oceanographic literature finds no universally-accepted method for control or estimation of open boundary conditions with a primitive equation model. Almost all previous studies have used simplified versions of the equations of motions to study open boundaries (Chareney et al. 1950; Robinson and Haidvogel 1980; Bennett and Kloeden 1981; Gunson and Malanotte-Rizzoli 1996a,b). With the quasi-geostrophic equations, for instance, open boundary conditions were successfully nudged toward desired results (Malanotte-Rizzoli and Holland 1986). Nudging is undesirable for the present research because it is dynamically inconsistent with the physics of the ocean and it also comprehensively removes a whole range of the wavenumber spectrum. Soon thereafter, Schröter et al. (1993) used an artificial recirculation zone surrounded by walls to simulate and control open boundaries. Seiler (1993) estimated open boundary conditions with a quasi-geostrophic ocean box model and its complementary adjoint model[6]. The

[6]Adjoint models are detailed in Section 3.2.

technical experience gleaned from the simplified equations of motion was applied to a primitive equation model only recently (Zhang and Marotzke 1999; Ferron and Marotzke 2003).

Two major difficulties have confronted previous attempts to control and estimate open boundary conditions in a primitive equation model. One, estimated open boundary conditions frequently are not physically reasonable. Zhang and Marotzke (1999) took a first look at this problem. Two, open boundary estimation is often very inefficient when many other control variables are present. Ferron and Marotzke (2002) resorted to a process that separately estimated open boundary conditions after other control variables had been optimized. In the next sections, this thesis offers two novel approaches to remedy the problems first seen by previous investigators.

Physical constraints on the open boundaries

Reasonable open boundary conditions have a few general characteristics: interior-boundary consistency, geostrophic balance, and nearly vanishing net volume flux. Open boundary estimation is formulated here with many additional constraints, which leads to an extension of the technique devised by Zhang and Marotzke (1999). A *hard constraint* is an equation that must be satisfied exactly; the model equation (Equation (3.2)) represents the collection of all hard constraints. A *soft constraint* is an equation that need not be satisfied exactly, but its inequality is penalized in the cost function. Therefore, soft constraints are satisfied with an arbitrary precision determined by their weight.

Open boundary control with the primitive equations

The boundary conditions in the GCM require the complete specification of the state: temperature, salinity, meridional and zonal velocity (see Appendix A). The first-guess boundary conditions are from the ECCO 2° state estimate. There are very few choices for a time-varying open ocean velocity field to be used for this purpose. The ECCO estimate is interpolated up to 1/6° and varies monthly. Likewise, we will allow the adjustments to the boundary conditions to occur monthly; hence there are 12 sets of

adjustments for one year.

The penalty for adjusting the open boundary conditions is:

$$\sum_t^{12}(\mathbf{T}_{o\,b} - \mathbf{T}_{o\,b_{ECCO}})^T\ \mathbf{W}_{T_{o\,b}}\ (\mathbf{T}_{o.b.} - \mathbf{T}_{o.b._{ECCO}}) \tag{2.18}$$

$$+\quad \sum_t^{12}(\mathbf{S}_{o\,b.} - \mathbf{S}_{o.b._{ECCO}})^T\ \mathbf{W}_{S_{o.b}}\ (\mathbf{S}_{o.b} - \mathbf{S}_{o.b._{ECCO}}) \tag{2.19}$$

$$+\quad \sum_t^{12}(\overline{\mathbf{U}}^z_{o.b.} - \overline{\mathbf{U}}^z_{o.b\ ECCO})^T\ \mathbf{W}_{V_{bt}}\ (\overline{\mathbf{U}}^z_{o.b.} - \overline{\mathbf{U}}^z_{o.b._{ECCO}}) \tag{2.20}$$

$$+\quad \sum_t^{12}(\mathbf{U}'_{o.b.} - \mathbf{U}'_{o.b\ ECCO})^T\ \mathbf{W}_{V_{bc}}\ (\mathbf{U}'_{o.b} - \mathbf{U}'_{o.b._{ECCO}}) \tag{2.21}$$

$$+\quad \sum_t^{12}(\overline{\mathbf{V}}^z_{o.b.} - \overline{\mathbf{V}}^z_{o\,b\ ECCO})^T\ \mathbf{W}_{V_{bt}}\ (\overline{\mathbf{V}}^z_{o.b} - \overline{\mathbf{V}}^z_{o.b\ ECCO}) \tag{2.22}$$

$$+\quad \sum_t^{12}(\mathbf{V}'_{o.b} - \mathbf{V}'_{o.b._{ECCO}})^T\ \mathbf{W}_{V_{bc}}\ (\mathbf{V}'_{o\,b.} - \mathbf{V}'_{o.b._{ECCO}}) \tag{2.23}$$

$$+\quad \sum_t^{12}(\tfrac{\partial \mathbf{V}'_\perp}{\partial z} + \tfrac{g}{\rho_o f}\tfrac{\partial \boldsymbol{\rho}}{\partial l})^T\ \mathbf{W}_{ageos}\ (\tfrac{\partial \mathbf{V}'_\perp}{\partial z} + \tfrac{g}{\rho_o f}(\tfrac{\partial \boldsymbol{\rho}}{\partial l}) \tag{2.24}$$

$$+\quad \sum_t^{12}(\overline{\mathbf{V}}^z_\perp\ {}^T\mathbf{A}_{lz})^T\ \mathbf{W}_{volflux}\ (\overline{\mathbf{V}}^z_\perp\ {}^T\mathbf{A}_{lz}) \tag{2.25}$$

where *ECCO* refers to the ECCO state estimate, $\mathbf{T}_{o.b.}$ and $\mathbf{S}_{o.b}$ are open boundary temperature and salinity, $\overline{\mathbf{U}}^z_{o\,b.}$ and $\overline{\mathbf{V}}^z_{o.b}$ are depth-averaged or "barotropic" boundary velocity, $\mathbf{U}'_{o.b}$ and $\mathbf{V}'_{o\,b.}$ are the "baroclinic" velocity, $\mathbf{V}_\perp$ is the open boundary normal velocity, $\partial\boldsymbol{\rho}/\partial l$ is the gradient of density along the boundary, $\mathbf{A}_{lz}$ is a vector of the area of the open boundary grid-cell faces, and $\mathbf{W}$ refers to various diagonal weighting matrices.

The weighting matrices serve different purposes for the various terms of the cost function. For terms (2.18)-(2.19), we are using the ECCO state estimate as a first guess. Similar to the rationale in Section 2.3.1, the open boundary temperature and salinity will be given the same uncertainty as the Levitus fields. This is because the coarse resolution ECCO boundary conditions once again do not include a mesoscale eddy field. This is a conservative estimate of uncertainty because the ECCO state estimate was computed for our particular year of interest, 1992-93, unlike the Levitus climatology. On the other hand, very little is known about the uncertainty in open boundary velocities. Instead of pleading complete ignorance, the weights in terms (2.20)-(2.23) constrain the velocities to have an appropriate magnitude. The weights are split into barotropic and baroclinic components because they obey different dynamics, and they need to

be controlled separately. Term (2.24) is a penalty for open boundary velocities which deviate from thermal wind balance (see Section 2.4.1). Finally, the net volume flux (term 2.25) into the domain is expected to be nearly balanced (see Section 2.4.2).

2.4.1 Thermal wind balance

The velocity field is strongly coupled to the density field, and a reasonable estimate should reflect this fact. The eastern subtropical gyre has a Rossby number of approximately 0.1 and therefore the coupling is primarily explained by geostrophic balance. Together with hydrostatic balance and the Boussinesq approximation, the thermal wind equations state that the vertical velocity shear depends on horizontal density gradients (Pond and Pickard 1983):

$$\frac{\partial u}{\partial z} = \frac{g}{\rho_0 f}\frac{\partial \rho}{\partial y}, \quad \frac{\partial v}{\partial z} = -\frac{g}{\rho_0 f}\frac{\partial \rho}{\partial x} \tag{2.26}$$

where u is velocity in the x direction, v is velocity in the y direction, g is gravity, f is the Coriolis parameter, and ρ_0 is a reference density. In the interior, the coupling is explicitly calculated by the general circulation model. On the open boundary, the ocean state is prescribed and does not necessarily follow the thermal wind equations. Unbalanced open boundary conditions create spurious gravity waves which cause deterioration in the boundary conditions' ability to control the model interior in a believable way. The estimation and control of open boundary conditions demand thermal wind balance.

Stevens's method: a hard constraint

The ocean state on the open boundaries can be kept in geostrophic balance by modifying the model equations. Stevens (1991) solved for the baroclinic normal velocity on the boundary by linearizing the momentum equation of a primitive equation model. The linearized momentum equation reduced to thermal wind balance to first order. To restate, only temperature and salinity were prescribed on the boundary and the baroclinic velocity was then diagnosed. The depth-integrated, or "barotropic", velocity is an

extra variable to be prescribed. Therefore, the open boundary normal velocity, $v_\perp$, is calculated

$$v_\perp(z) = \frac{g}{\rho_0 f} \int_H^z (\frac{\partial \rho}{\partial l})\, dz + v_0 + v_{bt} \tag{2.27}$$

where H is ocean depth, l is distance along the boundary, v_0 is an integration constant, and v_{bt} is the barotropic velocity. The integration constant is consistent with the definition of the barotropic velocity as the depth-weighted average velocity:

$$v_{bt} = \frac{1}{H} \int_H^0 v_\perp(z)\, dz. \tag{2.28}$$

Two problems exist with this method. First, thermal wind balance should only hold to the extent that geostrophic balance holds. The Rossby number for the eastern subtropical gyre is 0.1, which means that the ageostrophic current is roughly 10% of the geostrophic current. Furthermore, the mixed-layer and fronts have significantly larger Rossby numbers and stronger ageostrophic currents. The open boundary velocity should not exactly follow the geostrophic relation or else any information about the ageostrophic flow will be lost. Second, the calculation of Equation (2.27) is noisy due to the horizontal gradient. Zhang and Marotzke (1999) showed that practical implementation is frequently corrupted by noise. Based on these results, another method to constrain the open boundaries to thermal wind balance is sought.

Soft constraint method

The cost function can serve a dual purpose; not only can it constrain the model to observations, it can penalize the model's deviation from dynamical balance. A soft constraint (see Section 2.4) is ideal for thermal wind balance on physical grounds because it should not be satisfied perfectly. The extra term in the cost function is:

$$\sum_t^{12} (\frac{\partial \mathbf{V'}_\perp}{\partial z} + \frac{g}{\rho_0 f} \frac{\partial \boldsymbol{\rho}}{\partial l})^T\, \mathbf{W}_{ageos}\, (\frac{\partial \mathbf{V'}_\perp}{\partial z} + \frac{g}{\rho_0 f} \frac{\partial \boldsymbol{\rho}}{\partial l}) \tag{2.29}$$

where the cost function is summed over 12 months, $\mathbf{V}'_{\perp}$ is a vector of the monthly-averaged, open boundary normal baroclinic velocity, $\partial\boldsymbol{\rho}/\partial l$ is a vector of the gradient of density along the boundary, and $\mathbf{W}_{ageos}$ is a diagonal weighting matrix. The weights are appropriate for a Rossby number of 0.1 below 100 meters depth, and are zero for anything above 100 meters depth. Therefore, 10% of the magnitude of the velocity is the expected error. In practice, the model easily conforms to this soft constraint because the control variables completely control the size of this term. The use of soft constraints reveals the power of the least-squares problem; our formulation here is easy to apply to the previously-existing machinery and works well.

2.4.2 Estimating net volume flux

A convenient assumption is that the net mass flux into a region is perfectly zero, but observations from tide gauges (Wunsch and Gill 1976) and the TOPEX/POSEIDON altimeter (Stammer et al. 2000; Fu et al. 2001) do not always support this statement. Wunsch and Gill (1976) showed large mass flux convergences in the tropical Pacific tide gauge network. The TOPEX/POSEIDON altimeter mission showed surprisingly strong barotropic motions at high latitudes with timescales of 1-10 days (Stammer et al. 2000). The sea surface height variations due to these motions imply rapid, large-scale, depth-integrated movements of water. Recently, a 25-day period, large-scale oscillation was detected in the Argentine Basin (Fu et al. 2001). The wave could be explained by a basin mode with a depth-integrated transport of 50 Sv. These observations all suggest that there are timescales over which the net mass flux into a region of the ocean is nonzero. Ideally, the domain-wide mass flux convergence would be an estimated quantity from this thesis.

The distribution and movement of mass in the ocean is not understood fully. This is illustrated by Munk's (2003) assertion that global sea level rise can not be properly attributed to either eustatic or steric effects. Recent measurements of the global sea level trend (Munk 2002; Cazenave 2002) must be due to melting of land-bound ice (eustatic effect) or due to the expansion of warmed seawater (steric effect), but our best

estimates today are not capable of closing the budget. In a regional model, sea surface height observations are affected primarily by two analogous effects: the heat content of a column of water and the net influx of mass into the domain. If the net mass flux into a region were fixed to zero, the information content of the sea surface observations could be diminished or misinterpreted. Such concerns are probably not warranted in the Subduction Experiment region, but it is still a good opportunity to prepare the techniques for use in other regions.

In a Boussinesq model such as the MIT GCM, conservation of mass is exchanged for conservation of volume because of an inconsistency between the equation of state and the statement of nondivergent flow (Adcroft (1994), p.22). Ideally, the net volume flux into a region is not fixed to zero,

$$\oint_{bdy} \overline{\mathbf{V}}^z_\perp \; H(l) \; dl \neq 0, \tag{2.30}$$

but large imbalances are not allowed either. In discrete space and time, an imbalanced, i.e. nonzero, volume flux can be penalized by a soft constraint in the cost function:

$$\sum_t^{12} (\overline{\mathbf{V}}^z_\perp \; ^T \mathbf{A}_{lz})^T \; \mathbf{W}_{volflux} \; (\overline{\mathbf{V}}^z_\perp \; ^T \mathbf{A}_{lz}) \tag{2.31}$$

where $\overline{\mathbf{V}}^z_\perp$ is a vector of the depth-integrated velocity normal to the boundary, $\mathbf{A}_{lz}$ is a vector of the corresponding open boundary cross-sectional area, and $\mathbf{W}_{volflux}$ is a scalar weight. The weight is determined by physical reasoning; a 50 Sv imbalance like that reported by Fu et al. (2001) in the Argentine Basin could be considered an upper limit on volume imbalance. In that case, $\mathbf{W}_{volflux} = 1/(50\ Sv)^2$. Although 50 Sv seems like a very large number, this amounts to only a 3 mm/s horizontal inflow around the domain of the model. The addition of a soft constraint is a necessary step for any volume flux convergence estimates.

Ill-conditioning of the volume flux estimation problem

Estimating volume flux is difficult even with a linear system because of the physical processes involved and the associated mathematical ill-conditioning. A toy channel model with only two control parameters already displays the ill-conditioning. Consider a steady, rotating, zonal channel with constant inflow and outflow (Figure 2-5). The mean sea surface height trend in the channel and the meridional sea surface slope in the center of the channel are observed; these two quantities could be derived from TOPEX/POSEIDON satellite altimetry fields. The goal is to estimate the inflow and outflow of water into the channel. An imbalance of inflow and outflow makes a mean sea surface height trend due to the conservation of volume:

$$\frac{d\overline{\eta}}{dt} = \frac{A_{xy}}{A_{yz}}(u_{out} - u_{in}) \tag{2.32}$$

where A_{xy} is the sea surface area, A_{yz} is the cross-sectional area of the channel, and u is velocity in the zonal direction. The meridional sea surface slope is also observed; it is related to the channel velocity by geostrophic and hydrostatic balance:

$$\frac{d\eta}{dy} = \frac{f}{g}\overline{u} = \frac{f}{g}\frac{(u_{out} + u_{in})}{2}. \tag{2.33}$$

In this example, the problem is linear. In matrix form, the problem is restated:

$$\begin{pmatrix} -\frac{A_{xy}}{A_{yz}} & \frac{A_{xy}}{A_{yz}} \\ \frac{f}{2g} & \frac{f}{2g} \end{pmatrix} \begin{pmatrix} u_{in} \\ u_{out} \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} = \begin{pmatrix} \frac{d\overline{\eta}}{dt} \\ \frac{d\eta}{dy} \end{pmatrix}. \tag{2.34}$$

Knowledge of the right hand side can be used to invert for the flow field. However, this matrix is ill-conditioned in most oceanographic applications because of the values of the physical constants. For the Subduction Experiment model, $\frac{A_{xy}}{A_{yz}}$ is roughly 1000, and $\frac{f}{2g}$ is approximately 5 x 10^{-6} m^{-1}. Inversion of the matrix will lead to large errors because it is nearly singular (Strang 1996). A common strategy to better condition the

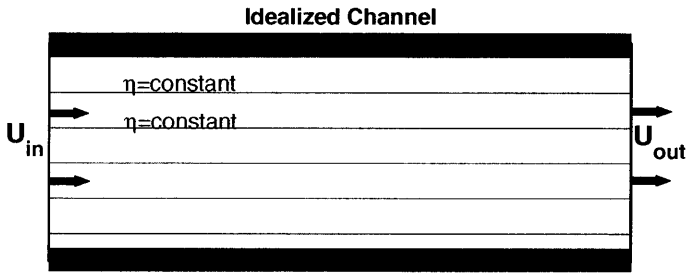

Figure 2-5: Schematic of idealized channel. Uniform and constant velocity enters and leaves the channel, which leads to a zonally uniform sea surface height under geostrophy. Any difference between the volume influx and outflux makes the mean sea surface height change with time.

matrix is row scaling, as discussed by Wunsch (1996, p. 121); however, the rows in this problem have already been scaled by the observational accuracy, n_1 and n_2, which is nearly equal in both rows. A second approach is column scaling; this recognizes that there is information in the expected solution covariance, $\mathbf{R}_{xx}$. The solution must reflect that the inflow and outflow are negatively correlated to conserve volume. Rescaling and rotating the input and output velocities,

$$\begin{pmatrix} u_{in}{}' \\ u_{out}{}' \end{pmatrix} = \mathbf{R}_{xx}^{-T/2} \begin{pmatrix} u_{in} \\ u_{out} \end{pmatrix}, \tag{2.35}$$

makes Equation (2.34) well-conditioned and easily invertible. Column scaling makes explicit the expectation that the difference between inflow and outflow is small.

Application to the general circulation model

In the general circulation model, ill-conditioning of the optimization is eliminated by nondimensionalization of the open boundary velocity controls, which is equivalent to the column scaling method above. For the GCM, nondimensionalization is numerically implemented term-by-term, which is analogous to a diagonal $\mathbf{R}_{xx}$, because large matrix multiplication is not possible. Unfortunately, a diagonal matrix does not resolve the ill-conditioning, because of the strong covariance between inflow and outflow. To resolve the

problem, the control parameters are re-chosen; this amounts to a rotation and rescaling of the controls. Originally, the barotropic normal velocities around the domain was chosen as the control parameters. Instead, one may add the domain-averaged imbalanced velocity as a control parameter itself. Then, the normal velocity has three components:

$$\overline{\mathbf{V}}_{\perp}^{z}(x,z) = V_{ECCO}(x,z) + V_{bt}(x) + V_{imbalance} \tag{2.36}$$

where V_{ECCO} is the first-guess barotropic velocity from the ECCO global state estimate, V_{bt} is the barotropic control adjustment, and $V_{imbalance}$ is another barotropic control adjustment which is evenly applied to all boundary points. In this particular form, the controls do not specify a unique open boundary velocity field because $V_{imbalance}$ can compensate for changes in V_{bt}. For uniqueness, a hard constraint[7] is added to the original barotropic control adjustments:

$$\oint_{bdy} V_{bt}(l)\ H(l)\ dl = 0, \tag{2.37}$$

the original barotropic adjustments are the domain-balanced part of the total barotropic adjustments. With this formulation of the problem, the net volume flux is estimated without a problem in the general circulation model.

In many of the early results of this thesis, the general circulation model is run with a hard constraint on the net volume flux. The constraint of zero net volume flux is appended to the model equations (Equation (3.2)). For the Subduction Experiment region, the estimated volume flux into the basin is nearly zero anyway, so the early results with a hard constraint are not significantly altered from later results.

2.5 Chapter summary

The observations of the Subduction Experiment do not provide enough coverage to adequately form budgets and analyze dynamical balances of the mesoscale ocean cir-

[7]The actual implementation is a discrete sum, but the meaning is more easily seen in the continuous formulation.

culation. Here, a combination of the observations with a state-of-the-art, 1/6° ocean general circulation model provides an estimate that has sufficient resolution in time and space. The concept is to find a model trajectory that fits the observations within their uncertainties. The cost function unambiguously describes the "goodness" of a particular model trajectory; it is the squared misfit between the model and the Subduction Experiment moorings and the TOPEX/POSEIDON satellite altimeter (as well as many other terms). The model trajectory is controlled by varying uncertain model parameters: the initial conditions, the surface forcing, and the open boundary conditions. Despite the high complexity, the combination of a model and observations here is just a large least-squares problem.

Chapter 3

Eddy-Resolving State Estimation

3.1 Overview of chapter

The search for an eddy-resolving model trajectory that fits observations is a challenge due to the nonlinear nature of the model itself. The method of Lagrange multipliers (Section 3.2) uses the gradient of the cost function to search for a model trajectory within the uncertainty of observations, but will the gradients derived from a nonlinear, eddy-resolving ocean model be useful? Nonlinear models potentially produce multiple stationary points in the cost function, and gradient-search methods may have difficulty in finding a solution to the least-squares problem. For example, optimization studies with geostrophic turbulence models (Tanguay et al. 1995) and basin-wide ocean models (Lea et al. 2000; Köhl and Willebrand 2003) converged to local minima that were not the true solution. In addition, ocean models have thresholds and switches which are further examples of nonlinearity. Local gradients do not give any information about thresholds, and may miss important features of the dynamics.

Despite these concerns, the intrinsic dynamics of the realistic eastern subtropical gyre model used here are more linear than the extreme models of previous studies that gave problematic results. A large supply of data (as shown on Page 43) and an excellent first guess of the controls from a coarse resolution model promise to help the search for a viable state estimate here. Under these conditions, the gradients of the eddy-resolving

primitive equation ocean model do help find a consistent solution between model and data for the eastern subtropical gyre. In the process of finding the model estimate, the dynamical behavior of the eddy-resolving model is quantified, with implications for predictability of the ocean. The final product of this chapter is an eddy-resolving state estimate to be used in Chapter 4 for the study of subduction.

3.2 Method of Lagrange multipliers

The method of Lagrange multipliers solves a constrained least-squares problem and is shown to be a logical choice for ocean state estimation. Although the term *Lagrange multiplier* is familiar to physicists, the method has been called many other names, most notably *the adjoint method* (Hall et al. 1982; Thacker and Long 1988; Tziperman and Thacker 1989), *the Pontryagin Principle* (Wunsch 1996), and *4D-Var* (LeDimet and Talagrand 1986; Talagrand 1997). The method is well-suited for oceanographic datasets where all the measurements have been collected and compiled. Then, the data can be used all at once – a *whole domain* approach (Figure 3-1). The method of Lagrange multipliers also saves computation; large covariance matrices are not calculated. Another feature is the utility of intermediate results; sensitivity information is a by-product of the optimization problem. The method of Lagrange multipliers is therefore an attractive choice for solving the ocean state estimation problem.

The method is potentially limited by strong nonlinearity in the model, the lack of uncertainty information, and the difficulty of hand-coding an adjoint model. Here, the goal is to extend the method to nonlinear systems. The lack of uncertainty information has been remedied in small-dimensional systems by use of the *Hessian matrix* (Thacker 1989). In addition, the adjoint of the MIT GCM is obtained with relative ease through an adjoint translator (Giering and Kaminski 1998). In hindsight, the traditional limitations of the method of Lagrange multipliers do not deter the investigation here; in fact, some of the drawbacks serve as motivation.

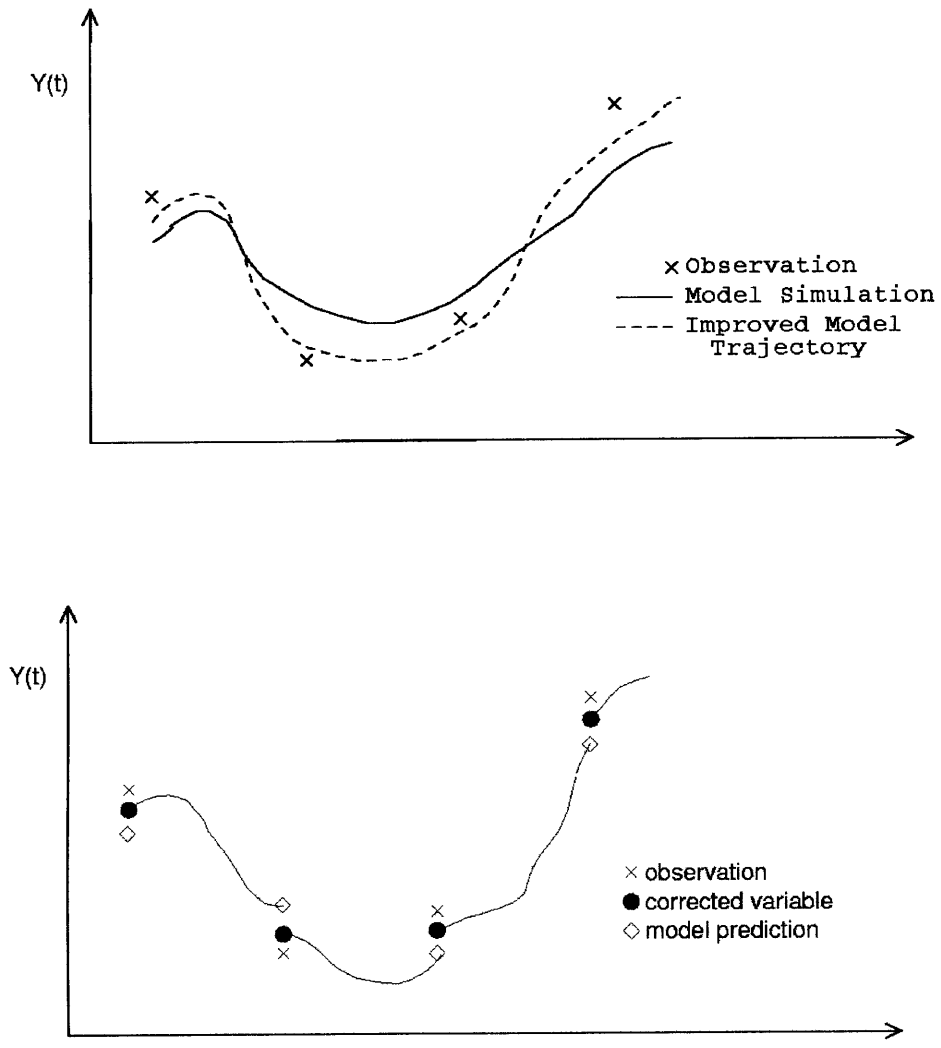

Figure 3-1: Pictorial view of two state estimation techniques. The method of Lagrange multipliers (*top*) is a whole-domain method used in this thesis. Whole-domain methods use observations over the entire time domain at once to fit the model. A more detailed picture of whole domain state estimation is given in Figure 2-1. In contrast, the Kalman Filter (*bottom*) is a sequential method which uses observations in sequential steps and incorporates incoming data. The Kalman Filter/Smoother (*not pictured*) improves the Kalman Filter solution, yielding the same solution as the method of Lagrange multipliers in a linear system. The Kalman Filter/Smoother is both a sequential and whole-domain method. From Giering and Kaminski (1998).

3.2.1 Appending Lagrange multipliers

The method of Lagrange multipliers finds a least-squares solution subject to a constraint. Mathematically, the method works by appending extra terms to the cost function. The original, constrained optimization problem[1] is transformed into an unconstrained one where special structure inherent in the equations allows efficient solution techniques. For example, a generic and condensed cost function is minimized using Lagrange multipliers below.

The goal is restated:

$$\textit{minimize} \quad \tilde{J} = \sum_{t=0}^{t_f} \left[\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)\right]^T \mathbf{W}(t) \left[\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)\right] + \sum_{t=0}^{t_f-1} \mathbf{u}(t)^T \mathbf{Q}(t) \mathbf{u}(t) \tag{3.1}$$

$$\textit{subject to the constraint} \qquad \mathbf{x}(t+1) = \mathcal{L}[\mathbf{x}(t),\ \mathbf{Bq}(t),\ \mathbf{\Gamma u}(t)] \tag{3.2}$$

where $\mathbf{x}(t)$ is the state vector of temperature, salinity, and velocity,
$\mathbf{y}(t)$ is the observations and $\mathbf{E}(t)\mathbf{x}(t)$ is the model estimate of those observations,
$\mathbf{u}(t)$ is the control vector of external forcing and boundary conditions,
$\mathbf{\Gamma u}(t)$ is the effect of control adjustments and model error on the model trajectory,
$\mathbf{Bq}(t)$ is the known forcing,
$\mathcal{L}$ represents the nonlinear model operator,
and $\mathbf{W}(t)$ and $\mathbf{Q}(t)$ are weighting matrices.
The time units have been nondimensionalized so that the timestep is one unit, $\Delta t = 1$.

The first term of the cost function is the squared misfit between model and observations. To relate this to Chapter 2, this generic term subsumes the first eight terms of the cost function, (2.1a-2.1g). The second term bounds the size of the control terms, which represent unknown boundary conditions, surface forcing errors, and model error. This term is a succinct way of writing terms (2.1h)-(2.1q) of the cost function. (Terms (2.1r)-(2.1s) have no analogue in the present example, but the mathematics would follow in a

[1]Optimization and minimization are used interchangeably. Optimization is a more general term encompassing both maximization and minimization problems.

similar way.) The constraint is the nonlinear model. For all minima of the cost function, $\partial\tilde{J}/\partial\mathbf{x}(t)$ and $\partial\tilde{J}/\partial\mathbf{u}(t)$ vanish. For a state of size M and controls of size N, $M+N$ equations need to be satisfied ($\partial\tilde{J}/\partial\mathbf{x}_i(t) = 0,\ 1 \leq i \leq M$, and $\partial\tilde{J}/\partial\mathbf{u}_k(t) = 0,\ 1 \leq k \leq N$). However, the state vector, $\mathbf{x}(t)$, directly depends on the control vector, $\mathbf{u}(t)$, by the model dynamics. Now, less than $M + N$ independent variables are available to satisfy the $M + N$ constraints for a minimum. This overdetermined system typically does not have a solution for arbitrary $\mathbf{x}(t)$ and $\mathbf{u}(t)$ because the model constraint is violated. Instead, the solution method should search for a stationary point while simultaneously satisfying the model constraint.

In the late 1700's, the Italian-French mathematician Lagrange suggested appending new terms to the cost function to solve the constrained minimization problem. Following his advice, the new function is

$$\begin{aligned} J = \textstyle\sum_{t=0}^{t_f} & \left[\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)\right]^T \mathbf{W}(t) \left[\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)\right] \\ & + \textstyle\sum_{t=0}^{t_f-1} \mathbf{u}(t)^T \mathbf{Q}(t)\mathbf{u}(t) \\ - \textstyle\sum_{t=0}^{t_f-1} & \boldsymbol{\mu}(t+1)^T \{\mathbf{x}(t+1) - \mathcal{L}[\mathbf{x}(t), \mathbf{B}\mathbf{q}(t), \boldsymbol{\Gamma}\mathbf{u}(t)]\} \end{aligned} \tag{3.3}$$

where $\boldsymbol{\mu}(t)$ is a vector of Lagrange multipliers. The number of Lagrange multipliers, M, is equal to the size of the state. For every state variable, there is a corresponding Lagrange multiplier. In this form, the appended cost function is sometimes called the *Lagrangian function*, in analogy to classical mechanics. The last term is always zero if the model constraint holds, so the numerical value of the appended cost function is the same as the original cost function. The Lagrange multiplier term is appended as a mathematical device so that all the variables, $\mathbf{x}(t)$, $\mathbf{u}(t)$, and now $\boldsymbol{\mu}(t)$, can be treated as independent variables (Strang 1996). This works because the Lagrange multipliers take values that make the partial derivatives ($\partial J/\partial\mathbf{x}_i(t),\ 1 \leq i \leq M$) vanish. The underlying mathematical machinery exploits the explicit relationship between the controls and state, as embodied in the forward model. If there are N controls, the original constrained minimization problem in the space of the state and the controls had dimension

$M+N$. The Lagrange multipliers reduce the problem to an unconstrained minimization problem of dimension N in the space of the controls alone.

After resolving the dependency between the controls and the state, all derivatives of the cost function all must independently equal zero for the constrained minimum. Taking these three sets of derivatives yields the three sets of normal equations (the analogue of the continuous-time *Euler-Lagrange equations* (LeDimet and Talagrand 1986)):

$$\frac{\partial J}{\partial \boldsymbol{\mu}(t)} = 0 \Rightarrow \qquad \mathbf{x}(t+1) = \mathcal{L}[\mathbf{x}(t),\ \mathbf{B}\mathbf{q}(t),\ \boldsymbol{\Gamma}\mathbf{u}(t)] \tag{3.4}$$

$$\frac{\partial J}{\partial \mathbf{x}(t)} = 0 \Rightarrow \quad \boldsymbol{\mu}(t) = \left(\frac{\partial \mathcal{L}}{\partial \mathbf{x}(t)}\right)^T \boldsymbol{\mu}(t+1) + \mathbf{E}(t)^T\ \mathbf{W}(t)\ [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t))] \tag{3.5}$$

$$\frac{\partial J}{\partial \mathbf{u}(t)} = 0 \Rightarrow \qquad \mathbf{u}(t) = -\mathbf{Q}(t)\left(\frac{\partial \mathcal{L}}{\partial(\boldsymbol{\Gamma}\mathbf{u})}\right)^T \boldsymbol{\Gamma}^T \boldsymbol{\mu}(t+1) \tag{3.6}$$

The first equation is the nonlinear model, the MIT GCM in this project. The second equation is the adjoint model. In this equation, the transpose of the tangent linear model (to be defined in Section 3.3.1) acts upon the Lagrange multiplier vector. The model-observation misfit, $\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)$, forces the adjoint model. The third equation relates the Lagrange multipliers and the controls. Recently, the study of the set of normal equations has been popularly called *adjoint modeling*. Considering all three sets of equations, there are $2M+N$ equations and $2M+N$ unknowns. Mathematically, this is a formally just-posed problem. In the case of linear constraints, solution is possible by matrix inversion – except for the large dimension of the problem. In any case, the method of Lagrange multipliers explicitly accounts for all constraints, and provides machinery to find a constrained minimum.

3.2.2 Solution method for the normal equations

For nonlinear constraints, the normal equations (3.4)-(3.6) are not directly solvable, but their special structure can be exploited. One procedure, used in this thesis, is:

- **1) Forward sweep**. Make a first guess of the controls, usually $\mathbf{u}^{(0)}(t) = 0$, and use the forward model (3.4) to get a first estimate of the state, $\mathbf{x}^{(0)}(t)$ (the

superscript (0) refers to the first-guess trajectory that evolves through time). If the misfit between the model and observations is within the expected error, the model trajectory is the state estimate. Otherwise, proceed to item 2.

- **2) Backward sweep.** $\mathbf{Ex}^{(0)}(t) - \mathbf{y}(t)$ can be evaluated and used to drive the adjoint model (3.5). Use $\boldsymbol{\mu}(t_f + 1) = 0$ as the initial conditions to the adjoint model because no observations are present after t_f. Integrate backwards in time (as detailed in Section 3.2.3 below).

- **3) Update controls.** Unless $\boldsymbol{\mu}(t) = 0$ for all times t, the third set of normal equations (3.6) will not be satisfied. $\boldsymbol{\mu}(t) = 0$ is not desirable, because then the model fits the observations exactly, which is not reasonable for observations with noise. Instead, use Equation (3.6) to give a new estimate, $\mathbf{u}^{(1)}(t)$ (to be explained in detail in Section 3.2.4). Return to item 1 and iterate the procedure.

3.2.3 Adjoint model integration

Step 2 above shows that the adjoint model can be integrated backwards in time when given the initial conditions, $\boldsymbol{\mu}(t_f + 1) = 0$. During the adjoint integration, the forward model trajectory is needed, but in reverse order. The transpose of the tangent linear model is linearized about the forward model state, as seen in (3.5). The time-evolving forward model state, however, is too large to be stored in memory at once. Checkpointing schemes are an efficient numerical tool for recalculating the forward model trajectory during an adjoint model run (Griewank and Walther 2000). At evenly-spaced checkpoint times, the forward model state is saved to disk for use in the adjoint model. In this way, neighboring forward model states can be recalculated with a short model run instead of the full model run from the initial time. Checkpointing works as a tradeoff that reduces memory requirements by adding computation. This technical advance from computer science makes the solution method of Section 3.2.2 possible.

3.2.4 Gradient descent

The third step above, "update controls", is not nearly as straightforward as previously presented; in fact, whole textbooks have been written on the subject of optimization theory (Luenberger 1984; Gill et al. 1986). To review, the problem here is analogous to navigating a mountain range and looking for the deepest hole (but in a many million-dimensional space!). The method of Lagrange multipliers calculates a gradient, $\partial J/\partial \mathbf{u}(t)$, to help search the control space. Is this a computationally efficient way to find a solution? The answer is apparent after comparing search methods which do not use gradients, and those that do.

Methods that do not use gradients, such as simulated annealing (Metropolis et al. 1953; Press et al. 1992 p. 443; Barth and Wunsch 1990) and the simplex method (Dantzig et al. 1955), have been used for many years with success. Genetic algorithms (Holland 1975; Davis 1991), another class of search methods, promise to improve the performance of non-gradient optimization methods, but they have rarely been tested in oceanographic applications (Barth 1992; Hernandez et al. 1995). How many forward model runs are necessary to find a solution with these non-gradient methods? In the region of a minimum in control space, the least-squares form of the cost function gives a quasi-parabolic topology,

$$J(\mathbf{u}) \approx \mathbf{u}^T\mathbf{B}^T\mathbf{B}\mathbf{u} - \mathbf{g}^T\mathbf{u} + c. \tag{3.7}$$

This assumption is proved in Section 3.3.1 with a linear model. The number of parameters that describe the shape of J is equal to the number of free parameters[2] in the matrix $\mathbf{B}$, the vector $\mathbf{g}$, and the scalar c, which is $N^2 + N + 1$ when N is the number of control variables. All of these parameters can change the location of the minimum. In a worst case scenario, $\mathcal{O}(N^2)$ pieces of information must be collected. This could be done by N^2 forward model integrations. For our case, it is impractical to run the forward model that many times.

[2]Precise accounting yields $(1/2)N(N+1)$ parameters. Because $\mathbf{B}^T\mathbf{B}$ is positive definite, $\mathbf{B}$ is an upper-triangular matrix by the Cholesky decomposition.

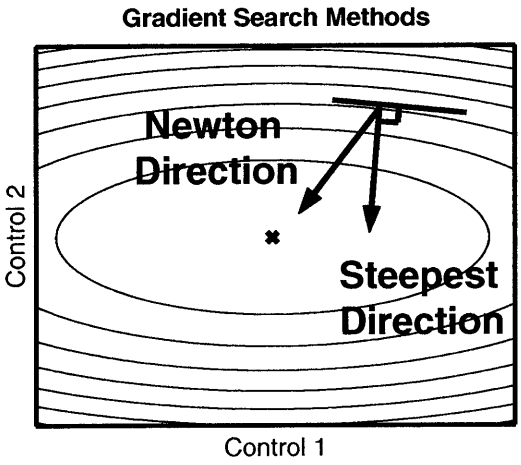

Figure 3-2: A schematic of a paraboloidal cost function topology with respect to two control directions in phase space. For an anisotropic paraboloid, contour lines of constant cost function trace an ellipse (*thin lines*, "J-isolines"). In this case, the direction perpendicular to the J-isolines (*Steepest Direction*) no longer points to the minimum. Using information from second derivatives, the direction to the minimum is calculated (*Newton Direction*).

Knowledge of the gradient increases the efficiency of a search algorithm and makes large-dimensional optimization possible. In contrast to the forward model, each integration of the adjoint model yields N independent pieces of information that help in the search for a minimum. The gradient of the cost function is a vector in N dimensions. As long as the adjoint model can be computed with less cost than N forward model integrations, the gradient gives a great amount of guidance in optimization, without an inordinate number of forward model integrations. In the case at hand, the adjoint of the MIT GCM calculates the gradient with a computational cost of six forward model integrations. In large-dimensional problems, calculation of the gradients from the adjoint model makes optimization possible.

A näive search would simply change the controls in the direction given by the gradient, but better gradient descent (or *direction set*) methods have been discovered. The method of steepest descent described by Press et al. (1992) "greedily" adjusts the controls in the direction of the gradient. This method is plagued by difficulties when "narrow valleys" are present; that is, when partial derivatives are very nearly zero in

some directions, but very large in others. The method always finds the local minimum, but it inefficiently searches in a zig-zag path near the bottom of a valley (see Press et al. (1992), Fig. 10.5.1, p. 407). Quasi-Newton methods[3] are superior because they take the second derivative, or curvature, into account. Suppose a cost function is well-approximated by Equation (3.7) and a first guess of the controls $\mathbf{u^{(0)}}$. With an evaluation of the cost function and gradient at $\mathbf{u}^{(0)}$, the underlying topology of the cost function is approximated by:

$$J(\mathbf{u}) = J(\mathbf{u}^{(0)}) + \nabla J(\mathbf{u}^{(0)})^T(\mathbf{u} - \mathbf{u}^{(0)}) + (\mathbf{u} - \mathbf{u}^{(0)})^T\mathbf{B}^T\mathbf{B}(\mathbf{u} - \mathbf{u}^{(0)}). \tag{3.8}$$

The gradient of Equation (3.8) gives

$$\nabla J(\mathbf{u}) = \nabla J(\mathbf{u}^{(0)}) + 2\mathbf{B}^T\mathbf{B}(\mathbf{u} - \mathbf{u}^{(0)}). \tag{3.9}$$

The local stationary point occurs where the gradient is zero. Therefore, the direction of the minimum is actually

$$(\mathbf{u}^{(MIN)} - \mathbf{u}^{(0)}) = -\frac{1}{2}(\mathbf{B}^T\mathbf{B})^{-1}\nabla J(\mathbf{u}^{(0)}) \tag{3.10}$$

where the steepest descent direction $\nabla J(\mathbf{u}^{(0)})$ is modified by $(\mathbf{B}^T\mathbf{B})^{-1}$ (Figure 3-2). This matrix is usually called the *Hessian*, $\mathbf{H} = (\mathbf{B}^T\mathbf{B})^{-1}$, and it contains second-derivative information. The variable storage quasi-Newton method of Gilbert and Lémarechal (1989) uses differences of the first derivatives to form an approximate Hessian. Hence, the storage of the Hessian is done without large use of computer memory. In summary, the variable storage quasi-Newton search accounts for many lessons learned in optimization theory, yet is computationally feasible for large problems. Gradient search using the method of Gilbert and Lémarechal (1989) is used in this thesis.

[3]Quasi-Newton methods are a type of *variable metric* optimization method which only approximates the Hessian matrix.

3.2.5 Interpretation of Lagrange multipliers

The Lagrange multipliers serve two completely different purposes; they are useful for optimization problems as shown above, but they also have a physical interpretation. The partial derivative of Equation (3.3) with respect to $\mathbf{u}$ gives:

$$\frac{\partial J}{\partial \mathbf{u}(t)} = \left(\frac{\partial \mathcal{L}}{\partial(\mathbf{\Gamma u})}\right)^T \mathbf{\Gamma}^T \boldsymbol{\mu}(t+1). \tag{3.11}$$

From this equation, the Lagrange multipliers give the gradient of the cost function with respect to the control variables. In an optimization context (like most of this thesis), J includes the data-model misfit, and hence, the Lagrange multipliers give the direction to change the controls in order to minimize J. This is the underlying principle behind the "update controls" step above. In this case, the Lagrange multipliers are directly related to the gradient that is used for optimization. Another fundamental equation relates the Lagrange multipliers to the gradient of the cost function with respect to the state (see Appendix B).

In addition to optimization applications, Lagrange multipliers supplement the dynamical understanding of a model. In cases where J represents a physical quantity, the form of Equation (3.11) will differ, but the Lagrange multipliers still give the gradient of the cost function with respect to various parameters. The Lagrange multipliers therefore represent *sensitivity* (Hall et al. 1982; Schröter and Wunsch 1986). This sensitivity has a physical significance in its own right and has been used to interpret the physics of the ocean (Marotzke et al. 1999; Bugnion 2001; Hill et al. 2004). The double nature of the Lagrange multipliers is an added benefit of the method.

3.3 Model dynamics and optimization

Model dynamics affect the shape of the cost function in control space through the model-data misfit, the first term in the generic cost function (Equation (3.1)). In the case of a linear model, the least-squares formulation has a global parabolic shape, as previously

assumed (see Equation (3.7)). More complicated shapes emerge when the model predictions, $\mathbf{Ex}(t)$, depend nonlinearly on the controls, $\mathbf{u}(t)$. Some cost function topologies make the search for a minimum more difficult, usually because the gradient with respect to the controls has little use. The emergence of many local minima in a cost function is one troublesome scenario, as gradient search methods do not distinguish between local and global minima. In previous studies (Lea et al. 2000; Köhl and Willebrand 2002), eddy-resolving ocean models based on the nonlinear equations of motion gave rise to many local minima. Models with thresholds are another example of nonlinearity. Gradients give a local measure of the cost function shape, but may not be accurate when extrapolated to a finite region of phase space with a dynamical regime change. In summary, the difficulties of nonlinear optimization are due to the model dynamics; specific cases are illustrated here, and then compared to the general circulation model problem.

3.3.1 Linear versus nonlinear models

In this section, the *recovery* or *initialization* problem of control theory is used to illustrate how the cost function shape differs when computed with a nonlinear model versus a linear model. Consider the goal of estimating the initial model state given one observation of the state at a later time. Successful recovery of the initial conditions depends on the length of time between the observation and the requested estimate, $t_f - t_0$. The results of this sample problem can be generalized to the case with many observations; hence, the arguments presented below are applicable and relevant to a wider variety of situations. The problem is restated as a least-squares minimization of the function:

$$\tilde{J} = [\mathbf{x}(t_f) - \mathbf{x}^{obs}(t_f)]^T \ \mathbf{W}(t_f) \ [\mathbf{x}(t_f) - \mathbf{x}^{obs}(t_f)], \tag{3.12}$$

where $\mathbf{x}(t)$ is the model state, $\mathbf{x}^{obs}(t)$ is an observation of the state, and $\mathbf{W}(t_f)$ is a weighting matrix. If all the observations are independent and weighted equally, $\mathbf{W}(t_f)$ is the identity matrix; for simplicity, we take this approach and drop $\mathbf{W}(t_f)$ hereafter.

The problem is solved by searching over the possible initial states. Therefore, knowl-

edge of the dependence of $\tilde{J}$ on the initial state, $\mathbf{x}_0$, is required. The model is part of this dependence:

$$\mathbf{x}(t_f) = \mathcal{L}_n \circ \ldots \circ \mathcal{L}_2 \circ \mathcal{L}_1 \circ \mathbf{x}(t_0) = \mathcal{R}(t_f, t_0)\ \mathbf{x}(t_0) \tag{3.13}$$

where n is the number of model timesteps between the initial and final time, $\circ$ is the composition operator, and the resolvent, $\mathcal{R}(t_f, t_0)$, is shorthand for the string of possibly-nonlinear model steps. In the unconstrained search space, the cost function is now:

$$\tilde{J}[\mathbf{x}(t_0)] = [\mathcal{R}(t_f, t_0)\ \mathbf{x}(t_0) - \mathbf{x}^{obs}(t_f)]^T\ [\mathcal{R}(t_f, t_0)\ \mathbf{x}(t_0) - \mathbf{x}^{obs}(t_f)], \tag{3.14}$$

There is a model trajectory that gives the minimum of $\tilde{J}$; the initial state of this trajectory is designated $\mathbf{x}^*(t_0)$. In the case of a perfect model and observation, the model with initial condition $\mathbf{x}^*(t_0)$ exactly predicts the observation:

$$\mathbf{x}^{obs}(t_f) = \mathcal{R}(t_f, t_0)\ \mathbf{x}^*(t_0). \tag{3.15}$$

The perfect model-data assumption clarifies the discussion, but is not necessary. Next, we wish to find the shape of the cost function around the minimum.

Before proceeding, the *tangent linear model* is defined. A perturbed nonlinear model trajectory can be integrated with the formula (Miller et al. 1994):

$$\mathcal{L}[\mathbf{x}(t) + \delta\mathbf{x}(t)] = \mathcal{L}[\mathbf{x}(t)] + \left(\frac{\partial \mathcal{L}}{\partial \mathbf{x}(t)}\right)\ \delta\mathbf{x}(t) + \delta\mathbf{x}(t)^T \left[\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}^2(t)}\right] \delta\mathbf{x}(t) + \ldots\ , \tag{3.16}$$

where the second-order term contains a third-order tensor. Subtracting the baseline nonlinear model trajectory and neglecting terms higher than order one, a perturbation to the state, $\delta\mathbf{x}(t)$, follows the dynamics of the so-called tangent linear model:

$$\delta\mathbf{x}(t+1) = \left(\frac{\partial \mathcal{L}}{\partial \mathbf{x}(t)}\right)\ \delta\mathbf{x}(t). \tag{3.17}$$

The matrix, $\partial\mathcal{L}/\partial\mathbf{x}(t)$, is sometimes called the *Jacobian* matrix. It is formed by the

derivatives of the nonlinear equations, $\mathcal{L}_1$, $\mathcal{L}_2$, etc., with respect to the state:

$$\left(\frac{\partial \mathcal{L}}{\partial \mathbf{x}(t)}\right) = \left(\begin{array}{cccc} \partial\mathcal{L}_1/\partial\mathbf{x}_1(t) & \partial\mathcal{L}_1/\partial\mathbf{x}_2(t) & \ldots & \partial\mathcal{L}_1/\partial\mathbf{x}_m(t) \\ \partial\mathcal{L}_2/\partial\mathbf{x}_1(t) & \partial\mathcal{L}_2/\partial\mathbf{x}_2(t) & \ldots & \partial\mathcal{L}_2/\partial\mathbf{x}_m(t) \\ \vdots & \vdots & \ddots & \vdots \\ \partial\mathcal{L}_m/\partial\mathbf{x}_1(t) & \partial\mathcal{L}_m/\partial\mathbf{x}_2(t) & \ldots & \partial\mathcal{L}_m/\partial\mathbf{x}_m(t) \end{array} \right)_{\mathbf{x}(t)} . \tag{3.18}$$

The model is always re-linearized about the changing model state, explicitly noted by the subscript $\mathbf{x}(t)$. Extending over many time steps, the final perturbation is related to the initial perturbation by

$$\delta\mathbf{x}(t_f) = \mathbf{R}(t_f, t_0)\ \delta\mathbf{x}(t_0), \tag{3.19}$$

where $\mathbf{R}$ is a linear resolvent made of a string of linear matrix multiplications. The validity of the tangent linear model to approximate the nonlinear dynamics is addressed more fully below.

The cost function, Equation (3.14), reduces to a quadratic form for linear models or for nonlinear models well-approximated by a tangent linear model (Figure 3-3, left side).

$$\begin{aligned} \tilde{J}[\mathbf{x}^*(t_0) + \delta\mathbf{x}(t_0)] &= [\mathbf{R}(t_f, t_0)\ \delta\mathbf{x}(t_0)]^T\ [\mathbf{R}(t_f, t_0)\ \delta\mathbf{x}(t_0)] & (3.20) \\ &= \delta\mathbf{x}(t_0)^T\ \mathbf{R}(t_f, t_0)^T \mathbf{R}(t_f, t_0)\ \delta\mathbf{x}(t_0) & (3.21) \end{aligned}$$

In contrast, the cost function is no longer globally quadratic and many local minima appear when the tangent linear model fails to well-approximate the nonlinear model. In that case, perturbations to the initial state are influenced by higher order terms. The cost function topology around the minimum is not purely quadratic:

$$\tilde{J}[\mathbf{x}^*(t_0) + \delta\mathbf{x}(t_0)] = \delta\mathbf{x}(t_0)^T\ \mathbf{R}(t_f, t_0)^T \mathbf{R}(t_f, t_0)\ \delta\mathbf{x}(t_0) + \mathcal{O}(\delta\mathbf{x}(t_0)^{2m}) \tag{3.22}$$

where m includes integers greater than one. Higher order terms destroy the parabolic nature of the function, and the original minimum is not necessarily unique. As seen in

this generic example, nonlinearity in a model is responsible for local minima.

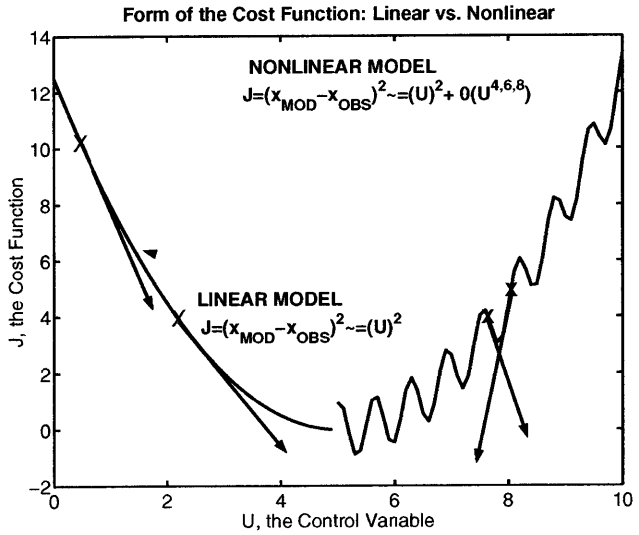

Figure 3-3: Schematic of the cost function with a linear versus nonlinear model. The linear model (*left*) gives a cost function with paraboloidal shape because of the least-squares formulation. A nonlinear model (*right*) potentially gives a much more complicated shape; discontinuities and multiple local minima are possible.

The preceding section hints at the role of model dynamics in the least-squares problem. Specifically, a nonlinear model can distort the simple, parabolic form of the sum of squares. However, the results of the previous section are strengthened by considering the physics of a simple dynamical system. The pendulum is chosen for study because it can be implemented as a nonlinear or linear set of equations, and it can also be stable, unstable, or chaotic.

3.3.2 Case study: Single pendulum

Is it possible to determine the angle and velocity of a pendulum at initial time with one observation at a later time? Like the previous section, this is a statement of the recovery problem of control theory. The simple formulation of this problem isolates the effect of the model dynamics on the optimization problem. In this case, the damped, single

pendulum of fixed length is used (following the textbook of Baker and Gollub (1990), see Figure 3-4). The motion of the pendulum is described by the equation:

$$\frac{d^2\theta}{dt^2} + q\,\frac{d\theta}{dt} + sin\ \theta = 0, \tag{3.23}$$

where θ is the displacement angle from vertical and q is a damping coefficient. To numerically implement this system, the angular velocity, ω, is included as part of the state and the system is discretized with a forward Euler timestep of time Δt. The discrete-time state space realization is:

$$\begin{pmatrix} \omega(t+\Delta t) \\ \theta(t+\Delta t) \end{pmatrix} = \begin{pmatrix} (1-q\Delta t)\ \omega(t)\ -\ \Delta t\ sin\ \theta(t) \\ \Delta t\ \omega(t)\ +\ \theta(t) \end{pmatrix}. \tag{3.24}$$

The tangent linear model, according to Equation (3.18), is:

$$\begin{pmatrix} \delta\omega(t+\Delta t) \\ \delta\theta(t+\Delta t) \end{pmatrix} = \begin{pmatrix} 1-q\Delta t & -\Delta t\ cos\ \theta(t) \\ \Delta t & 1 \end{pmatrix} \begin{pmatrix} \delta\omega(t) \\ \delta\theta(t) \end{pmatrix}. \tag{3.25}$$

The cost function, Equation (3.12), is rewritten for the pendulum:

$$\tilde{J} = [\theta(t_f) - \theta^{obs}(t_f)]^2 + [\omega(t_f) - \omega^{obs}(t_f)]^2 \tag{3.26}$$

where θ^{obs} and ω^{obs} are observations. We next consider the linear pendulum with stable and unstable dynamics, then contrast the cost function shape with stable and unstable nonlinear dynamics.

Linear, stable pendulum

Although the full equations of motion for the pendulum are nonlinear, a traditional approach is to make the small-angle approximation. The dynamics of the pendulum are

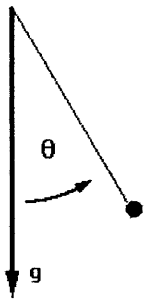

Figure 3-4: Diagram of the single pendulum. Large angles, θ, are allowed in the nonlinear system. The pendulum has a massless rod of fixed length.

linearized around the state of zero displacement, $\theta = 0$.

$$\begin{pmatrix} \omega(t+\Delta t) \\ \theta(t+\Delta t) \end{pmatrix} = \begin{pmatrix} 1-q\Delta t & -\Delta t \\ \Delta t & 1 \end{pmatrix} \begin{pmatrix} \omega(t) \\ \theta(t) \end{pmatrix}. \tag{3.27}$$

Equation (3.27) is the discrete-time form of the continuous-time equation:

$$\frac{d^2\theta}{dt^2} + q\,\frac{d\theta}{dt} + \theta = 0, \tag{3.28}$$

with the linear term of the Taylor series expansion, $sin\ \theta \approx \theta$, replacing sine in the nonlinear equation (3.23). The linearized pendulum dynamics should not be confused with the dynamics of the tangent-linear model, although they are related. The linearized pendulum of this section is always linearized around zero displacement, but the tangent linear model is re-linearized around a changing nonlinear model trajectory.

To examine the shape of the cost function, consider an "identical twin" experiment. The observation is generated by running the model with initial displacement of $-\pi/6$ *radians* and zero velocity. Assuming a perfect model and observation, the shape of the cost function is generated by changing the initial conditions and evaluating $\tilde{J}$. One-dimensional slices of the cost function are made by varying the initial displacement angle and by keeping the initial velocity fixed to zero. Regardless of the elapsed

time between the initial conditions and the observation, a slice of the cost function is a parabola (Figure 3-5,*upper left panel*). The cost function becomes less steep if the elapsed time between initial and final time, $t_f - t_0$, is increased. A flat cost function is one where the model is relatively unconstrained. Thacker (1989) showed that the curvature around the minimum gives the uncertainty of the estimate; a deeper "hole" yields a more constrained estimate. In the pendulum example, the recovered initial conditions become more uncertain with time because of the dissipation of information by damping. The timescale of memory loss is roughly equivalent to $1/q$, or 100 s, in this particular example. The cost function tends to zero everywhere for time integrations longer than the damping timescale. In summary, a linear model gives a paraboloidal cost function, leading to a straightforward search for the minimum unless the memory of the initial conditions is lost.

The linearized pendulum has an equilibrium point at rest, $\theta = 0$, $\omega = 0$. The system is stable[4] if an arbitrary perturbation remains in a finite neighborhood of the equilibrium for all time and approaches the equilibrium as time goes to infinity. For an unforced, linear dynamical model, $\mathbf{x}(n\Delta t) = \mathbf{A}^n\ \mathbf{x}(0)$, decompose the initial state into the eigenmodes, $\mathbf{g}_i$, of $\mathbf{A}$:

$$\mathbf{x}(0) = \sum_{i=1}^{M} \alpha_i(0)\mathbf{g}_i, \tag{3.29}$$

where $\alpha(t)$ is the time-variable projection of the state onto a particular eigenmode, i. In the present case, the dynamical model does not vary in time, and hence, the eigenmodes are fixed. Therefore, the evolution of the state follows a simple modal form:

$$\mathbf{x}(n\Delta t) = \sum_{i=1}^{M} \lambda_i^n \alpha_i(0)\mathbf{g}_i, \tag{3.30}$$

where λ_i is the i-th eigenvalue. Division of the last two equations,

$$\frac{\alpha_i(t)}{\alpha_i(0)} = \lambda_i^n, \tag{3.31}$$

[4]Technically, this is the definition of *asymptotic stability.*

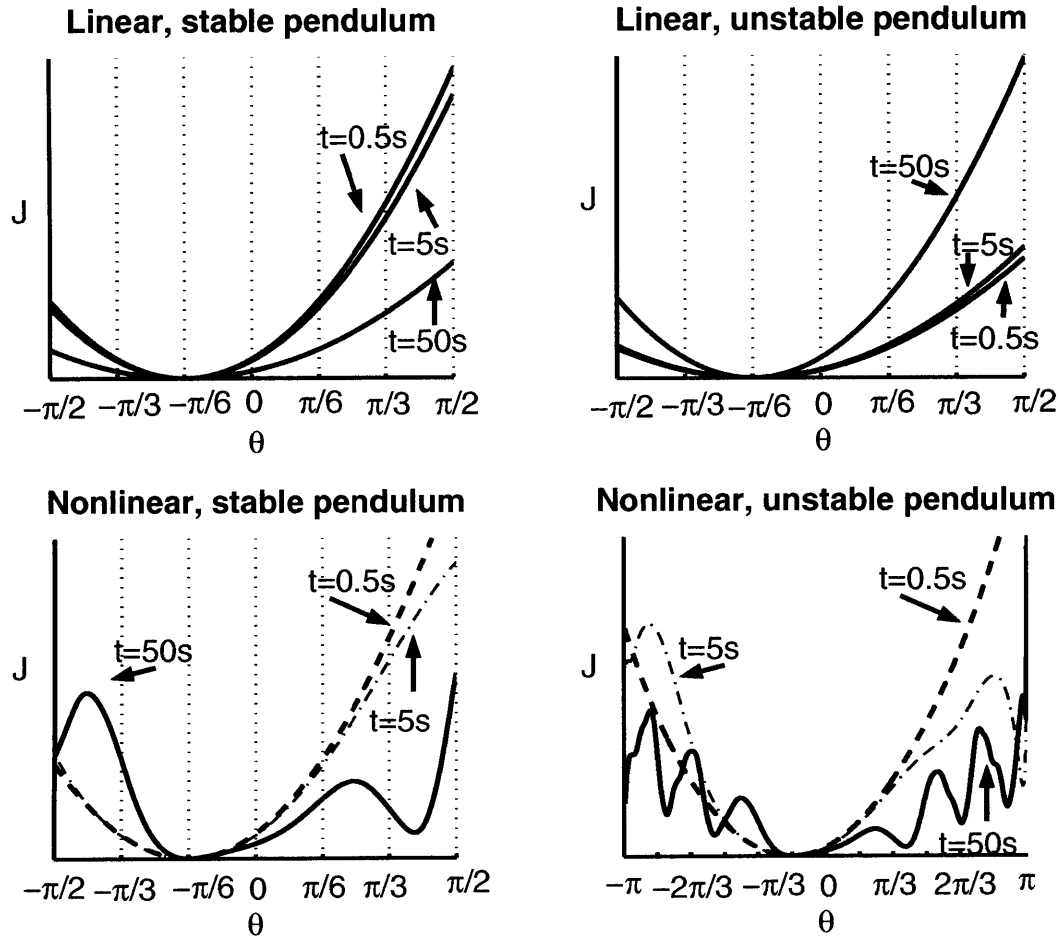

Figure 3-5: Cost function with respect to the initial pendulum angle. A synthetic observation was made from a model run with intial angle, $\theta = -\pi/6$. The time between the initial state and the cost function evaluation is 0.5, 5, or 50 seconds. *Upper left*: Linear, stable pendulum. *Upper right*: Linear, unstable pendulum *Lower left*: Nonlinear, stable pendulum. *Lower right*: Nonlinear, unstable pendulum. Notice the wider scale for θ in the lower, right panel. The pendulum's dynamical regimes are further explained in the text.

gives a stability criterion. Eigenvalues with magnitude greater than one grow exponentially with time. The discrete-time pendulum is stable for $0 < \Delta t < q$. Without forcing, the pendulum returns to rest for all initial conditions. Due to stability, the cost function magnitude decreases with increasing integration time; in other words, all model trajectories eventually converge. The case where the model varies with time leads to a slightly different interpretation of the stability criterion, and is discussed later in the section on nonlinear pendulum dynamics.

Linear, unstable pendulum

To explore the impact of instability, consider changing the sign of θ in the linearized pendulum equation, which is equivalent to linearizing the inverted pendulum:

$$\frac{d^2\theta}{dt^2} + q\,\frac{d\theta}{dt} - \theta = 0. \tag{3.32}$$

Physical justification is not sought for this change, but it is a simple way to render the problem unstable. Eigenvalue analysis shows that one unstable mode is present. A typical observation provides a strong constraint to the initial unstable mode, because an error in that mode grows with time. Consequently, the cost function becomes increasingly steep as the time between initial and final time is lengthened. As seen in Figure 3-5 (*upper right panel*), the model initial conditions are well-constrained. The shape of the cost function is still parabolic, so instability does not impede the search for the minimum.

The two previous examples with the linear pendulum appear straightforward; however, special situations should be mentioned. Parabolic cost functions with varying steepness in different directions result from ill-conditioned problems. As mentioned in Section 3.2.4, searches may be inefficient in this case. An extreme example is that of the banana-shaped valley, in which steepest-descent methods fail. Sums of independent parabolic terms in a cost function may yield such complicated forms. Another problem which may occur in linear models is the non-computability of the gradient. For the

unstable model, gradients grow exponentially with time, and they may be too large to be computed by numerical means. To summarize, linear dynamics, whether stable or unstable, give parabolically-shaped cost functions. In most cases, the search for a minimum of a paraboloid is efficient, but special circumstances do exist.

Nonlinear, stable pendulum

Stability of the nonlinear pendulum is determined by the linearized dynamics around each point in phase space. A global measure of stability is no longer possible. Stability of the tangent linear model is interpreted as the convergence of neighboring nonlinear trajectories. For the pendulum, the tangent linear matrix has eigenvalues greater than one when linearized about a state in the upper-half plane ($\theta < -\pi/2,\ \theta > \pi/2$, see Figure 3-6). Gravity accelerates a horizontal pendulum most strongly; in the upper-half plane, a pendulum perturbed towards the horizontal is more rapidly accelerated downwards: an unstable configuration. Conversely, the lower-half plane is stable. Even though the pendulum is not globally stable, the behavior of a stable, nonlinear model can be examined by looking at the lower-half plane alone. Hereafter, the nonlinear pendulum restricted to the lower-half plane is referred as the "nonlinear, stable pendulum."

The cost function computed with the nonlinear, stable pendulum has four stationary points, two local maxima and two local minima (Figure 3-5, *lower left panel*). The only difference in the dynamics is a nonlinear term. Gradient search methods find the nearest minimum, but no clear test exists to distinguish the global minimum from a local one. This example shows that nonlinear models, even those that are stable, can create local minima.

The tangent linear model well-approximates the nonlinear dynamics for a limited amount of time, the *nonlinear timescale* (Gauthier 1992; Miller et al. 1994). For example, consider the dynamics of the pendulum from the starting angle of $\theta(t_0) = 3\pi/8$, near a local minimum of the cost function. The pendulum trajectory can be computed by either the nonlinear model, or by the tangent linear model around the trajectory with the correct initial angle, $\theta^*(t_0) = -\pi/6$. After fifteen seconds, the tangent linear model

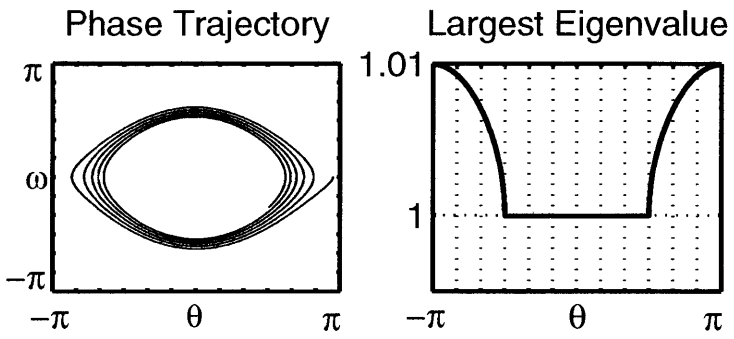

Figure 3-6: Characteristics of the nonlinear pendulum. *Left*: The nonlinear pendulum traces a damped quasi-periodic orbit in phase space. The pendulum was implemented with a timestep of 0.01 s and a damping coefficient of 0.01 s^{-1}. *Right*: The pendulum is unstable in the upper half plane where the magnitude of the greatest eigenvalue, $|\lambda|$, exceeds one. In the lower half plane, the pendulum is stable (nearly neutral) with the largest eigenvalue just less than one.

makes a different prediction than the nonlinear model (Figure 3-7). The inaccuracy of the tangent linear model has two causes. First, the pendulum frequency is a function of amplitude, but the tangent linear model is linearized around a trajectory with a smaller amplitude, and hence, an inaccurately-short period. Second, the tangent linear model predicts divergence of the two pendulums, as seen by the growth[5] of the envelope of $\Delta\theta$, even though two nonlinear trajectories converge. In summary, the length of time integration and the transient behavior of a system must be considered when assessing the validity of the tangent linear model.

The preceeding section does not claim that the tangent linear model is incorrect. Instead, the validity of the tangent linear model depends upon the size of the initial perturbation. For a sufficiently small perturbation, the tangent linear model does well-approximate the nonlinear dynamics; given the proper state to linearize about, the tangent linear model is successful. For the pendulum, the angle after fifty seconds is a

[5]Perturbation growth occurs in the nonlinear pendulum despite asymptotic stability. The state transition matrix is non-self-adjoint, and *non-normal growth* (Farrell 1989; Farrell and Moore 1993) is possible. In this system, non-normal growth occurs because two pendulums with slightly different initial conditions go out of phase, leading to large differences. Over long time periods, the decaying amplitude of oscillations ceases the divergence of trajectories, and non-normal growth is seen to be transient.

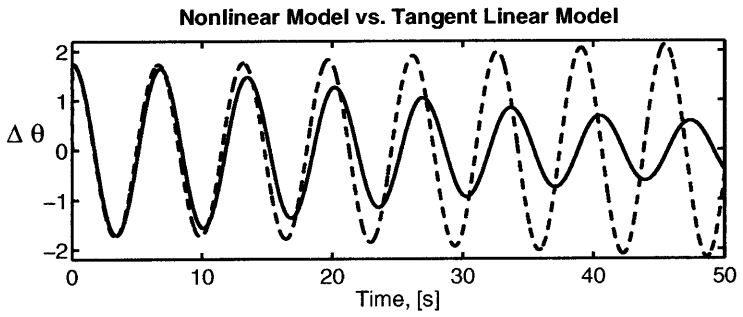

Figure 3-7: The difference of angle, $\Delta\theta$, between two model trajectories as computed by the nonlinear model (*solid line*) and the tangent linear model (*dashed line*). The two nonlinear trajectories are started with initial angle $\theta(t_0) = -\pi/6$ and $\theta(t_0) = 3\pi/8$. The tangent linear model is linearized about the former trajectory.

nonlinear function of the initial angle (Figure 3-8). The tangent-linear model is valid for an exceedingly-small region around $\theta(t_0) = -\pi/6$. In this small region, the cost function, according to arguments in Section 3.3.1, is locally parabolic.

Nonlinear, unstable pendulum

As mentioned above, the nonlinear pendulum is stable in the lower-half plane, and unstable in the upper-half plane. With initial conditions in the upper-half plane, the pendulum trajectory is episodically-unstable. For simplicity, any pendulum that enters the upper-half plane at any time is called a "nonlinear, unstable pendulum." A slice of the cost function contains many local minima with fifty seconds of elapsed time between initial and final state (Figure 3-5, *lower right panel*). Instability of the model dynamics is not a prerequisite for the emergence of local minima, but it exacerbates the problem. Neighboring nonlinear trajectories diverge in time, and knowledge of the correct state for linearization of the tangent linear model is lost with time. For the nonlinear, unstable pendulum, gradient search only yields the global minimum for short time intervals or with an excellent first guess.

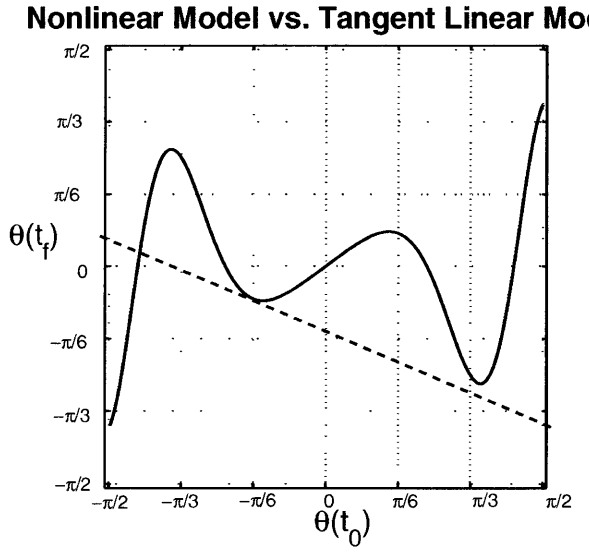

Figure 3-8: Pendulum angle at time $t_f = 50\ s$ as a function of initial angle. The function is computed by the nonlinear dynamics (*solid line*), and by the tangent linear model about the trajectory with initial angle, $\theta(t_0) = -\pi/6$ (*dashed line*).

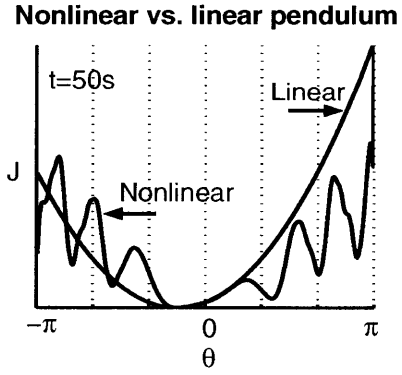

Figure 3-9: The form of the cost function of the pendulum with fifty seconds of elapsed time between initial conditions and the observation. The cost function is computed with both the nonlinear and linear model. Fifty seconds exceeds the nonlinear timescale, so local minima appear in the cost function.

Nonlinear, chaotic pendulum

With the addition of forcing, the single pendulum is chaotic in certain parameter ranges (see Appendix B for the equations of motion). Nonlinear, chaotic pendulums are a subset of nonlinear, unstable systems. The short-time dynamical behavior of the two classes of models are identical for our purposes. However, differences appear in the long-time behavior. Gradients of nonlinear, unstable models tend to zero with damping. In contrast, gradients computed by nonlinear, chaotic models grow exponentially for an indefinite amount of time despite damping. Therefore, sensitivity analysis with long-time integrations of nonlinear, chaotic systems have two problems: the potential non-computability of very large gradients, similar to unstable, linear models, and the emergence of many local minima, as seen in many nonlinear models.

3.3.3 Models with thresholds

Nearly all numerical models have thresholds due to physical or numerical reasons. Numerical programs necessarily include many switches, such as conditional *if* statements. One ocean process that depends upon a threshold is convection. To examine the impact of model thresholds on a cost function, consider a water column undergoing cooling at the surface (Figure 3-10, *left panel*). The simplest, discrete representation of the vertical stratification has two components, surface density, ρ_1, and abyssal density, ρ_2. In a numerical model, convection is typically implemented as two-step process. First, cooling is applied to the surface.

$$\rho_1(t_0 + 1) = \rho_1(t_0) + Q, \tag{3.33}$$

$$\rho_2(t_0 + 1) = \rho_2(t_0), \tag{3.34}$$

where Q is surface forcing. Second, if the surface density is greater than or equal to the abyssal density, the ocean convects and subsequently mixes.

$$\boxed{\text{if}}\ \rho_1(t_0+1) \geq \rho_2(t_0+1) \ \rightarrow \ \begin{cases} \rho_1(t_0+2) = (\rho_1(t_0+1) + \rho_2(t_0+1))/2 \\ \rho_2(t_0+2) = (\rho_1(t_0+1) + \rho_2(t_0+1))/2 \end{cases} \tag{3.35}$$

where the arrow represents fulfillment of the conditional statement. If the column is gravitationally stable, no convection happens.

$$\boxed{\text{else if}}\ \rho_1(t_0+1) < \rho_2(t_0+1) \ \rightarrow \ \begin{cases} \rho_1(t_0+2) = \rho_1(t_0+1) \\ \rho_2(t_0+2) = \rho_2(t_0+1) \end{cases} \tag{3.36}$$

Suppose an observation of the abyssal density is available at time $t_0 + 2$. Then, the squared data-model misfit is:

$$\tilde{J} = [\rho_2(t_0+2) - \rho_{obs}]^2. \tag{3.37}$$

The goal of the toy example is determine the correct amount of cooling in order to

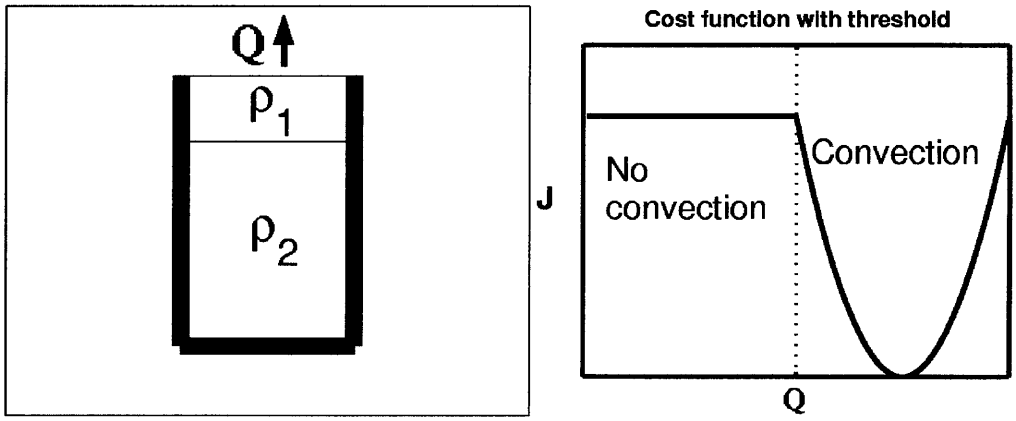

Figure 3-10: *Left panel*: Schematic of an oceanic water column with upper density, ρ_1, and abyssal density, ρ_2. Cooling, Q, is applied to the surface. *Right panel*: Data-model misfit as a function of cooling. Two dynamical regimes are present: a non-convective regime (*left half*), and a convective regime (*right half*), which greatly affect the cost function form.

reproduce the observed abyssal density. The cost function value with respect to cooling shows the impact of the threshold in the model dynamics (Figure 3-10, *right panel*). Cooling affects the observational site only when convection is happening. Therefore, the gradient in the non-convective regime is zero, and is very different than the gradient in the convective regime.

This example illustrates that the gradient information is local, and may not accurately predict the value of the cost function with a finite perturbation to the controls. Sensitivity studies, where only one adjoint calculation is performed, get only a linear picture of the model dynamics, and their applicability may be limited in a highly nonlinear model. In the minimization context, gradients are calculated at many different points in phase space, yielding some overlying picture of the nonlinearity of the model dynamics. Due to this fact, dynamical regime shifts are not expected to be a major problem in finding a solution to the least-squares problem here.

Differentiability of model dynamics

The cost function presented above has one special point at the threshold between the convective and non-convective states. This forces one to consider the *differentiability* of the model dynamics. Some investigators call any conditional statement nondifferentiable, but the previous paragraph shows that such statements can usually be handled by accurate linearization. With chaotic models, very large gradients have been attributed to nondifferentiable dynamics (Köhl and Willebrand 2002). Formally, there is a distinction between unstable and nondifferentiable dynamics. Unstable (or chaotic) dynamics are differentiable provided that the local neighborhood of examination is small enough. Machine precision is an eventual limit, at which point an unstable model is indistinguishable from a nondifferentiable system within numerical accuracy. Here, we use the formal definition of nondifferentiability. A numerical model statement is symbolically written, $x_{out} = g(x_{in})$, where g can be a nonlinear function, and x_{in} and x_{out} are continuous scalars. If $[\partial g/\partial x_{in}]_{x_{in}}$ does not exist, then the model is said to have a nondifferentiable point at x_{in}.

For the convective water column example, the gradient of the cost function with respect to the cooling is examined through the chain of model steps. Following the ideas of Marotzke et al. (1999), the cost function is written as:

$$\tilde{J} = f \circ \mathcal{L}_{1,2} \circ \mathcal{L}_{0,1} \circ \mathbf{x}(t_0) = f(\mathcal{L}_{1,2}(\mathcal{L}_{0,1}(\mathbf{x}(t_0)))) = f(\mathcal{L}_{1,2}(\mathbf{x}(t_0) + \mathbf{b}Q)) \tag{3.38}$$

where f maps the final state onto a scalar, $\circ$ is the composition operator, $\mathcal{L}_{t_1,t_2}$ represents the model step from $t = t_1$ to t_2, $\mathbf{x}(t_0)$ is the initial state, and $\mathbf{b}$ is the column vector, $[1\ ,\ 0]^T$. The derivative of the cost function with respect to Q is determined by the chain rule:

$$\frac{\partial \tilde{J}}{\partial Q} = f'(\mathcal{L}'_{1,2}(\mathbf{x}(t_0) + \mathbf{b}Q)') = f'(\mathcal{L}'_{1,2}(\mathbf{b})). \tag{3.39}$$

The derivative of f with respect to the state, f', is $[0\ ,\ 2(\rho(t_0+2) - \rho_{obs})]$, a well-defined quantity for all reasonable values of abyssal density. Likewise, the column vector, $\mathbf{b}$, is well defined. However, the tangent-linear model, $\mathcal{L}'$, depends upon the physical regime for linearization. For all convecting states, the tangent linear model is:

$$\mathcal{L}'_{1,2} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \tag{3.40}$$

On the other hand, the tangent linear model for nonconvecting states is:

$$\mathcal{L}'_{1,2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}. \tag{3.41}$$

Evaluation of the gradient is now a series of vector and matrix multiplications. There is a particular amount of cooling, $Q_{threshold}$, which leads to a homogeneous water column at $t = 1$, i.e. $\rho_1(t_0 + 1) = \rho_2(t_0 + 1)$. For an infinitesimal perturbation of cooling, $Q = Q_{threshold} + \epsilon$, the gradient of the cost function is:

$$\left[\frac{\partial \tilde{J}}{\partial Q}\right]_+ = [0\ \ 2(\rho_2(t_0+2) - \rho_{obs})] \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} [1\ \ 0]^T = \rho_2(t_0+2) - \rho_{obs}. \tag{3.42}$$

Approaching the threshold from the nonconvecting side,

$$\left[\frac{\partial \tilde{J}}{\partial Q}\right]_{-} = [0 \;\; 2(\rho_2(t_0+2) - \rho_{obs})] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} [1 \;\; 0]^T = 0. \tag{3.43}$$

Because the two limits do not agree, the gradient does not exist at this point. Differentiability has meaning in this discretized model because the density is expressed as a continuous variable. On the other hand, the derivative of the state with respect to time, $\partial\rho/\partial t$, or the vertical gradient of density, $\partial\rho/\partial z$, can not be well-defined in a continuous sense. To summarize, the convection threshold represents a nondifferentiable point, because the gradient does not formally exist.

Models with thresholds open the possibility that the gradient of the cost function may not exist at a point. However, the probability of landing exactly on this threshold is formally zero, because the forcing and the cost function are continuous scalars (Griewank 2000). In addition, the automatic adjoint code generator (TAF) still computes gradients at the nondifferentiable point, which are equivalent to one-sided gradients. The adjoint compiler handles conditional statements in the same way as other nonlinear statements – with linearization around the full forward model trajectory. Despite the formal difficulties with nondifferentiable points, they have not posed a problem in practice to this date.

Summary of the influence of model dynamics on J

- Nonlinear model dynamics give rise to the possibility of local minima in the cost function, and hence, multiple solutions.

- Local minima in the cost function are possible in nonlinear systems with locally-unstable trajectories, or even in a stable nonlinear model with transient growth of perturbations.

- A dynamical regime shift, such as those caused by model thresholds, is a situation where the adjoint-computed gradient differs from a finite-difference approxima-

tion. This is a problem for sensitivity studies, but is not a major deterrent for minimization of a cost function.

3.4 Coarse-resolution optimization

This study begins with a coarse-resolution state estimation problem for two reasons: to test the numerical machinery, and to potentially use the result as a new best guess for the eddy-resolving calculation. State estimation with a coarse resolution ocean model avoids many of the problems of an eddy-resolution estimate because the model is quasi-linear and the control space is much smaller. Table 3.1 summarizes the differences in the 2° and 1/6° estimation problems. The large values of friction necessary to numerically stabilize a coarse resolution model make the dynamics quasi-linear. Coarse resolution models have been brought into consistency with data by a number of past investigators (Marotzke and Wunsch 1993; Stammer et al. 2002). Computationally, the 2° estimation problem consumes a relatively small amount of resources. Large-scale biases in the forcing and regional model inadequacies can be accounted for in the coarse-resolution estimate. Correction of the biases is much more computationally efficient at coarse resolution. In summary, coarse resolution state estimation with the regional model takes a relatively small effort, but the potential benefits for the fine resolution estimation problem are great.

To implement the coarse-resolution regional estimate, all external forcings and boundary conditions are taken from the ECCO global estimate with the same resolution. The time period of the coarse-resolution estimate is identical to the fine-resolution one: June 1, 1992, to June 1, 1993. The cost function has the same form (Equation 2.1) as the fine resolution problem, but the weights are changed. A coarse-resolution model does not resolve motions at scales less than the grid spacing, and such information in the observations must be considered noise. The Zang-Wunsch spectrum is used to predict the energy at scales less than 400 km, the sub-gridscale and the diffusively-dominated

	2°	**1/6°**
Horizontal Resolution	$(167 - 218)$ km x 222 km	$(14.2 - 18.2)$ km x 18.5 km
Grid Points	20 x 16 x 23 vertical levels	192 x 168 x 23 vertical levels
Time Step	3600 s = 1 hr.	900 s = 15 min.
Lap. Horiz. Viscosity	5x10^4 m^2/s	0
Lap. Horiz. Diffusivity	1x10^3 m^2/s	0
Biharmonic Horiz. Vis./Diff.	0	2x10^{11} m^4/s
Vertical Viscosity	1x10^{-3} m^2/s	1x10^{-3} m^2/s
Vertical Diffusivity	1x10^{-5} m^2/s	1x10^{-5} m^2/s
Reynolds Number	≈ 1	≈ 25
State Vector	1.70 x 10^4 elements	3.14 x 10^6 elements
Control Vector	9.11 x 10^4 elements	5.49 x 10^6 elements
Model Input	7.68 x 10^5 forcing elements	7.98 x 10^7 forcing elements
Model Output	1.50 x 10^8 estimated elements	1.09 x 10^{11} estimated elements
Processors	1 processor	24-48 processors
Computational Time	2 cpu hours/iteration	400 cpu hours/iteration
Search Iterations	$\approx$ 40 iterations	$\approx$ 120 iterations
Total Computer Time	$\approx$ 80 hours (2.3 days)	$\approx$ 50,000 hours (5.7 years)

Table 3.1: Coarse and fine resolution state estimation

range near the grid spacing. Below 400 km, the model wavenumber spectrum is too steep; power decreases with wavenumber too rapidly due to the diffusive nature of the model. To restate, the same observations are used in both estimates, but much larger misfits are acceptable in the coarse-resolution problem. The coarse-resolution state estimate here differs from the ECCO estimate for two primary reasons: the open boundary formulation of the model, and the inclusion of new Subduction Experiment data in the cost function. The result, detailed below, is a regional state estimate at coarse resolution which is significantly improved for our particular study. The estimate is then used for the fine resolution problem by a linear-interpolation onto a finer grid.

The method of Lagrange multipliers brings the ocean circulation within observational uncertainty in fifty iterations of the forward and adjoint models (see left panel, Figure 3-11). Therefore, the control parameters chosen in Chapter 2 are capable of controlling and changing the interior ocean circulation. Furthermore, fifty iterations is extremely efficient considering the control vector of 100,000 elements (i.e., $N_{iterations} << N_{controls}$). The successive updates of the controls further illustrates the efficiency of the optimiza-

tion. The control variables quadratically converge upon the minimum of the cost function subject to the coarse resolution model (right panel, Figure 3-11); this is the theoretical rate of convergence for the quasi-Newton method with a parabolically-shaped cost function (Press et al. 1992). Indeed, when the two panels of Figure 3-11 are combined, the shape of the cost function in control space is a parabola (Figure 3-12). This topology is expected for a diffusive coarse-resolution ocean model. The solution for the control variables is within the prior estimated range of uncertainty. It is not surprising that the method works so well for a coarse resolution model, because it is a nearly-linear system.

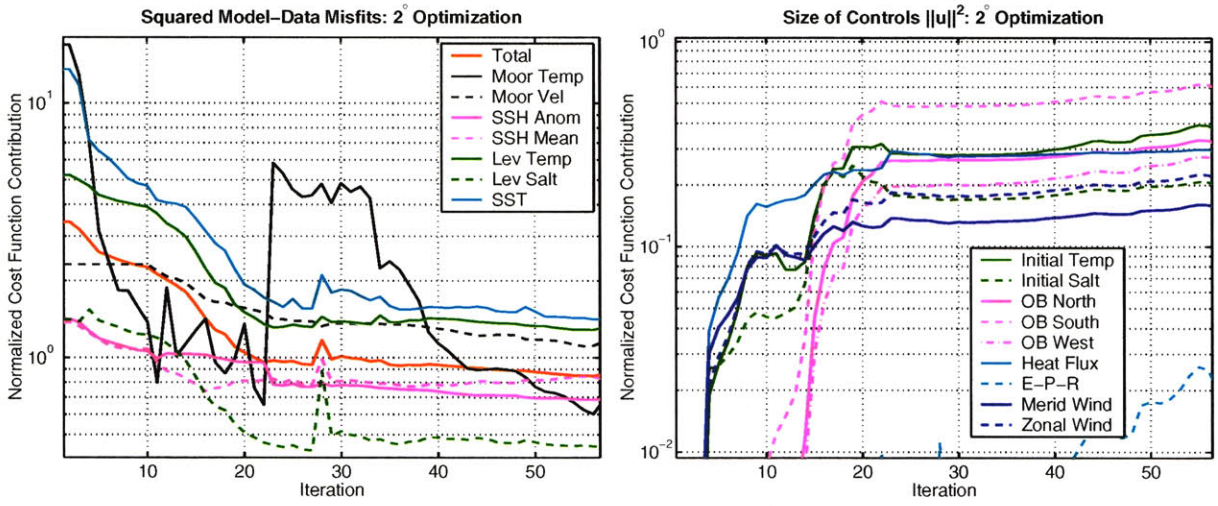

Figure 3-11: *Left panel*: Normalized model-data misfit as a function of iteration of the search method. A value of 1 (10^0) is expected. Irregularities are caused by improvements and changes in the numerical code; for example, the increase in the mooring temperature misfit occurred when the data-model mapping was improved in the numerical code. *Right panel*: The size of the control adjustments, $||u||^2$, for the same experiment.

3.4.1 Coarse-resolution misfits

The simulation with zero control adjustments has several large-scale hydrographic deficiencies which require adjustments in the controls. Sea surface temperatures approach $35°C$ in the northern basin ($30-40°N$). A southward shift of the semi-permanent Azores

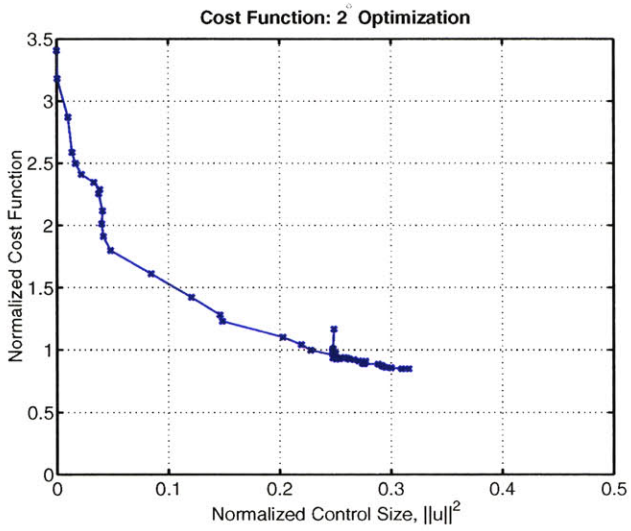

Figure 3-12: Magnitude of the cost function with respect to the size of the control adjustments. 55 cost function evaluations are plotted with "X"'s. A half-parabola emerges, consistent with a quasi-linear model. Irregular points are present because changes were made in the cost function weights.

High, with associated heat flux changes of 50 W/m^2, cools summertime SST's in this region. Overly-warm sea surface temperatures are also associated with a weakened Canaries Current in the simulation. The optimization shifts the open boundary southern velocity from north to south in order to accommodate more cold water advection along the coast. Abnormally warm SST is a ubiquitous problem of the ECCO state estimate[6]. Another major deficiency of the simulation is the meridional slope of the winter mixed layer base; the mixed-layer deepens to the south, reaching a depth of 220 m, at $22°N$. Observations and climatologies alike show that the mixed-layer shoals equatorward, a crucial feature for subduction (Woods 1985; Marshall et al. 1993). The western boundary fluxes too much heat away from the eastern subtropics between $20 - 30°N$. The optimization responds by both warming the western boundary at these latitudes, and by decreasing the westward exit flow. The optimized estimate of mixed-layer depth then shoals towards the south, and never reaches a depth greater then 170 m, in close ac-

[6]Here, we have used the original ECCO state estimate from the adjoint method, 1992-1997. Later estimates do not have the same preponderance of overly-heated sea surface temperatures (D. Stammer, pers. comm.) because of the addition of an explicit boundary layer scheme.

cordance with observations. Analysis of the individual misfit terms in the cost function further substantiates the large-scale deficiencies of the model.

Mooring misfit: 2°

Estimated temperature at the mooring sites reflects gradual cooling of the sea surface and subsequent deepening of the mixed-layer in winter, as observed. Simulated (no data constraint) temperature is deficient in many ways: an overly-deep mixed-layer, mistimed winter onset, and a too-weak seasonal cycle. The estimate adjusts the large-scale heat budget of the ocean ocean to improve the characteristics of the seasonal cycle. Mooring temperature is shown to be controllable in this study; upper ocean measurements are used which are used to directly estimate changes in surface forcing and initial conditions. Deep hydrographic measurements may not be controllable because of the long times needed for surface signals to propagate to the deep ocean. This question is open for future research.

TOPEX/POSEIDON misfit: 2°

TOPEX/POSEIDON satellite altimetry is unique in this study in that the observational uncertainty rivals the dynamical signal. For example, the background variability in this region approaches $10-20\ cm$ but the noise is around $5\ cm$. For this reason, the maximum misfit of the model is bounded at a fairly small value relative to the other data types. When considering only the large-scale observational signal, 58% of the SSH variance is noise because it is at wavelengths less than $400\ km$. The original SSH anomaly misfit is 30% greater than the expected value, but changes in the large-scale structure bring the estimated surface height field into consistency. The SSH mean field from TOPEX/POSEIDON is also used. It is also a relatively weak constraint on the model dynamics because of the uncertainty of the geoid at high wavenumbers. The general circulation model is consistent with the observed SSH mean field for all control variable values used here.

Climatological misfits: 2°

Ocean climatologies are difficult to fit, even within their larger uncertainty. For example, the state estimate differs from Levitus climatological temperatures in the eastern boundary current region off of Africa and the Iberian Peninsula. Iterations of the forward/adjoint model bring the estimate closer to the Levitus climatology, but not into complete statistical consistency. This is not because the climatology has been downweighted too much; numerically, its contribution to the cost function is the larger than all other terms. Inconsistency between the Levitus climatology and other datasets is not implied, as the *a priori* tests of Section 2.2.3 dismiss such a possibility. The last possibility is that the Levitus climatology is inconsistent with the equations of motion. Because it is a long-term, time-mean statistical average of various data sources, the latter explanation seems most likely.

A strict comparison of the modeled temperature and Reynolds SST data reveals inadequacies in the model dynamics. Surface layers of the model are too warm in the summer because the seasonal mixed-layer is not deep enough (Figure 3-13). The KPP boundary layer model parameterizes wind-stirred deepening of the mixed-layer, but does not flux enough heat downwards. Summertime errors are evident in the biased histogram of the model-data misfit. For statistical rigor, Gaussian errors are assumed, but inadequacy of the model dynamics breaks this posterior test. The standard deviation of the SST misfit is also slightly larger than the expected value, which is measured by the cost function. This paragraph is a call for model improvement.

3.4.2 Coarse-resolution control adjustments

Which controls are most important to bring the model into consistency with the observational signal? The gradient of the cost function with respect to the controls gives a quantitative answer. After nondimensionalizing each gradient by its data type and depth, the initial temperature and open boundary conditions are most important over the first year of integration. The memory of initial conditions extends well beyond one year – both forward model studies (Griffies and Bryan 1997) and adjoint studies

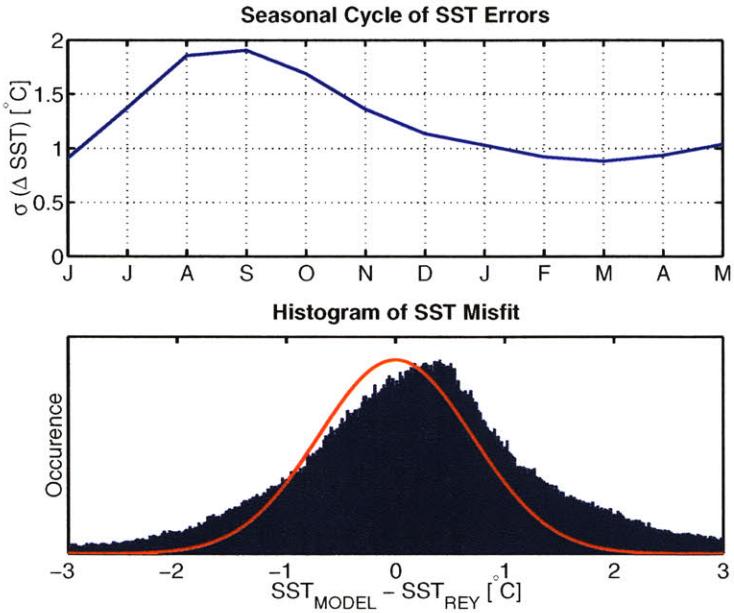

Figure 3-13: *Top panel*: Standard deviation of SST misfit as a function of month. *Lower panel:* Histogram (*blue*) of SST misfit, and the assumed prior error statistics (*red line*).

(Bugnion 2001) have shown a memory of at least ten years in the upper ocean.

Previous state estimation studies have seen the emergence of spurious small-scale noise in the control adjustments (Zhang and Marotzke 1999), but this is not a problem with the formulation here. The open boundary temperature and normal velocity fields play a similarly important role in controlling the ocean circulation.

Control Statistics

To satisfy *a priori* assumptions in the cost function, the magnitude of the control adjustments must be within an expected range. For uncorrelated control adjustments with a Gaussian distribution and a standard deviation of one, the squared controls should follow a chi-squared (χ_1^2) distribution with one degree of freedom (Wunsch 1996). The control adjustments for the eddy-resolving state estimate follow a chi-squared distribution, but are more strongly clustered around zero (Figure 3-14). This suggests that the controls are correlated, which is reasonable based on knowledge of typical geophysical

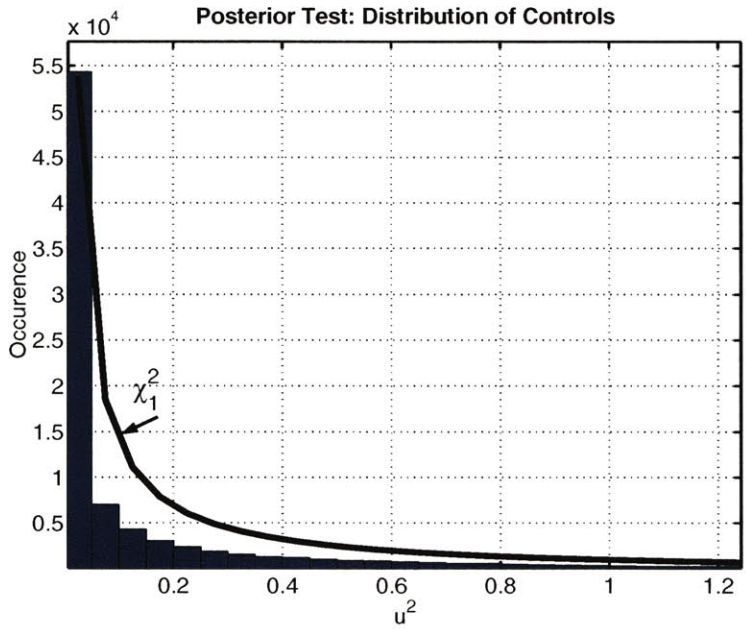

Figure 3-14: Distribution of squared control adjustments. The binned controls (*blue*) are compared to the prior error statistics, a chi-squared distribution with one degree of freedom (*black line*). The controls are correlated because the binned distribution is more strongly clustered around zero. Based on the knowledge of typical geophysical fields, the control variables should not be completely independent, and the posterior test seems acceptable.

fields. Atmospheric forcing fields, for example, should be correlated at large length-scales primarily due to the larger Rossby radius of deformation in the atmosphere. In conclusion, this posterior test successfully shows that the control adjustments have a reasonable size.

3.5 Fine-resolution optimization

Spatial resolution determines the degree of nonlinearity in many oceanographic models because frictional coefficients can be made smaller with higher resolution. Therefore, high-resolution simulations typically have a more-realistic Reynolds number and more-nonlinear dynamics. The arguments of Section 3.3 show that multiple solutions to the

least-squares problem are more likely to exist in this case. Multiple minima were seen in small-dimensional geophysical systems, like that of Ghil et al. (1991). In a quasi-geostrophic, three-layer double gyre model, the resolution directly affected the shape of the cost function (Köhl and Willebrand (2002), Figure 3-15). Coarse resolution models, which are nearly linear due to high viscosity, produced cost functions with a parabolic shape, but the high-resolution counterpart was irregularly shaped. Studies with geophysical turbulence (Tanguay et al. 1995) showed that small scales, where frequencies are highest, are likely to be most nonlinear in geophysical phenomena. Dynamics of different ocean regions also have distinct levels of nonlinearity. Assimilation of Gulf Stream eddies was successful over a three month window, but the optimization diverged for longer times (Schröter et al. 1993), which the authors prescribed to the model becoming "more nonlinear" with time. Prior to this thesis, the prospects for state estimation in the eastern subtropical gyre were unknown.

3.5.1 Chaos in geophysical systems

The quasi-geostrophic basin model of Lea et al. (2000), and the primitive equation model of Kohl and Willebrand (2003) were nonlinear and chaotic. Long time integrations reveal the differences between nonchaotic and chaotic nonlinear models. Nonlinear models generally lose sensitivity to initial conditions with increasing time, but chaotic models are exceptions. In addition to many local minima, cost functions from chaotic models behave like a discontinuous function (a Weierstrass function, McShane (1989)). Gradients do not give any useful information for a finite-sized neighborhood of phase space. Sensitivity of the initial conditions of a chaotic model remains indefinitely, but the conditions themselves are unrecoverable. For a successful optimization, long time integrations of chaotic models must be avoided.

Chaos in the Northeast Atlantic regional model?

A prerequisite for nonlinear chaos, as defined by Lea et al. (2000), is a model with instability. The Subduction Experiment region has relatively low levels of eddy energy and

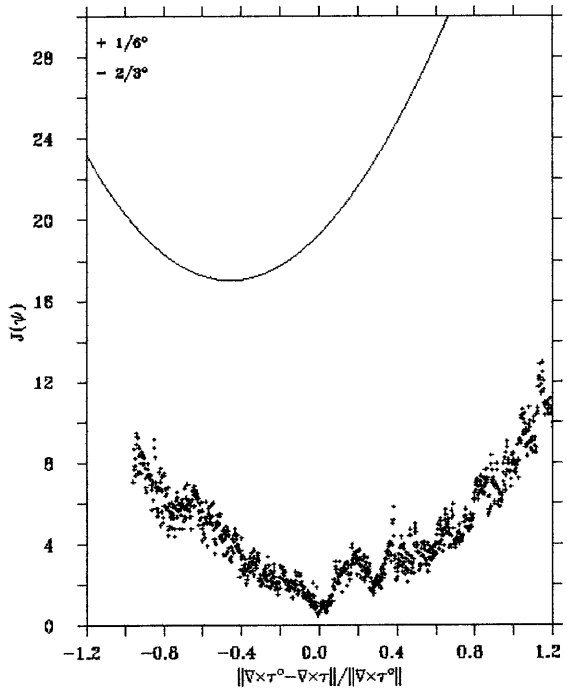

Figure 3-15: Cost functions from a fine and coarse-resolution quasi-geostrophic double gyre model with three levels. The cost function is the SSH misfit as a function of changes in the wind stress. The inference is that the 1/6° model is highly nonlinear. From Köhl and Willebrand (2002).

no western boundary currents; both imply a more linear dynamical regime. The general circulation model of this thesis is nonlinear and episodically unstable, as seen in two time integrations of the GCM with a small perturbation (Figure 3-16). A slight change in the control parameters leads to quasi-linear divergence of the model trajectories in phase space. 50-day episodes of exponential divergence suggest weak instability. Baroclinic instability is an intrinsically unstable element of ocean models which can explain the results here. With the weak nonlinearity of the Subduction Experiment region, it is unknown if the shape of the cost function with the eddy-resolving model is smooth or irregular.

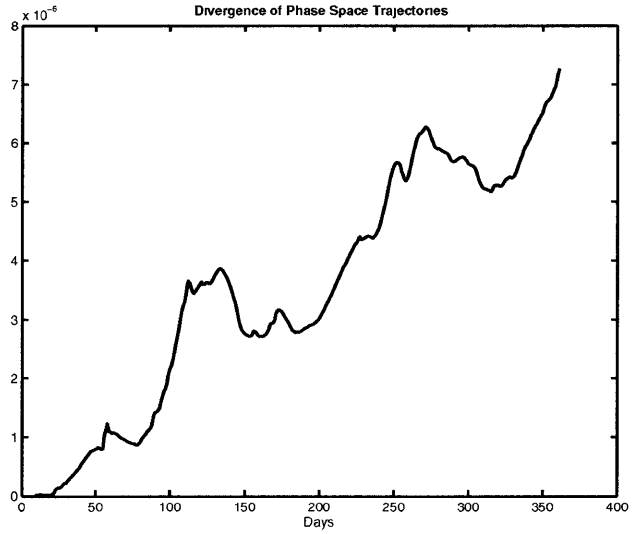

Figure 3-16: Two nonlinear integrations of the Subduction Experiment with a small perturbation to the control variables. The divergence between the two trajectories is defined as the sum of the squared difference in sea surface height, $\sum^{x,y}(SSH_1 - SSH_2)^2$. Exponential divergence occurs in short episodes, but the overall character is quasi-linear.

A direct check for the presence of chaotic dynamics can be done through the adjoint model. The adjoint model calculates the sensitivity to initial conditions; chaotic models have sensitivity which grows exponentially with increasing integration time. The time evolution of the Lagrange multipliers gives the time-evolution of the sensitivity. As shown in Section 3.2.5 and Appendix B, the Lagrange multipliers represent the sensi-

tivity to the model state at that particular time. Hence, the Lagrange multipliers three months before the final time represent the sensitivity of the initial conditions of a three month integration. The adjoint model has the same stability characteristics as the tangent linear model because the eigenvalues of a matrix and its transpose are the same. The reverse-time evolution of the Lagrange multipliers reflects this symmetry. The time evolution of the Lagrange multipliers of the regional GCM do not grow exponentially with time, instead they saturate in less than a year (Figure 3-17). The adjoint model is therefore stable over long time integrations and the dynamics of the system are not chaotic. For reference, the Lagrange multipliers of the episodically-unstable nonlinear pendulum have the same behavior (Figure 3-18). Based on this evidence, chaotic dynamics are not present in the Subduction Experiment model and the gradients of the cost function are calculable for long time integrations.

As previously mentioned, gradients computed from chaotic ocean models (Lea et al. 2000; Köhl and Willebrand 2003) had limited utility because of the nearly-discontinuous form of the cost function. Is the gradient, as computed by the eddy-resolving adjoint model of this study, relevant for finite perturbations of the control vector? This is a necessary condition for the gradient search method to succeed. The first iteration of the optimization can be used as a *gradient check*. By a Taylor series expansion:

$$J(\mathbf{u}) - J(\mathbf{u}^{(0)}) \;=\; \nabla J(\mathbf{u}^{(0)})^T \; (\mathbf{u} - \mathbf{u}^{(0)}). \tag{3.44}$$

The approximation potentially fails due to the parabolic and higher order terms in the cost function (Equation (3.8)), and also due to discontinuities in the neighborhood of $\mathbf{u}^{(0)}$. Also, use of a very small perturbation will be inaccurate due to cancellation errors and loss of significant digits (Griewank 2000). Using the first gradient, first cost function value, and a small perturbation of the controls, the expected cost function is calculable by (3.44). Then, the new controls are used to complete an integration of the nonlinear model and check the correspondence. Here, errors are usually 1%, although occasional point error values up to 50% occur. They are attributable primarily to the curvature of the cost function. Errors of this magnitude are acceptable, as the purpose of the

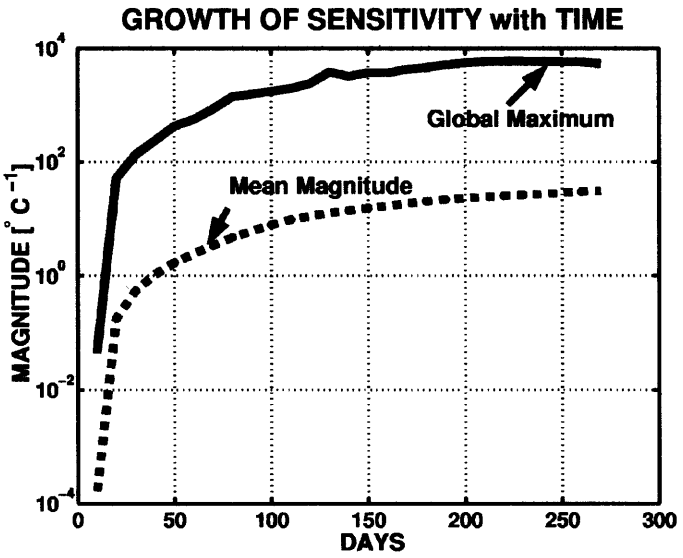

Figure 3-17: The evolution of the Lagrange multipliers (adjoint state) of the GCM with reversed time. The Lagrange multipliers are interpreted as the sensitivity of the initial conditions to a time integration of specific length. Both the maximum Lagrange mulitplier, $||\mu(t)||_\infty$, (*solid line*) and average magnitude of the Lagrange multipliers, $||\mu||_2$, (*dashed line*) are plotted.

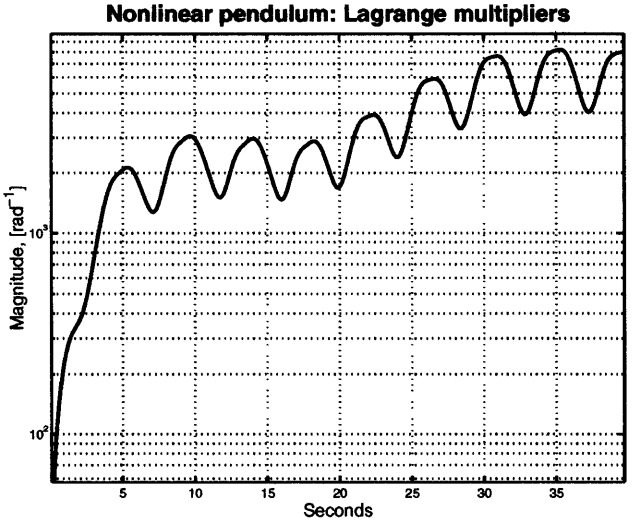

Figure 3-18: The evolution of the Lagrange multipliers of the nonlinear pendulum with reversed time. Here, the maximum Lagrange multiplier, $||\mu(t)||_\infty$, is plotted. The time evolution is not as steady as the GCM because no spatial averaging is possible with the nonlinear pendulum.

gradient is simply to point downhill. Exact gradient values are not necessary as long as a minimum is ultimately found. A successful gradient check shows that the eddy-resolving model of this region avoids some of the problems of previous chaotic models.

Another check of the gradient information is qualitative: does it look physically reasonable? In most cases, the gradients carry the signature of an adjoint Rossby wave traveling towards the eastern half of the basin. In addition, baroclinically-unstable bands appear to be more important, as inferred by Galanti and Tziperman (2002). This argument gives faith that the numerical machinery is accurately implemented.

In summary, the gradients calculated from the regional GCM contain useful information, but the cost function may have more than one stationary point. Therefore, local minima in the cost function still represent a concern which can slow or stop the convergence to an adequate solution to the least-squares problem. Because gradient descent finds the nearest minimum, the first-guess set of controls is extremely important. To remedy the possible convergence to an inadequate local minimum, the first guess of the controls comes from a regional, coarse resolution state estimate. Although the dynamics of the fine[7] resolution model are different, the coarse resolution estimate is expected to have some skill in predicting the ocean observations.

3.5.2 The first guess

Application of the 2° control adjustments to the 1/6° problem is hypothesized as a way to make a good first guess. But, do the coarse-resolution controls improve the eddy-resolving model? Two eddy-resolving model trajectories are compared: a run with zero control adjustments and another with coarse-resolution estimated controls. A comparison of the two cost function values (Table 3.2) shows the improvement by the coarse-resolution controls. These adjustments decrease the total observational cost function elements by 3%, primarily by bringing the model closer to the Levitus climatological temperature and Reynolds SST. This is a statement that the predictions

[7]Fine resolution and eddy resolution are used interchangeably to identify the 1/6° model and estimate.

Cost Function Element	Simulation	Coarse-resolution Controls
Mooring Temperature	2.24	2.01
Mooring Velocity	0.98	1.02
SSH Anomaly	1.32	1.24
SSH Mean	1.01	0.94
Levitus Temperature	2.06	1.82
Levitus Salinity	0.76	0.76
Reynolds SST	6.30	3.67

Table 3.2: Squared misfit of cost function terms normalized by their expected value. The expected value is computed by treating all small-scale motions as noise. Here, the comparison is made between two integrations of the eddy-resolving model, one with zero control adjustments (*Simulation*), the other with controls estimated from the coarse-resolution model (*Coarse-resolution controls*).

made by the coarse resolution model do carry over to the eddy-resolving case, and that the eddy-resolving model has some elements with quasi-linear dynamics. On the other hand, improvement of only 3% shows that the eddy field is still not simulated within observational uncertainty.

3.5.3 Fine-resolution cost and controls

Starting from coarse-resolution controls, the method of Lagrange multipliers is then applied to the eddy-resolving model. Improvement of the model trajectory comes at a slower pace due to the increased search space dimension. Nevertheless, the first-guess model run is near to the observations at the beginning, less than thirty iterations bring the large-scale state estimate within expected errors (Figure 3-19). The expected errors here include the entire eddy field; the observational terms are therefore downweighted. The first goal is to determine if any solution exists to the least squares problem. The solid red line reaches the normalized value of $J = 1$, corresponding a root mean square error that is equal to the *a priori* expected error. Therefore, the optimization finds a reasonable solution to the least-squares problem.

The fine-resolution optimization of the previous section fits the model to the large-scale data signal. In this case, mooring velocities offer much less information content

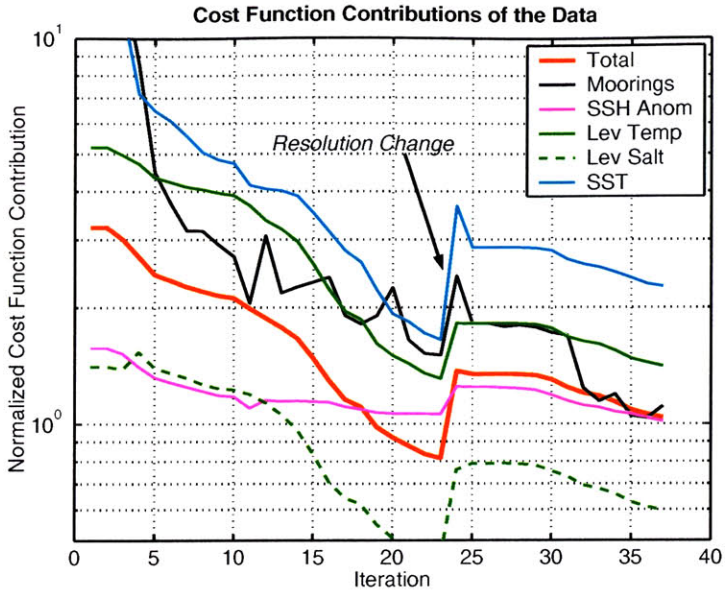

Figure 3-19: Normalized model-data misfit as a function of iteration of the search method for the coarse and eddy-resolving optimization. A value of 1 (10^0) is expected.

than the temperature profile. Velocity has a bluer spectrum than temperature, hence, much of the variability in velocity is explained by small-scale processes. Unresolved processes are considered noise here; in total, almost 50% of the signal in the velocity measurements is neglected here. Therefore, mooring velocities are a weak constraint which is easily predicted within observational uncertainty by the model. As seen here, the mooring temperature observations are much more valuable for constraining the large-scale circulation than point-measurements of velocity.

Fine-resolution controls

It is emphasized that the methodology here estimates all the control variables *simultaneously*, and no special means are necessary to control the open boundaries. This difference to the recent work of Ferron and Marotzke (2002) may be due to a better decomposition of open boundary velocity. The ill-conditioning present in open boundary estimation (as discussed in Section 2.4.2) is a possible cause for slow or stalled covergence to a solution.

Much like the coarse-resolution experiment (pictured in Figure 3-11), adjustments

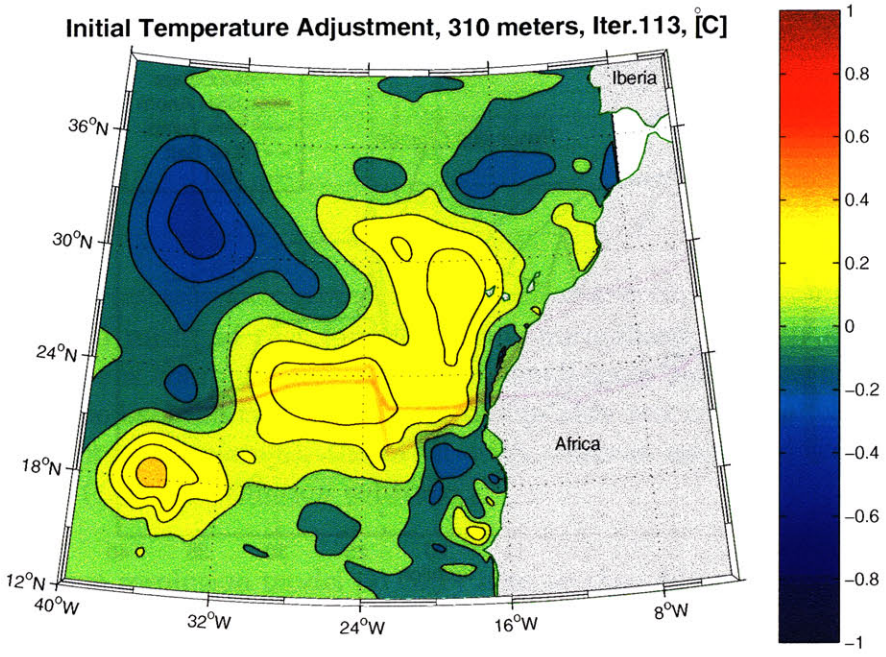

Figure 3-20: Initial temperature adjustment to bring the eddy-resolving estimate into consistency with the large-scale observational signal.

to the initial conditions and open boundaries have the most influence on the ocean circulation over one year. The estimated adjustment to the initial temperature is large-scale, and has a reasonable magnitude relative to the interannual variability of the ocean (Roemmich and Wunsch 1984) (Figure 3-20). The strong influence of the open boundary conditions is seen in a dye-release experiment in the forward model. Dye is constantly added at the lateral boundaries and allowed to advect and diffuse away. The result (Figure 3-21) is that almost half of the domain is affected by the boundaries in one year. Extrapolation suggests that the entire region would be covered by the passive tracer within three to five years. Therefore, the strong influence of the open boundary controls is expected.

Pitfalls in eddy-resolving optimization

Figure 3-19 shows a nearly monotonic decrease of the cost function with iteration. However, many intermediate steps failed due to numerical and physical problems; they are

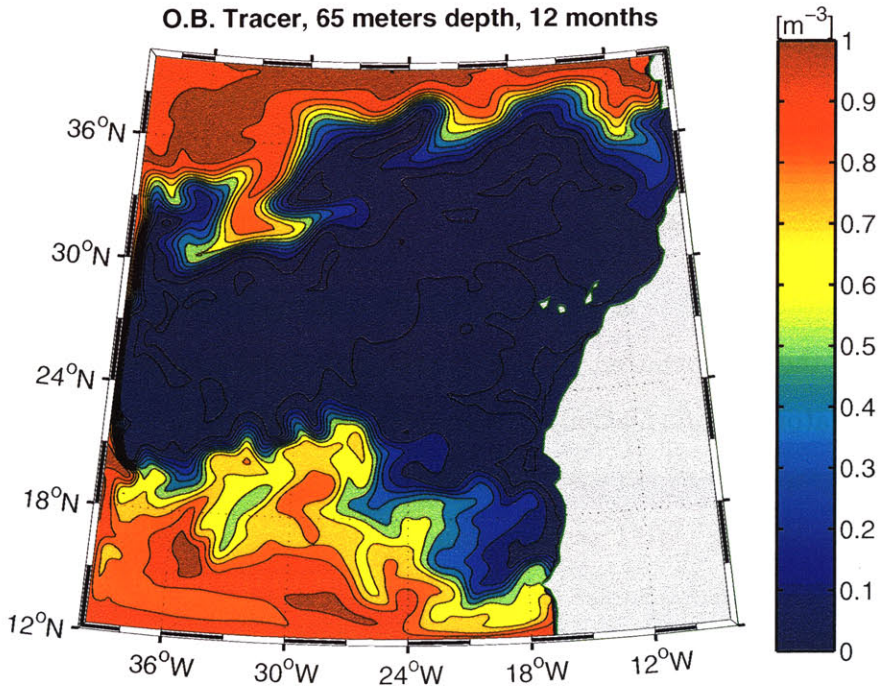

Figure 3-21: Tracer concentration $[m^{-3}]$ at 65 meters depth of a passive dye constantly released from the open boundaries with concentration 1 m^{-3}. This snapshot is taken one year after the initial release of dye. The contour interval is 0.1 m^{-3}.

catalogued in this section. Special cases arise when the gradient computed by the eddy-resolving adjoint model is less useful. The adjoint of the KPP boundary layer model is troublesome in shallow water and at depth due to the shear instability term. The solution here is to only use KPP in the forward model in the boundary layer, and to avoid simulating the shelf circulation. Also, the Hessian information calculated from the gradients is frequently not useful. For this reason, a steepest descent method periodically works better than the full quasi-Newton method. This is a clue that the underlying cost function topology is not well-represented by a paraboloid, in effect the topology is irregular. Most of these problems, now known, can be avoided in future optimizations.

3.5.4 Cross-validation

A stringent posterior test is to compare the state estimate with observations that were withheld from the optimization. *Cross-validation* tests the model's ability to be a dynamic interpolator: Is information accurately carried away from the observational sites? WOCE hydrographic sections exist in the same region and time as the Subduction Experiment. The WOCE AR11 section along $33°W$ was completed in November, 1992 (P.I. Joyce). The transect passes the western moorings at $19°N$ and $33°N$, but nearly 1500 km of ocean without hydrographic measurements separates the two.

The differences between the model simulation and the state estimate are biggest in the upper 100 meters (Figure 3-22). Because of the changes in upper ocean structure, the mixed-layer depth is deeper by 50-100 meters in the state estimate. The hydrography in other parts of the region also differs greatly between the two model runs. Mixed-layer depth is an essential quantity for subduction and must be modelled accurately to give any confidence in estimated subduction rates.

The state estimate visually appears to reproduce the observations to a greater extent, and error estimates confirm this assertion (Figure 3-23). In general, the upper layer hydrographic structure is significantly improved in the state estimate relative to the withheld WOCE hydrography; data-model misfits are no larger than $1 - 2$ $°C$. The unconstrained model simulation does not transport enough heat down into the water

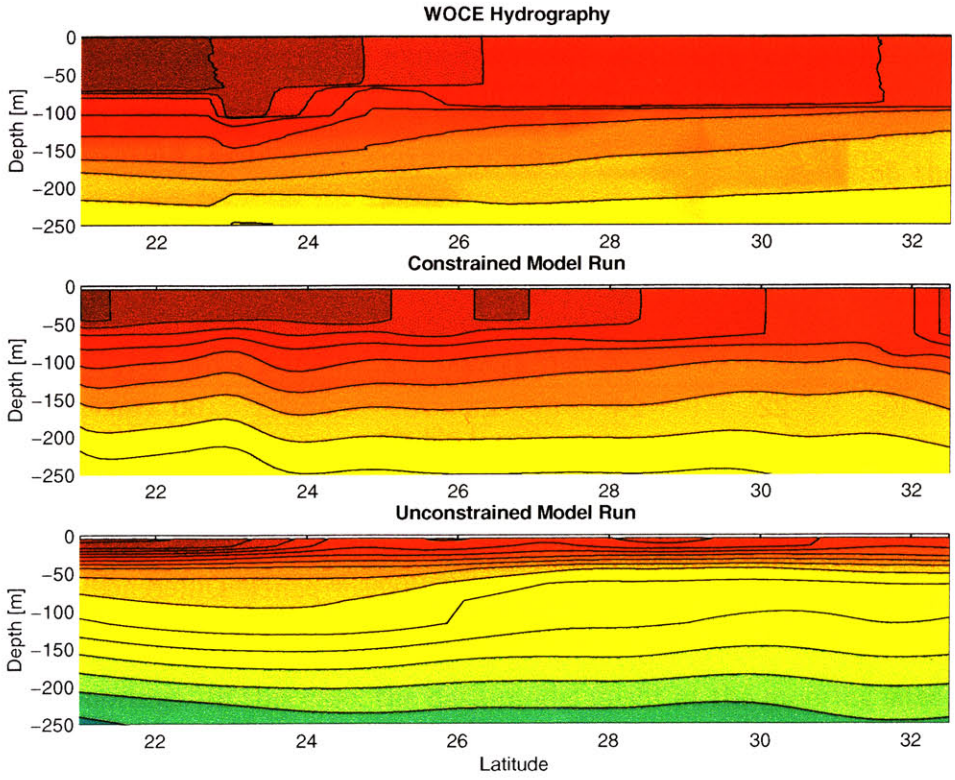

Figure 3-22: Meridional sections of potential temperature along the WOCE AR11 line ($33°W$) in November, 1992. *Top panel*: Observations from WOCE (courtesy of T. Joyce). *Middle panel*: Constrained model estimate. *Lower panel*: Unconstrained model simulation.

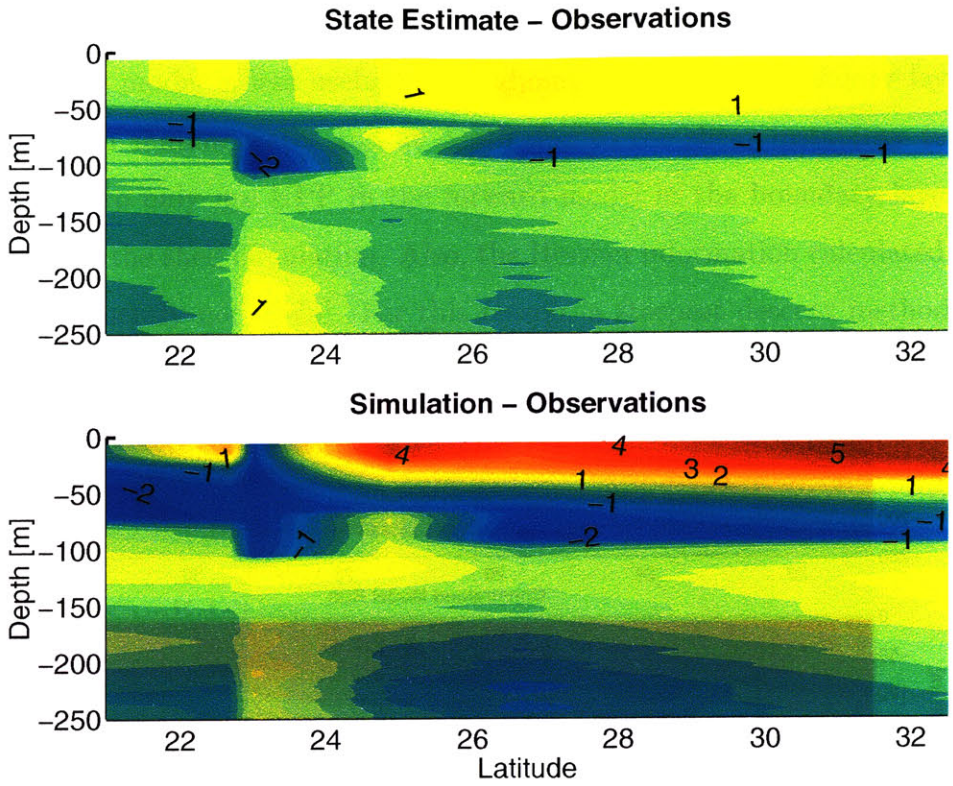

Figure 3-23: Error in potential temperature along the meridional WOCE AR11 line ($33°W$) in November, 1992. *Upper panel*: Difference between the state estimate (constrained model) temperature and observations. *Lower panel*: Difference between model simulated temperature (no data constraint) and observations.

column, and hence, is $4 - 5\ °C$ warmer than the observations at the surface. This success of the model in reproducing withheld data lends confidence to the state estimate throughout the entire domain, even away from sites of observations. Although the state estimate is an improvement, systematic errors do remain. Estimated surface temperature is as much as $1\ °$ warmer than observed, yet is erroneously cold at the base of the mixed layer ($50 - 100$ meters depth). The modelled physics of the mixed layer lead to this deficiency.

3.5.5 Tracking eddies

After finding a state estimate consistent with the large-scale observational signal, the next question is whether a model can be constrained to both the large and small-scale data signal. If the answer is affirmative, then individual eddies can be tracked, insofar as they are observed. The technical implementation of this new problem is very similar to the previous experiment, only the observational weights must be increased in order to correspond to the decrease in the expected errors. Although the mathematical transformation between the two problems is straightforward, the new least-squares problem poses a more stringent test than the original. Finding the model solution that fits the large scale signal alone is roughly equivalent to the study of Köhl and Willebrand (2002), where statistical characteristics were constrained. Tracking individual eddy trajectories is a more demanding task, and one in which the existence of any solution can not be determined *a priori*.

Optimization of the full cost function with stringent weights frequently stalls in control space. Changing the weights usually results in further improvements of the model trajectory. One particular change is to only weight the mooring terms in the cost function. This is a somewhat simpler test for the method: Fit the full observational signal of the moorings. In this case, the data-model misfit at the mooring sites decreases from 7.6 σ to 1.8 σ where σ is the expected error (Figure 3-24). The gradient information looks plausible and a slow rate of convergence is kept. Approximately 150 iterations of the forward-adjoint model are probably needed for complete consistency. Physically, the state estimate time-series at the Central mooring resembles the results of Spall et al. (2000); vertical diffusion transfers the warm summertime surface temperature to greater depth after a few months.

Estimates of the initial eddy field

What control adjustments allow eddies to be tracked away from the moorings? Analysis of the adjoint-calculated gradient shows two bands of increased sensitivity to the cost function: the Azores Current and the North Equatorial Current. Previous studies,

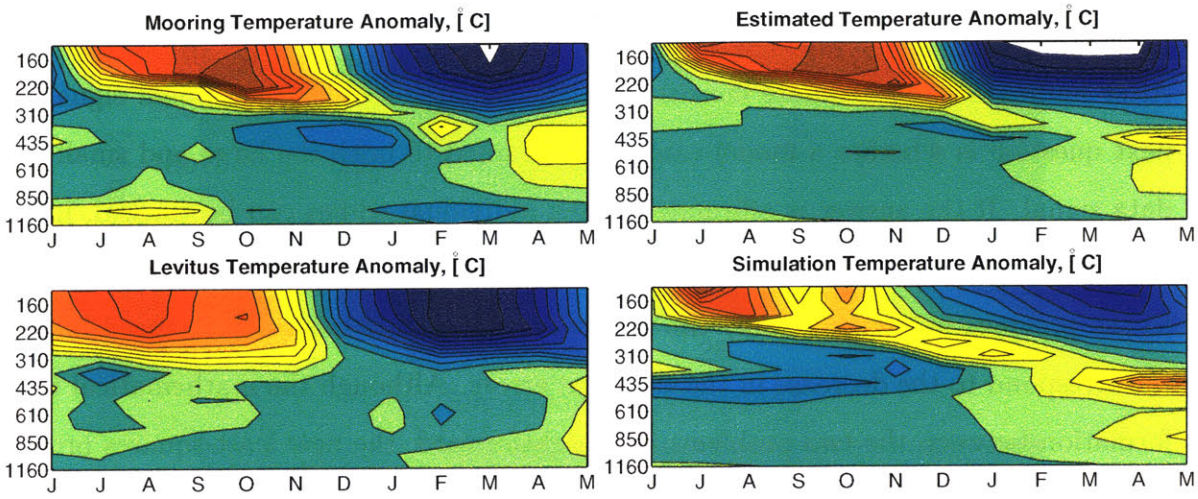

Figure 3-24: Four depth-time diagrams of potential temperature at the Central Mooring site from June 1, 1992, to June 1, 1993. *Top left*: Mooring observations, *bottom left*: Levitus climatology, *top right*: Constrained model estimate, *bottom right*: Unconstrained model simulation. The constrained model estimate accurately depicts the timing of and magnitude of the seasonal cycle, unlike the unconstrained model.

including Gill et al. (1974), have shown the basic state North Equatorial Current to be baroclinically-unstable. The Azores Current is also a source of eddy energy, as seen in the TOPEX/POSEIDON altimeter measurement. Baroclinic instability is theorized to increase the sensitivity of these regions (Galanti and Tziperman 2003), because eddies can grow and transport information away from their formation site. In the optimization here, small perturbations in the initial conditions lead to large changes in the eddy field at later times (Figure 3-25). Furthermore, the mooring contribution of the cost function is most sensitive to initial temperature. Finally, it should be noted that the estimated eddy field still has low levels of kinetic energy during the first two months; the spin-up problem has not been completely solved by state estimation.

3.6 Summary

There is no fundamental obstacle to constraining an eddy-resolving model to observations in this region of the ocean. Here, a state estimate consistent to the large-scale

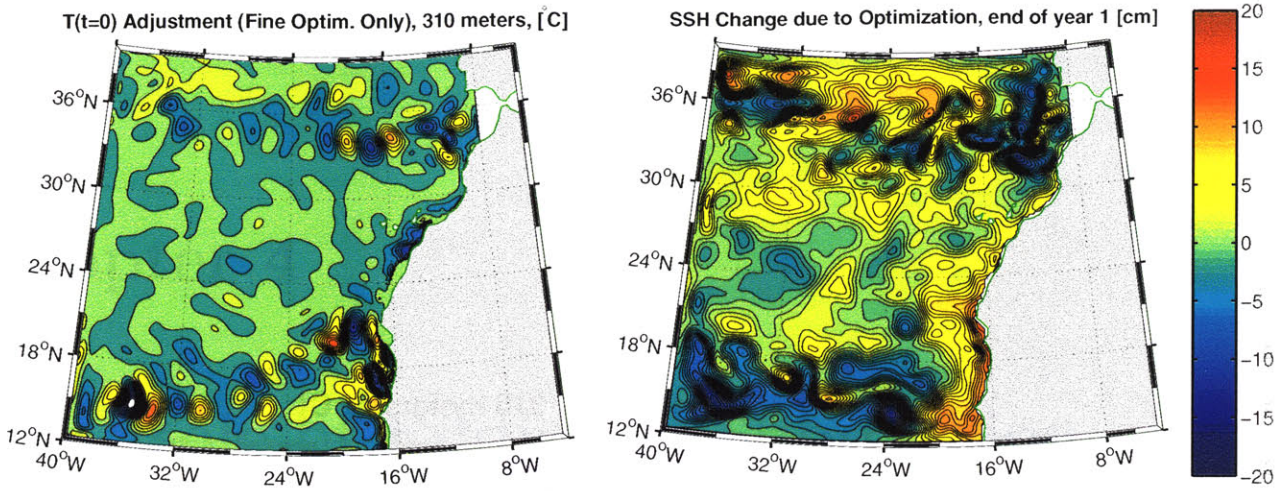

Figure 3-25: *Left*: Initial temperature adjustment from the optimization of the small-scale observational signal. *Right*: Rearrangement of the sea surface height field after one year by the initial temperature adjustment.

signal in all observations is found. Furthermore, small-scale motions observed by the moorings are capably reproduced by the state estimate, as well. Individual eddies are tracked insofar as they influence the mooring sites. The search for these state estimates is helped by the following conditions:

- The eastern subtropical gyre is more quiescent than the western boundary of the basin, where strongly nonlinear features exist.
- A coarse-resolution model skillfully simulates much of the large-scale ocean circulation, and can be used to eliminate major biases in an eddy-resolving model.

The result is a time-evolving, three-dimensional estimate of the ocean circulation which reasonably fits a wide variety of available information and exactly follows the dynamics of the MIT General Circulation Model (Figure 3-26). In addition, we now have improved estimates of the initial eddy field, open boundary conditions, wind stresses, and air-sea fluxes. The state estimate is ideal for the study of the role of eddies in subduction because it is dynamically consistent and it explicitly resolves eddy-scale motions.

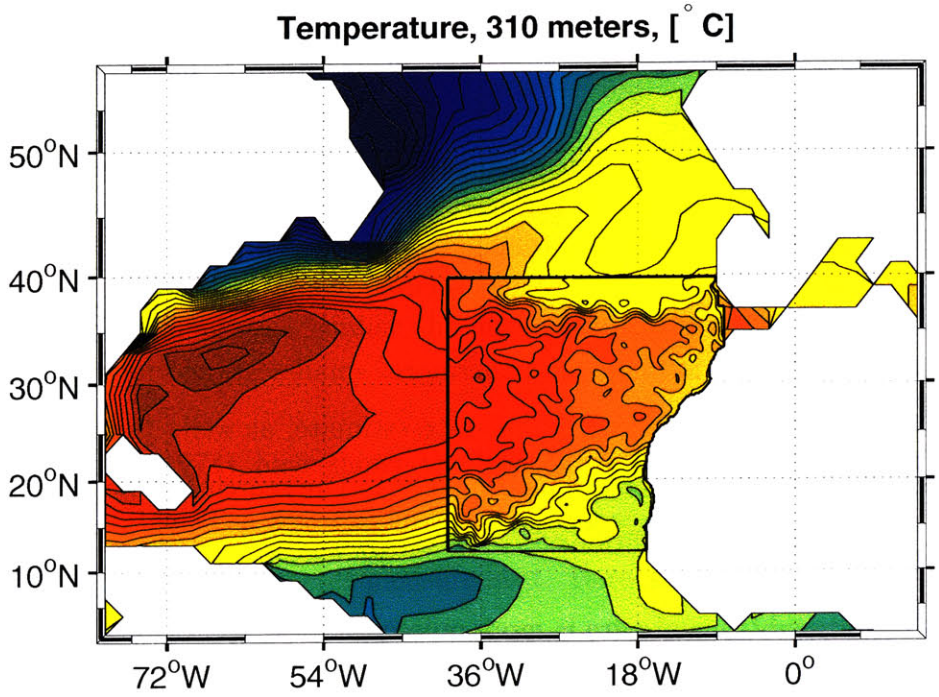

Figure 3-26: Nested view of the 1/6° regional state estimate inside the 2° ECCO state estimate. Potential temperature at 310 meters depth, with a contour interval of 1°C, is shown. The boundary between the two estimates (*thick black line*) is discontinuous in temperature because of the open boundary control adjustments.

Chapter 4

The Role of Eddies in Subduction

4.1 Overview

Chapter 4 quantifies the many processes that subduct water in the eastern North Atlantic Ocean. Vertical velocity in the upper ocean, usually attributed to Ekman pumping (Montgomery 1938), is a logical candidate to transport surface waters into the main thermocline. Vertical velocity, in conjunction with the seasonal cycle of mixed-layer depth, biases the properties of subducted water to late-winter values through the so-called "mixed-layer demon" (Stommel 1979). A less obvious process, at least until the work of Woods (1985), is the subduction of water through a sloping mixed-layer base, termed "lateral-induction". Relatively little is known about its importance of mesoscale eddies and the associated "eddy subduction" in the eastern half of the subtropical gyre, although both theory (Marshall 1997) and studies of other regions (Hazeleger and Drijfhout 2000) suggest that eddies are important. In short, subduction can be caused by a combination of Ekman pumping, surface buoyancy forcing, horizontal flow across a sloping surface, and mesoscale eddies, but their relative regional importance has not been quantified.

The state estimate of Chapter 3 is ideal for this study because it is dynamically self-consistent and physical mechanisms can be associated with the movement of subducted water. In addition, eddy-scale motions are resolved, and hence, eddy subduction can be

explicitly diagnosed. Observations alone do not provide the necessary spatial coverage to diagnose all the processes that lead to subduction. Models do provide adequate coverage of spatial fields, but previous studies have either been in an idealized context (Marshall 1997; Hazeleger and Drijfhout 2000), or they have not explicitly resolved eddy motions (Marshall et al. 1999; Spall et al. 2000). Furthermore, no study of subduction has ever been conducted with a model-observation synthesis, such as the product of Chapter 3.

After reviewing the physical process that subduct water (Section 4.2), Chapter 4 diagnoses subduction rates in the state estimate. To understand the impact of subduction on particular water masses, calculations are primarily done in a density coordinates. The challenge of isopycnal analysis of a level-coordinate information source is addressed in Section 4.2.4. Density-coordinate calculations, in conjunction with spatial maps of subduction, show that the Azores Current and the North Equatorial Current are prospective sites of significant eddy subduction.

4.2 Kinematics of subduction

Subduction is the transfer of fluid from the mixed layer into the thermocline, contingent that it does not become re-entrained to the mixed layer later in the same year. This kinematic definition of subduction is inherently Lagrangian; water parcels are followed throughout a seasonal cycle. In this way, subducted water only refers to permanently subducted water; that is, water must pass below the maximum mixed-layer depth of that particular year. The philosophical bias of this study is that subduction is a quantity of interest only over long time periods, such as an integer number of seasonal cycles. This is in contrast to some previous authors (i.e., Cushman-Roisin 1987) who do not make any distinction between detrainment and subduction.

Mathematically, subduction is a intimately related to entrainment and detrainment of the mixed layer. The instantaneous rate of water exchange between the mixed layer

and the underlying stratum is quantified as an *entrainment velocity*, w_*:

$$w_* = \frac{\partial h}{\partial t} + w_h + \mathbf{u}_h \cdot \nabla h \tag{4.1}$$

where h is the depth of the mixed layer, w_h is the vertical velocity at depth h, and $\mathbf{u}_h$ is the two-component horizontal velocity at depth h. The entrainment velocity gives the rate of volume transfer per unit horizontal area, usually expressed in units of meters per year. The symbol, w_*, follows the convention for cross-interface volume flux from Pedlosky (1996), although he mostly dealt with isopycnal surfaces, and the mixed-layer base need not correspond to one. Positive entrainment velocity represents entrainment, and conversely, negative w_* is detrainment. Note that entrainment can occur without any vertical velocity or any physical movement of water. One example is a resting ocean with a shoaling mixed layer. The volume of the mixed layer decreases and water is detrained, but individual water particles do not move. Horizontal velocity through a sloping mixed-layer base, termed *lateral induction*, also entrains or detrains water. From above, detrainment from the mixed layer is not completely controlled by Ekman pumping, nor is it a purely vertical process.

The volume of detrained water is the integral of the entrainment velocity over space and time:

$$V \equiv \int^t \int^{\mathcal{A}} (-w_*)\, d\mathcal{A}\, dt, \tag{4.2}$$

where the script $\mathcal{A}$ is the horizontal area[1] of interest. If the area of integration is restricted to regions of detrainment, then V will be positive, indicating a volume of water transferred out of the mixed layer. Without the specification that $w_* < 0$ in (4.2), the volume of detrained water could be negative, corresponding to a conversion of water from the main thermocline to the mixed layer. Next, the distinction between detrainment and subduction is elucidated.

When integrating over an integral number of seasonal cycles, (4.2) calculates the *net volume of subducted water*. Because a full seasonal cycle is considered, only water

[1]Areas are always denoted by a script $\mathcal{A}$ in this thesis in order to reserve A for cross-isopycnal advective flux (Garrett et al. 1995).

that escapes the mixed layer without being re-entrained is counted in V. Therefore, the result of competition between detrainment and entrainment yields net subduction.

4.2.1 Subduction and the seasonal cycle

To the first order, the seasonal cycle of the mixed layer controls subduction. The rate of change of the mixed-layer depth, $\partial h/\partial t$, typically dominates the terms in (4.1). As a consequence, most of the water detrained from the mixed-layer is re-entrained later in the year by the deepening mixed layer. Only water which is detrained in the late winter and early spring, when the mixed layer is deepest, remains in the main thermocline. Stommel (1979) first pointed out this effect, now termed the "mixed-layer demon", which explains the bias of the main thermocline to late-winter mixed-layer properties (also recall Figure 1-1). Williams et al. (1995) confirmed that this process is active in a primitive equation model at coarse resolution. In the state estimate here, a first step of Section 4.4 is to check if the mixed-layer demon is operating.

4.2.2 Water-mass subduction rates

One goal of this thesis is to determine how subduction sets the water-mass distribution of the main thermocline. A "water mass" is defined as the water in a particular potential-density class. Here, the rate of injection of a water mass into the main thermocline is categorized as a function of σ, the potential density referenced to the surface. Following the notation of Marshall (1997), the net volume flux across the mixed-layer base at a density less than σ is $S(\sigma, t)$:

$$S(\sigma, t) \equiv \int^{\mathcal{A}(\sigma,t)} (-w_*)\, d\mathcal{A}, \tag{4.3}$$

where $\mathcal{A}$ is the surface area of the mixed-layer base at density[2] less than σ. The advantage of density-coordinate analysis is that any water-mass can be isolated. For example, net volume flux in an infinitesimal density band, σ to $\sigma + \delta\sigma$, is δS:

$$\delta S = \frac{\partial S}{\partial \sigma}\delta\sigma, \tag{4.4}$$

the divergence of $S(\sigma, t)$ with respect to density. From (4.4), the net volume flux in a finite density interval b to a is given by the difference, $S(a,t) - S(b,t)$. As seen above, the power of density-space calculations is that any water mass can be individually examined.

Equation (4.3) is based upon the instantaneous detrainment rate, but previous investigators (Marshall and Marshall 1995; Hazeleger and Drijfhout 2000) have emphasized that entrainment is not always equal to subduction because of the intricacies of the seasonal cycle. However, when averaging over at least one seasonal cycle, the net volume flux into the main thermocline is equal to the volume of subducted water. Therefore, the *water-mass subduction rate*, $\overline{S}(\sigma)$, is the average volume flux:

$$\overline{S}(\sigma) \equiv \frac{\int^t \int^{\mathcal{A}(\sigma,t)} (-w_*)\; d\mathcal{A}\; dt}{\int^t dt}, \tag{4.5}$$

where t is an integer number of seasonal cycles. Water-mass subduction rates are easily compared to other oceanographic quantities. It is a volume transfer per unit time, and is large enough over an ocean basin to justify the use of Sverdrups ($1\ Sv = 10^6\ m^3/s$) for units[3]. In this thesis, subduction rates are only defined as an average over an entire seasonal cycle, and the time integral in (4.5) must be sufficiently long.

[2]There is a slight distinction between the definition of $S(\sigma, t)$ here, and $M(\sigma, t)$ of Marshall et al. (1999). Marshall et al. (1999) define $M(\sigma, t)$ as the volume flux between potential density σ_1 and σ. Here, based on the practical approach of setting σ_1 to a very small value, $M(\sigma, t)$ is defined as the volume flux at all potential densities less than σ.

[3]Upper-case variables are used for volume fluxes with units of Sverdrups. Volume fluxes per unit horizontal area, in units of m/s or m/yr, are denoted with lower-case variables.

4.2.3 Eddy contributions to subduction

Water is subducted by both mean and time-variable components of the circulation, although this was not explicitly noted in Equation (4.2) or Equation (4.5). Here, we derive the explicit contribution of eddies to the water-mass subduction rate. Following Marshall (1997), Equation (4.5) is rewritten and expanded into mean and time-variable components of the circulation:

$$\overline{S}(\sigma) = - \overline{w_* \mathcal{A}} = - \overline{(\overline{w}_* + w'_*)\ (\overline{\mathcal{A}} + \mathcal{A}')}, \tag{4.6}$$

where the overbar defines a running mean over one or more seasonal cycles. By the rules of averaging,

$$\overline{S}(\sigma) = - \ (\ \overline{w}_* \overline{\mathcal{A}} + \overline{w'_* \mathcal{A}'}\ ). \tag{4.7}$$

The mean and deviation of entrainment velocity must be carefully defined. Specifically,

$$\overline{w}_* = \frac{\partial \overline{h}}{\partial t} + \overline{w}_h + \overline{\mathbf{u}}_h \cdot \nabla \overline{h} + \overline{\mathbf{u}'_h \cdot \nabla h'}, \tag{4.8}$$

$$w'_* = w_* - \overline{w}_*. \tag{4.9}$$

Notice that the mean entrainment velocity includes a contribution from the correlation between mixed-layer depth and velocity variations. Expanding all the terms, the total water-mass subduction rate is:

$$\overline{S}(\sigma) = - \left\{ \frac{\partial \overline{h}}{\partial t} + \overline{w}_h + \overline{\mathbf{u}}_h \cdot \nabla \overline{h} + \overline{\mathbf{u}'_h \cdot \nabla h'} \right\}\ \overline{\mathcal{A}} - \overline{w'_* \mathcal{A}'}. \tag{4.10}$$

In particular, the third term in the lateral induction term, first emphasized by Woods (1985). The fourth term is an eddy thickness flux across the moving mixed-layer base. The last term on the right hand side, originally noted by Marshall (1997), shows that correlations between the local subduction rate and the surface outcrop area also subduct water. As seen above, the rate at which water is transferred into the main thermocline need not be set by the Eulerian mean quantities alone.

To isolate the role of eddies, Marshall (1997) defined the eddy component of subduction by grouping all the time-variable terms in (4.10):

$$S_{eddy}(\sigma) = -(\overline{\mathbf{u}_h' \cdot \nabla h'})\ \overline{\mathcal{A}} - \overline{w_*' \mathcal{A}'}, \tag{4.11}$$

where the entire term represents a bolus flux by eddies. In an eddying circulation, subduction by the mean flow may not be the only relevant quantity for understanding water mass distributions. Follows and Marshall (1994) estimated that eddy fluxes across typical oceanic fronts drive subduction with magnitudes comparable to the mean flow. In the Antarctic Circumpolar Current, the subduction of Antarctic Intermediate Water is not adequately captured by mean subduction rates (Marshall 1997). Furthermore, subduction in the Gulf Stream system is dominated by eddy-scale motions with rates up to 150 m/yr (Hazeleger and Drijfhout 2000). These are all cases where *eddy subduction* is a non-negligible component of the total subduction.

4.2.4 Surface layer volume budget

The water-mass subduction rate, introduced above, is only one part of a larger volume budget. The complete surface layer volume budget is performed for two reasons. One, the influence of eddy motions on subduction can be estimated in an independent way through other terms in the budget. Two, isopycnal budgets allow a connection between kinematics and thermodynamics by the combined usage of conservation of volume and buoyancy. Through a density-space approach, the similarity between subduction and *transformation*, the flow of water across isopycnals, is clearly seen. In summary, estimates of subduction through isopycnal budgets give another way to quantify the importance of eddy motions in individual density classes.

The "surface layer" is defined as the seasonally-varying part of the ocean – everywhere shallower than the maximum mixed-layer depth. The volume of surface layer with potential density less than σ can be changed by volume flux through four boundaries, $\mathcal{A}_{th}$, $\mathcal{A}_\sigma$, $\mathcal{A}_B$, and $\mathcal{A}_S$ (Figure 4-1). The four boundaries are the base of the mixed layer,

an isopycnal, the domain boundary, and the sea surface, respectively. Therefore, the volume budget of the surface layer over one year is:

$$\frac{\partial V(\sigma,t)}{\partial t} = M_B(\sigma,t) - A(\sigma,t) - S(\sigma,t), \tag{4.12}$$

where $V(\sigma,t)$ is the density-class volume, $M_B(\sigma,t)$ is the volume flux through the open boundaries, and $A(\sigma,t)$ is the diapycnal volume flux (to be explicitly defined in the next section). Volume flux through the surface by evaporation and precipitation is much smaller than the other fluxes, and neglected in (4.12). For a steady-state ocean basin, Marshall et al. (1999) reduced (4.12) to show the direct relationship between subduction and diapycnal advective fluxes:

$$\overline{S}(\sigma) = -\overline{A}(\sigma). \tag{4.13}$$

This thesis explores the extent to which the above relationship holds in the Northeast Atlantic. Equation (4.13) is important because the eddy component of $\overline{A}(\sigma)$ must have some relation to $S_{eddy}(\sigma)$.

Due to the mixed-layer demon, only water which moves out of the seasonally-varying ocean into the main thermocline is permanently subducted. The boundary between the seasonal and main thermocline is defined as the maximum mixed-layer depth in the year, usually occurring in late winter. Therefore, we choose the bottom boundary of the "surface layer" to be fixed at the maximum mixed-layer base over one year.

Terms of the volume budget

When considering the water-mass subduction rate across a fixed bottom boundary, the problem is simplified dramatically. The area-normalized volume flux across the deepest mixed-layer base is usually called the *annual subduction rate* (Cushman-Roisin 1987; Nurser and Marshall 1991; Marshall et al. 1993), and is defined:

$$s_{ann} \equiv \frac{V_{ann}}{\int^{1\ yr}\int^{\mathcal{A}} d\mathcal{A}\ dt} = \frac{-(\overline{w}_H + \overline{\mathbf{u}}_H \cdot \nabla H)\cdot(\overline{\mathcal{A}})\cdot(1\ yr)}{(\overline{\mathcal{A}})\cdot(1\ yr)}, \tag{4.14}$$

where V_{ann} is the volume of subducted water, H is the depth of the maximum mixed-layer depth, $\overline{\mathbf{u}}_H$ and $\overline{w}_H$ are the mean velocities at depth H, and the overbar represents a one-year mean. Simplifying (4.14), a relationship of diagnostic utility results:

$$s_{ann} = -(\overline{w}_H + \overline{\mathbf{u}}_H \cdot \nabla H). \tag{4.15}$$

Therefore, the net volume flux across a fixed lower boundary is:

$$S(\sigma, t) = -\int^{\mathcal{A}_{th}} (\overline{w}_H + \overline{\mathbf{u}}_H \cdot \nabla H)\, d\mathcal{A}, \tag{4.16}$$

by use of Equation (4.3) and Equation (4.14).

The next term of the volume budget, Equation (4.12) is the open boundary mass source, M_B:

$$M_B(\sigma, t) = \int^{\mathcal{A}_B(\sigma,t)} \mathbf{v}_B \cdot \hat{\mathbf{n}}_B \, d\mathcal{A}, \tag{4.17}$$

where $\mathcal{A}_B(\sigma, t)$ is the surface area of the open boundary at density less than σ, $\mathbf{v}_B$ is the open boundary velocity, and $\hat{\mathbf{n}}_B$ is the direction normal to the boundary. The addition of open boundary terms to the works of Walin (1982) and Speer and Tziperman (1992) is a novel development of this thesis. Conceptually, the lateral boundary can be thought of as a special case of the bottom boundary where $H(x, y) = 0$ outside the domain of interest. This technique is used to derive the open boundary modifications.

The diapycnal advective volume flux, $A(\sigma, t)$, is defined as:

$$A(\sigma, t) = \int^{\mathcal{A}_\sigma(\sigma,t)} (\mathbf{v} - \mathbf{v}_\sigma) \cdot \hat{\mathbf{n}}_\sigma \, d\mathcal{A}, \qquad \hat{\mathbf{n}}_\sigma \equiv \nabla\sigma / |\nabla\sigma|, \tag{4.18}$$

where $\mathbf{v}$ is the fluid velocity, $\mathbf{v}_\sigma$ is the isopycnal velocity, and $\hat{\mathbf{n}}_\sigma$ is the direction normal to the isopycnal. The diapycnal volume flux is calculated following the meandering isopycnal. A is positive for flow across isopycnals to higher density. The cross-isopycnal advective flux gives the sum influence of the interior ocean dynamics on the water-mass volume budget.

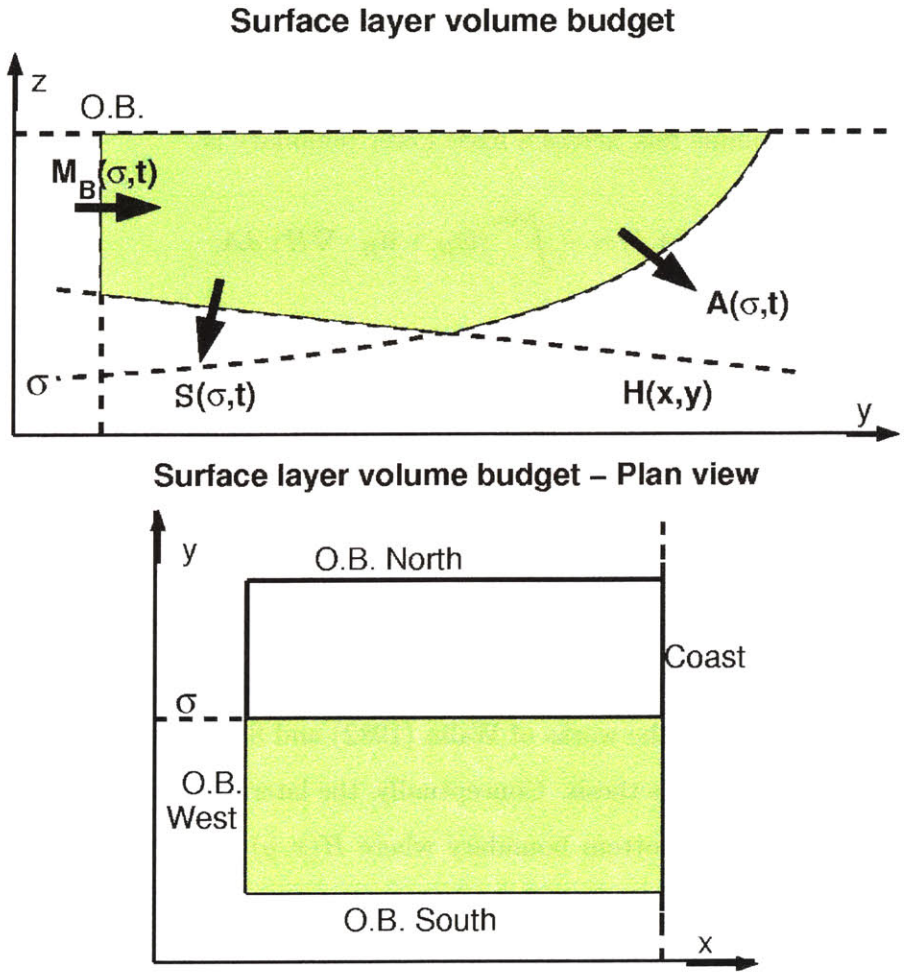

Figure 4-1: The shaded fluid is bounded by an isopycnal, the maximum mixed-layer depth, $H(x, y)$, the sea surface, and the regional boundary. These surfaces have areas $\mathcal{A}_\sigma(\sigma, t)$, $\mathcal{A}_{th}(\sigma, t)$, $\mathcal{A}_S(\sigma, t)$, and $\mathcal{A}_B(\sigma, t)$. The diapycnal volume flux through the isopycnal σ is $A(\sigma, t)$, the volume flux across $H(x, y)$ into the interior of the ocean is $S(\sigma, t)$, and the open boundary volume flux is $M_B(\sigma, t)$.

4.2.5 Eddy diapycnal fluxes

The goal of this section is to determine the role of eddies in subduction in individual density classes. The previous arguments have shown, through the conservation of volume, that there is a direct connection between subduction and cross-isopycnal flow in the surface layer. In this framework, eddies contribute to cross-isopycnal volume fluxes, and thus, affect subduction. An estimate of the importance of eddies in the volume budget of the upper ocean is presented next.

The effect of time-variable motions, excluding the seasonal cycle, is isolated by defining a term, A_{eddy}:

$$A_{eddy}(\sigma,t) = \int^{\mathcal{A}_\sigma(\sigma,t)} (\mathbf{v} - \mathbf{v}_\sigma)\cdot\hat{\mathbf{n}}_\sigma \; d\mathcal{A} - \int^{\overline{\mathcal{A}}_\sigma(\sigma,t)} (\overline{\mathbf{v}} - \overline{\mathbf{v}}_\sigma)\cdot\overline{\hat{\mathbf{n}}}_\sigma \; d\overline{\mathcal{A}} \tag{4.19}$$

where the overbar represents a running mean of one month. Although other definitions of A_{eddy} are possible, this definition closely isolates the effect of eddies that must be parameterized in a coarse-resolution model. Furthermore, A_{eddy} is another way to study the eddy contribution to the water-mass subduction rate – an indirect way to calculate S_{eddy} of Section 4.2.3.

4.2.6 Surface layer buoyancy budget

Water crosses isopycnals only in the face of diabatic processes (Figure 4-2). Consequently, there is a direct connection between kinematics and thermodynamics. This thesis introduces the thermodynamics of the surface layer for two reasons. One, diapycnal fluxes, computed earlier in the kinematic section, may also be inferred by complementary buoyancy fluxes. An independent check on the diagnostic method is then possible. Two, the addition of thermodynamics allows a more complete picture of the ocean processes at work.

Transformation rates

The buoyancy budget of the surface layer control volume follows the same geometry of Figure 4-1. In an ocean where the only diabatic forcing is at the surface, the buoyancy flux, $\mathcal{B}$ is comprised of freshwater flux and heat flux:

$$\mathcal{B}(x,y) = \frac{g}{\rho_0}\left(\frac{\alpha}{C_w}H_Q + \rho_0 \beta S\ H_F\right), \tag{4.20}$$

where H_Q is heat flux, H_F is freshwater flux, α is the thermal expansion coefficient, β is the haline expansion coefficient, and C_w is the specific heat capacity of seawater. Based on the conservation of volume and simple thermodynamics, the diapycnal advection of buoyancy is balanced by the surface forcing in an isopycnal layer:

$$A(\sigma,t) = F(\sigma,t) \equiv \frac{\partial B(\sigma,t)}{\partial \sigma}, \quad where \quad B(\sigma,t) = \int^{\mathcal{A}_s} \mathcal{B}(x,y)\ d\mathcal{A}, \tag{4.21}$$

and $F(\sigma,t)$ is the average *water-mass transformation rate.* The water-mass transformation rate is the buoyancy convergence in a particular density band, which can be interpreted as the rate that fluid moves across an isopycnal due to surface buoyancy forcing. Walin (1982) used (4.21) to diagnose the "poleward drift" of the upper ocean from a climatology of surface fluxes. Speer and Tziperman (1992) later calculated $F(\sigma,t)$ with climatological datasets. They used this definition to identify the water-mass formation rates of the North Atlantic Ocean.

In oceanographic datasets, diagnosed transformation rates, $F(\sigma,t)$, differ largely from diapycnal fluxes, $A(\sigma,t)$ (Speer and Tziperman 1992; Garrett et al. 1995), primarily because interior ocean dynamics are neglected in the buoyancy budget. In the surface layer region, non-advective buoyancy input is balanced by the advective component. However, the presumed balance in (4.21) is not complete because diffusion is ignored. Non-advective buoyancy input is due to surface buoyancy forcing, B, diffusion across isopycnals, D_σ, diffusion across the mixed-layer base, D_{ml}, and, in the regional case, diffusion across the open boundaries, D_B. In this way, the total supply of buoyancy by

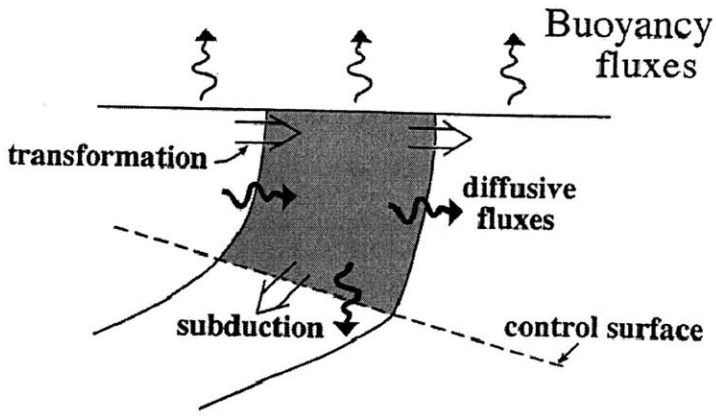

Figure 4-2: Diagram of a fixed control region of the mixed layer bounded by the sea surface and two isopycnals. Both volume and buoyancy budgets can be formed in the shaded region, leading to a set of diagnostics in density space. Subduction is the volume flux across the fixed lower control surface. Buoyancy fluxes at the sea surface and diffusive fluxes in the interior transform water masses from one density class to another. From Marshall et al. (1999).

diffusion has three components:

$$D(\sigma,t) = D_\sigma(\sigma,t) + D_{ml}(\sigma,t) + D_B(\sigma,t). \tag{4.22}$$

The advective supply of buoyancy is by diapycnal advection or by advection across the open boundaries. For an enclosed region of the ocean, Garrett et al. (1995) give a detailed derivation of (4.21), and Nakamura (1995) independently derived this relation for atmospheric tracers. With these new definitions, conservation of buoyancy can be rewritten in a final form:

$$A(\sigma,t) = F(\sigma,t) - \frac{\partial D(\sigma,t)}{\partial \sigma}. \tag{4.23}$$

The diapycnal volume flux above also contains a contribution due to the heat transport by horizontal motions across isopycnals, although we find later that it is a small contribution. In summary, an explicit relationship between the kinematics and thermodynamics

of subduction has been found. This relationship will be used in Section 4.4 to check the kinematic estimates of subduction, and could be used to provide some interpretation of the dynamical processes.

4.3 Regional circulation and subduction pathways

Kinematically, two quantities are necessary to calculate the subduction rate: $\mathbf{u}$, the ocean velocity field, and h, the mixed-layer depth. First, the characteristics of the circulation implied by the state estimate of the eastern subtropical gyre are discussed. Later, the seasonal cycle of the mixed layer and its role as a rectifier of subducted water is detailed. After this introduction, subduction rates are diagnosed throughout the basin. Specifically, the role of eddies in subduction is kinematically estimated.

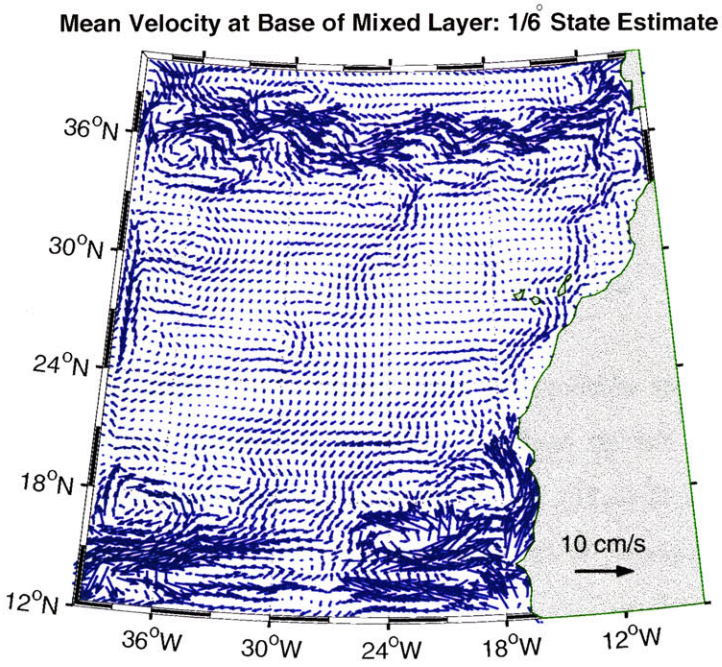

Figure 4-3: Mean circulation over one year at the depth of the deepest mixed layer.

The mean circulation of the upper ocean is dominated by the Azores Current with speeds up to 20 cm/s (Figure 4-3). The current transports about 12 Sv of water eastward

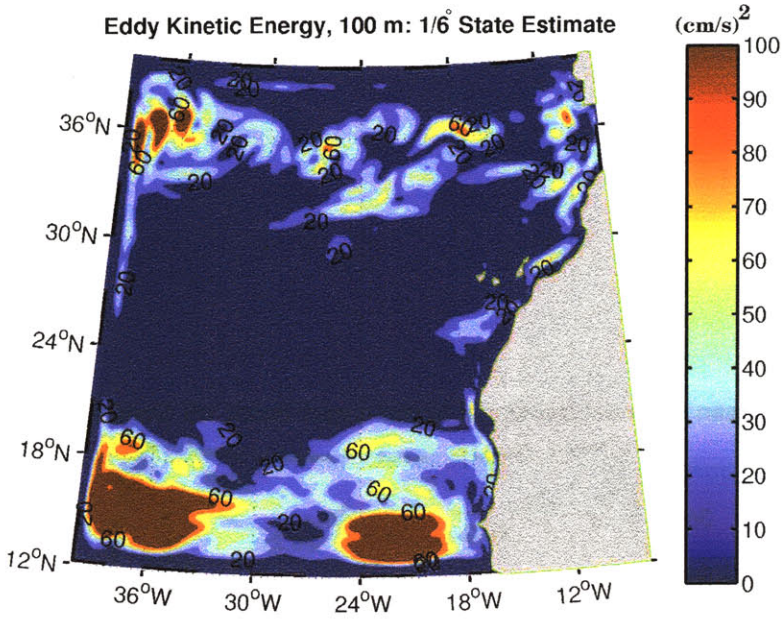

Figure 4-4: Eddy kinetic energy at 100 meters depth.

at $40°W$, which diminishes to roughly 3 Sv near the Mediterranean outflow. No formal error bars have been calculated, but many numerical simulations have been performed. In the simulations that fit the observations in a reasonable way, the strength of the current varies by no more than 2 Sv in the west, and 1 Sv in the east. The Azores Current of the state estimate has similar width and transport found in surveys by research vessels (Rudnick and Luyten 1996; Joyce et al. 1998). Most modeled fronts are considerably weaker than the strong current found in this state estimate (Jia 2000; New et al. 2001). The position of the east-west axis is $36°N$, which is farther north than the climatological position by $1 - 3°$ of latitude, but consistent with the recent synthesis of Weller et al. (2004) for the years 1991-1993. Upon impinging on the eastern boundary, almost 1 Sv of downwelling occurs in the Gulf of Cadiz. Roughly two-thirds of the this downwelled water retroflects to the south, with the remaining portion going north. The causes of the Azores Current, and its effect on subduction, are discussed next.

Solution of the least-squares problem of Chapter 3 offers information about the pro-

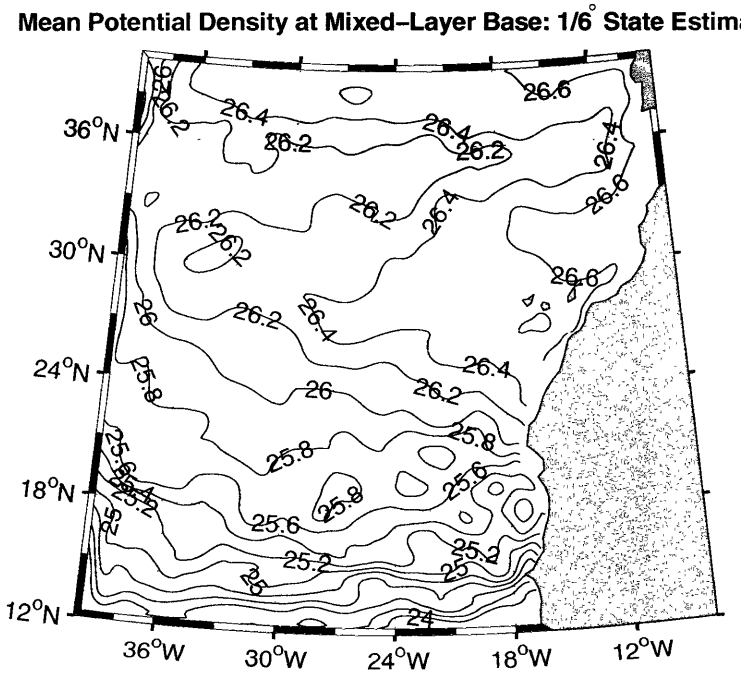

Figure 4-5: Mean potential density at the maximum mixed-layer depth. The spatial density structure serves as a new basis function to examine and understand subduction.

cesses that form the Azores Current. The ECCO 2° global state estimate (Stammer et al. 2002) models the Mediterranean Sea and includes an Azores Current with reasonable transport. However, the use of the ECCO state estimate as a open boundary condition in the 1/6° model here does not drive a realistic Azores Current (Figure 4-6). Instead, the inflow of water on the western open boundary meanders southward, and does not follow a tight, zonal trajectory across the domain. The model simulation may be inaccurate because the 2° ECCO Azores Current is too broad and too warm.

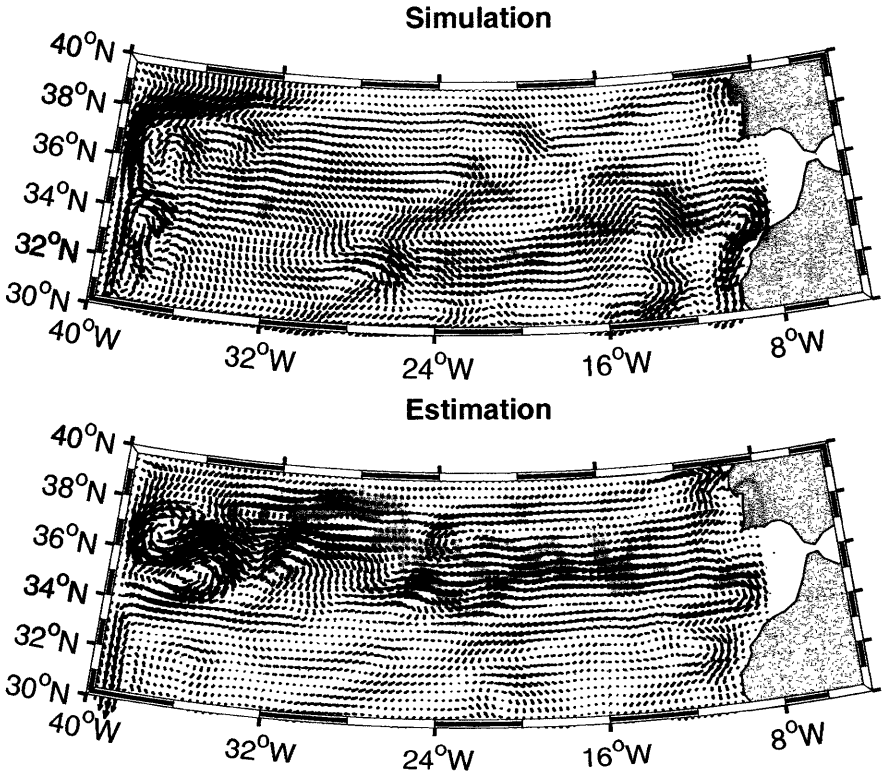

Figure 4-6: Velocity snapshot in the Azores Current subregion. Without any data constraint, the model simulation (*upper panel*) has an Azores Current that meanders southeastward. The state estimate (*lower panel*) shows a tight, zonal current in accordance with the observations of 1991-93. The maximum velocity vector is 10 cm/s.

Although the placement of the Subduction Experiment moorings intentionally avoided the Azores Current, the TOPEX/POSEIDON altimeter still provides information on the proper ocean circulation in this subregion. Observations of sea surface height anomaly by

TOPEX/POSEIDON demonstrate that SSH variance is too weak in the model simulation (Figure 4-7). To correct this inconsistency with observations, the western boundary inflow of Azores Current water is shifted southward and intensified into a narrower jet in the state estimate. Therefore, the northern and western boundary condition has a strong influence on the formation and maintenance of a realistic Azores Current over one to two years.

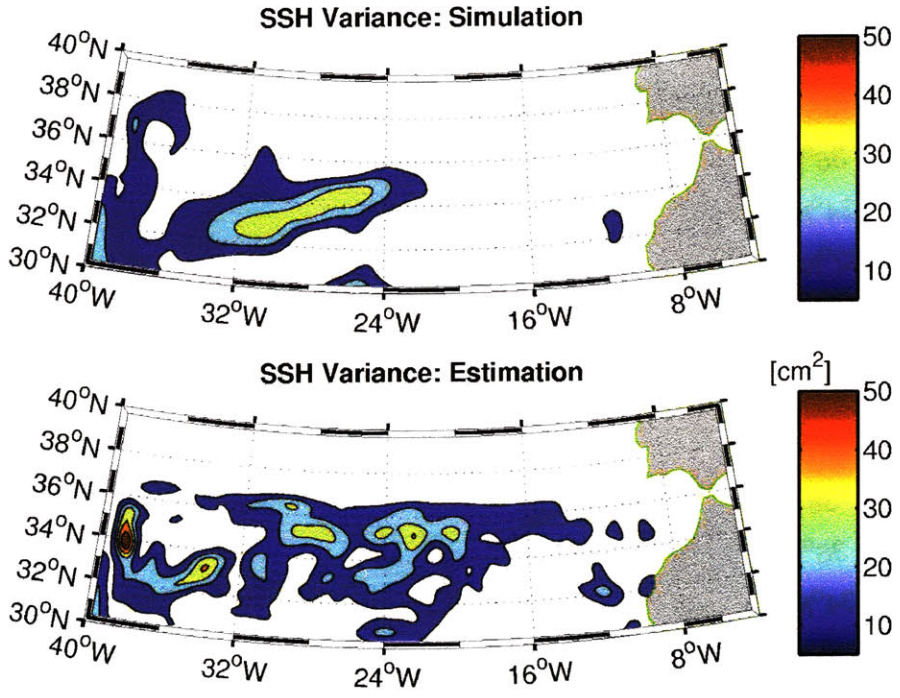

Figure 4-7: Sea surface height variance in the Azores Current subregion as seen in the model simulation (*upper panel*) and the state estimate (*lower panel*). The lower panel has the same general spatial structure of TOPEX/POSEIDON observations, and roughly 60% of the energy. The state estimate is generally in accordance with observations while the model simulation is not. The contour interval is 10 cm^2.

Previous studies have shown the strong sensitivity of the modeled Azores Current to model formulation (Jia 2000; New et al. 2001). Specifically, isopycnal coordinate models were found to give a stronger current than their z-coordinate counterparts. The inference is that water mass transformation in the Gulf of Cadiz and the simulation

of the Mediterranean outflow was crucial to a realistic simulation in the open ocean. Furthermore, Jia (2000) showed in an initial value problem that the modeled Azores Current formed from the east and extended westward with the speed of a baroclinic Rossby wave. Therefore, spinup of a current of realistic strength took $15 - 20$ years. Here, the state estimate does not include the Mediterranean Sea, and the total model duration is only two years.

The importance of the Mediterranean in the present study can be determined by performing a sensitivity study with the adjoint model. Sensitivity information is very efficiently calculated by the adjoint model, and presents a second major advantage of the methodology of this thesis (recall Section 3.2.5). Using the same cost function as Chapter 3, the magnitude of the gradient with respect to the open boundaries gives the relative importance of each boundary. The eastern boundary, located in the Mediterranean outflow, is not more important than the other boundaries. In fact, the western boundary and northwest corner of the domain do appear to be most important. This result was checked by changing the open boundary conditions of the forward model; the Mediterranean outflow was opened and the fluxes were prescribed by the ECCO global state estimate. The open-ocean circulation was not significantly changed by this modification to the forward model. The results do not necessarily conflict with the findings of previous investigators. Here, the Azores Current is present in some form in all runs, and the sensitivity study is measuring the stability of the already-formed current. The initial value description of Jia (2000) addresses a decidedly different problem. In addition, the timescale of analysis here, one to two years, is much shorter than the decadal timescales of other studies.

The Azores Current has a profound influence on subduction. The most apparent effect is the distortion of the mean streamlines of the upper ocean into a zonal jet, away from a southwestward flow (Figure 4-8). Ocean theories which depend upon Sverdrup balance alone do not explain such a deviation in upper-ocean circulation. Consequently, the ventilated thermocline theory of Luyten et al. (1983) and the extension by Huang and Russell (1995) have upper-ocean subduction pathways which are different from those

observed. Because of the lack of an advective pathway across the Azores Current, the theory of Rhines and Young (1982) predicts a a region of homogenized potential vorticity behind the front. Robbins et al. (1998) searched for such a region, but it was not present in observations. These results suggest that some subduction mechanisms in the region of the Azores Current are not included in classical theory.

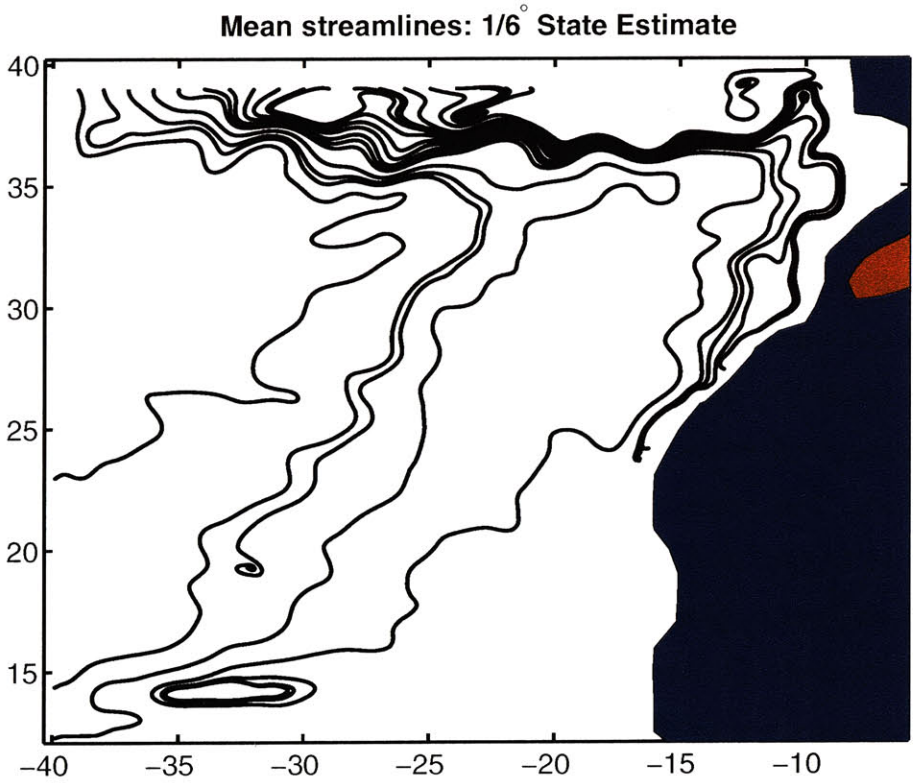

Figure 4-8: Streamlines of the mean horizontal velocity field at the depth of the maximum mixed layer. The streamlines start at the western end of the northern boundary with a spacing of 1°, from $39°W$ to $21°W$. The flow carries water into the Azores Current, then in a general southwestward trajectory that is consistent with the drift of SOFAR floats (Sundermeyer and Price 1988).

4.4 Subduction in the state estimate

4.4.1 Seasonal cycle of entrainment

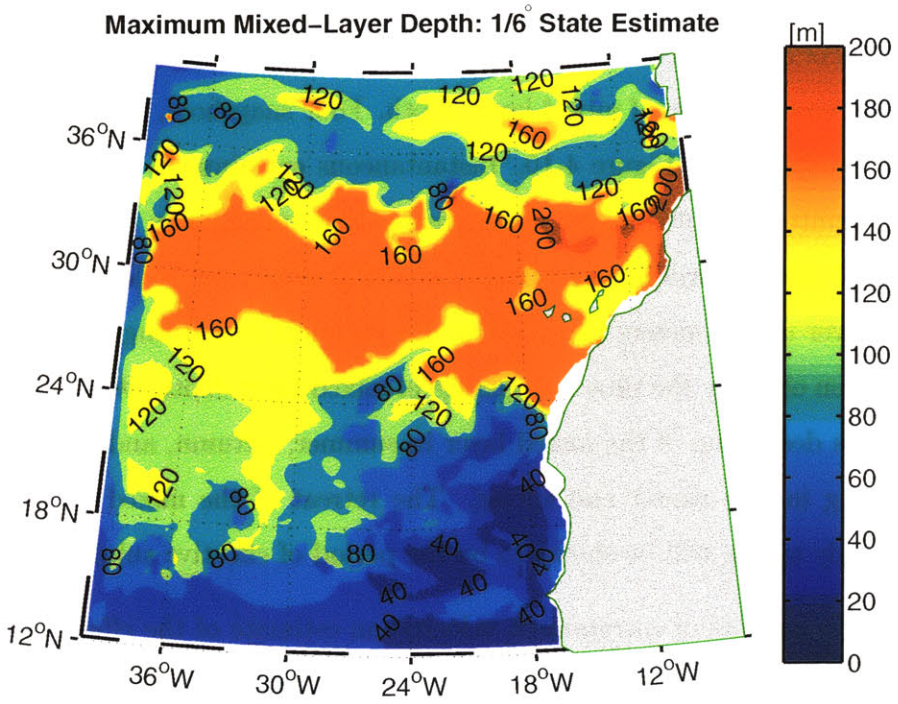

Figure 4-9: Maximum mixed-layer depth over one seasonal cycle. The mixed layer is defined as the region where the density difference to the surface is less than 0.025 kg/m^3.

Does the mixed-layer demon operate in the state estimate? The primary requirement is that the seasonal variation of mixed-layer depth is larger than the vertical displacement of water parcels. Maximum mixed-layer depths in February reach 200 meters (Figure 4-9). Wintertime mixed-layer depth shoals equatorward of 25° N, in accordance with climatologies (Marshall et al. 1993; Levitus and Boyer 1994) and traditional thinking. However, the region between 25° N and 35° N does not have an equatorward gradient of mixed-layer depth, which is surprising but in accordance with the observational synthesis of Weller et al. (2004). The mixed layer is shallower in the Azores Current, due to the input of relatively buoyant water throughout the year. Because the summertime

mixed layer shoals everywhere to $20-30$ meters depth, seasonal changes of mixed layer approach 150 meters. In comparison, vertical motions displace water parcels no more than 30 meters in one year. The magnitudes of these two processes suggest that the mixed-layer demon is important in this region.

Snapshots of entrainment velocity, calculated by Equation (4.1), quantify the importance of the seasonal cycle of mixed-layer depth. Four snapshots, representing the four seasons, are displayed in Figure 4-10. Instantaneous entrainment velocities frequently exceed a magnitude of 1000 m/yr, much greater than any Ekman pumping rates. The largest magnitude of entrainment velocity is -3500 m/yr: equivalent to a shoaling of the mixed layer of 100 meters in ten days. Entrainment has strong seasonality due to the domination of w_* by the time rate change of mixed-layer depth, $\partial h/\partial t$. The seasonal cycle includes deepening of the mixed layer in summer, autumn, and early winter, and rapid shoaling in late winter and spring. The retreat of the mixed layer toward the surface in early spring will be shown to be the period of effective subduction.

From the snapshots of entrainment velocity, an estimate of the effective subduction period is possible. This period begins at the time of maximum mixed-layer depth, and ends when the volume of detrained water equals the volume of subducted water for the whole year. In other words, water may be detrained after the end of the effective period, but it will be re-entrained later. The time of maximum mixed-layer depth can be defined two ways: the time of maximum volume of the mixed layer (here, February 20), or the median time of maximum mixed-layer depth throughout the region (March 15). Weller et al. (2004) remarked that the deepest mixed layers occur in February in the north of the domain, and in March in the south, in close accordance with the state estimate. Figure 4-11 integrates the volume of detrained water after February 20. Over the entire domain and over one year, $2.1 \cdot 10^{14}$ m^3 of water is detrained, which amounts to a subduction rate of 27 m/yr over the entire domain. By this method, the effective subduction period is 53 days, because an equivalent amount of water is detrained in this time. As a measure of the error in the diagnostics, integration from March 15 instead of February 20 yields a period of 45 days. Other studies have estimated that subduction

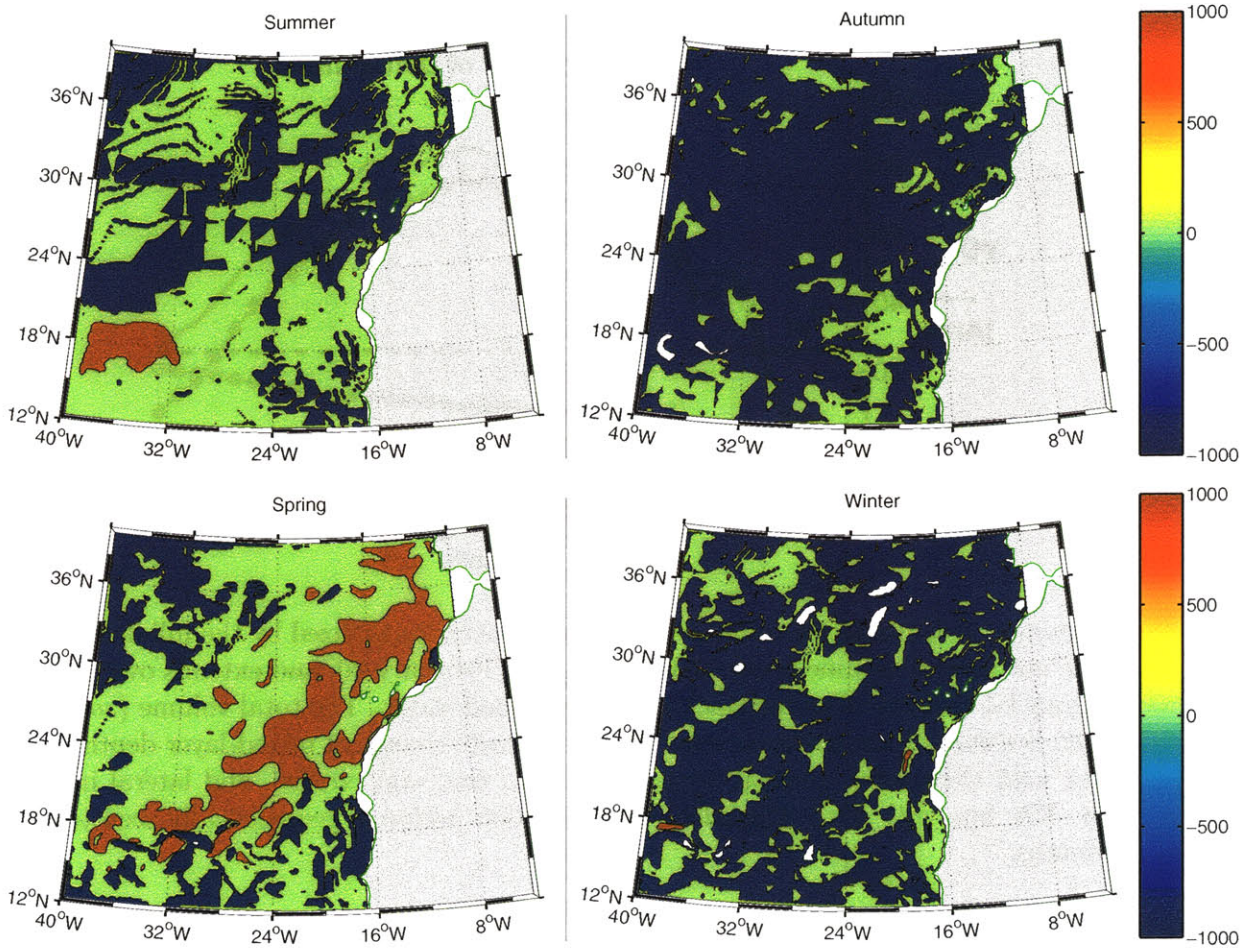

Figure 4-10: The seasonal cycle of detrainment from the mixed layer. The seasonal cycle proceeds clockwise. Blue areas represent large values of entrainment, $w_* < -1000\ m/yr$. Dark red areas represent large values of detrainment, $w_* > 1000\ m/yr$, and are potential sites of subduction. Green areas include all intermediate values of entrainment velocity. As the entrainment velocities are dominated by local values of $\partial h/\partial t$, the patterns are not associated with any frontal structures.

occurs over 1.8 to 2.2 months in the North Atlantic (Marshall et al. 1993; Hazeleger and Drijfhout 2000). The short time of subduction shows that Stommel's mixed-layer demon stroboscopically regulates the passage of water into the main thermocline of the state estimate.

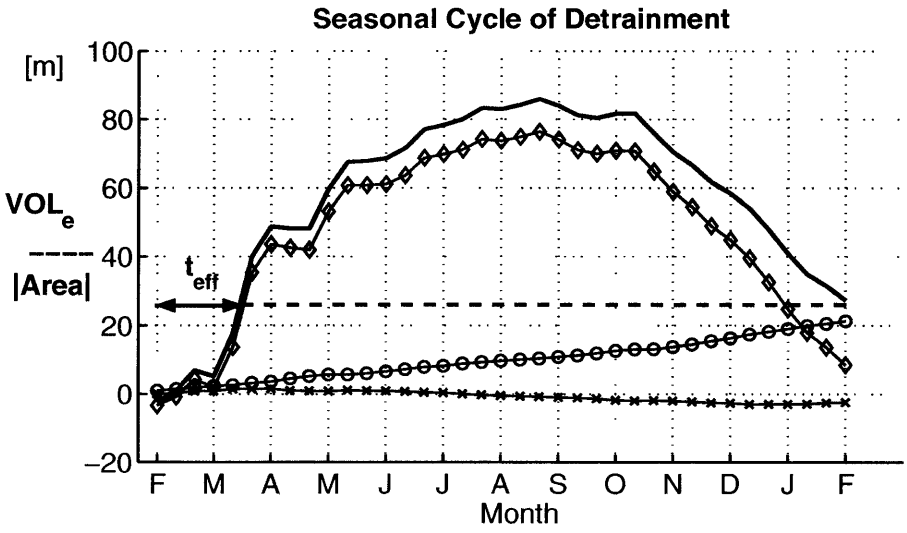

Figure 4-11: Cumulative volume of detrained water, normalized by the full domain area. The normalized volume is calculated beginning February 20, and extends over one year. It may be thought of as the thickness of detrained water. The total volume (*solid, black line*) is a sum of the contributions by the time rate change of mixed-layer depth ($\partial h/\partial t$, *line with diamonds*), vertical displacement (w, *line with circles*), and lateral induction ($\mathbf{u} \cdot \nabla h$, *line with X's*). The effective subduction period is geometrically seen to be 1.5 months.

4.4.2 Estimated subduction rates

Before calculating the water-mass subduction rate, Equation (4.3), we wish to understand the geographic distribution of subduction. Using the exact form of (4.15) without smoothing any of the fields, the annual subduction rate is calculated for the state estimate at eddy-resolution (Figure 4-12). Small-scale variations in the maximum mixed-layer depth (recall Figure 4-9) and the horizontal circulation field (Figure 4-3) lead to locally-intense volume fluxes. These volume fluxes are oriented horizontally across the

sloping mixed-layer base. The inherently-noisy nature of the gradient of H is responsible for subduction rates up to 300 m/yr. In contrast, the vertical velocity field at the mixed-layer base is predominantly large-scale. The small-scale intense subduction rates here are not an artifact of the diagnostic scheme; the definition of annual subduction is responsible. A one-year time average is only two or three baroclinic life-cycles: not long enough to eliminate small-scale features in the mean circulation of an eddy-resolving state estimate.

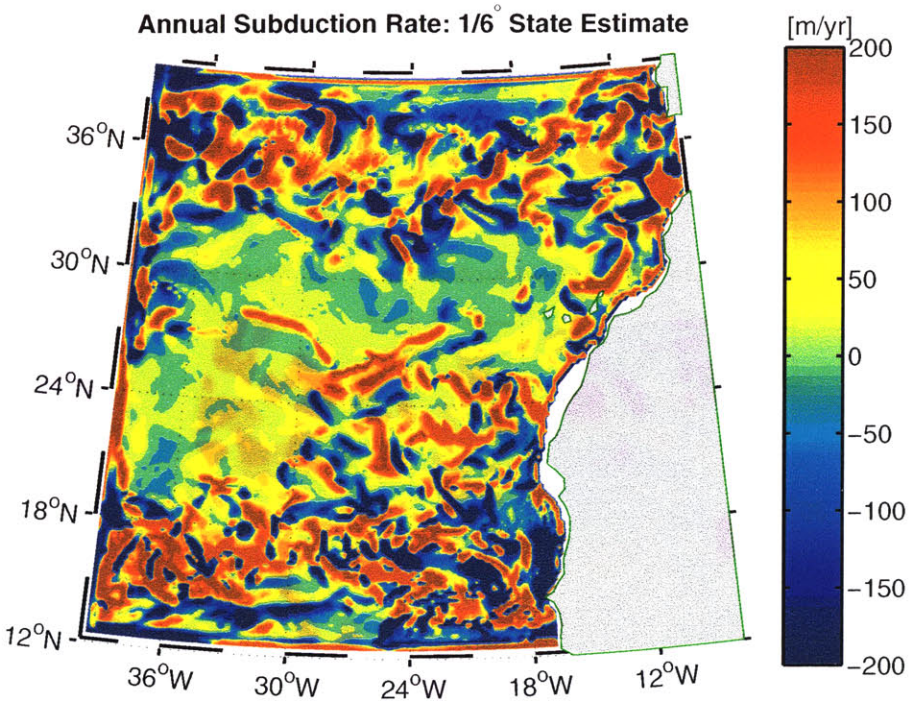

Figure 4-12: Annual subduction rate, $s_{ann} = -(\overline{w}_H + \overline{\mathbf{u}}_H \cdot \nabla H)$. The small-scale, intense subduction rates are due to lateral induction by the mean circulation.

Although local subduction rates are intense, small-scale features do not necessarily lead to net subduction when integrated over the domain. The domain-averaged annual subduction rate is approximately 6 Sv. The error is estimated as less than 1 Sv by the sensitivity of the model. For comparison, a large-scale subduction rate can be defined by evaluating (4.15) with the coarse-grained fields. Coarse-graining is accomplished by a 2° running mean on the velocity and mixed-layer depth fields. The new definition of the

large-scale annual subduction rate is closer to the quantity calculated by previous studies, as the maximum mixed-layer depth field was usually smoothed by other authors. The large-scale annual subduction rate (Figure 4-13) gives a domain-averaged subduction of roughly 5.5 Sv for the state estimate. The patterns of subduction differ due to small-scale features still present in the mean circulation fields. However, the magnitude of subduction is relatively unchanged despite the fact that gradients are less sharp in the coarse-grained fields. Further identification of the role of eddies in subduction requires the explicit study of the small-scale, time-varying fields. Unfortunately, time-variability does not enter Equation (4.15) because the base of the mixed-layer is chosen to be fixed with time.

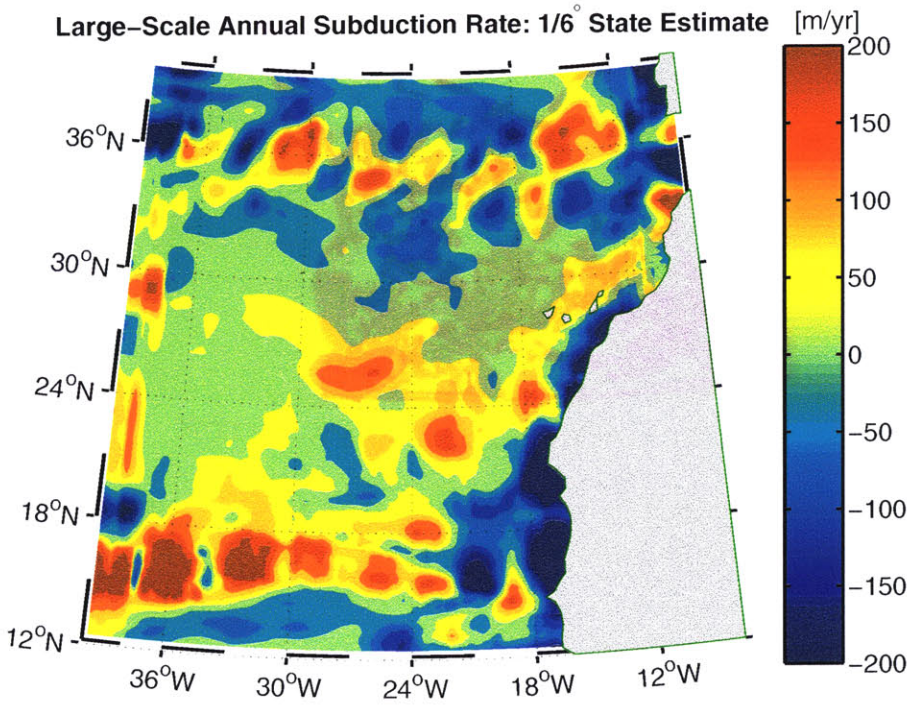

Figure 4-13: Large-scale, annual subduction rate $s_{lrgscl} = -(\overline{w}^*_{H^*} + \overline{\mathbf{u}}^*_{H^*} \cdot \nabla H^*)$, where a star indicates a field that has been coarse-grained. The domain-integrated subduction rate is not significantly altered by the coarse-graining.

Estimates of eddy subduction

Eddy subduction in an Eulerian frame of reference reduces to the eddy volume flux, $\overline{\mathbf{u}_h' \cdot \nabla h'}$, across the moving mixed layer, $h(t)$. This is the first term of the eddy contribution to the water-mass subduction rate, Equation 4.11. One subtlety in diagnosing the state estimate is the definition of the mean circulation. Our definition of the mean is actually a monthly mean, so that time-variability of the seasonal cycle is not grouped with "eddy" variability.

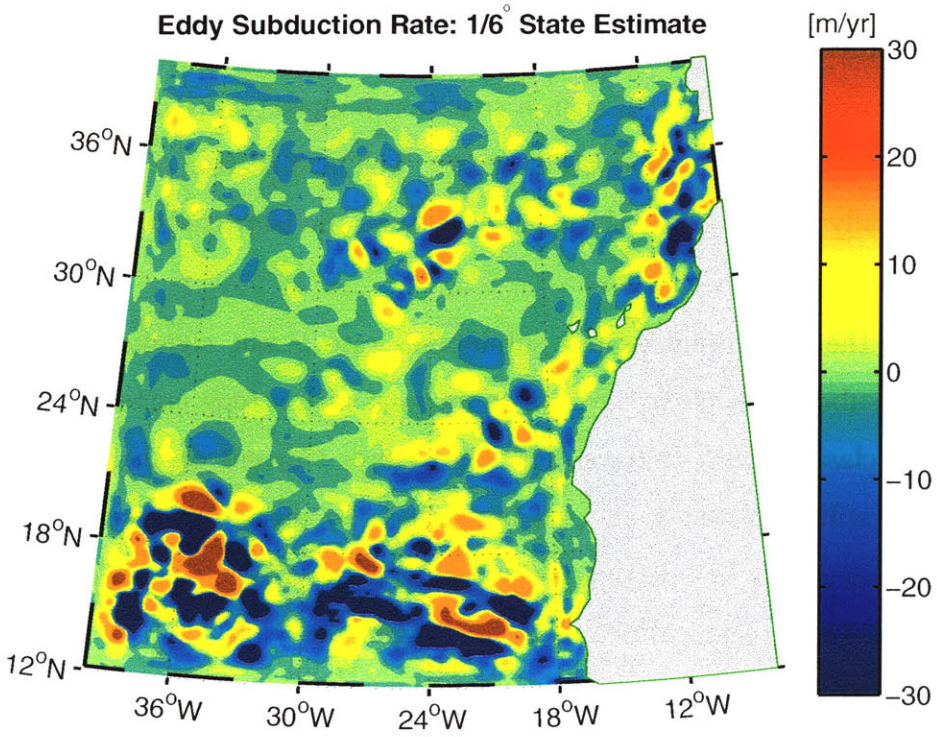

Figure 4-14: Eddy subduction rate, computed as the eddy-thickness flux across the time-variable mixed-layer base. Note the color scale ranges from -30 m/yr to 30 m/yr, a smaller range than Figure 4-12. Eddy subduction is largest in regions with enhanced eddy kinetic energy.

Local values of eddy subduction approach 40 m/yr in the North Equatorial Current and in parts of the Azores Current (Figure 4-14). In these frontal regions, eddy subduction is locally non-negligible in comparison to Ekman pumping rates of only $20-30$ m/yr. The magnitude of the eddy component of subduction is about 15% of the

annual subduction in select locations. However, the Eulerian field of eddy subduction does not allow a good evaluation of the net effect of eddies on water mass formation. When averaging over an area larger than the eddy length-scale, the net contribution to subduction nearly vanishes. This is a limitation of the Eulerian definition of subduction; a more-Lagrangian viewpoint is needed to isolate the net impact of eddies. One attempt to visualize the impact of the eddies is to overlay the mean density contours with the eddy subduction rate (Figure 4-15). This isolates the first term of s_{eddy}. In general, no clear pattern is evident. In some cases, such as the bullseye in the Azores Current, subduction is positioned near a "wide mouth" in the isopycnals, a place with greater than average spacing. If such a situation happens more often that subduction near packed isopycnals, net subduction occurs. The next section of this thesis attempts to systematically evaluate the relationship between the isopycnals and subduction in a way that can not be done visually.

Although the domain-integrated subduction rate is not significantly modified by eddies, eddy subduction is strongest in subregions with strong currents. This suggests that density classes which outcrop near the Azores Current or the North Equatorial Current may still be strongly affected by eddies. To check this proposition, a second perspective is available by isolating the impact of subduction in particular water masses (see Appendix D for technical details of the diagnostics).

Estimated water-mass subduction rates

The water-mass subduction rate is directly estimated from the velocity field at the mixed-layer base (Figure 4-16). A domain-integrated 4 Sv of subduction occurs in the domain. Due to the mixed-layer demon, all of the subducting water is in the density range $\sigma > 25.0$, corresponding to the densities that outcrop in the late winter. The eddy subduction rate, $S_{eddy}(\sigma)$, is nonzero and indicates that net subduction due to eddies is occurring. Eddies act to subduct water in the late-winter density classes, but obduct water at $\sigma < 25.2$. Through this density-coordinate analysis, the net impact of eddies is suggested. To further understand the processes that cause subduction, the surface layer

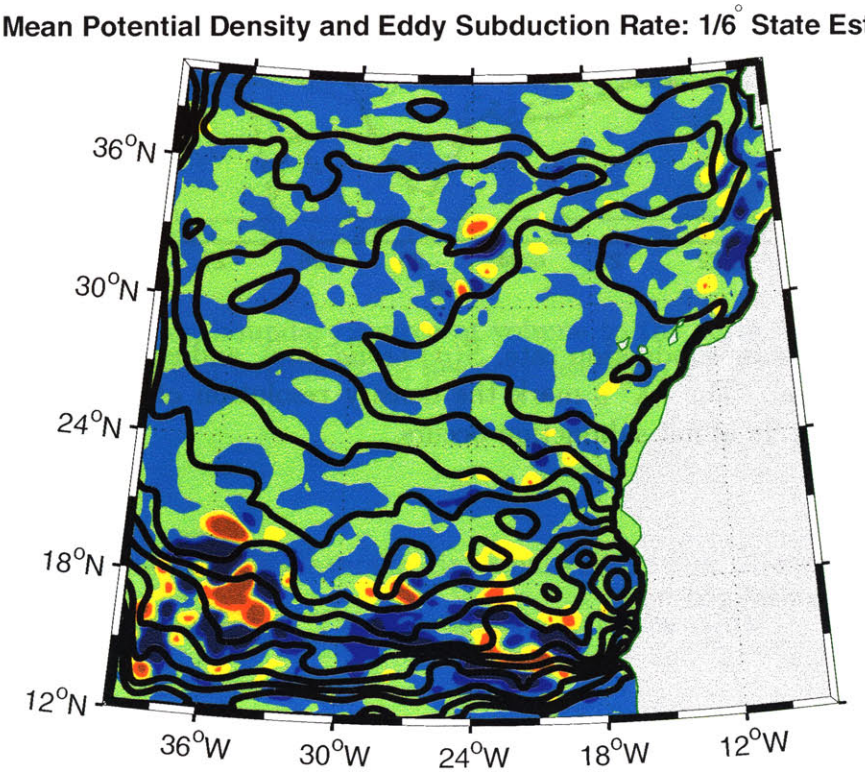

Figure 4-15: The eddy subduction rate overlaid with the mean density contours at the mixed-layer base. In some instances, subduction occurs in regions with widely-spaced isopycnals.

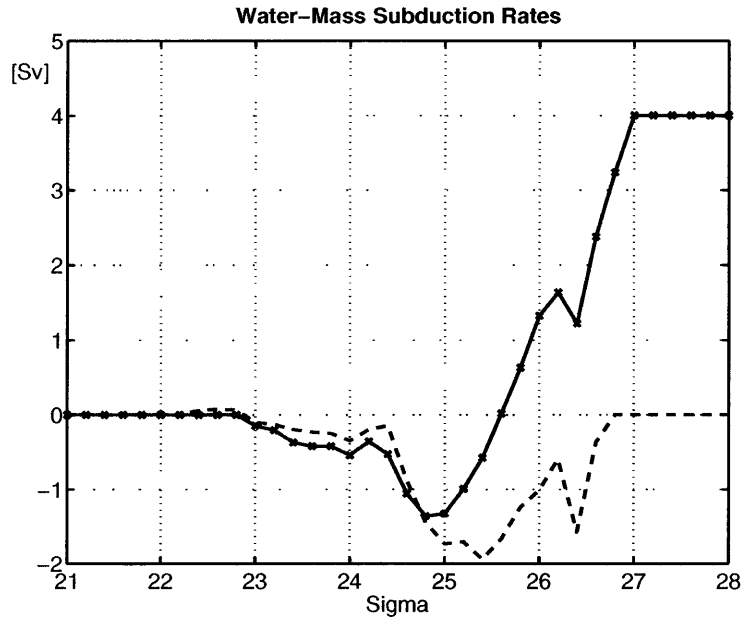

Figure 4-16: Water-mass subduction rate, $\overline{S}(\sigma)$ (*solid line*), and eddy subduction rate, $S_{eddy}(\sigma)$ (*dashed line*). The "X"'s mark the density resolution of the diagnostics. A domain-integrated 4 Sv of net subduction occurs.

volume budget is considered next.

4.4.3 Estimated surface layer volume budget

The surface layer volume budget allows the determination of the relative importance of the open boundaries versus the interior dynamics in setting the water mass properties of subducted fluid. The annual-mean open boundary volume flux, $\overline{M}_B(\sigma)$, is calculated with monthly average fields, and density bins of $\Delta\sigma = 0.2$ (Figure 4-17). $\overline{M}_B(\sigma)$ shows that some of light ($\sigma < 24.2$) water is expelled from the basin. This happens primarily in the uppermost 50 meters near the North Equatorial Current. The majority of incoming water (≈ 5 Sv) is in the density class $24.2 < \sigma < 26.5$. This indicates that the typical subtropical mode water classes are laterally recirculating in the subtropical gyre. Over the entire domain, the surface-layer open boundary flux is a net source of water; i.e., $\overline{M}_B(\sigma = 27) = 4$ Sv. This excess water must be subducted, which is shown below.

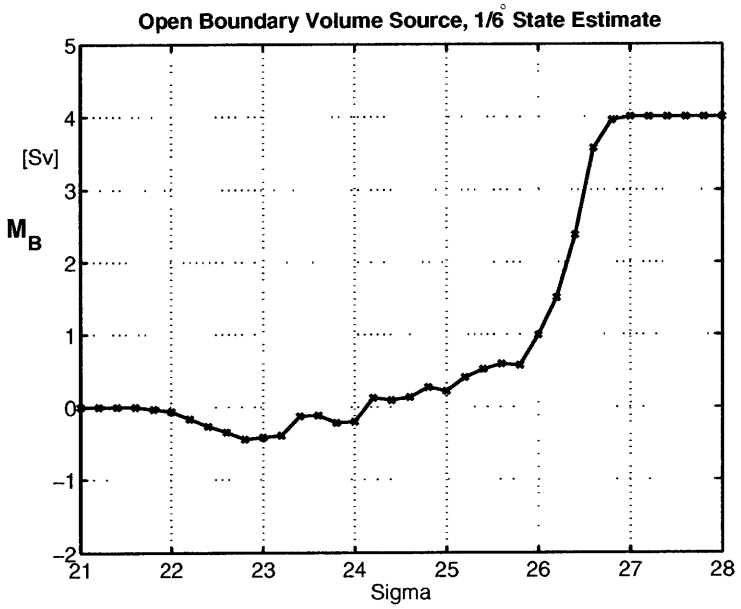

Figure 4-17: The open boundary source of volume to the surface layer. $\overline{M}_B(\sigma)$ is the annual average source of volume at all densities less than σ.

Figure 4-18 shows the time-average diapycnal volume flux, $\overline{A}(\sigma)$, in the surface layer as defined by Equation (4.18). Two components comprise $\overline{A}(\sigma)$:

$$\overline{A}(\sigma) = \overline{A}_E(\sigma) - \overline{\frac{\partial V(\sigma)}{\partial t}}, \tag{4.24}$$

Over much of the domain, water flows across isopycnals toward higher density because $\overline{A}(\sigma) > 0$. However, water is formed only in the range $26.0 < \sigma < 27.0$ where the diapycnal flux is convergent. In an integral sense, water leaves the lighter classes of water, and is made more dense in the domain. The high values of $\overline{\partial V(\sigma)/\partial t}$ show that the model's isopycnals have been displaced over one year.

The domain-integrated subduction rate, $\overline{S}(\sigma > 27) = 4\ Sv$, is given by the water-mass subduction rate at the maximum potential density in the surface layer. This value is effectively set by $\overline{M_B}(\sigma > 27)$, as any excess water that enters through the open boundaries is subducted across the mixed-layer base by volume conservation.

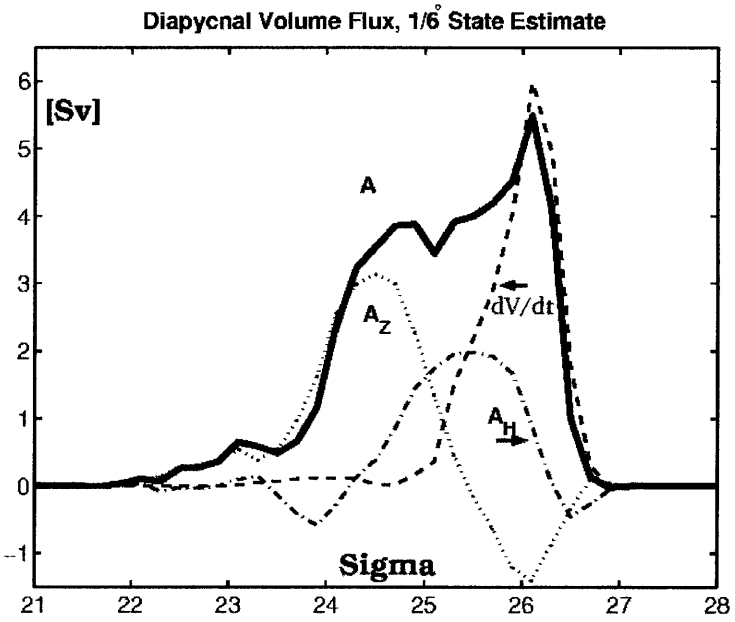

Figure 4-18: Diapycnal volume fluxes in the surface layer. All overbars have been dropped in the figure. Together, $\overline{A}_H(\sigma)$ and $\overline{A}_Z(\sigma)$ represent the Eulerian component of volume flux through the horizontal and vertical velocity fields. $\overline{\partial V(\sigma)/\partial t}$ represents storage in an isopycnal which results from the net displacement of a density surface.

The combination of $\overline{M_B}(\sigma)$ and $\overline{A}(\sigma)$ determine $\overline{S}(\sigma)$, and hence, the properties of the subducted water (Figure 4-16). When considering the relative importance of these two terms, three different regimes in the domain emerge. The shallow, summertime density range, $\sigma < 24.2$, is dominated by a throughflow into the surface layer from below and then expulsion out of the open boundaries. It is surprising that such light water would flow from the interior to the surface layer, but this is similar to the results of Marshall et al. (1999) for the entire North Atlantic. The light density-classes are characterized by obduction despite a downward Ekman velocity. In the intermediate density range, $24.2 < \sigma < 26.0$, water outcrops in the southern half of the domain in winter. Diapycnal fluxes work against the open boundary source and reduce the amount of subduction as water is transported to greater density. Still, large volumes are subducted, and the open boundary source is the dominant player. In the densest density range, $26.0 < \sigma < 27.0$, diapycnal fluxes have the dominant impact on the subducted water properties. Water accumulates in the surface layer at these densities due to the diapycnal flux, and little subduction occurs despite the additional source of open boundary water.

In the density range $25.5 < \sigma < 26.5$, high-frequency motions produce a maximum diapycnal flux of 1 Sv, nearly as large as the flux of 2.5 Sv by mean fluid velocity. This density range encompasses the region of the Azores Current (recall Figure 4-5). Formation of water masses depends upon the convergence of the diapycnal fluxes; $\overline{A}_{eddy}(\sigma)$ is convergent throughout most density classes, yielding a net formation of water by eddy processes. In general, the derivative of $\overline{A}_{eddy}(\sigma)$ rivals that of $\overline{A}_E(\sigma)$. Because diapycnal volume flux is directly related to subduction, this calculation quantifies the contribution of eddies to subduction.

Kinematic error estimates/Sensitivity analysis

The surface layer volume budget is self-consistent and perfectly closed. The water-mass subduction rate can be computed directly, or through a combination of open boundary and interior terms; the results match exactly. However, errors are present due to the use of a discretized density coordinate. In the GCM, it is assumed that the density is

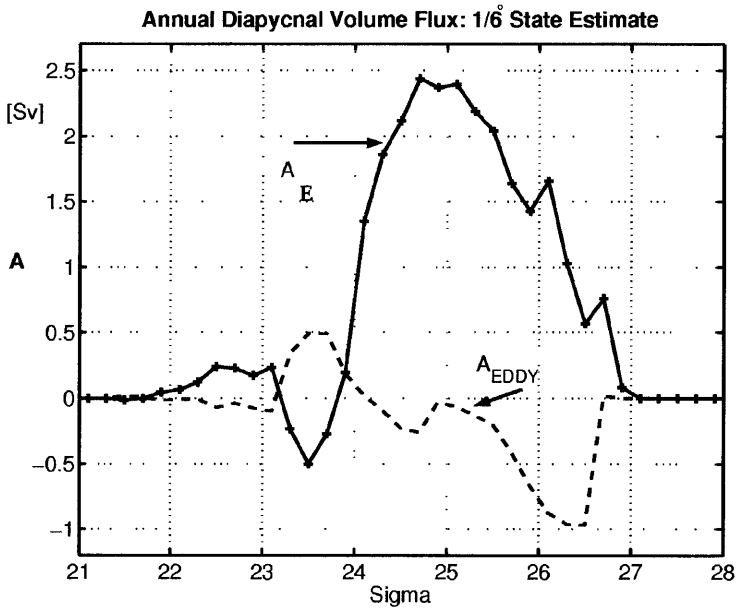

Figure 4-19: Annual diapycnal volume flux due to the Eulerian velocity, $\overline{A}_E(\sigma)$, versus the time-averaged eddy volume flux, $\overline{A}_{eddy}(\sigma)$. The contribution to net formation of water by eddies is exceeds 1 Sv.

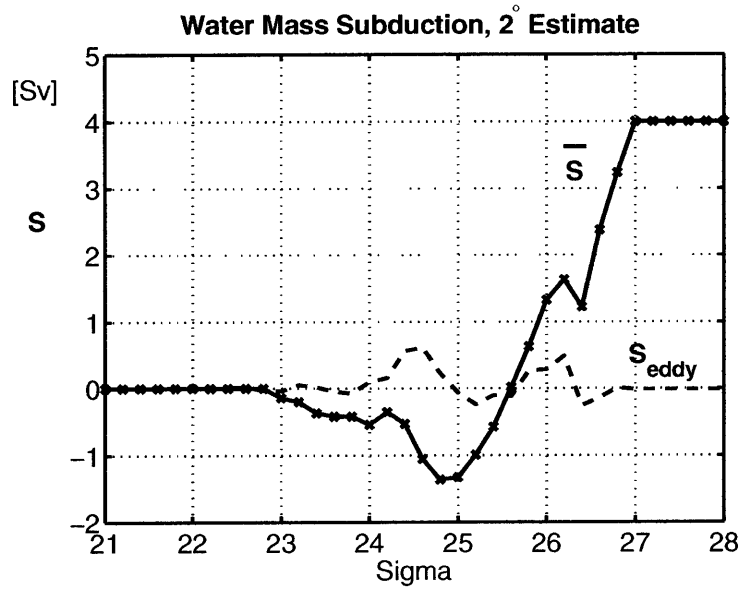

Figure 4-20: Water-mass subduction rate, $\overline{S}(\sigma)$ (*solid line*), and eddy subduction rate, $S_{eddy}(\sigma)$ (*dashed line*), for the 2 degree state estimate. The "X"'s mark the density resolution of the diagnostics. The subduction rate is computed about a time-variable mixed-layer base. As the 2° state estimate has little eddy variability, $S_{eddy}(\sigma)$ represents an estimate of the error of the diagnostic scheme.

constant along an entire grid face, and the error can then be computed (Marshall et al. 1999). In a fine resolution state estimate, the assumption of constant density along a grid face is good because of the small area of an individual grid cell. To get an idea of the maximum level of error due to density discretization, we examine the diagnostics of the 2° state estimate. This coarse resolution estimate does not contain energetic eddy motions, and this can be checked through diagnosis of $S_{eddy}(\sigma)$. Figure 4-20 shows that eddy subduction has a root mean square value of 0.2 *Sv*. Most of this "eddy subduction" is actually error in the diagnostic scheme. In this way, an upper bound of 0.2 *Sv* is estimated for the density-coordinate analysis. Thermodynamic budgets (discussed later) have much larger sources of error.

The decomposition of the circulation into mean and eddy components is troublesome in this region because of the lack of a stable mean velocity field. Considering the zonal

velocity time-series from 1° North Atlantic state estimate (Stammer et al. 2002), at least ten years of data is needed for a stable mean in the interior of the Subduction Experiment region. Figure 4-21 shows a typical velocity timeseries from the ECCO state estimate where the determination of the mean velocity depends upon the averaging interval throughout the years 1992-2002. Near the NE mooring, the situation is even worse. A mooring deployed by the Insitut für Meereskunde (Mooring Kiel 276) with a ten-year time series has still failed to produce statistically-significant mean velocities. Müller and Siedler (1992) have commented that variability in the 4-6 year frequency band is responsible for the unstable means. In the calculation of the eddy subduction rates here, interannual variability causes error in the estimates that swamp any other source.

Another source of error in the subduction rates is due to the state estimate error itself. Although methods have been developed to gain the error statistics of the state estimate (Thacker 1989), this is computationally unfeasible for the present problem. However, the sensitivity of the results can be estimated by considering the multiple forward model runs that have been performed in the optimization process. For example, the transport of the Azores Current is consistently 12 Sv in the perturbed forward run, with a typical deviation of less than 0.2 Sv. The path of the current varies more widely, with differences up to 200 km. Error in the displacement of a feature is difficult to represent in a simple error bar, as it usually leads to non-Gaussian statistics (Lawson and Hansen 2004). Based on the sensitivity of the forward model, the integrated subduction rate over the domain has errors on the order of 0.5 Sv, and local subduction rates near the Subduction Experiment moorings are accurate within 15 m/yr. Away from the explicit data constraint of the eddies by the moorings, the eddy pattern of subduction is sometimes shifted, leading to errors of $30-40$ m/yr which is the magnitude of the eddy subduction itself. As seen above, a sense of the errors in subduction rates is possible through a sensitivity study, even though formal error bars are difficult to estimate.

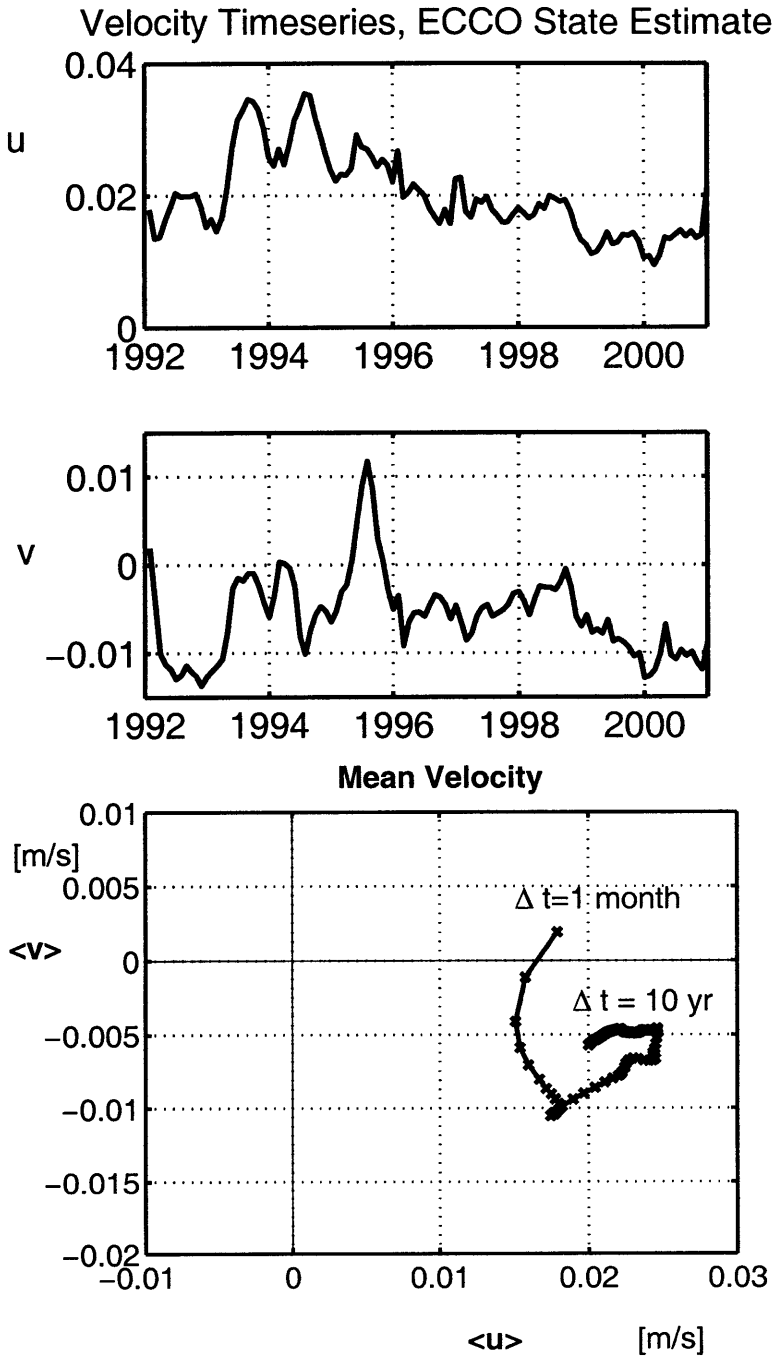

Figure 4-21: Characteristics of the ECCO 1° state estimate velocity timeseries at (35° N, 30° W). *Top two panels*: Time series of zonal and meridional velocity at 222 meters depth. *Bottom panel*: Mean velocity as a function of averaging interval, starting from one month and extending to the average of the entire timeseries, 10 years. These plots show that interannual variability is a dominant component of the time series, and stable means are only obtained after long averaging periods.

4.4.4 A thermodynamic check on the diagnostics

Cross-isopycnal advective flow must be accompanied by thermodynamic forcing. Therefore, we can compute thermodynamic transformation rates as a check on the previous kinematic diagnostics. In the Subduction Experiment region, transformation and subduction have similar trends in density-space (compare $\overline{F}(\sigma)$ of Figure 4-22 and $\overline{A}(\sigma)$ of Figure 4-18). The differences between the air-sea transformation rate, $\overline{F}(\sigma)$, and the diapycnal volume flux, $\overline{A}(\sigma)$, however, are due to mixing in the ocean. Marshall et al. (1999) have shown that the diffusion terms, $D(\sigma, t)$, of the surface layer reconcile the differences, and close the budget, in a numerical model. Likewise here, diffusion reduces the residuals in the thermodynamic budget.

The residual of the thermodynamic budget (4.23) is shown in Figure 4-23. Marshall et al. (1999) point out two sources of error in the budget: discretization error and unresolved variability. The size of the residuals, $1 - 2$ Sv, is due to our lack of ability to accurately reconstruct the buoyancy equation offline. The diffusion term, D, was approximated in the diagnostics by using a constant background diffusivity, although the true diffusivity varied as a function of space and time. Notice that the pattern of the diapycnal fluxes by diffusion closely resembles the thermodynamic residual. This is suggestive that proper diagnosis of the diffusion terms in the state estimate would completely close the budget. However, the thermodynamic residuals of $1-2$ Sv presently rival the size of the eddy subduction signal. Therefore, a check of eddy subduction rates through purely thermodynamic means is postponed at this time.

4.5 Summary

- The state estimate confirms that the mixed-layer demon, originally formulated by Stommel, operates in the eastern North Atlantic, and allows effective subduction during $45 - 60$ days of the late-winter.

- Annual subduction rates have locally-intense subduction and obduction up to 200 m/yr. Lateral induction by the small-scale mean circulation is responsible

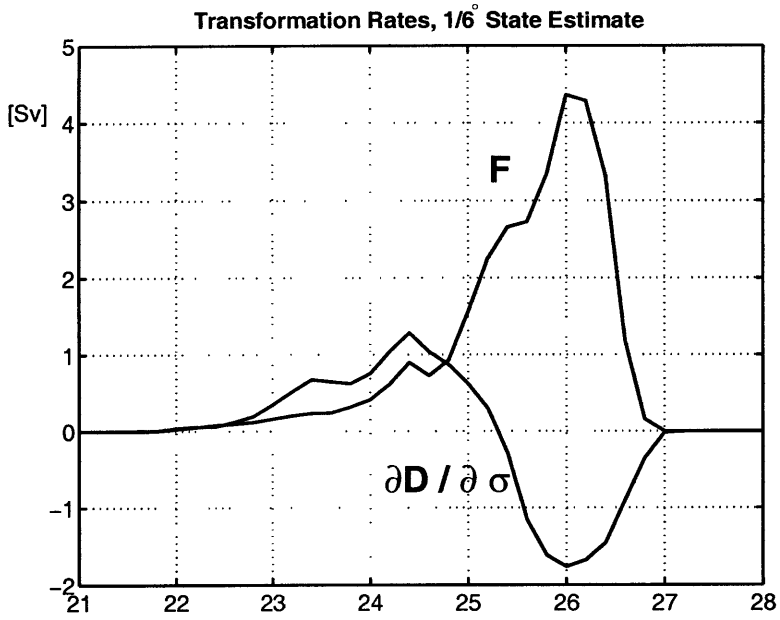

Figure 4-22: Annual air-sea transformation rate, $\overline{F}(\sigma)$.

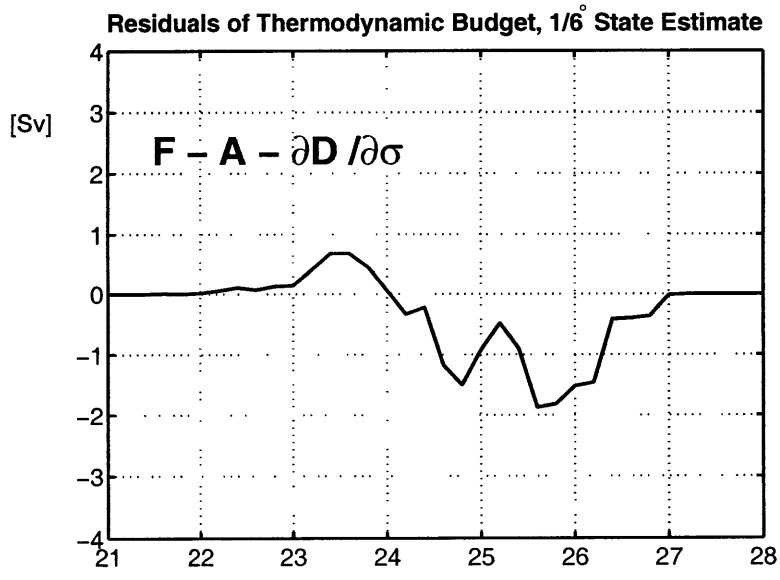

Figure 4-23: Residual of the thermodynamic budget.

for such intense subduction.

- Eulerian maps of eddy subduction show local volume fluxes up to 40 m/yr, significantly higher than seen in parameterizations of coarse-resolution models (10 m/yr, Spall et al. 2000).

- Isopycnal analysis suggests that eddy subduction is as large as mean subduction in the density class $25.8 < \sigma < 26.2$. Parameterizations which do not include this effect will produce biases in water-mass properties.

Chapter 5

Conclusion

A review of the thesis follows, with an emphasis on directing the interested reader to the detailed sections. After the recap, connections are discussed between the results in an effort to answer questions of a wider scope. Finally, a number of future projects and unresolved questions are suggested.

5.1 Review of results

Quantitative review of the classic theory with observations reveals that our previous view of subduction is incomplete. A first step here is to quantify the basic pattern of subduction in the eastern North Atlantic by using the field measurements of the Subduction Experiment and the TOPEX/POSEIDON satellite. In particular, relatively little is known about the role of eddies in controlling subduction in the eastern half of the subtropical gyre. Chapter 1 reviews the state of the science and poses two questions:

- **What is the magnitude and pattern of subduction in the eastern North Atlantic?**
- **Does eddy subduction significantly affect the total subduction rate?**

The measurements from the Subduction Experiment moorings have inadequate spatial coverage to diagnose subduction. To remedy the problem, an eddy-resolving state

estimate was made by bringing a 1/6° North Atlantic regional model into consistency with data, thus yielding an estimate of oceanic fields at the fine resolution of the model. Chapter 2 shows that the problem of combining a model and observations is a giant least-squares problem (detailed in Sections 2.2-2.4). Fine resolution and open boundaries both lead to special considerations in the mathematical formulation of the problem (Sections 2.2.1, 2.4.1, and 2.4.2). Here, the defining and novel nature of the problem is the high-dimension of the state and the nonlinearity of the model.

For oceanographic datasets, the method of Lagrange multipliers is the ideal choice to solve the constrained least-squares problem and form a state estimate (Section 3.2). The method hinges upon the availability and usefulness of the adjoint to the model. In contrast to previous studies at eddy-resolution, the information from the adjoint model is useful for finding a consistent solution between the model and observations (Section 3.5). Individual eddies that are observed by the Subduction Experiment moorings can be estimated (Section 3.5.5). The incorporation of a good first guess from a coarse-resolution model is crucial for the success of the method (Section 3.4). The findings of Chapter 3 can be summarized as:

- **No fundamental obstacle exists to constraining an eddy-resolving model to observations in this region.**

- **Eddies observed by the Subduction Experiment mooring array can be tracked in the state estimate.**

The state estimate is a dynamically-consistent, high-resolution information source that allows diagnosis of both total subduction and eddy subduction. The effective subduction period is roughly 50 days in late-winter (Section 4.4.1). After accounting for the mixed-layer demon, approximately 5 Sv is subducted into the main thermocline. Fine resolution estimates (1/6°) of the annual subduction rate are dominated by the small-scale subduction signal of magnitude up to 200 m/yr locally. Eddy subduction is calculated as the volume flux of water across the moving mixed-layer base; eddy subduction rates as high as 40 m/yr are common (Section 4.4.2). To gauge the net effect

of eddies, subduction is integrated in density-space (Section 4.4.3). The contribution of eddies to subduction is a comparable to the total subduction in the density class $25.5 < \sigma < 26.5$, which includes isopycnals that outcrop in the Azores Current in late winter (Section 4.2.5). The results of Chapter 4 are:

- **Eddy subduction rates are frequently 15% of the local subduction rate in the eastern North Atlantic.**
- **Eddy subduction is a contributor to water mass formation, and the combination of Eulerian and density-space calculations suggest that the frontal regions, such as the Azores Current and the North Equatorial Current, play a large role.**

5.2 Discussion

This thesis suggests the importance of eddies even in a region that does not include the western boundary of the basin. Previously, Marshall (1997) hypothesized that eddy subduction was important in western boundary currents, the Antarctic Circumpolar Current, and deep convection sites. Hazeleger and Drijfhout (2000) quantified the eddy contribution to subduction at 150 m/yr in an idealized Gulf Stream model. The independent Subduction Experiment synthesis of Weller et al. (2004) likewise concluded that the explicit study of eddies was necessary to close budgets and understand dynamical processes, although they did not attempt such a study. Here, frontal regions in the eastern half of the subtropical gyre have locally significant rates of eddy subduction. Relative to the energetic eddy regions suggested by Marshall (1997), eddy subduction rates of the eastern subtropical gyre are small, but give a non-negligible contribution to water mass transformation rates. Away from fronts, subduction due to eddies is negligible in the Subduction Experiment region. Nevertheless, the overall picture is one of an ocean with ubiquitous mesoscale energy that can not be ignored *a priori*.

Eddy subduction rates are locally large, but do eddies have any net impact? The annual subduction rate (Section 4.4.2, Figure 4-13) has small-scale structures due to

lateral induction by the mean circulation. A spatial average of the small-scale signal in the Azores Current region gives a subduction rate of 29 m/yr, nearly equivalent to the non-frontal value of 27 m/yr south of the front. The explicit calculation of eddy subduction (Section 4.4.2, Figure 4-14) is dominated by an alternating pattern of entrainment and detrainment. Spatial averages on scales larger than an individual eddy tend to zero by the cancellation of the dipoles. These results show that eddies do not have a net volume flux into the thermocline over the domain. Instead, eddies only significantly affect the water mass formation rates, as seen in a density-space calculation.

Model resolution and subduction

What resolution is necessary to adequately model subduction? An advantage of this thesis is that we have used two complementary models, one at 1/6° that explicitly resolves eddies, and another at 2° with the Gent-McWilliams (GM) eddy-parameterization scheme. The annual subduction rate in the coarse-resolution state estimate is very similar to the large-scale subduction rate of the fine-resolution state estimate (compare Figure 5-1 to Figure 4-13). Subduction by the mean circulation is well-captured in the eastern subtropical gyre by a coarse-resolution model.

When considering the necessary resolution of a model run, the coarse-resolution simulation of Spall et al. (2000) serves as another comparison. Spall et al. (2000) examined the output of an ocean model with 2° resolution in the Subduction Experiment region. They were able to estimate the eddy-subduction rate even though eddies were not present in the simulation. Marshall (1997) showed that because eddy subduction is equivalent to a transport by the bolus velocity, an eddy parameterization scheme (i.e., (Gent et al. 1995)) should also parameterize eddy subduction. Spall et al. (2000) estimated eddy subduction rates of no larger than 10 m/yr (Spall et al. 2000) by GM, while the explicit calculation of our fine-resolution estimate ranged to 40 m/yr. The eddy-parameterization scheme may underestimate subduction because of the inherent two-dimensional picture upon which it is based. The Azores Current, in particular, has a strong retroflection at the Mediterranean Outflow and a strong countercurrent. These

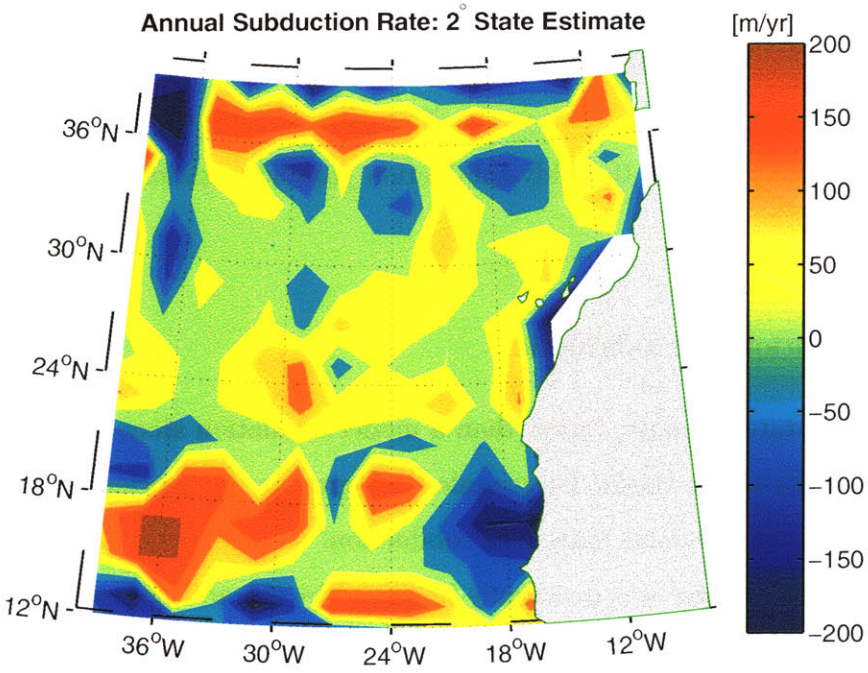

Figure 5-1: Annual subduction rate calculated from a coarse-resolution state estimate.

features are complications not considered in the theoretical underpinnings of GM. From these results, the three-dimensional pathway of water subducted by eddies in the Azores Current seems difficult to parameterize.

Because the global run of Spall et al. (2000) underrepresented the effect of eddy subduction, large-scale hydrographic biases are expected over time. The timescale of error growth depends on a ratio between the eddy subduction rate and the total volume inventory in a particular density band. This scaling argument indicates that biases will become large after ten to twenty years of model integration. Spall et al. (2000) take pains to show the similarity of their model simulation to the mooring data of the Subduction Experiment over two years. Nevertheless, a two-year observational record is not long enough to test the eddy-subduction parameterization in a coarse-resolution model.

State estimation and subduction

The use of a state estimate, rather than a model simulation alone, greatly affected the scientific results of this thesis. Large-scale hydrographic deficiencies gave a model simulation with an unreasonable pattern of subduction. Of utmost importance, the seasonal cycle of the mixed layer was drastically improved by the addition of observations. The large-scale slope of the mixed-layer base reversed orientation under an observational constraint. Estimates of lateral induction were most improved by state estimation. Transformation rates (Figure 5-2) are also adjusted by the improved air-sea flux fields. At the very least, the methodology of this thesis removed the major sources of error in the formulation of the regional model. No reasonable diagnostics of subduction would have been possible in the model simulation without data constraints.

Besides large-scale hydrographic changes, eddy-resolving state estimation changed the spatial pattern of eddy kinetic energy. To a large extent, relatively high values of kinetic energy are a prerequisite for significant eddy subduction. Thus, the inclusion of observations in the eddy-resolving model allowed an improved determination of the regions where eddy subduction is important, such as the Azores Current and the North

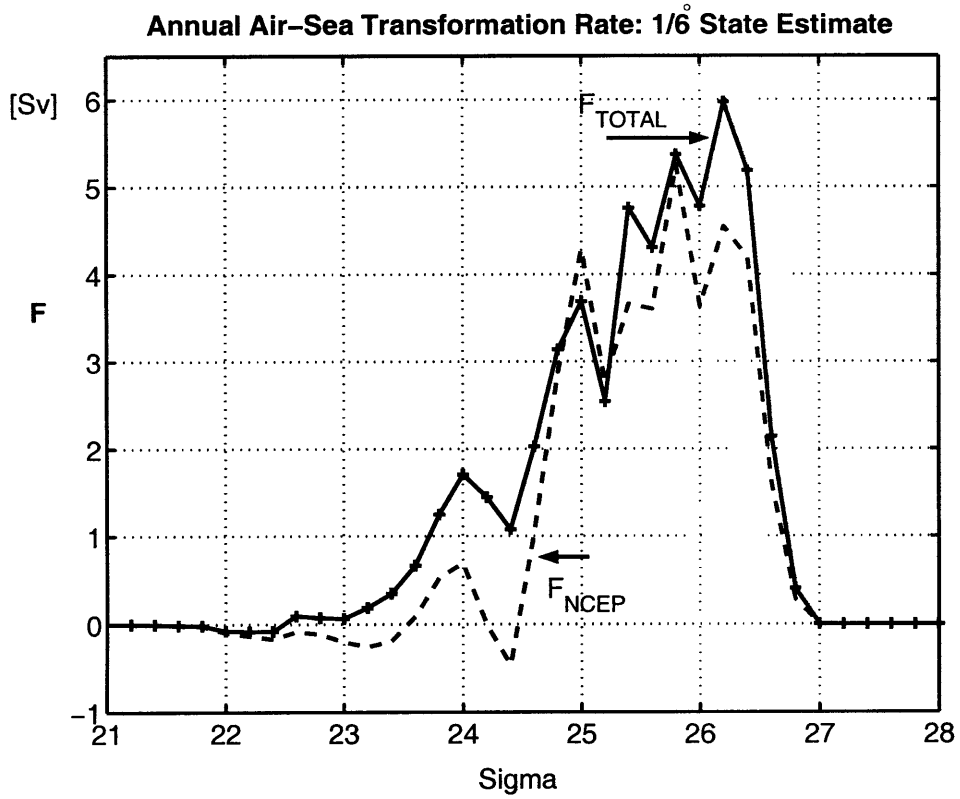

Figure 5-2: Annual air-sea transformation rate, F, of a nearly-optimized state estimate. The transformation rate in a simulation forced by NCEP heat and freshwater fluxes is labeled F_{NCEP}. The state estimate, where surface fluxes have been adjusted, in labeled F_{TOTAL}. Adjustments by state estimation fine-tune the water mass transformation characteristics.

Equatorial Current, and those where it is not.

5.3 Future work

The diagnosis of eddy subduction is limited in this thesis by the short duration of the state estimate. One year constitutes only three to four eddy lifecycles, and averages over this time interval still contain small-scale structures. Idealized model studies of subduction (i.e. Hazeleger and Drijfhout (2000)) usually average over twenty or more years of model integration for a statistical steady state. A logical next step is the diagnosis of eddy subduction in a realistic model simulation with many seasonal cycles. Given the present computational resources, it is feasible to run a realistic model at eddy-resolution over twenty years in a regional configuration.

As seen in Section 5.2, the state estimate has an improved large-scale hydrographic structure relative to the model simulation. Hence, subduction is captured more realistically when observations are taken into account. A long-term, eddy-resolving state estimate is the ultimate tool to study the role of eddies in subduction. The computational burden of finding a state estimate is perhaps 100 times greater than a model simulation. Nevertheless, the ECCO Group has already discussed a ten-year, 1/4° state estimate, so the future may not be far away.

Besides the computational requirements of state estimation, a long record of observations is also necessary. For a nonlinear system, it is likely that a dense supply of data is required to keep the state estimate on track. The continuous supply of sea surface height data from first TOPEX/POSEIDON, and now the JASON satellite, is a major boost to the data stream. Field campaigns have increasingly focused on new forms of instruments, such as the global array of floats in the ARGO experiment. With new forms of measurements, new methods may be needed to incorporate the information into a state estimate. For the future, the observational design problem must be explicitly addressed if the ocean is to be monitored over a wide range of space and time scales.

The methodology of state estimation will be tested in the case of long-duration eddy-

resolving estimates. Regional models are increasingly controlled by the open boundary conditions over long times, but it is difficult to model open boundaries accurately. Global models also serve as a test because of differing physical regimes in different regions. Western boundary currents, open ocean deep convection, and sea-ice formation are nonlinear processes and probably represent necessary components of a realistic global ocean model. In the face of these strong nonlinear features, the usefulness of adjoint-computed gradients can be questioned (Lea et al. 2000; Köhl and Willebrand 2002). The method of Lagrange multipliers may not be applicable blindly.

The scientific field of state estimation evolved with low-dimensional, linear problems. Modifications are necessary for the oceanographic setting because of the high-dimension and nonlinearity of the equations of motion. Two preliminary adjustments were implemented in this thesis. One, the size of the control space was effectively reduced by nondiagonal weighting matrices. Spatial and temporal correlation in the control fields means that the effective degrees of freedom, and hence, size of the search space, is lessened. Two, Section 3.5.3 used a cost function which enforced the model to follow only the large-scale observational signal. This could be called a *multiscale method.* The knowledge of spatial and temporal correlations could be used more fully in state estimation, and the steps taken here are just a start. As a final note, control theory was the original source of the state estimation methodology. It is logical to continue to apply ideas borrowed from control theory to the high-dimensional, nonlinear world of fluid dynamics and climate science.

Appendix A

MIT General Circulation Model equations

This is a brief introduction to the formulation of the primitive equations used by the MIT General Circulation Model (MIT GCM). Marshall et al. (1997a,b) described the model in greater detail. Here, the model is used with the hydrostatic form of the primitive equations under the Boussinesq assumption. The model conserves horizontal and vertical momentum, volume, heat and salt. With the equation of state and the free surface equation, seven equations constitute the core dynamics of the numerical model.

$$\frac{D\mathbf{u}}{Dt} = -\frac{\nabla p}{\rho_0} - 2\Omega \times \mathbf{u} + \frac{1}{\rho_0}\nabla \cdot \tau_h + \nu_h \nabla^4 \mathbf{u} + \frac{\partial}{\partial z}\nu_z \frac{\partial \mathbf{u}}{\partial z} \tag{A.1}$$

$$\partial_z p = -g\rho \tag{A.2}$$

$$\partial_z w = -\nabla \cdot \mathbf{u} \tag{A.3}$$

$$\frac{D\theta}{Dt} = \kappa_h \nabla^4 \theta + \frac{\partial}{\partial z}\kappa_z \frac{\partial \theta}{\partial z} + \nabla \cdot H_Q \tag{A.4}$$

$$\frac{DS}{Dt} = \kappa_h \nabla^4 S + \frac{\partial}{\partial z}\kappa_z \frac{\partial S}{\partial z} + \nabla \cdot H_F \tag{A.5}$$

$$\rho = \rho(\theta, S, z) \tag{A.6}$$

$$\partial_t \eta = -\nabla \cdot \int_{-H}^{0} \mathbf{u}\, dz + (P - E) \tag{A.7}$$

In the horizontal momentum equation (A.1), $\mathbf{u}$ is the 2-component horizontal velocity, p is the deviation of the pressure from a resting ocean of density ρ_0, Ω is the rotation rate of Earth, ∇ is a horizontal operator, τ represents wind stress at the surface, and ν is viscosity. The vertical momentum equation (A.2) reduces to hydrostatic balance, where g is gravity. Conservation of mass becomes conservation of volume (A.3) under the Boussinesq assumption, equally expressed as nondivergence of the three-dimensional flow. Heat, directly related to potential temperature, θ, is conserved in the absence of diffusion, κ, and external heating, H_Q (A.4). Salinity, S, is also conserved in the absence of diffusion and freshwater forcing, H_F (A.5). The equation of state (A.6) is a nonlinear polynomial in which density depends on temperature, salinity, and depth. The sea surface height η evolution, described by (A.7), introduces a new prognostic equation in the hydrostatic PE's. $(P - E)$ is the volume input by excess precipitation over evaporation. The general circulation model is nonlinear due to the equation of state, as well as the advection terms hidden in the total derivatives, $\frac{D}{Dt}$. In sum, there are 7 dependent variables and 7 equations for their evolution.

The KPP model (as discussed in Section 1.2.2) is appended to the model. It diagnoses turbulent viscosity and diffusivity which is then used in the prognostic model equations.

$$\nu_z = \nu_z(x, y, z, \rho, H_Q, H_F) \tag{A.8}$$

$$\kappa_z = \kappa_z(x, y, z, \rho, H_Q, H_F) \tag{A.9}$$

Solution Method

The hydrostatic primitive equations are discretized on a staggered grid, the C grid of Arakawa (1977). The bottom boundary has no-slip conditions, but the lateral solid boundaries have slip conditions. Potential temperature, salinity, horizontal velocity, and sea surface height are prognostic quantities, stepped forward in time by an Adams-Bashforth discretization. Vertical velocity and density are diagnostic, calculated by Eqs. (A.2,A.6).

Pressure is also a diagnostic variable, but it is not explicitly described by the previous

equations. To form an explicit equation, split pressure into two components, surface and hydrostatic:

$$p(x,y,z) = p_S(x,y) + p_H(x,y,z) \approx g\rho_0\eta + \int_z^0 g\rho \, dz, \tag{A.10}$$

where hydrostatic balance makes a direct connection between surface pressure and sea surface height. Nevertheless, the surface pressure still does not have an explicit equation which will guarantee a nondivergent flow. To remedy this problem, the horizontal momentum equation is written in a simplified form, where pressure is split into two components, and $\mathbf{G}_u$ takes the place of the extra right hand terms:

$$\frac{\partial \mathbf{u}}{\partial t} = -\frac{\nabla p_S}{\rho_0} + \mathbf{G}_u = -g\nabla\eta + \mathbf{G}_u, \tag{A.11}$$

with use of the linearized definition of surface pressure (Equation A.10). Now, substituting the previous equation into the time derivative of the free surface equation (Equation A.7) yields an elliptic equation

$$\nabla \cdot (gH\nabla\eta) + \frac{\partial^2 \eta}{\partial t^2} = -\nabla \cdot \int_H^0 \mathbf{G}_u \, dz + \frac{\partial}{\partial t}(P - E) \tag{A.12}$$

where the next to last term should vanish. In practice, the new velocity $\mathbf{u}$ is not perfectly nondivergent, so the term is kept and it leads to adjustment in the pressure field. Equation (A.12) is discretized with a backwards implicit scheme, and is solved iteratively when the boundary is irregular.

The boundary conditions for the elliptic operator are modified with open boundaries. With a closed boundary, the operator has homogeneous Neumann boundary conditions, $\nabla p \cdot n = \mathbf{G} \cdot n = 0$. When the domain boundaries are open, one new term is added to the elliptic equation boundary conditions; it is a term that allows for a change in total volume inside the domain. Zhang et al. (1999) gave a full discussion of the modification, but in a technical sense, it is a very small change to the numerical code.

Appendix B

Relationship of the forward and adjoint state

The Lagrange multipliers are frequently called the *adjoint state* because they are stepped backwards in time by the adjoint model, in analogy with the variables stepped by the forward model, the *forward state*. The analogy is made more complete when considering that the number of Lagrange multipliers equals the number of state variables at any time. Equivalently, the adjoint model has the same dimension as the forward model. Algorithmically-differentiated numerical code makes explicit the connection between the forward and adjoint state; for example, Marotzke et al. (1999) show that the adjoint state has sensitivity information directly related to the corresponding forward state. With the mathematical formulation of the Lagrangian function, equation (3.3) of Chapter 3, a tight relationship still exists but is not clearly seen.

The first difference between the numerical adjoint code and the formal mathematics is the status of the adjoint state at time $t = 0$. In the numerical code, $\boldsymbol{\mu}(0)$ exists, but it is not defined for the equations of Chapter 3. With a few extra definitions, the adjoint state can be extended formally to $t = 0$. Consider the first time step of the model:

$$\mathbf{x}(1) = \mathcal{L}[\mathbf{x}(0), \mathbf{B}\mathbf{q}(0), \boldsymbol{\Gamma}\mathbf{u}(0)]. \tag{B.1}$$

Upon closer inspection, this time step has two parts: the specification of the initial conditions, then the forward step of the model dynamics. The specification of the initial conditions of the forward model could be written as a separate model step:

$$\mathbf{x}(0) = \mathbf{q}_i(0) + \mathbf{u}_i(0) \tag{B.2}$$

where $\mathbf{q}_i(0)$ is the first guess of the initial conditions, and $\mathbf{u}_i(0)$ is a control adjustment to the initial conditions. Furthermore, this statement could be added explicitly to the Lagrangian function with a preceding Lagrange multiplier:

$$\begin{aligned} J = \textstyle\sum_{t=1}^{t_f} \; & [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)]^T \mathbf{W}(t) \; [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)] \\ & + \textstyle\sum_{t=0}^{t_f-1} \; \mathbf{u}(t)^T \mathbf{Q}(t) \mathbf{u}(t) \\ - \textstyle\sum_{t=0}^{t_f-1} \; & \boldsymbol{\mu}(t+1)^T \{\mathbf{x}(t+1) - \mathcal{L}[\mathbf{x}(t), \mathbf{B}\mathbf{q}(t), \boldsymbol{\Gamma}\mathbf{u}(t)]\} \\ & - \boldsymbol{\mu}(0)^T \{\mathbf{x}(0) - \mathbf{q}_i(0) - \mathbf{u}_i(0)\}. \end{aligned} \tag{B.3}$$

where $\boldsymbol{\mu}(0)$ will be shown to be a judicious choice for the new Lagrange multiplier. The adjoint equation for the timestep from $t = 1$ to $t = 0$ is slightly changed. This adjoint model timestep is recovered by setting the derivative of J with respect to $\mathbf{x}(0)$ equal to zero:

$$\frac{\partial J}{\partial \mathbf{x}(0)} = -\boldsymbol{\mu}(0) + (\frac{\partial \mathcal{L}}{\partial \mathbf{x}(0)})^T \boldsymbol{\mu}(1) = 0, \tag{B.4}$$

and rearranging,

$$\boldsymbol{\mu}(0) = (\frac{\partial \mathcal{L}}{\partial \mathbf{x}(0))})^T \boldsymbol{\mu}(1). \tag{B.5}$$

With a backwards sweep of the adjoint model, the Lagrange multiplier at $t = 0$ is computable. The meaning of $\boldsymbol{\mu}(0)$ is seen as:

$$\frac{\partial J}{\partial \mathbf{q}_i(0)} = \boldsymbol{\mu}(0), \tag{B.6}$$

the sensitivity of the cost function with respect to the initial conditions. This relationship, derived through the formal mathematics, is easily seen in the numerical code.

A similar idea can be used at any timestep. Consider an additive perturbation to the state as a hypothetical control variable $\mathbf{u}^*(t)$, which is not penalized in the Lagrangian function and not part of the original controls, $\mathbf{u}(t)$. The additive perturbation is applied in the model:

$$\mathbf{x}(t+1) = \mathcal{L}[\mathbf{x}(t), \mathbf{B}\mathbf{q}(t), \mathbf{\Gamma}\mathbf{u}(t)] + \mathbf{u}^*(t+1) \tag{B.7}$$

The Lagrangian function is now rewritten as:

$$\begin{aligned} J = &\textstyle\sum_{t=1}^{t_f} \; [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)]^T \mathbf{W}(t) \; [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)] \\ &+ \textstyle\sum_{t=0}^{t_f-1} \; \mathbf{u}(t)^T \mathbf{Q}(t) \mathbf{u}(t) \\ &- \textstyle\sum_{t=0}^{t_f-1} \; \boldsymbol{\mu}(t+1)^T \{\mathbf{x}(t+1) - \mathcal{L}[\mathbf{x}(t), \mathbf{B}\mathbf{q}(t), \mathbf{\Gamma}\mathbf{u}(t)] - \mathbf{u}^*(t+1)\} \end{aligned} \tag{B.8}$$

The meaning of the Lagrange multipliers for times $1 < t < t_f$ is elucidated:

$$\frac{\partial J}{\partial \mathbf{u}^*(t)} = \boldsymbol{\mu}(t). \tag{B.9}$$

The Lagrange multiplier is the sensitivity of the cost function to an additive perturbation of the state at its respective time. In other words, the Lagrange multiplier gives the influence of each state element as if it were independently adjustable.

The previous result is important when interpreting the time history of the adjoint state (i.e., Figure 3-17). The adjoint state is interpreted as the sensitivity of J to $\mathbf{x}(t)$. This sensitivity has the same magnitude as the sensitivity to initial conditions of a model trajectory of $t_f - t$ time units.

Appendix C

Chaotic dynamics of the forced, nonlinear pendulum

The forced, nonlinear pendulum is a simple system that is chaotic. If periodic forcing is added to the nonlinear pendulum, its dynamics are governed by:

$$\frac{d^2\theta}{dt^2} + q\,\frac{d\theta}{dt} + sin(\theta) = \mathcal{F} \tag{C.1}$$

where $\mathcal{F} = g\ cos(\omega_d t)$, and q is a damping coefficient. The continuous-time state space realization of the system is:

$$\frac{d\omega}{dt} = -q\,\omega - sin(\theta) + g\ cos(\phi) \tag{C.2}$$

$$\frac{d\theta}{dt} = \omega \tag{C.3}$$

$$\frac{d\phi}{dt} = \omega_d \tag{C.4}$$

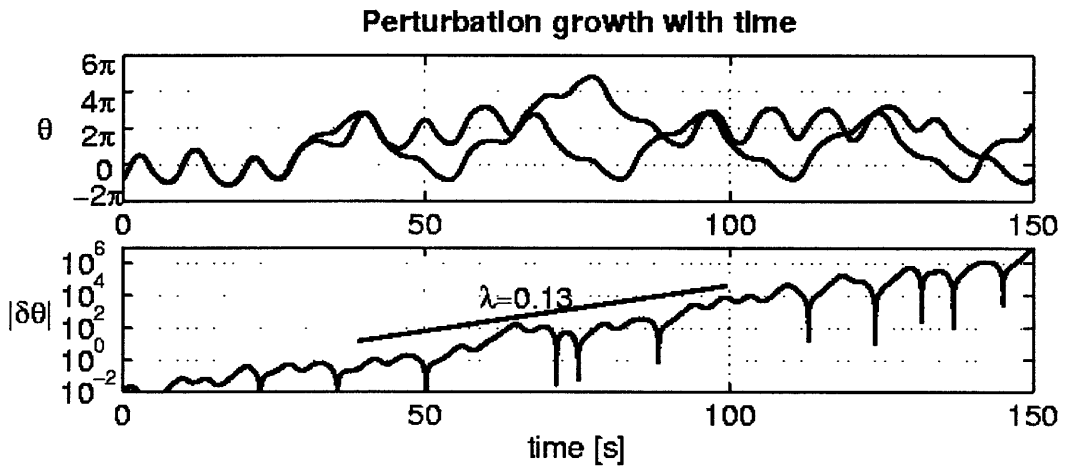

Figure C-1: Growth of an initial perturbation in the forced, nonlinear pendulum. *Upper*: The time evolution of the displacement angle, θ, for two pendulums with initial angle separation of 0.01 *radians*. The model trajectories diverge after fifty seconds due to chaotic dynamics. *Lower*: The time evolution of the magnitude of an infinitesimal perturbation of θ as calculated by the tangent linear model. The growth is exponential for an indefinite period of time, $|\delta\theta(t)| = 0.01e^{\lambda t}$, characteristic of a chaotic system.

where the state includes θ as the displacement angle, ω the angular velocity, and ϕ the phase of the forcing. The linearized continuous-time propagator is:

$$\left(\frac{\partial \mathcal{L}}{\partial \mathbf{x}(t)}\right) = \begin{pmatrix} -q & -cos(\theta) & -g\ sin(\phi) \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{C.5}$$

which has a maximum eigenvalue of $\lambda = 0.67$ for the sample point $[\omega, \theta, \phi] = [1.29, -2.96, 333.35]$. The dynamics are unstable at this point and many other points in phase space.

As an aside, the previous argument assumes that the nonlinear model is linearized around a fixed state, but this is not accurate for a dynamic model. Instability of the linearized model is not a sufficient condition for chaos and unbounded exponential growth of perturbations. The magnitude of the difference between two trajectories is actually determined by the greatest singular value of the linearized model. For the pendulum, this explains why the perturbation's magnitude grows with an exponent of $\lambda = 0.13$ (Figure C-1, lower panel) instead of the largest eigenvalue, $\lambda = 0.67$, at our sample

point. Furthermore, Lyapunov exponents are related to the singular values of the matrix over long integration times, and they describe the exponential divergence of neighboring particles around the entire state space (Palmer 1996). The nonlinear pendulum does have a positive Lyapunov exponent, a more exact test for chaotic dynamics.

Nearly-nondifferentiable dynamics can result from chaos. Here, chaos is defined as the sensitive dependence on initial conditions (Lorenz 1963; Gauthier 1992). A slight perturbation to the state, $[\omega, \theta, \phi]$, means that the model never returns to the original trajectory. Although the cost function itself has a physical bound, no bound exists for the gradients of the cost function of a long time-integration of a chaotic model (Figure C-2). Eventually, the Lagrange multipliers are so large that they are incalculable by a numerical implementation of the adjoint equations; the model is nearly nondifferentiable.

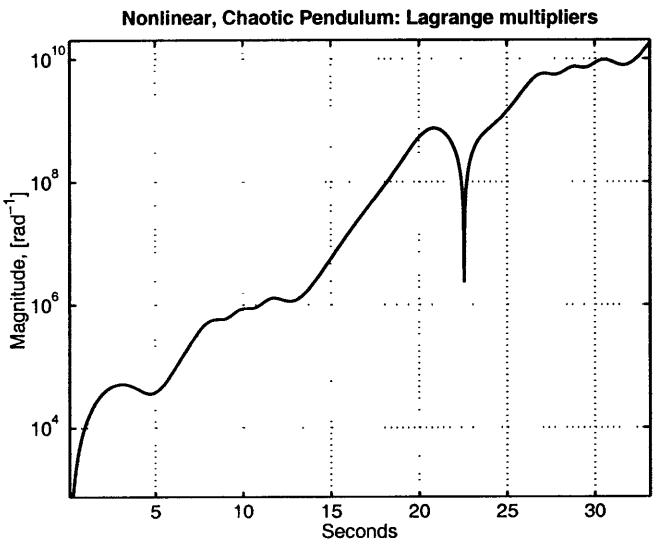

Figure C-2: The evolution of the Lagrange multipliers of the nonlinear, chaotic pendulum ($g = 1.15$) with reversed time. The maximum Lagrange multiplier, $||\mu(t)||_\infty$, increases exponentially. Compare to Figure 3-18.

Appendix D

Isopycnal analysis in a z-coordinate state estimate

Isopycnal analysis of the z-coordinate (or *level coordinate*) state estimate follows the numerical model analysis in Appendices A and B of Marshall et al. (1999). A few modifications have been made to the analysis here, in an effort to reduce diagnostic errors and in order to apply the treatment to a regional model.

One extension to the work of Marshall et al. (1999) is discussed next. The water-mass subduction rate is never explicitly calculated in Marshall et al. (1999). Here, the exact diagnosis is more complicated due to open boundary sources. The recipe for calculating $M(\sigma, t)$ follows. Define the base of the control volume, $H(x, y)$, to be the depth of the mixed layer. Outside of the regional boundaries, set $H(x, y) = 0$. The volume flux at density less than σ across the surface defined by $H(x, y)$ is $M(\sigma, t)$. There are two sources to $M(\sigma, t)$:

$$M(\sigma, t) = M_B(\sigma, t) - S(\sigma, t), \tag{D.1}$$

the volume flux across the lateral boundary, $M_B(\sigma, t)$, and the volume flux across the horizontally-varying bottom boundary, $S(\sigma, t)$. When diagnosed on the C-grid of the MIT GCM and state estimate,

$$
\begin{aligned}
& M(\sigma,t) \\
&= \Sigma_{ijk} \quad u(i,j,k,t)\cdot a_{yz}(i-1/2,j,k)\cdot \Pi_{ML_u}(\sigma_n,i,j,k,t) \\
&+\Sigma_{ijk} \quad v(i,j-1/2,k,t)\cdot a_{xz}(i,j-1/2,k)\cdot \Pi_{ML_v}(\sigma_n,i,j,k,t) \\
&+\Sigma_{ijk} \quad w(i,j,k-1/2,t)\cdot a_{xy}(i,j)\cdot \Pi_{ML_w}(\sigma_n,i,j,k,t),
\end{aligned} \tag{D.2}
$$

where $a_{xy,xz,yz}$ is the area of the respective grid face and $\Pi_{ML_{u,v,w}}$ is a boxcar function. The velocity is defined on a staggered grid relative to the tracer and density fields. Hence, coordinates with 1/2 refer to grid faces, not the center of grid cells. Density values must be interpolated to grid faces, and a simple linear scheme is used here. From above, the boxcar function, Π_{ML_u}, for example, is defined by:

$$
\Pi_{ML_u}(\sigma_n,i,j,k,t) = \begin{cases}
1 \quad if \begin{cases} \sigma(i-\frac{1}{2},j,k,t) < \sigma_n \\ H(i-1,j) \le z(k) < H(i,j) \end{cases} \\
-1 \quad if \begin{cases} \sigma(i-\frac{1}{2},j,k,t) < \sigma_n \\ H(i-1,j) > z(k) \ge H(i,j) \end{cases} \\
0 \qquad otherwise
\end{cases} \tag{D.3}
$$

Boxcar functions for the other components of velocity follow in a similar way.

$S(\sigma,t)$ must still be isolated from $M(\sigma,t)$. Replace the full velocity field with the open boundary velocity field, $(u,v,w) = (u_B, v_B, 0)$, and reevaluate Equation (D.2) to estimate the open boundary volume flux, $M_B(\sigma,t)$. Then, the water-mass subduction rate is deduced by subtraction (Equation (D.1)).

To eliminate any linear interpolation in density space, the analysis of the diapycnal advective flux has been modified. Instead of computing the flux across the bounding isopycnals of a density bin, we compute the advective flux across the same set of density contours that are used in Equation (D.2). Then, the equation for $A(\sigma,t)$ is identical to

Appendix A of Marshall et al. (1999), except the boxcar function is redefined:

$$\Pi_u(\sigma_n, i, j, k, t)) = \begin{cases} 1 \quad if \begin{cases} \sigma(i-1,j,k,t) < \sigma_n < \sigma(i,j,k,t) \\ z(k) < H(i,j) \\ z(k) < H(i-1,j) \end{cases} \\ -1 \quad if \begin{cases} \sigma(i,j,k,t) < \sigma_n < \sigma(i-1,j,k,t) \\ z(k) < H(i,j) \\ z(k) < H(i-1,j) \end{cases} \\ 0 \qquad otherwise \end{cases} \tag{D.5}$$

References

Adcroft, A., and D. Marshall, 1998: How slippery are piecewise-constant coastlines in numerical ocean models?, *Tellus*, **50 (1)**, 95–108.

Adcroft, A., 1994: Numerical algorithms for use in a dynamical model of the ocean, Ph.D. thesis, Imperial College London.

Arakawa, A., and V. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model, *Methods in Computational Physics*, **17**, 174–267.

Baker, G. L., and J. P. Gollub, 1990: *Chaotic Dynamics: An Introduction*, Cambridge University Press.

Barth, N., 1992: Oceanographic experiment design II: Genetic algorithms, *J. Atm. Oceanic. Tech.*, **9 (4)**, 434–443.

Barth, N., and C. Wunsch, 1990: Oceanographic experiment design by simulated annealing, *J. Phys. Oceanogr.*, **20**, 1249–1263.

Bennett, A., and P. E. Kloeden, 1981: The ill-posedness of open ocean models, *J. Phys. Oceanogr.*, **12**, 1004–1018.

Bennett, A. F., 2002: *Inverse Modeling of the Ocean and Atmosphere*, Cambridge University Press, 234 pp.

Brink, N. J., K. A. Moyer, R. P. Trask and R. A. Weller, 1995: The Subduction Experiment: Mooring field program and data summary, Woods Hole Oceanographic Institution, *Tech. rep.*

Broecker, W., 1991: The great ocean conveyor, *Oceanography*, **4**, 79–89.

Bugnion, V., 2001: Driving the ocean's overturning: An adjoint sensitivity study, Ph.D. thesis, Massachusetts Institute of Technology.

Cazenave, A., 2002: Global trends in sea level, Presented at *WOCE and Beyond*, San Antonio, TX, 21 November 2002.

Chang, P., L. Ji, H. Li and M. Flügel, 1996: Chaotic dynamics versus stochastic processes in El Niño-Southern Oscillation in coupled ocean-atmosphere models, *Physica D*, **98**, 301–320.

Charney, J. G., R. Fjortoft and J. von Neumann, 1950: Numerical integration of the barotropic voriticity equation, *Tellus*, **2**, 237–254.

Chelton, D. B., S. K. Esbensen, M. G. Schlax, N. Thum, M. H. Freilich, F. J. Wentz, C. L. Gentemann, M. J. McPhaden and P. S. Schopf, 2001: Observations of coupling between

surface wind stress and sea surface temperature in the eastern Tropical Pacific, *J. Climate*, **14**, 1479–1498.

Cushman-Roisin, B., 1987: Subduction, in *Aha Huliko'a Proceedings of the University of Hawai'i*, edited by P. Muller, and D. Henderson, pp. 181–196.

Czaja, A., and C. Frankignoul, 2002: Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation, *J. Climate*, **15**, 3280–3290.

Dahleh, M. A., and I. Diaz-Bobillo, 1999: *Control of Uncertain Systems: A Linear Programming Approach*, Pergamon Press.

Dantzig, G. B., A. Orden and P. Wolfe, 1955: Generalized simplex method for minimizing a linear form under linear inequality restraints, *Pacific J. Math.*, **5**, 183–195.

Davis, L., 1991: *Handbook of Genetic Algorithms*, p. 385, Van Nostrand Reinhold, New York.

Defant, A., 1936: *The Troposphere: Scientific Results of the German Atlantic Expedition of the Research Vessel 'Meteor' 1925-1927*, Verlag von Walter de Gruyter and Co., Berlin and Leipzig.

Deser, C., and M. L. Blackmon, 1995: On the relationship between tropical and North Pacific termperature variations, *J. Climate*, **8**(6), 1677–1680.

Farrell, B. F., 1989: Optimal excitation of baroclinic waves, *J. Atmos. Sci.*, **46**, 1193–1206.

Farrell, B. F., and A. M. Moore, 1993: An adjoint method for obtaining the most rapidly growing perturbations to oceanic flows, *J. Phys. Oceanogr.*, **22**, 338–349.

Ferron, B., and J. Marotzke, 2003: Impact of 4D-variational assimilation of WOCE hydrography on the meridional circulation of the Indian Ocean, *Deep Sea Research Part II: Topical Studies in Oceanography*, **50**, 2005–2021.

Follows, M. J., and J. C. Marshall, 1994: Eddy driven exchange at ocean fronts, *Ocean Modelling*, **102**, 5–9.

Follows, M. J., and J. C. Marshall, 1996: On models of bomb ^{14}C in the North Atlantic, *J. Geophys. Res.*, **101**, 22577–22582.

Foster, I., 1995: *Designing and Building Parallel Programs*, Addison-Wesley, Also published on the Web as Designing and Building Parallel Programs (Online).

Fu, L.-L., B. Cheng and B. Qiu, 2001: 25-day period large-scale oscillations in the Argentine Basin revealed by the TOPEX/Poseidon altimeter, *J. Phys. Oceanogr.*, **31**, 506–516.

Fu, L.-L., E. Christensen, C. A. Yamarone, M. Lefebvre, Y. Menard, M. Dorrer and P. Escudier,

1994: TOPEX/POSEIDON mission overview, *J. Geophys. Res.*, **99 (C12)**, 24369–24381.

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.

Galanti, E., and E. Tziperman, 2003: A mid-latitude ENSO teleconnection mechanism via baroclinically unstable Long Rossby Waves, *J. Phys. Oceanogr.*, **33**, 1877–1888.

Garrett, C., K. Speer and E. Tragou, 1995: The relationship between water mass transformation and the surface buoyancy flux, with application to Phillips' Red Sea model, *J. Phys. Oceanogr.*, **25**, 1696–1705.

Gauthier, P., 1992: Chaos and quadri-dimensional data assimilation: A study based on the Lorenz model, *Tellus*, **44A**, 2–17.

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

Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography, in *Advances in Geophysics*, vol. 33, pp. 141–266, Academic Press.

Giering, R., and T. Kaminski, 1998: Recipes for adjoint code construction, *ACM Trans. Math. Software*, **24 (4)**, 437–474.

Gilbert, J. C., and C. Lemaréchal, 1989: Some numerical experiments with variable-storage quasi-Newton algorithms, *Math. Program.*, **45**, 407–435.

Gilcrest, B., and G. Cressman, 1954: An experiment in objective analysis, *Tellus*, **6**, 308–318.

Gill, A. E., J. S. A. Green and A. J. Simmons, 1974: Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies, *Deep-Sea Res.*, **21**, 499–528.

Gill, P. E., W. Murray and M. H. Wright, 1986: *Practical Optimization*, p. 401, Academic Press.

Griewank, A., and A. Walther, 2000: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation, *ACM Trans. Math. Software*, **26 (1)**.

Griewank, A., 2000: *Evaluating Derivatives: Principles and Techniques of Alogorithmic Differentiation*, p. 369, Society for Industrial and Applied Mathematics, Philadelphia.

Griffies, S. M., and K. Bryan, 1997: Predictability of North Atlantic interdecadal variability, *Science*, pp. 181–184.

Gu, D., and S. G. H. Philander, 1997: Interdecadal climate fluctuations that depend on exchanges between the tropics and extratropics, *Science*, **275 (7)**, 805–807.

Gunson, J. R., and P. Malanotte-Rizzoli, 1996a: Assimilation studies of open-ocean flows 1. Estimation of initial and boundary conditions, *J. Geophys. Res.*, **101 (C12)**, 28,457–28,472.

Gunson, J. R., and P. Malanotte-Rizzoli, 1996b: Assimilation studies of open-ocean flows 2. Error measures with strongly nonlinear dynamics, *J. Geophys. Res.*, **101 (C12)**, 28,473–28,488.

Haidvogel, D., and A. Beckmann, 1999: *Numerical Ocean Circulation Modeling*, p. 344, Imperial College Press, London.

Hall, M. C. G., D. G. Cacuci and M. E. Schlesinger, 1982: Sensitivity analysis of a radiative-convective model by the adjoint method, *J. Atmos. Sci.*, **39**, 2038–2050.

Hartmann, D. L., 1994: *Global Physical Climatology*, Academic Press, 411 pp.

Hazeleger, W., and S. S. Drijfhout, 2000: Eddy subduction in a model of the subtropical gyre, *J. Phys. Oceanogr.*, **29**, 266–288.

Hernandez, F., P.-Y. Le Traon and N. H. Barth, 1995: Optimizing a drifter cast strategy with a genetic algorithm, *J. Atm. Oceanic. Tech*, **12**, 330–345.

Hill, C., V. Bugnion, M. Follows and J. Marshall, 2004: Evaluating carbon sequestration efficiency in an ocean circulation model by adjoint sensitivity analysis, *J. Climate*, Submitted.

Holland, J., 1975: *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.

Huang, R. X., 1990: On the three-dimensional structure of the wind-driven circulation in the North Atlantic, *Dyn. Atmos. Oceans*, **15**, 117–159.

Huang, R. X., and S. Russell, 1995: Ventilation of the subtropical North Pacific, *J. Phys. Oceanogr.*, **24**, 2589–2605.

Iselin, C. O., 1939: The influence of vertical and lateral turbulence on the characteristics of the waters at mid-depths, *Trans. Am. Geophys. Un.*, **20**, 414–417.

Jenkins, W. J., 1987: ^3H and ^3He in the beta triangle: Observations for gyre ventilation and oxygen utilization rates, *J. Phys. Oceanogr.*, **17**, 763–783.

Jenkins, W. J., 1988: Studying subtropical thermocline ventilation and circulation using Tritium and ^3He, *J. Geophys. Res.*, **103**, 15817–15813.

Jia, Y., 2000: Formation of an Azores Current due to Mediterranean overflow in a modelling

study of the North Atlantic, *J. Phys. Oceanogr.*, **103**, 2342–2358.

Joyce, T. M., J. R. Luyten, A. Kubryakov, F. B. Bahr and J. S. Pallant, 1998: Meso- to large-scale structure of subducting water in the subtropical gyre of the eastern North Atlantic ocean, *J. Phys. Oceanogr.*, **28**, 40–61.

Kalnay, E., and coauthors, 1996: The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.*, pp. 77431–77437.

Köhl, A., and J. Willebrand, 2002: An adjoint method for the assimilation of statistical characteristics into eddy-resolving ocean models, *Tellus*, **54 (4)**, 406–425.

Köhl, A., and J. Willebrand, 2003: Variational assimilation of SSH variability from TOPEX/POSEIDON and ERS1 into an eddy-permitting model of the North Atlantic, *J. Geophys. Res.*, **108 (C3)**, 3092.

Large, W., J. C. McWilliams and S. C. Doney, 1994: Oceanic vertical mixing: A review and model with nonlocal boundary layer parameterization, *Rev. of Geophys.*, **32**, 363–403.

Lawson, W. G., and J. A. Hansen, 2004: Displacement error models and ensemble-based data assimilation, Submitted to *Mon. Wea. Rev.*

Lazar, A., T. Inui, P. Malanotte-Rizzoli, A. J. Busalacchi, L. Wang and R. Murtugudde, 2002: Seasonality of the ventilation of the tropical Atlantic thermocline in an ocean general circulation model, *J. Geophys. Res.*, **107 (C8)**, 10.1029/2000JC000667.

Lea, D., 2001: Joint assimilation of sea surface temperature and sea surface height, Ph.D. thesis, University of Oxford.

Lea, D. J., M. R. Allen and T. W. N. Haine, 2000: Sensitivity analysis of the climate of a chaotic system, *Tellus*, **52A**, 523–532.

LeDimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, *Tellus*, **38A**, 97–110.

Ledwell, J. R., A. J. Watson and C. S. Law, 1993: Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment, *Nature*, **364**, 701–703.

Lemoine, F., D. E. Smith, L. Kunz, R. Smith, E. C. Pavlis et al., 1997: *The Development of the NASA GSFC and NIMA Joint Geopotential Model*, IAP Symposium, Springer-Verlag, in Proc. Int. Symp. Gravity, Geoid and Marine Geodesy.

Levitus, S. R., and T. P. Boyer, 1994: *World Ocean Atlas 1994: Temperature*, U. S. Dept. of Commerce.

Levitus, S. R., R. Burgett and T. Boyer, 1994: *World Ocean Atlas 1994: Salinity*, U. S. Dept. of Commerce.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194.

Lorenz, E. N., 1963: Deterministic, nonperiodic flow, *J. Atmos. Sci.*, **20**, 130–141.

Luenberger, D. G., 1984: *Linear and Nonlinear Programming*, Addison-Wesley, 491 pp.

Luyten, J. R., J. Pedlosky and H. Stommel, 1983: The ventilated thermocline, *J. Phys. Oceanogr.*, **13**, 292–309.

Malanotte-Rizzoli, P., K. Hedstrom, H. Arango and D. B. Haidvogel, 2000: Water mass pathways between the subtropical and tropical ocean in a climatological simulation of the North Atlantic ocean circulation, *Dyn. Atmos. Oceans*, **32**, 331–371.

Malanotte-Rizzoli, P., and W. R. Holland, 1986: Data constraints applied to models of the ocean general circulation, Part I: The steady case, *J. Phys. Oceanogr.*, **16**, 1665–1687.

Marchesiello, P., J. C. McWilliams and A. Shchepetkin, 2001: Open boundary conditions for long-term integration of regional oceanic models, *Ocean Modelling*, **3**, 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**, 529–547.

Marotzke, J., and C. Wunsch, 1993: Finding the steady state of a general circulation model through data assimilation: Application to the North Atlantic Ocean, *J. Geophys. Res.*, **98**, 20149–20167.

Marshall, D., 1997: Subduction of water masses in an eddying ocean, *J. Mar. Res.*, **55**, 201–222.

Marshall, D., and J. Marshall, 1995: On the thermodynamics of subduction, *J. Phys. Oceanogr.*, **25**, 138–151.

Marshall, J., A. Adcroft, C. Hill, L. Perelman and C. Heisey, 1997a: A finite volume, incompressible Navier-Stokes model for studies of the ocean on parallel computers, *J. Geophys. Res.*, **102**, 5753–5766.

Marshall, J., C. Hill, L. Perelman and A. Adcroft, 1997b: Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling, *J. Geophys. Res.*, **102**, 5733–5752.

Marshall, J., D. Jamous and J. Nilsson, 1999: Reconciling thermodynamic and dynamic meth-

ods of computation of water-mass transformation rates, *Deep-Sea Res.*, **46**, 545–572.

Marshall, J. C., and A. J. G. Nurser, 1991: A continuously stratified thermocline model incorporating a mixed layer of variable thickness and density, *J. Phys. Oceanogr.*, **21**, 1780–1792.

Marshall, J. C., A. J. G. Nurser and R. G. Williams, 1993: Inferring the subduction rate and period over the North Atlantic, *J. Phys. Oceanogr.*, **23**, 1315–1329.

McCreary, J. P., and P. Lu, 1994: Interaction between the subtropical and equatorial ocean circulations: The subtropical cell, *J. Phys. Oceanogr.*, **24**, 466–497.

McShane, E. J., 1989: The calculus of variations from the beginning through to optimal control theory, *SIAM J. Control Optim.*, **27**, 916–939.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, 1953: Equations of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087–1092.

Miller, R. N., M. Ghil and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear systems, *Journal of the Atmospheric Sciences*, **51**, 1037–1056.

Montgomery, R. B., 1938: Circulation in upper layers of southern North Atlantic deduced with use of isentropic analysis, *Papers Phys. Oceanogr. Met.*, **6 (2)**.

Moyer, K. A., and R. A. Weller, 1995: Observations of surface forcing from the Subduction Experiment: A comparison with global model products and climatological datasets, *J. Climate*, **10**, 2725–2742.

Müller, T. J., and G. Siedler, 1992: Multi-year time series in the eastern North Atlantic Ocean, *J. Mar. Res.*, **50**, 63–98.

Munk, W., 2002: Twentieth century sea level: An enigma, *P. Natl. Acad. Sci. U.S.A.*, **99 (10)**, 6550–6555.

Munk, W., 2003: Ocean freshening, sea level rising, *Science*, **300 (5628)**, 2041–2043.

Nakamura, N., 1995: Modified Lagrangian-mean diagnostics of the stratospheric polar vortices. Part I. Formulation and analysis of GFDL,SKYHI GCM, *J. Atmos. Sci.*, **52**, 2096–2108.

New, A. L., Y. Jia, M. Coulibaly and J. Dengg, 2001: On the role of the Azores current in the ventilation of the North Atlantic Ocean, *Prog. Oceanogr.*, **48**, 163–194.

Nurser, A. J. G., and J. C. Marshall, 1991: On the relationship between subduction rates and diabatic forcing of the mixed layer, *J. Phys. Oceanogr.*, **21**, 1793–1802.

O'Brien, J. J., 1986: *Advanced Physical Oceanography Numerical Modelling*, p. 608, D. Reidel,

Dordrecht.

Oliger, J., and A. Sundström, 1978: Theoretical and practical aspects of some initial value boundary in fluid dynamics, *SIAM J. App. Math.*, **35**, 419–446.

Orlanski, I., 1976: A simple boundary condition for unbounded hyperbolic flows, *J. Comput. Phys.*, **21**, 251–269.

Pallant, J. S., F. B. Bahr, T. M. Joyce, J. P. Dean and J. R. Luyten, 1995: Subduction in the subtropical gyre: Seasoar cruises data report, Woods Hole Oceanographic Institution, *Tech. rep.*

Palmer, T. N., 1996: *Predictability of the Atmosphere and Oceans: From Days to Decades*, chapter 3, pp. 83–156, NATO ASI Series, Springer, in Decadal Climate Variability: Dynamics and Predictability.

Parrilla, G., A. Lavin, H. Bryden, M. Garcia and R. Millard, 1994: Rising termperatures in the subtropical North Atlantic over the past 35 years, *Nature*, **369**, 48–51.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics*, Springer, 710 pp.

Pedlosky, J., 1996: *Ocean Circulation Theory*, Springer, 464 pp.

Peixoto, J., and A. H. Oort, 1992: *Physics of Climate*, p. 520, AIP Press.

Pond, S., and G. L. Pickard, 1983: *Introductory Dynamical Oceanography*, p. 329, Pergamon Press.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992: *Numerical Recipes: The Art of Scientific Computing*, p. 963, 2nd ed. Cambridge University Press.

Price, J. F., 2001: Subduction, in *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, pp. 357–371, Academic Press.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation, *J. Climate*, **7**, 929–948.

Rhines, P. B., and W. R. Young, 1982: Homogenization of potential vorticity in planetary gyres, *J. Fluid Mech.*, **122**, 347–367.

Robbins, P. E., J. F. Price, W. B. Owens and W. J. Jenkins, 2000: The importance of lateral diffusion for the ventilation of the lower thermocline in the subtropical North Atlantic, *J. Phys. Oceanogr.*, **30**, 67–89.

Robinson, A. R., and D. Haidvogel, 1980: Dynamical forecast experiments with a barotropic open ocean model, *J. Phys. Oceanogr.*, **10**, 1909–1928.

Robinson, A. R., and H. Stommel, 2000: The ideal fluid thermocline, *J. Phys. Oceanogr.*, **30**, 67–89.

Roemmich, D., and C. Wunsch, 1984: Apparent change in the climatic state of the deep North Atlantic Ocean, *Nature*, **307**, 447–450.

Rudnick, D. L., and J. R. Luyten, 1996: Intensive surveys of the Azores Front. Part I: Tracers and dynamics, *J. Geophys. Res.*, **101 (C1)**, 923–939.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis, *Mon. Wea. Rev.*, **98**, 875–883.

Sausen, R., K. Barthel and K. Hasselmann, 1988: Coupled ocean-atmosphere models with flux corrections, *Clim. Dyn.*, **2**, 154–163.

Schröter, J., U. Seiler and M. Wenzel, 1993: Variational assimilation of GEOSAT data into an eddy-resolving model of the Gulf Stream Extension area, *J. Phys. Oceanogr.*, **23**, 925–953.

Schröter, J., and C. Wunsch, 1986: Solution of nonlinear finite difference ocean models by optimization methods with sensitivity and observational strategy analysis, *J. Phys. Oceanogr.*, **16**, 1855–1874.

Seiler, U., 1993: Estimation of open boundary conditions with the adjoint method, *J. Geophys. Res.*, **98**, 22855–22870.

Sheinbaum, J., and D. L. T. Anderson, 1990: Variational assimilation of XBT data. part I, *J. Phys. Oceanogr.*, **20**, 672–688.

Siedler, G., J. Church and J. Gould, 2001: *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, Academic Press, 750 pp.

Spall, M. A., 1990: Circulation in the Canary Basin: A model/data analysis, *J. Geophys. Res.*, **95**, 9611–9628.

Spall, M. A., 1995: Frontogenesis, subduction, and cross-front exchange at upper ocean fronts, *J. Geophys. Res.*, **100**, 2543–2557.

Spall, M. A., R. A. Weller and P. W. Furey, 2000: Modeling the three-dimensional upper ocean heat budget and subduction rate during the Subduction Experiment, *J. Phys. Oceanogr.*, **30**, 26151–26166.

Speer, K., and E. Tziperman, 1992: Rates of water mass formation in the North Atlantic Ocean, *J. Phys. Oceanogr.*, **22**, 93–104.

Stammer, D., 1997: Global characteristics of ocean variability estimated from regional

TOPEX/Poseidon altimeter measurements, *J. Phys. Oceanogr.*, **27**, 1743–1769.

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 (C9)**, 3118.

Stammer, D., C. Wunsch and R. Ponte, 2000: De-aliasing of global high frequency barotropic motions in altimeter observations, *Geophys. Res. Letters*, **27**, 1175–1178.

Stevens, D. P., 1991: The open boundary condition in the United Kingdom fine-resolution Antarctic model, *J. Phys. Oceanogr.*, **21**, 1494–1499.

Stommel, H., 1979: Determination of water mass properties of water pumped down from the Ekman layer to the geostrophic flow below, *Proc. Natl. Acad. Sci.*, **76**, 3051–3055.

Strang, G., 1996: *Introduction to Applied Mathematics*, Harcourt Brace Jovanovich, 414 pp.

Sundermeyer, M. A., and J. F. Price, 1998: Lateral mixing and the North Atlantic Tracer Release Experiment: Observations and numerical simulations of Lagrangian particles and a passive tracer, *J. Geophys. Res.*, **103 (C10)**, 21481–21497.

Tai, C. K., and J. Kuhn, 1995: Orbit and tide error reduction for the first 2 years of TOPEX/POSEIDON, *J. Geophys. Res.*, **100 (C12)**, 25353–25363.

Talagrand, O., 1997: Assimilation of observations: An introduction, *J. Met. Soc. Jpn.*, **75**, 191–209.

Tanguay, M., P. Bartello and P. Gauthier, 1995: Four-dimensional data assimilation with a wide range of scales, *Tellus*, **47A**, 974–997.

Tarantola, A., 1987: *Inverse Problem Theory. Methods for Data Fitting and Model Parameter Estimation*, Elsevier, 232 pp.

Tennekes, H., 1973: The logarithmic wind profile, *J. Atmos. Sci.*, **30**, 234–238.

Thacker, W. C., 1989: The role of the Hessian matrix in fitting models to measurements, *J. Geophys. Res.*, **94**, 6177–6196.

Thacker, W. C., and R. B. Long, 1988: Fitting dynamics to data, *J. Geophys. Res.*, **93**, 1227–1240.

Tziperman, E., and W. C. Thacker, 1989: An optimal-control/adjoint-equations approach to studying the oceanic general circulation, *J. Phys. Oceanogr.*, **19**, 1471–1485.

Van Trees, H. L., 1968: *Detection, Estimation, and Modulation Theory. Part I.*, J. Wiley, New York, 437 pp.

Walin, G., 1982: On the relation between sea-surface heat flow and thermal circulation in the ocean, *Tellus*, **34**, 187–195.

Welander, 2000: The ideal fluid thermocline, *J. Phys. Oceanogr.*, **30**, 67–89.

Weller, R. A., 2003: Subduction, in *Aha Huliko'a Proceedings of the University of Hawai'i*, edited by P. Muller, and D. Henderson, pp. 171–176.

Weller, R. A., P. W. Furey, M. A. Spall and R. E. Davis, 2004: The large-scale context for oceanic subduction in the Northeast Atlantic, *Deep-Sea Res.*, In press.

Williams, R. G., J. Marshall and G. Nurser, 1995: Does Stommel's mixed layer "demon" work?, *J. Phys. Oceanogr.*, **25**, 3089–3102.

WOCE Synthesis and Modelling Working Group, 1997: WOCE Analysis, Interpretation, Modelling and Synthesis (AIMS) Strategy, World Ocean Circulation Experiment, *Tech. rep.*

Woods, J. D., 1985: The physics of thermocline ventilation, in *Coupled Ocean-Atmosphere Models*, edited by J. C. J. Nihoul, pp. 543–590.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem*, p. 437, Cambridge University Press, New York.

Wunsch, C., 1997: The vertical partition of horizontal kinetic energy, *J. Phys. Oceanogr.*, **27**, 1770–1793.

Wunsch, C., and A. E. Gill, 1976: Observations of equatorially trapped waves in Pacific sea level variations, *Deep-Sea Res.*, **23**, 371–390.

Wunsch, C., and D. Stammer, 1998: Satellite altimetry, the marine geoid, and the oceanic general circulation, *Annu. Rev. Earth Planet Sci.*, **26**, 219–53.

Wüst, G., 1935: Schichtung und zirkulation des Atlantischen Ozeans. Die stratosphare., in *Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs- und Vermessungsschiff "Meteor" 1925-1927 (reprinted as "The Stratosphere of the Atlantic Ocean"*, edited by W. J. Emery, pp. 1–180.

Yu, L., and P. Malanotte-Rizzoli, 1996: Analysis of the North Atlantic climatologies using a combined OGCM/adjoint approach, *J. Mar. Res.*, **54**, 867–913.

Zang, X., and C. Wunsch, 2001: Spectral description of low-frequency oceanic variability, *J. Phys. Oceanogr.*, **31**, 3073–3095.

Zhang, K. Q., and J. Marotzke, 1999: The importance of open-boundary estimation for an Indian Ocean GCM-Data synthesis, *J. Mar. Res.*, **57**, 305–334.