Assimilation of Altimetry Data in a High-Resolution Limited-Area Primitive Equation Model of the Southern Ocean

by

Andrew Woolf, BSc (Hons I)

Submitted in fulfilment of the requirements

for the Degree of

Doctor of Philosophy

University of Tasmania

October, 2002

Declaration

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution. To the best of my knowledge and belief this thesis contains no material previously published or written by another person except where due acknowledgement is made in the text.

ANDREW WOOLF

Authority of access

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968.*

ANDREW WOOLF

University of Tasmania

October 2002

Abstract

An eddy-permitting (44 km at 50°S) numerical model of the open ocean south of Australia is combined with satellite altimetry data in an assimilation study of Southern Ocean dynamics. The domain chosen is from 110 to 190° East and 30 to 70° South, encompassing both the Southeast Indian and Macquarie Ridge complexes. This region is thought to play an important role in the momentum balance of the Antarctic Circumpolar Current (ACC) and meridional heat and freshwater exchanges.

The Hamburg Ocean Primitive Equation (HOPE) model (Wolff et al. 1997) is an implicit free-surface primitive-equation model discretised on an Arakawa E-grid. It is modified to operate in an open boundary configuration. For the barotropic mode, the open boundary condition is based on characteristic variables and extends the condition proposed by Flather (1976). The boundary condition preserves water volume while allowing the transmission of shallow-water gravity waves, an important feature with sequential data assimilation. Volume transports normal to the boundaries are required. Across the ACC, these are determined adaptively through a thermal wind calculation. On the northern boundary, prescribed transports model the Leeuwin Current and East Australia Current. For the baroclinic velocities, a modified Orlanski-type (Orlanski 1976) radiation condition is applied on the eastern boundary, while a zero-gradient condition is applied on the northern boundary. Because of the staggered grid in HOPE, the baroclinic velocities on the western boundary can be calculated prognostically as in the interior. For tracers a combination of relaxation towards climatology for inflow and upstream advection on outflow is applied along all the open boundaries. The model undergoes considerable drift in a forty-year spinup run. This results from a lack of sea-ice in the model, and

inadequate surface forcing and deep-water formation.

The reduced-order optimal interpolation System for Ocean Forecasting and Analysis (SOFA) of De Mey (1999) is integrated with HOPE to perform the assimilation. The order-reduction applies a transformation in the vertical between observed sea-level anomalies and the internal density structure, preserving both water-mass properties and potential vorticity on isopycnals (Cooper and Haines 1996). Such a model of water-column variability is consistent with that observed in repeat hydrographic sections of the WOCE SR3 line through the model domain. A dynamical analysis of the vertical projection scheme in an assimilation context shows that it excites primarily barotropic topographic Rossby waves.

One year of data from the TOPEX/POSEIDON satellite mission is used in a series of trial assimilation runs. The assimilation system produces reasonable analyses of sea-level anomaly, and improved estimates of meridional eddy heatflux. A comparison with a WOCE hydrographic section succesfully captures a significant eddy feature, but at a reduced level. A trial is made of a modification to the vertical projection scheme that includes a degree of barotropic variability. This modification provides small but significant improvements to statistics of forecast performance and patterns of sea-level variability.

Acknowledgements

I would first like to thank my supervisors Nathan Bindoff and Richard Coleman for their valuable advice, helpful suggestions and great patience. Jörg Wolff supplied a version of the Hamburg Ocean Primitive Equation (HOPE) model code and assistance with setting it up and unravelling its mysteries.

Three months spent at the Toulouse laboratory of the Centre National d'Études Spatiales with Pierre De Mey was crucial to successfully interfacing the System for Ocean Forecasting and Analysis (SOFA) with the HOPE model. I thank Pierre for his assistance and hospitality, and for making available the SOFA code.

An analysis of the application of a uniform vertical displacement model to WOCE hydrographic data was first performed by Nathan Bindoff (personal communication) and extended here in section 5.2.4 to include comparison against altimetric measurements.

The work in this thesis was financially supported primarily through a Commonwealth Australian Postgraduate Award and a grant from the National Greenhouse Advisory Committee. Some additional funding was provided by the Antarctic Cooperative Research Centre and CSIRO Division of Marine Research.

My period of candidature was extended, but also enhanced, by activity with the Tasmanian University Postgraduate Association and the Council of Australian Postgraduate Associations. I shall always value the friendships made and experience gained through these organisations.

I am very grateful for the good humour and friendship of fellow students at the Institute of Antarctic and Southern Ocean Studies, particularly those of 28 and 30 Earl St. I shall miss the Friday cake days.

Finally, and above all, I wish to thank my mother for her enduring and en-

dearing support throughout my long years of learning. Through all the difficult years she has never failed in her encouragement at every step. Her parents — my grandparents — have also provided more inspiration than they might have realised[1]. It is to her, and them, that I dedicate this thesis.

[1]Unfortunately, my grandfather, who always maintained an interest in the progress of this thesis, passed away just a few months before its completion.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is the attempt to understand the Earth's global climate system that ultimately motivates much current oceanographic research. The oceans play a pre-eminent role in this system. Of the solar radiation incident on the planet, the atmosphere absorbs between 19 and 25% (Wallace and Hobbs 1977; Peixoto and Oort 1992; Moran and Morgan 2002) and the hydrosphere around 38% (Sørenson 1979, Figure 97). The annual average internal heat content of the oceans is several orders of magnitude greater than the atmosphere[1], and the annual cycle about this average is five times greater ($\sim 8 \times 10^{22} J$) than for the atmosphere ($\sim 1.5 \times 10^{22} J$). Satellite radiometer observations of the annual variation in the global net radiation flux at the top of the atmosphere are almost entirely explained by the change in oceanic heat storage (Ellis et al. 1978). In its 1995 Second Assessment Report, the Intergovernmental Panel on Climate Change noted the crucial role of the oceans in climate change scenarios (IPCC 1995):

> Because of the thermal inertia of the oceans, only 50-90% of the eventual equilibrium temperature change *[under various greenhouse gas emission scenarios]* would have been realized by 2100 and temperature would continue to increase beyond 2100, even if concentrations of greenhouse gases were stabilized by that time.

Given the very large reservoir of heat contained in the oceans, it is no surprise that the storage and transport of heat by ocean currents gives rise to significant climate and weather effects.

Higher temperatures in equatorial regions with respect to the poles sets up a net poleward heat transport. This process may be regarded as a thermodynamic heat engine (Peixoto and Oort 1992), with the resulting work producing the weather and current dynamics which are such important features of the biosphere. Though difficult to measure directly, particularly in the oceans, estimates of the annual mean poleward heat transport is roughly the same for the oceans as the atmosphere (see figure 1.1). In the Northern Hemisphere, the oceans dominate at low latitudes, transporting almost 4 PW

[1] Peixoto and Oort (1992, Chapter 13) give the respective global average heat content per unit surface area as $4.52 \times 10^{12} Jm^{-2}$ for the oceans and $1.8 \times 10^{9} Jm^{-2}$ for the atmosphere. As well, the surface area of the world's oceans is around 70% of the planet.

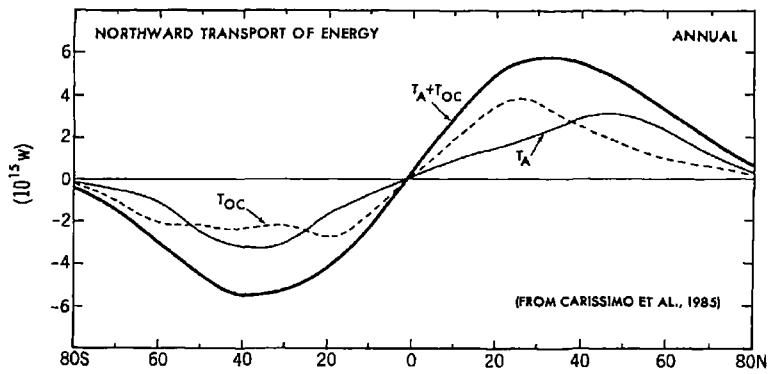


Figure 1.1: Meridional profiles of zonally-averaged northwards heat transport. The three curves show the atmospheric (T_A) and oceanic (T_{OC}) components, and the total *(from Peixoto and Oort (1992)).*

($4 \times 10^{15} W$) of heat northwards at 25°N — about double that transported by the atmosphere at the same latitude (Carissimo et al. 1985). The oceans also dominate in the Southern Hemisphere at latitudes of the Antarctic Circumpolar Current (ACC), around 60°S, where again the heat carried by the oceans is roughly double that carried by the atmosphere.

Broecker (1991) in his now famous account described the global scale thermohaline circulation of the ocean as a giant conveyor (see Figure 1.2, and also Schmitz (1995), Rossby (1996), Macdonald and Wunsch (1996), and Broecker (1997)). The Gulf Stream and its North Atlantic Current extension transports warm salty waters north to the Labrador, Greenland and Norwegian Seas. Along-path cooling and deep convection transforms these warm surface waters into the cold and dense North Atlantic Deep Water (NADW) which subsequently spreads south in a deep western boundary current before filling much of the deep South Atlantic basin. At high southern latitudes NADW is entrained into the ACC, where mixing produces Circumpolar Deep Water (CDW). Very cold and salty Antarctic Bottom Water (AABW) is formed along the Antarctic coast during wintertime sea-ice growth. AABW mixes with CDW in the circumpolar region and spreads north into each of the major ocean basins. Some of this deep water upwells in the south-eastern Pacific before turning westwards and entering the Indian Ocean via the Indonesian Throughflow (Schmitz 1995). Both throughflow waters and deep waters upwelled in the Indian Ocean return to the Atlantic around the southern tip of Africa — the so-called "warm-water path" for return flow into the North Atlantic. This closes the meridional overturning circulation through renewal of NADW. A second renewal path (the "cold-water path") is via Drake Passage.

This present-day thermohaline circulation has significant influence on local climatic conditions. Without the warm Gulf Stream, for instance, the climate of most of western Europe would be substantially cooler (Broecker 1991; Rahmstorf 1997; Broecker 1997; Bunyard 1999). Numerous studies (Marotzke and Willebrand 1991; Weaver and Sarachik 1991; Power and Kleeman 1993; Maier-Reimer et al. 1993), however, have indicated the likelihood of multiple stable thermohaline circulation patterns. Stommel (1961) found two distinct

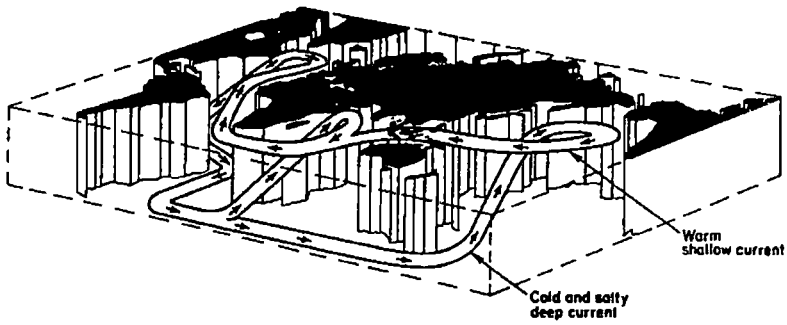


Figure 1.2: The great ocean conveyor of Broecker (1991) *(graphic from Wolff et al. (1991))*

solutions even under the same surface boundary conditions. It is likely that some decadal to centennial climatic variability is due to changes in the thermohaline circulation of the oceans.

1.1 The Southern Ocean and the global thermohaline circulation

The Southern Ocean plays an important role in the global thermohaline circulation. The ACC is the only current joining all three major basins. Both the cold- and warm-water routes for renewal of NADW rely on the Southern Ocean for their existence. In addition, deep and bottom waters formed in the circumpolar region are crucial components of the global conveyor. Hoffert (1990) estimates that an increase in global mean temperature of four degrees Celsius could shut down the seasonal sea-ice pump around Antarctica, and therefore AABW formation and the global thermohaline circulation. Less dramatically, AABW and Antarctic Intermediate Water (AAIW) together affect, through mixing and circulation, over 50% of the world's ocean volume (Worthington 1981; England 1993; Marsh et al. 2000)); both of these water masses are formed in the Southern Ocean.

1.2 Eddy heat flux

In addition to the large-scale quasi-stationary thermohaline circulation, the dynamics of the Southern Ocean play a very significant role in the climate system. Nowlin and Klinck (1986) (from the work of de Szoeke and Levine (1981)) noted that the mean geostrophic heat flux across a circumpolar path is zero, whereas that associated with eddy processes is likely to dominate the overall poleward heat budget. This budget consists of a balance between the ocean-to-atmosphere heat exchange south of the ACC (Q), and the poleward oceanic fluxes associated with Ekman transport (F_{Ekman}), mean geostrophic flow (F_{geost}), eddy processes (F_{eddy}), and deep boundary currents (F_{bc}):

$$Q = F_{Ekman} + F_{geost} + F_{eddy} + F_{bc}. \tag{1.1}$$

For the left-hand-side of equation (1.1), de Szoeke and Levine (1981) noted an estimate by Gordon of 0.3 PW. This was a revision of an earlier published estimate (Gordon and Taylor 1975) of 0.4 PW. A later publication (Gordon and Owens 1987) maintained the figure of 0.3 PW. The oceanic heat loss at high latitudes should exactly balance the net heat gain of lower latitude waters. Hastenrath (1980) had calculated the latter as 0.35 PW at 53°S, the mean latitude of the Polar Front. Alternatively, Trenberth (1979) had inferred a poleward oceanic heat flux of 1 PW in the 50–60°S latitude band from atmospheric data. Nowlin and Klinck thus used Gordon's estimate for Q of 0.3 PW, with large uncertainty (±30–50%).

For the right-hand-side of equation (1.1), de Szoeke and Levine (1981) had estimated the total (baroclinic plus barotropic) mean geostrophic heat flux (F_{geost}) to be 0 ± 0.33 PW. This was calculated along a path that closely followed the mean position of the Polar Front. For the Ekman component (F_{Ekman}), they had estimated -0.15 PW (that is, directed equatorwards) with a large, but uncalculated, error. For the total balance, therefore, Nowlin and Klinck (1986) estimated that together eddy processes and deep boundary currents, $F_{eddy} + F_{bc}$, must account for a mean poleward heat flux of 0.45 PW. de Szoeke and Levine's equivalent figure included an error estimate of perhaps ±0.3 PW. As an extreme estimate of the contribution by deep boundary currents, they calculated 0.15 PW for 20 Sv of bottom-water production with an average temperature of -0.5°C. Despite the considerable uncertainties involved, this leaves eddy processes as the largest component (between 0.3 and 0.45 PW) in the above heat budget.

1.2.1 In-situ observations

To date there have been only a few in-situ observational studies of eddy heat fluxes across the ACC: in Drake Passage (Bryden 1979; Sciremammano 1980; Nowlin et al. 1985), southeast of New Zealand (Bryden and Heath 1985) and, recently, south of Australia (Phillips and Rintoul 2000; Phillips 2000).

Bryden (1979) analysed data from an array of six moorings spread across Drake Passage, collected in 1975 and 1976 during the International Southern Ocean Studies (ISOS) program of the International Decade for Ocean Exploration (IDOE). From the measurements, he found the eddy heat flux at 2700 m to be 6.7 kW/m^2 directed poleward, and statistically significant. If this value was extended throughout the average water depth (4000 m), and around the circumpolar length of the ACC (20 000 km), it would lead to an average poleward heat flux of 0.5 PW. Sciremammano (1980) extended Bryden's time-series by two years, including additional moorings in the central part of Drake Passage, and found results largely consistent with Bryden's for the common moorings. The additional central moorings, however, recorded highly variable heat fluxes. The range over three years for depths greater than 1000 m was 9–28 kW/m^2 with a mean and standard deviation of (17 ± 7) kW/m^2. These values were significantly higher than the deep measurements of Bryden. Bryden performed a spectral analysis to find the fluxes were due to motions with timescales longer than ten days. Sciremammano found most of the heat flux

was due to discrete events with timescales of 5–60 days, occurring several times per year. A third study in Drake Passage was published by Nowlin, Worley and Whitworth (1985). They analysed data collected from an array of moorings in place during 1979. They noted several potential problems with previous analyses: mooring "blowover" could lead to heat flux over-estimates as great as 20%, long-period "non-eddy" signals could contaminate the estimates, and the local analysis direction should be "cross-frontal" rather than "poleward" due to the changing direction of the front. Applying solutions to these problems, they calculated an average across-stream eddy heat flux for all instruments of 3.7 kW/m^2 poleward, with a range of 1 to 17 kW/m^2. Extending this along the entire ACC yields a total heat flux of 0.3 PW.

The measurements taken southeast of New Zealand (Bryden and Heath 1985) were, in part, motivated by a concern that the relatively high meridional temperature gradients in Drake Passage might lead to greater than average eddy heat fluxes. They obtained variable results, as high as 35 kW/m^2, but none significantly different from zero, statistically. Due to the long timescales involved, they concluded longer records would be needed.

Very recently, Phillips and Rintoul (2000) have analysed mooring data collected south of Australia during 1993 and 1994. They corrected for tides and mooring motion, and used a local cross-frontal co-ordinate system to analyse eddy heat fluxes. They found statistically significant depth-averaged across-stream fluxes of 11 kW/m^2 poleward for variability with timescales between two and 90 days. This is three times larger than the bandpassed Drake Passage result of 3.7 kW/m^2 (Nowlin et al. 1985). For the all-frequency variability, they obtained an across-stream eddy flux of 41 kW/m^2, which again is much larger than either of the non-filtered Drake Passage results (6.7 kW/m^2 and 17 kW/m^2) or the results southeast of New Zealand.

In all of these results, observed eddy heat flux has been directed poleward, and has been sufficient to balance the ocean-to-atmosphere heat loss south of the ACC, if extended circumpolarly. Wunsch (1999) concluded from a range of data that the eddy heat flux in the Southern Ocean was moderate compared to the mid-latitude western boundaries, but, due to the ACC's zonal extent, probably dominated the Southern Ocean heat balance.

As will be seen below, the meridional eddy heat flux is intimately connected with the other great question of Southern Ocean dynamics — that of the momentum balance of the ACC. Essentially, this assumes that the ACC is in a quasi-steady state so that there exists a dynamical balance between its momentum sources and sinks. The most significant source of momentum is the eastwards wind which circles the globe at these latitudes. The question of the momentum sinks, on the other hand, is less clear.

1.3 Momentum balance of the ACC

Nowlin and Klinck (1986) provide, amongst other things, a detailed history of theories and models of the ACC. The summary here contains only the key developments.

By 1950, the theory of the wind-driven circulation of the ocean was largely known, through the work of Sverdrup (1947), Munk (1950) and Stommel (1948). In essence the theory provides that a net meridional transport is induced to preserve absolute vorticity due to vortex-tube shrinking or stretching associated with Ekman pumping:

$$\rho V = \frac{(\nabla \times \boldsymbol{\tau}).\boldsymbol{k}}{\beta}, \tag{1.2}$$

where β is the meridional gradient of Coriolis parameter, $\boldsymbol{\tau}$ is applied windstress, and the mass transport ρV is called the *Sverdrup transport.* Sverdrup dynamics apply in a closed basin, with continuity requiring a compensating non-Sverdrup flow. This occurs as a narrow western boundary current where friction dominates.

Two early attempts to model the ACC regarded the current as purely zonal with no blocking topography — as is the case at latitudes of the Drake Passage. Thus, Sverdrup dynamics were not held to apply.

1.3.1 Hidaka's dilemma

Hidaka and Tsuchiya (1953) used a steady-state geostrophic channel model with both horizontal and vertical viscous mixing, forced by a constant surface windstress. For reasonable values of applied windstress and horizontal viscosity, they obtained a mass transport almost two orders of magnitude too large. The value of the vertical viscosity did not much affect the calculation. To obtain reasonable transports, a horizontal mixing parameter unrealistically large was required. These results have come to be called "Hidaka's dilemma" (Wolff et al. 1991). In Hidaka and Tsuchiya's results, horizontal velocity vanished at the bottom.

1.3.2 Topographic form drag or Sverdrup dynamics?

An earlier model proposed by Munk and Palmén (1951), explicitly relied on the bottom velocity being non-zero to limit the transport to reasonable values. Their model, like that of Hidaka and Tsuchiya, assumed a balance between surface windstress and lateral friction, but was applied to the vertically integrated flow. Like Hidaka and Tsuchiya after them, Munk and Palmén obtained a mass transport one hundred times too large, for reasonable values of horizontal viscosity. They offered two solutions to the discrepancy. The first was the same as the solution of Hidaka and Tsuchiya — to assume an unreasonably large value of lateral viscosity. They preferred a second solution which has come to be called *topographic form drag.* In net terms, this mechanism proposes that the circumpolar eastwards surface windstress, rather than being balanced by lateral friction, instead is exactly balanced by the zonal integral of zonal pressure gradients across topographic features along the path of the ACC. This happens if, on the average, the bottom pressure on the up-current side of a topographic obstacle exceeds the pressure on the lee-side, and corresponds to a transfer of momentum to the solid earth (see Figure 1.3). They

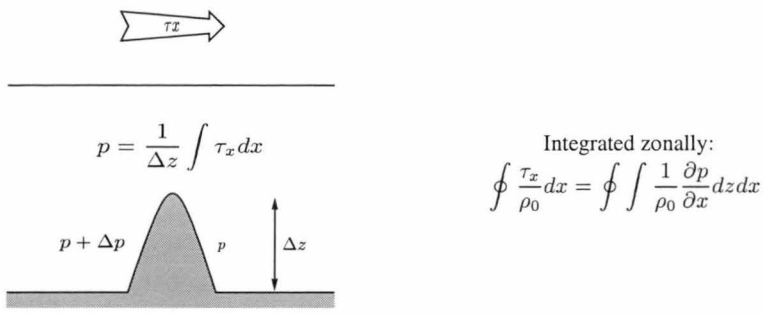


Figure 1.3: Topographic form drag proposes a balance between windstress and pressure gradient over topography *(after Johnson and Bryden (1989)).*

calculated that an average pressure difference equivalent to four centimetres of dynamic height across each of the four major ridges encountered by the ACC (*i.e.* the Kerguelen, South Pacific and Macquarie ridges, and the South Antillean arc) is sufficient to balance the observed surface windstress. The authors did not offer an estimate of the strength of the ACC based on their theory.

Unlike these two earlier studies, Stommel (1957), and later Baker (1982), Godfrey (1989) and Chelton et al. (1990a), proposed that canonical Sverdrup dynamics could be used to estimate the mass transport of the ACC. They noted that, rather than progressing directly eastwards along its zonal path, the ACC instead veers south by around 10° of latitude in its course from the western South Atlantic to the eastern South Pacific. They noted that this is entirely consistent with a Sverdrup balance with the prevailing negative windstress curl. The required frictional western boundary current was held to be the strong northwards flowing Falkland Current, off the eastern coast of South America. Integrating the Sverdrup relation (equation (1.2)) along a zonal path excluding the Drake Passage-Falkland Current zone of return flow, transports very close to that observed through Drake Passage are obtained. For instance, Baker (1982) obtained transports of 173 and (190 ± 60) Sv for two different sets of wind data, along latitude 55°S. Godfrey (1989), along 54°S, obtained 128 Sv, and Chelton et al. (1990a) obtained 114 Sv along 55°S. For comparison, current meter records in Drake Passage (Whitworth et al. 1982) recorded an average transport of 130 Sv.

Recently, vigorous debate has occurred (Warren et al. 1996; Hughes 1997; Warren et al. 1997; Olbers 1998; Warren et al. 1998) on whether the form drag mechanism of Munk and Palmén or a Sverdrup balance is the more correct description both of the strength of the ACC and its overall momentum balance.

As noted by Munk and Palmén (1951), topographic form drag requires that the stress input at the ocean surface by the wind somehow be manifested at the ocean floor in order that it be balanced by pressure differences across submarine ridges. They offered two suggestions as to how this might occur. A net northwards transport in the surface layers (to around 1000 m depth, say) would export absolute angular momentum (due to planetary rotation) away from the circumpolar region, exactly balancing the angular momentum input by the eastwards wind. (A northwards flow roughly equivalent to that observed, due primarily to the Benguela and Peru Currents, could effect such

a balance.) Mass conservation then demands an equivalent volume of flow southwards at depth. For absolute angular momentum to be conserved for this southwards flow, there must also be eastwards flow at depth (increasing with latitude). It is the topographic form drag acting against this eastwards flow that completes the overall stress balance. Properly seen, and in the context of the form drag *vs* Sverdrup balance debate, this seems rather to be a statement of planetary-scale absolute angular momentum conservation than a constraint on the size of the ACC. The absolute angular momentum of the Southern Ocean associated with planetary rotation is at least two orders of magnitude greater than that associated with ACC transport (see, for example Ponte and Rosen (1994) and references therein). Indeed, proponents of a Sverdrup balance claim (Warren et al. 1996) that form drag "is really just a statement that northward Ekman transport in the circumpolar Drake Passage zone is compensated by deep southward geostrophic flow" and therefore "is actually irrelevant to the magnitude of the *[Antarctic Circumpolar Current]* itself".

Those who favour form drag have developed further the second suggestion offered by Munk and Palmén for the mechanism by which a surface windstress might be manifested at the seafloor. This mechanism is akin to topographic form drag but acts on isopycnal surfaces down through the water column rather than solid topography. As proposed by Munk and Palmén, "[e]ach layer induces, by turbulent interchange, motion in the layer beneath, and in this manner the wind stress is transmitted to the sea bottom". Johnson and Bryden (1989) developed the theory in more detail and pointed out the equivalence between a downwards flux of eastward momentum and a poleward eddy heat flux. The time-mean zonal integral of this *interfacial form stress* at some depth is given by:

$$\oint \overline{p' \frac{\partial \zeta'}{\partial x}} dx,$$

where p' is the pressure perturbation at an isopycnal height perturbation ζ', and the contour integral is performed around a circumpolar path. Integrating by parts, this is equivalent to:

$$-\oint \overline{\zeta' \frac{\partial p'}{\partial x}} dx.$$

Assuming geostrophy, $(1/\rho_0)\partial p'/\partial x = fv'$, and small isopycnal displacements so that $\zeta' = -\theta'/\overline{\theta}_z$, the above equation can be rewritten:

$$-\oint \overline{\zeta' \frac{\partial p'}{\partial x}} dx = \oint \rho_0 f \frac{\overline{v'\theta'}}{\overline{\theta}_z} dx$$

Thus, a downwards flux of eastwards momentum (the interfacial form stress) is equivalent to a poleward eddy heat flux.

Johnson and Bryden (1989) assumed that the surface windstress was transmitted in this manner undiminished to the seafloor (to be balanced there by topographic form stress):

$$\rho_0 f \frac{\overline{v'\theta'}}{\overline{\theta}_z} = \tau^x$$

They tested the implied balance between meridional eddy heat flux and surface windstress using current meter data collected in Drake Passage during the ISOS program. The measured eddy heat fluxes were larger than the mean windstress, though generally within the errors. Spatial variations of eddy heat flux along the ACC, and divergence of Reynolds stresses, $\frac{\partial \overline{<u'v'>}}{\partial y}$, were suggested as possible contributors to the discrepancy.

Compared with other regions of the world oceans, there is a significant paucity of hydrographic data for the Southern Ocean (Levitus and Boyer 1994). This is due to the logistic difficulty and expense of conducting dedicated marine science cruises in this region, and the lack of merchant shipping routes providing platforms from which to collect upper-ocean measurements. As noted earlier, in-situ timeseries measurements have been limited to a few studies in only three locations. Consequent to this lack of in-situ data, satellite measurements and numerical modelling have become pivotal tools in studying Southern Ocean dynamics.

1.4 Satellite altimetry

The development of satellite remote-sensing technology has provided valuable new sources of data for investigating the oceans. Ocean colour sensors give biologists information on chlorophyll concentration and biological activity; infrared radiometers provide measurements of sea-surface temperature; microwave instruments measure surface winds (scatterometers) and albedo (passive microwave radiometers). Satellite altimeters provide regular accurate measurements of sea-level on a global basis.

Unlike sea-surface temperature measurements, altimetric measurements of sea-level are not compromised by the presence of atmospheric cloud cover. In addition, the altimeter is unique in providing an integrated picture of the ocean — sea-level responds to both barotropic and baroclinic dynamics. Satellites are limited to observing surface properties of the ocean; sea-level alone contains the signature of sub-surface processes. In addition, the gradients of sea-level relative to the earth's gravitational equipotential (the *geoid*) provides directly the surface geostrophic flow field[2].

The books by Stewart (1985), Jones (1993), Robinson (1994), and Fu and Cazenave (2000) provide broad introductions (and extensive reference lists) to the use of satellite remote-sensing, including altimetry, in oceanography. The volume edited by Rummel and Sansò (1993) concentrates on altimetry. Fu and Cheney (1995) reviewed the application over one decade of satellite altimetry to circulation studies (including the western boundary currents and tropical oceans, as well as the general circulation) and ocean variability on both the large-scale and meso-scale. Wunsch and Stammer (1998) recently have reviewed the technology and its use in determining the ocean circulation, both stationary and time-variable. They make particular reference to the TOPEX/POSEIDON mission.

[2]Surface Ekman flow, which directly balances windstress via the Coriolis force in steady-state (Gill 1982, §9.2), has no sea-level signature except through Ekman pumping.

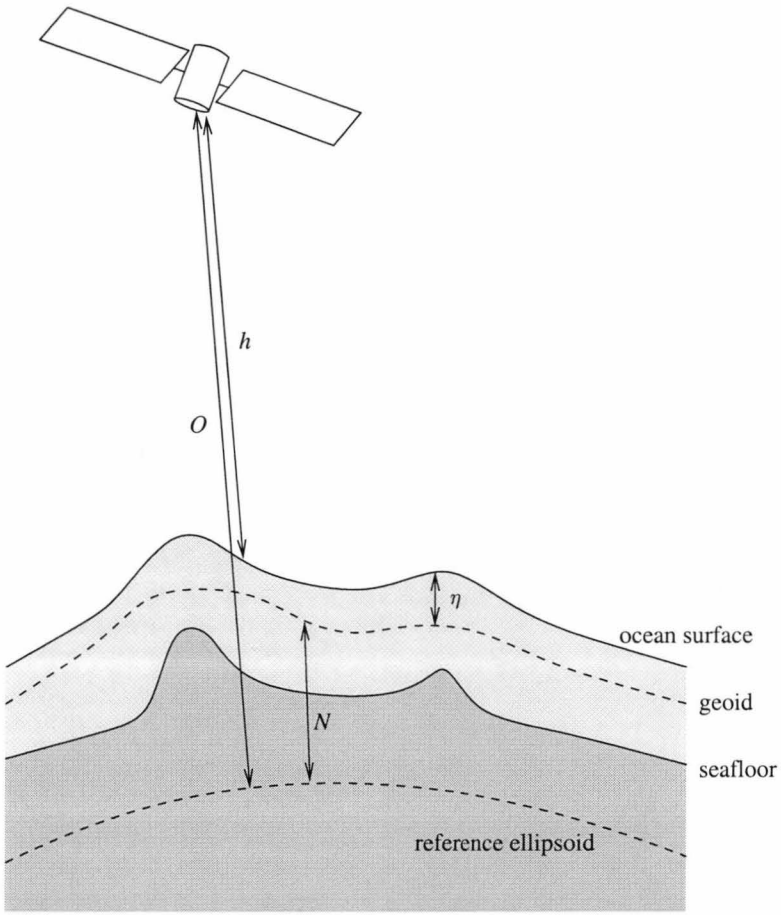


Figure 1.4: Schematic of satellite altimeter geometry. The altimeter measures the distance h to the sea-surface. From knowledge of the satellite's orbit, O, and the geoid, N, the dynamic topography η may be determined.

In principle, the altimeter's operation is straightforward (Figure 1.4): timed radar pulses are used to measure the distance, h, between the satellite platform and the instantaneous sea surface. With knowledge both of the satellite's orbit relative to the earth's reference ellipsoid, O, and the geoid, N, the sea-surface height (SSH) relative to the geoid is then known:

$$\eta = O - N - h$$

To be useful, the SSH should be determined to $\mathcal{O}$(cm) accuracy, placing severe requirements on geoid knowledge, the altimetric radar, and knowledge of the satellite's orbit.

As mentioned above, the geoid is the earth's gravitational equipotential. It is that surface that would be made by a uniform-density ocean at rest and in the absence of tides, currents, winds and atmospheric pressure variations. The most accurate geoid compiled to date is the Earth Geopotential Model 96 (EGM96), with an RMS accuracy of better than 12 cm over water at spatial scales larger than a few thousand kilometres (Lemoine et al. 1997).

To achieve sufficient altimeter accuracy, several corrections must be made to the measured SSH. Delays in the altimeter signal are caused both by tropo-

spheric water vapour and ionospheric free-electron content, both of which vary in space and time and must be corrected for. The radar pulse transmitted from the altimeter has a "footprint" diameter of several kilometres. Over this area, the reflected signal is affected by the sea-state: maximum energy in the return pulse generally does not correspond to reflection from the mean sea-surface. These so-called electromagnetic bias effects contribute significant errors that are difficult to correct. Local atmospheric pressure loading on the ocean produces changes in sea-level of the order of one cm per millibar. This "inverse barometer" effect must be compensated accordingly. Finally, the ocean's tides are generally ignored for oceanographic problems, so must be modelled and removed from the measured SSH signal.

Two of the earliest satellite altimeters were GEOS-3 and SEASAT, with combined altimeter and orbit accuracies of $\mathcal{O}$(m). These missions greatly improved knowledge of the marine geoid, but were not useful for oceanographic purposes (Wunsch and Stammer 1998). The US Navy launched Geosat in 1985 in a classified mission to improve the geoid for military purposes. Two and a half years of the mission, however, was run in an unclassified mode. Initial orbit errors were around 2 m, but this was reduced to around 10 cm following release by the US Navy in 1993 of additional tracking data (Fu and Cheney 1995). The total Geosat error budget was reduced to around 15 cm. The European Space Agency in 1991 launched ERS-1, the first satellite altimeter dedicated to marine science. Though failure of an onboard PRARE tracking system initially hampered the mission, laser tracking of the satellite later enabled orbit errors to be reduced to around 15 cm (Fu and Cheney 1995; Scharoo and Visser 1998). An identical satellite, ERS-2, was launched in 1995. The joint French-US mission, TOPEX/POSEIDON, was launched in August 1992, and has achieved a remarkable orbit accuracy of around 2 cm.

1.4.1 Altimetry and the Southern Ocean

See the reviews by Fu and Cheney (1995) and Wunsch and Stammer (1998), and references therein, for specific ocean studies utilising altimeter measurements. The following is a selection of work relevant to Southern Ocean and ACC dynamics.

Chelton et al. (1990b) used 26 months of unclassified Geosat data to examine variability of SSH and surface currents in the Southern Ocean. Both mean flow and variability were found to be strongly controlled by bathymetry. By calculating empirical orthogonal functions (EOFs), large-scale structure of the variability was investigated. They found only 33% of variance explained by the first three modes, implying a generally local rather than circumpolar pattern to variability.

Morrow et al. (1992) also used Geosat data to calculate components of surface velocity variance at satellite crossover points in the Southern Ocean, averaged over $3° \times 3°$ bins. They found a largely isotropic eddy momentum flux, except in regions where the mean flow interacts with topography, for example over the Macquarie Ridge complex. A higher resolution analysis southeast of Australia showed that the principal axes of velocity variance were strongly

steered by topography. A calculation of surface Reynolds stresses $\overline{u'v'}$ showed several regions of convergence around the circumpolar band — regions where eddies were acting to accelerate the mean flow. The zonally averaged eddy momentum flux was divergent and roughly 5 cm^2/s, far smaller than the 100 cm^2/s required to balance the eastwards wind stress by laterally divergent momentum fluxes alone. These results have been confirmed by other authors (Johnson et al. 1992; Morrow et al. 1994; Wilkin and Morrow 1994; Gille 1995; Gille 1997).

Stammer (1998) used TOPEX/POSEIDON altimeter data to estimate global kinetic energy, and eddy variability length and time scales. Using a baroclinic instability model for the variability, the estimates were used to estimate meridional eddy fluxes of salt and heat. For the Southern Ocean, a zonally integrated poleward eddy heat flux of around 0.28 PW at 40°S was found.

Gille and Kelly (1996) used Geosat data to estimate length and time scales of Southern Ocean sea-level variability, and to investigate large-scale coherence of these fluctuations. They found spatial decorrelation scales of around 85 km and timescales of 34 days. Empirical orthogonal functions calculated for the time-varying surface transport provided no evidence for coherent global-scale variability that might be expected from large-scale changes in wind forcing. By contrast, Mestas-Nuñez et al. (1992) found a significant correlation between sea-level observed by altimeter and that calculated from satellite scatterometer winds assuming a barotropic Sverdrup model for the South Pacific.

Jacobs et al. (1993) used two years of Geosat altimetry in the Pacific Ocean to extract evidence of Rossby waves. They fitted the Geosat data to the known dispersion relation for both barotropic and first-order baroclinic Rossby waves and found statistically significant results throughout the basin. Elevated levels of Rossby wave variability were detected in the ACC, as well as the Kuroshio and East Australian Currents.

Hughes (1995) examined TOPEX/POSEIDON data for evidence of Rossby waves in the Southern Ocean. He noted that uncertainties in the mean flow hindered the use of dispersion relations to identify Rossby waves. Instead he used a complex principal components technique capable of analysing travelling disturbances. Extracting wave speeds, periods and wavelengths, he obtained results consistent with the existence of Rossby waves in parts of the Southern Ocean, with wave structure able to be resolved in the southeast Pacific.

1.5 Computational modelling

A second widely used tool for studying ACC dynamics is numerical modelling. With increasing computer power and more sophisticated techniques, computational results are becoming more realistic and are being used with growing confidence to verify physical models. Clearly, any problem that can be formulated in mathematical terms is a candidate for being solved with the help of computers. Applications in geophysical fluid dynamics have consistently been

among the foremost users of the latest high-end computing technology[3]. In oceanography, a numerical model can act as proxy for the real ocean which is impossible to sample as easily in time and space. Ocean models can provide insight into oceanic processes on various scales, or play an important role (together with atmospheric circulation models) in coupled models of the earth's climate. Combined with real observations, ocean models may aid in estimating fluxes that are difficult to measure, or enable predictions of synoptic circulation patterns.

Semtner (1995) reviews the history of computational modelling of ocean circulation. The volume edited by O'Brien (1986a) provides extensive coverage of the state of the art in 1985. The monographs by Kowalik and Murty (1993) and Kantha and Clayson (2000) are excellent references on ocean models and the numerical techniques used. Marchuk and Sarkisyan (1988) provide a more mathematical review, with extensive references to the Russian literature. A recent review by Griffies et al. (2000) provides an excellent summary of the major ocean modelling trends and most models in widespread use.

Some of the earliest ocean modelling efforts were those undertaken at the Geophysical Fluid Dynamics Laboratory (GFDL) in the United States by Kirk Bryan and Michael Cox in the 1960s (Bryan 1969). These models discretised the equations of motion and tracer (heat and salt) conservation on a finite-difference grid with tracer gridpoints offset from velocity points. Both the hydrostatic and Boussinesq approximations were made. The finite-difference advection formulations conserved momentum and second-order moments of velocity and tracers. The time differencing for the momentum equations was an explicit leap-frog scheme (with an occasional Euler step) for the pressure, nonlinear and viscous terms, and implicit for the Coriolis term. This allowed a timestep longer than the inertial period. A centred-in-space explicit scheme was used for tracer advection and diffusion. For the external mode, a rigid-lid assumption was used, leading to a prognostic Poisson equation for transport streamfunction being solved each timestep. The vertical velocity was diagnosed from the continuity equation. This early model has become a mainstay of ocean modelling with derivatives (the "GFDL model" (Cox 1984) and the "Modular Ocean Model", MOM (Pacanowski et al. 1991)) still being widely used today. One significant modification was the substitution of a free-surface formulation in place of the rigid-lid (Killworth et al. 1991; Dukowicz and Smith 1994).

The Bryan-Cox model is an example of a *primitive-equation (PE) model*, so called because it is a finite-difference implementation of the *primitive equations*. (These are the basic equations of motion and tracer conservation under the hydrostatic assumption, and are an approximation to the fundamental Navier-Stokes equations of fluid flow. See section 2.1.1.) These models incorporate sub-grid scale effects via diffusive parametrizations. Typically, a

[3]Since June 1993, each of the bi-annual lists of the top 500 supercomputers (Meuer et al. 2001) maintained at the internet address `http://www.top500.org` has ranked *Weather* among the top eight application areas, by number of installations; beating such other areas as defence, electronics and energy. In addition, the broad field of *Geophysics* has rated similarly highly. Ranked by total maximal computing power, *Weather* has been the top application area since November 1996.

very large amount of computer memory is required in order to model basin- or global-scale circulations at eddy resolution. The eddy parametrizations of coarser resolution primitive equation models fail to accurately reproduce the effects of eddies: currents are too diffuse, and meridional fluxes are incorrect. In order to better resolve these eddy processes with limited-memory computers, *quasi-geostrophic (QG) models* were developed - notably by Bill Holland (Holland 1978; Holland 1986). Pedlosky (1964; 1987) develops the QG theory in detail. The QG model is formulated by expanding the primitive equations in powers of the Rossby number. The zeroth-order terms provide the geostrophic approximation. Retaining in addition the first-order terms leads to an equation for conservation of potential vorticity. The model can be extended to a three-dimensional stratified fluid, with a series of layers in the vertical. Conservation of potential vorticity in each layer provides a prognostic equation for the quasi-geostrophic streamfunction. The QG model filters out gravity waves, and can be used only if assumptions of small Rossby number, small bottom slope and small isopycnal slopes are satisfied.

Most ocean general circulation models today are either of the primitive-equation or quasi-geostrophic varieties. However, shallow-water models are often used for modelling tides and storm surges where barotropic dynamics are most important (Kantha and Clayson 2000, Chapters 6 and 7). Other major differences relate to co-ordinate systems, finite-difference schemes, parametrizations, and surface forcing.

1.5.1 Co-ordinate systems

Numerous co-ordinate systems, in both the horizontal and vertical, have been used in ocean models. The so-called z-level models use layers of fixed thickness. Unless special provision is made, these models represent topography only at the resolution of the layer depths. They also typically represent diffusion separably in the horizontal and vertical. *Isopycnal models* resolve the vertical dimension along isopycnal surfaces (see for example the *Ocean isoPYCnal model* (OPYC) (Oberhuber 1993)). Such models better represent diffusion, which occurs preferentially along density surfaces (Redi 1982; Gent and McWilliams 1990). *σ-coordinate models* resolve the vertical in a terrain-following co-ordinate (*e.g.* the Princeton Ocean Model (POM) (Blumberg and Mellor 1987) and the Spectral Primitive Equation Model (SPEM) (Haidvogel et al. 1991)) and have particular advantage for regions of steep topography. In the horizontal, the earliest models used Cartesian co-ordinates based on latitude and longitude. Further development led to curvilinear co-ordinates which could be aligned to follow local coastlines, for instance. Conformal mappings allowed the construction of Cartesian grids which place the poles over land, thus avoiding singularities associated with convergence of meridians in the Arctic Ocean. These grids may also be chosen to provide enhanced resolution over regions of interest within the model domain. Murray (1996) discusses horizontal co-ordinate systems.

The numerical schemes used to implement various processes in ocean models are not unique to these models. Many of them are straightforward applications of known results in numerical methods.

1.5.2 Finite-difference schemes

Explicit (*forward Euler*) time-differencing, for instance, is known to be unstable (Press et al. 1992, §16.6) under certain circumstances. The well-known Courant-Friedrichs-Lewy (CFL) stability criterion provides an upper-limit to the allowable timestep for an advective or wave process. In free-surface ocean models, which resolve surface gravity waves with phase velocity $(gH)^{1/2}$ (where H is water depth), timesteps for the external mode must be an order of magnitude smaller than for the internal modes. Alternatively, an implicit formulation of the external mode sacrifices accurate representation of surface gravity waves for stability at longer timesteps. Implicit differencing schemes, however, demand considerable additional complexity in their implementation. Either an iterative solution must be calculated, or a very large sparse matrix must be solved. Predictor-corrector schemes (Press et al. 1992) provide a compromise, first using an explicit (predictor) step to calculate new spatial derivatives, which are then substituted in the right-hand-side for a second (corrector) step. Consider the following partial differential equation for a variable φ for example:

$$\frac{\partial \varphi}{\partial t} = \mathcal{F}(\varphi_x),$$

where $\mathcal{F}$ is some operator. Then the n-th predictor-corrector time step may be written as follows:

$$\begin{aligned} \varphi' &= \varphi^n + \Delta t\, \mathcal{F}(\varphi_x^n) \\ \varphi^{n+1} &= \varphi^n + \Delta t\, \mathcal{F}(\varphi_x'), \end{aligned}$$

where Δt is the timestep. (In fact, predictor-corrector schemes are explicit and therefore prone to instability, but are more accurate than forward Euler schemes (Press et al. 1992, §16.7)). Time-differencing schemes are discussed in more detail by Mesinger and Arakawa (1976) and O'Brien (1986b).

Like time-differencing, space-differencing may be implemented in many different ways. Five different spatial arrangements of grid variables were analysed by Arakawa and Lamb (1977) for their error properties under geostrophic adjustment. The vast majority of ocean models today use one or other of these five grids, identified universally according to Arakawa's labelling ("A" through "E"). For example, the GFDL series of models use the B-grid, while the Princeton Ocean Model (POM) (Blumberg and Mellor 1987) uses the C-grid. Arakawa also developed momentum advection formulations that preserve both kinetic energy and enstrophy (Arakawa 1966). The requirement to accurately represent advective processes has led to a variety of schemes being developed for tracer advection. Whilst only first-order accurate, the physically-motivated *upwind* differencing scheme, for instance, is known to perform better than centred-in-space differencing for advective processes.

1.5.3 Sub-grid-scale parametrizations

One of the largest areas of active research in ocean modelling is the attempt to better parametrize unresolved processes. The basic equations describe motions from the microscopic to the ocean-basin scale. As well, stochastic and

nondeterministic processes exist at all scales. In practice, physical processes on various time and space scales remain unresolved in any given ocean model. Due to nonlinearities, interactions occur across the entire range of scales in the real world, so that unrepresented processes may have serious consequences for the degree of realism achieved. Gargett (1986) and Wolff (1999) review some of these parametrization efforts. A recent volume (Chassignet and Verron 1998) provides a broad summary of progress to date. For primitive equation models, sub-grid scale turbulent processes are parametrized as eddy diffusivities for momentum and tracers in the horizontal and vertical. The simplest parametrization uses Fickian eddy diffusivities and viscosities which are constant in space and time. More sophisticated closure schemes[4] represent the effects of unresolved processes by diffusivities and viscosities which vary according to properties of the resolved flow, such as shear (Pacanowski and Philander 1981). For z-level models, rotated diffusion tensors better represent isopycnal and diapycnal mixing (Redi 1982; Gent and McWilliams 1990). Sub-grid scale parametrization in quasi-geostrophic models focusses on the representation of eddy fluxes of quasi-geostrophic potential vorticity (Wolff 1999).

1.5.4 Model forcing

Ocean models are generally forced at the surface by both winds and buoyancy fluxes. While the precise mechanisms of wind forcing are not perfectly understood, and certainly depend on sub-grid scale processes (Wolff 1999), wind forcing is generally represented as a body force in the upper layer of the model. The wind stress is related to the square of the wind speed with the use of a drag coefficient (Gill 1982, §2.4). Typically, winds are not used directly, but rather windstresses previously diagnosed from atmospheric analyses. Surface buoyancy forcing in the simplest case may be performed by relaxing temperature and salinity in the upper layer to climatological values (known as Newtonian, or modified Haney (after Haney (1971)), relaxation). While this is physically reasonable in the case of temperature, there is no physical basis for the relaxation of salinity (it is not salt that is physically fluxed across the ocean surface). Alternatively a freshwater flux may be specified as the excess of precipitation over evaporation, independent of surface salinity. Relaxation of surface temperature, with prescribed freshwater fluxes for surface salinity forcing, is referred to as *mixed boundary conditions.* Bryan (1986) showed that a model under restoring boundary conditions may be unstable upon a transition to mixed boundary conditions. Other authors (Marotzke and Willebrand 1991; Weaver and Sarachik 1991; Tziperman et al. 1994) have investigated the problem and the "multiple equilibria" of the thermohaline circulation associated with it.

[4]So-called because the unknown second-order correlations of the sub-grid scale turbulent fluctuations must be given some prescribed form in order to produce a closed set of equations. See, for example, Pedlosky (1987, §4.2). An attempted analytical expression for the second-order moments will, because of nonlinearities, require an infinite hierarchy of third- and higher-order terms.

1.5.5 Limited-area modelling

The restrictions of limited computational resources have necessitated the development of techniques for *limited-area modelling.* These represent domains smaller than the whole globe — basin-scale or smaller regional domains, for instance. Such models inevitably have one or more boundaries as open ocean, rather than land. Consequently, the straightforward *free-slip*[5] boundary conditions applicable at land boundaries must be replaced by *open boundary conditions* (OBCs) for all prognostic model variables. There is a considerable literature on OBCs in ocean models, though ultimately many of the approaches developed are *ad-hoc* mechanisms applicable to the specific problem at hand. Oliger and Sundström (1978) have shown that the primitive equations are ill-posed under local pointwise open boundary conditions. (In practice, stable OBCs can be found for PE models, perhaps at the expense of foregoing the hydrostatic assumption, or some other relation, at the boundaries. Theoretical ill-posedness, however, underscores the difficulties.) For QG models, a commonly used OBC is that developed by Charney et al. (1950), where streamfunction is prescribed along all open boundaries, and vorticity at boundary inflow points. Clearly the first requirement for an OBC is that it remain stable throughout the model integration. Beyond that, OBCs may be required to satisfy additional constraints, depending on the problem at hand. They may, for instance, be required to be non-reflective to certain wave processes, to conserve water volume (or other integral properties) within the model domain, or to specify certain tidal or other known mass transports. Røed and Cooper (1986) reviewed a range of OBCs used in ocean models. Palma and Matano reviewed and evaluated OBCs for both the barotropic (Palma and Matano 1998) and baroclinic (Palma and Matano 2000) components of a PE model.

1.5.6 Southern Ocean modelling

Applications of modelling in oceanography extend across the range of spatial scales from convection processes (Marshall and Schott 1999), to regional models (*e.g.* De Mey and Robinson (1987), Bryan and Holland (1989), Wolff et al. (1991), Griffiths (1995)), and on to global climate models (*e.g.* Anderson and Willebrand (1992), Wolff (1994)). For the Southern Ocean and ACC, modelling may be central to answering some of the most interesting questions; for example the role of eddies in the meridional heat transport and zonal momentum balance. Some large-scale phenomena (*e.g.* Antarctic Circumpolar Wave, topographic effects etc.) may only be understood through modelling, because of the impracticality of taking measurements as densely as required in time and space. Conversely, ACC modelling may have particular benefits for model development. For example, Killworth et al. (2000) have shown that annual mean surface temperatures and heat fluxes are significantly biased in regions of strong advection like the ACC, under surface tracer relaxation boundary conditions. Similarly, Jiang et al. (1999) examined equilibrium solutions of

[5] Only the normal component of velocity is specified to be zero. The alternate "no-slip" boundary condition for viscous flow sets both components of velocity to zero at the boundary.

a coarse resolution GFDL model under several surface boundary conditions and diffusion parametrizations. They show that ACC poleward heat transports are too large with a conventional diffusion scheme, but were realistic if a Gent-McWilliams scheme (1990) was used.

Numerous ACC modelling activities have focussed on questions of dynamics; particularly the interaction of the mean flow with topography, and the role of eddies in both momentum balance and heat transport. A selection of ACC modelling studies is reviewed below.

McWilliams et al. (1978) employed a wind-driven two-layer quasi-geostrophic model at eddy-resolution to investigate eddy effects in a periodic zonal channel. The initial configuration was a 1000 km square zonal periodic flat-bottom channel, forced by an eastwards sinusoidal windstress. The grid resolution was 20 km. An equilibrium volume transport of over 900 Sv was reached after 1000 days. The addition of a partial meridional barrier (with a gap of 310 km) caused this to drop to 400-600 Sv, still much larger than observed values. The channel was then lengthened to 2000 km, retaining the meridional barrier, with no change in equilibrium volume transport. The final experiment was the inclusion of a topographic ridge beneath the gap in the meridional barrier. The height of the ridge had to be sufficient to close the f/H contours for the lower layer, yet still satisfy the quasi-geostrophic constraints. A height of 500 m was chosen. The volume transport decreased by a factor of five to less than 100 Sv when the ridge was included. Thus, topography was found to play a major role in regulating ACC transport, consistent with the topographic form drag theory of Munk and Palmén (1950). The authors also analysed the momentum balance and found that the momentum input by the wind in the upper layer was transferred to the lower layer by interfacial form drag where it was balanced by bottom friction, or topographic form drag in the case of the topographic ridge. In addition the eddies were found to concentrate the upper layer jet.

Wolff et al. (1991) used a similar model, but extended it to include experiments with various topographic obstacles, including one with a realistic bathymetry representing the Macquarie Ridge Complex. The two-layer QG model used a gridsize of 20 km in a periodic zonal channel of length 4000 km and width 1500 km. No meridional barrier was included. The authors carried out detailed analyses of the momentum balance for all their experiments. They found an overall balance the same as that of McWilliams et al. (1978): momentum input by the wind in the upper layer was transferred to the lower layer through interfacial form stress, and was removed by either bottom friction over a flat bottom or topographic form stress over topography. The location of topographic obstacles determined whether standing or transient eddies dominated in contributing to interfacial form stress. In the case of realistic topography, the standing eddies dominated. The authors also found that Reynolds stresses in the upper layer acted to strengthen the main jet.

A two-layer QG model was recently employed by Witter and Chelton (1998) to further investigate the effects of topography on the spatial distribution of the eddies, and eddy-mean flow interaction. The model was a periodic zonal channel of length 4320 km and width 1200 km, with 20 km grid resolu-

tion, forced by an eastwards sinusoidal windstress. A uniform zonal ridge of width 800 km and height 400 m extended the length of the channel. In two additional experiments, first the ridge width and then both width and height were varied midway along its length. The latter was intended to represent an idealised model of the Australian-Antarctic Discordance of the Southeast Indian Ridge south of Australia. The time-mean flow consisted of a zonal jet which was steered by the topography in the cases of variable ridge height and width. An analysis of eddy kinetic energy showed very significant enhancements downstream of the variation in topography for the second two experiments. A detailed analysis of instability processes revealed an unexpected discrepancy - the flow was most unstable where time-mean baroclinic shear was smallest. This was resolved by showing that topography has a strong stabilizing effect on the flow. Specifically, stability is increased where topographic slope increases the ambient potential vorticity gradient (*e.g.* along the northern flank of a zonal ridge in the southern hemisphere). The authors also carried out analyses of energy and vorticity balances, and concluded that improved eddy parametrizations may result if based on considerations of potential vorticity and topography.

A number of other studies have used QG models in idealised channel studies of ACC dynamics, *e.g.* McWilliams and Chow (1981), and Treguier and McWilliams (1990).

Primitive-equation models have also been used to investigate ACC dynamics. Principal amongst these is the Fine Resolution Antarctic Model (FRAM), developed collaboratively by 17 researchers under the joint title of the "FRAM Group" (The FRAM Group 1991; Webb et al. 1991). The model extended around the globe southwards of 24°S, with a resolution of 0.5° in longitude and 0.25° in latitude, and with 32 levels in the vertical. The open boundary condition of Stevens (1991) was used along the northern boundary. The model was initialised with uniform cold (-2°C), salty (36.69 psu) water which was relaxed to Levitus climatology during the first six years of the run – the so-called "robust diagnostic scheme" of Sarmiento and Bryan (1982). Wind-forcing was initially zero, but was linearly increased during the third year to equal the Hellerman and Rosenstein (1983) annual mean wind climatology for the next three years. After six years, seasonal wind forcing was introduced, as well as surface buoyancy forcing through relaxation to annual mean Levitus surface temperature and salinity. After the six year robust diagnostic period, the model was free-run for another ten years. The numerical output was made available to the broader oceanographic community, and a large number of authors have published analyses of ACC dynamics from FRAM output. Killworth (1992) found a large barotropic character in both time-mean and eddy components of flow in FRAM. Thompson (1983) calculated meridional heat transport and found an eddy component of 0.12 PW southwards. Both the Ekman and mean geostrophic components were negligible in the model. The momentum budget of the ACC in FRAM has been analysed by several authors (Killworth and Nanneh 1994; Ivchenko et al. 1996; Stevens and Ivchenko 1997). To leading-order, these analyses all have confirmed the balance of Munk and Palmén of windstress being transmitted downwards through the water column

by interfacial form stress to be balanced on the bottom by topographic drag. Ivchenko et al. (1997) analysed the energy budget in FRAM. They found that the main balance was between kinetic energy input by the wind and mean potential energy stored in the density structure. Eddy kinetic energy arose through internal instability, primarily baroclinic.

With increasing computer memory and power, recent primitive equation models have modelled the global circulation at a higher resolution than FRAM. For example, Stammer et al. (1996) extended the 1/2° Semtner-Chervin model (Semtner and Chervin 1992) to 1/4° globally in the "Parallel Ocean Climate Model" (POCM). POCM output, together with FRAM output, was analysed by Hughes et al. (1999) to look for evidence in models of a proposed large-scale barotropic mode for ACC transport. Webb et al. (1997) have formulated the 1/4° global "Ocean Circulation and Climate Advanced Modelling" project (OCCAM) model. Maltrud et al. (1998) have developed a free-surface GFDL-type model at an average resolution of 1/6° globally, the "Parallel Ocean Program" (POP) model. Output from this model, together with FRAM output, was analysed by Best et al. (1999) in an investigation of ACC eddy dynamics in models.

There have been very few regional ACC modelling studies with realistic bathymetry. Wolff et al. (1991) was mentioned above. Vogeler and Schröter (1995; 1999) have used a limited-area QG model in an assimilation study located in the African sector of the Southern Ocean. Several studies have included the Atlantic sector of the ACC in basin-scale simulations of the South Atlantic (Ezer and Mellor 1997; Gan et al. 1998; Barnier et al. 1998; Treguier et al. 1999; Treguier et al. 2000).

In the work presented here, a limited-area model of the Southern Ocean is developed. Such a model has the advantage of allowing higher resolution than would otherwise be possible for given computational resources. The model's primitive equation formulation includes realistic bathymetry, allowing better representation of the topographic interactions known to be important in the Southern Ocean.

1.6 Data assimilation

The paucity of data from the Southern Ocean was mentioned earlier. Logistical complexity and expense prevent the taking of measurements sufficient in time and space to describe fully the dynamics of the ACC along its circumpolar path. Despite already having played a significant role in developing our understanding of these dynamics, computational models nevertheless remain imperfect. They may suffer from unresolved physics (*e.g.* sub-grid scale processes), imperfect physics (*e.g.* simplified equation of state, rigid-lid approximation), poor forcing (*e.g.* surface relaxation), or inaccurate numerics (*e.g.* inferior advection schemes, inappropriate grid dispersion properties). Certainly they cannot, unconstrained, diagnose or predict the synoptic ocean state.

Data assimilation attempts to combine data and models to produce estimates of the ocean state that are dynamically consistent, and consistent

also with the measured data. Under the broadest definition, the term "data assimilation" may be used to describe a wide range of activities. Insofar as inverse methods attempt to optimise a steady solution subject both to data and constraints (model), they may be regarded as data assimilation techniques. Indeed, general circulation models (GCMs) have been used together with climatological data to obtain optimised steady circulations (Marotzke and Wunsch 1993; Lee and Marotzke 1997). The robust diagnostic method of Sarmiento and Bryan (1982) has been referred to as a "very simple data assimilation methodology" (Malanotte-Rizzoli and Tziperman 1996). Generally speaking, however, data assimilation in oceanography is used to refer to techniques which combine GCMs and synoptic data to produce dynamical estimates of the ocean state. This more limited definition should be understood in the discussion that follows.

In data assimilation, the model may be regarded as dynamically interpolating the data, and the data may partially compensate for model inadequacies. The applications are very broad and include nowcasting[6] and/or forecasting, model improvement, and state estimation (for process studies, for example). Of course, there is a long history of data assimilation in meteorology where it is a fundamental component of numerical weather prediction systems (Bengtsson et al. 1981).

For the Southern Ocean, there is great potential for data assimilation in the detailed investigation of synoptic variability. The first baroclinic Rossby radius is typically smaller than 1/5° throughout the Southern Ocean (Chelton et al. 1998). Thus, model variability is unlikely to approach that of the real ocean with resolutions coarser than 1/10°. Data assimilation, therefore, may be able to increase the level of variability in Southern Ocean models, and therefore improve estimates of eddy heatflux and processes of zonal momentum balance. In order to achieve this, the assimilated data needs to have a high density in both space and time. Spaceborne instruments are the most promising candidates for data sources. Currently both sea-level and sea-surface temperature are measured, though the latter suffers through cloud cover.

1.6.1 Assimilation studies

Ezer and Mellor (1994) used Geosat altimeter data and a high-resolution model of the Gulf Stream area to nowcast mesoscale Gulf Stream dynamics during 1987-1988. While the system improved estimates of the 500 m temperature only by around 15-30%, estimates of the Gulf Stream axis location were improved by up to 75%. Stammer (1997a) also used Geosat data with a quasi-geostrophic model of the eastern North Atlantic to estimate synoptic fields in 1988-1989. He found a significant correlation between estimated fields and hydrographic observations from the same period. In a similar region, Gavart et al. (1999) used TOPEX/POSEIDON and ERS-1 altimeter data together with an open ocean primitive equation model of the Azores Current at high resolution. Data from 1993 were assimilated and significant correlations were obtained

[6]Nowcasting refers to forming an accurate estimate of the synoptic ocean state at the present time.

between estimated velocity fields and surface drifter data. In an evaluation of forecast ability, Mellor and Ezer (1991) obtained good forecast skill out to 10-20 days by assimilating simulated Geosat altimetry data. Carnes et al. (1996) described a sophisticated assimilation system for real-time nowcasts and forecasts of the north Pacific Ocean. The model is a six-layer high-resolution primitive equation model and data assimilated include IR imagery, altimetry and XBT data. The system ultimately is intended for operational naval use.

A particularly promising application of data assimilation is to improve ocean general circulation models. Schröter (1989) demonstrated how it is possible to retrieve diffusion and other parameters of a simple 1-D advection-diffusion model through assimilation of noisy observations of tracer concentration. Later, Schröter et al. (1993) used a two-layer quasi-geostrophic model of the Gulf Stream extension area to assimilate Geosat altimeter data. In a series of preliminary experiments, they were able to retrieve fairly well the friction parameter of an earlier model run by assimilating simulated Geosat SSH (sea-surface height) observations. Assimilation runs with real data produced viscosities which improved the model's tracking of the data, but which were mostly negative. However, they suggested that physically reasonable values may result from longer assimilation periods. Fu et al. (1993) were able to obtain a slight improvement to the synoptic winds used to drive an equatorial wave model by assimilating Geosat altimeter observations. Oke et al. (2000) have assimilated surface velocity data derived from coastal radar arrays into a very high resolution shelf model of the Oregon coast. By examining the corrections to the model induced by the assimilation scheme, they were able to determine that both surface winds and bottom stress may have been inadequate.

The estimation of ocean state using data assimilation is extremely powerful, allowing models to dynamically interpolate/extrapolate measurements to locations and times void of data. Assimilation of altimetry data is a prime example, with surface measurements being used to estimate subsurface fields (Hurlburt 1986; De Mey and Robinson 1987; Haines 1991; Morrow and Mey 1995). Ishikawa et al. (1996) used drifting buoy and altimetric data to correct small-scale mean SSH estimates — otherwise limited to geoid accuracy — in an idealised model of the North Pacific. This is particularly useful in the region of strong western boundary currents where spatial changes of mean SSH may be as large as changes in time. Although dynamical processes can be investigated using models alone, there is added benefit through performing analysis on the four-dimensional fields estimated via data assimilation. These should better characterise the evolution of the real ocean.

1.6.2 Data sources

A range of data has been assimilated into ocean models. Mentioned already have been climatological temperature and salinity (Marotzke and Wunsch 1993; Lee and Marotzke 1997), drifting buoy data (Ishikawa et al. 1996), and surface velocity data (Oke et al. 2000). Also used for assimilation has been deep float data (De Mey and Robinson 1987), and simulated surface temper-

ature (Cooper and Haines 1996), airborne XBT (Moore 1991), and acoustic tomography (Fukumori and Malanotte-Rizzoli 1995) measurements. Unique among data sources for its global and homogeneous density of coverage in space and time is the satellite altimeter. It is no surprise that satellite altimetry is therefore one of the most prevalent and promising data sources for combining with models through assimilation techniques. Most of the studies already mentioned have used altimetry data.

The verification of assimilation studies is most important, particularly if it is assumed that the assimilation improves model output realism in an objective sense. One important verification is a type of compatibility check: to ensure that the assimilation can at least improve estimates under the assumption of a perfect model. To this end, the so-called "twin experiment" paradigm has proven useful. In such an experiment, a free-running model is first used to generate artificial data representative of a particular sensor or measurement technique. The simulated data are then used in a second run with the same model to test the assimilation scheme. The assimilation run will generally use a different initialisation or different model parameters so that the model trajectory would otherwise differ from the sampled run. A useful assimilation technique will reduce this difference. In addition to many of the authors mentioned above, the twin experiment technique has been used to validate assimilation methodologies by Haines (1994), Dombrowsky and De Mey (1992), and Fukumori et al. (1993). A more salient test is to compare estimated fields from an assimilation of actual data with independent observations. Thus, altimetry assimilation results have been tested against independent hydrographic (Oschlies and Willebrand 1996; Dombrowsky and De Mey 1992), current meter (Fukumori 1995; Stammer 1997a), and subsurface drifter (Schröter et al. 1993; Gavart et al. 1999) measurements. A third important verification method is to monitor the norms of the innovation vectors (*i.e.* the differences between as-yet-unassimilated observations and model equivalents). Any useful assimilation scheme will reduce these differences. This statistic is examined in the assimilation experiments reported in chapter 5.

A very thorough review of assimilation techniques in meteorology and oceanography was provided by Ghil and Malanotte-Rizzoli (1991). Anderson et al. (1996) review techniques and applications to tropical oceanography, the Gulf Stream, acoustic tomography and for improving model parameters. They also cover assimilation of altimetry data, and for wave modelling. Fu and Cheney (1995) cite numerous applications of altimetry assimilation in their review of the use of satellite altimetry in oceanography. The volumes edited by Anderson and Willebrand (1989), Brasseur and Nihoul (1994) and Malanotte-Rizzoli (1996) include both reviews and numerous applications of data assimilation in oceanography. Fukumori (2000) provides a detailed introduction to altimetry assimilation in a recent volume. For meteorology, see Bengtsson et al. (1981). In addition, the special volume of *Dynamics of Atmospheres and Oceans* (1989, **13**, No. 3-4) was dedicated to data assimilation.

Brief descriptions of the major assimilation techniques will now be given. These have in common the objective of combining some form of data with a dynamical circulation model. Broadly, assimilation techniques can be classified

either as *sequential* (sometimes *statistical*) or *variational*. The former applies corrections to the model state at specific *analysis* times during a model run while the latter attempts to optimise an entire model run to match observations. In fact, the methods are strongly related to each other and may produce the same solutions (Ghil and Malanotte-Rizzoli 1991; Fukumori et al. 1993). The method chosen is largely dependent on practical considerations.

1.6.3 Sequential assimilation

Sequential assimilation techniques may be modelled symbolically (at analysis time) as follows:

$$\boldsymbol{w}^a = \boldsymbol{w}^f + \boldsymbol{K}(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}), \tag{1.3}$$

where $\boldsymbol{w}$ represents the model state vector (all gridpoint model variables). The superscript a refers to the corrected model at analysis time, f refers to the *model forecast* prior to correction. The observations are represented by $\boldsymbol{y}^o$ and $\boldsymbol{H}$ is an operator which generates equivalent observations from the model state. $\boldsymbol{K}$ is a weight matrix. The various sequential assimilation techniques differ primarily in the form of $\boldsymbol{K}$. The analysis, $\boldsymbol{w}^a$, forms the initial state of the model in a subsequent integration (or *forecast*) until the next analysis time, when the evolved model state is used as the model forecast, $\boldsymbol{w}^f$. This analysis-forecast cycle is continued throughout the assimilation period. In practice, the model-data misfits, $\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}$, may be accumulated throughout the forecast synoptically with the data.

Direct insertion and nudging

The simplest sequential assimilation technique is *direct insertion*, where model variables are substituted directly for their observed values ($\boldsymbol{K}$ is effectively the identity or a trivial spatial interpolation matrix). See Hurlburt (1986), Berry and Marshall (1989) and Haines (1991) for examples.

A more useful technique is *nudging*, where every model timestep is effectively an analysis time, with model variables continually relaxed towards observations. Equation (1.3) is more conventionally cast as an addition of forcing terms to the model prognostic equations:

$$\frac{\partial \boldsymbol{w}}{\partial t} = (\text{physics}) + \boldsymbol{K}(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}). \tag{1.4}$$

Examples of nudging can be found in Holland and Malanotte-Rizzoli (1989), Haines (1991), Verron (1992) and Stammer (1997a).

Successive corrections

The *successive corrections* technique applies corrections (1.3) at intermittent analysis times with weights in $\boldsymbol{K}$ that decrease with increasing distance in space and time between model variables and data location. The scheme may be used iteratively each analysis time, with successively smaller spatial scales used in $\boldsymbol{K}$ each iteration. An example of this technique is the study by Moore et al. (1987). Daley (1991) described successive corrections methods in detail.

Optimal interpolation

None of the sequential assimilation schemes described so far explicitly take into account either model or observation errors. *Optimal interpolation* (OI), on the other hand, assumes that both the observations, $\boldsymbol{y}^o$, and the model forecast, $\boldsymbol{w}^f$, will be in error with respect to the true ocean state, $\boldsymbol{w}^t$ say. Denote the covariances of these errors by $\boldsymbol{C}^o$ and $\boldsymbol{C}^f$ respectively:

$$\begin{aligned} \boldsymbol{C}^o &= E[(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}^t)(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}^t)^{\mathrm{T}}] \\ \boldsymbol{C}^f &= E[(\boldsymbol{w}^f - \boldsymbol{w}^t)(\boldsymbol{w}^f - \boldsymbol{w}^t)^{\mathrm{T}}] \end{aligned}$$

Then optimal interpolation weights the model forecast, $\boldsymbol{w}^f$, and the observations, $\boldsymbol{y}^o$, in inverse proportion to their respective errors, at analysis time. The weight matrix takes the form:

$$\boldsymbol{K} = \boldsymbol{C}^f \boldsymbol{H}^{\mathrm{T}} (\boldsymbol{H}\boldsymbol{C}^f \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}^o)^{-1}. \tag{1.5}$$

The technique is optimal in the sense that it minimizes the analysis error $E[\|\boldsymbol{w}^a - \boldsymbol{w}^t\|^2]$. (Derivations may be found in Gelb (1974), Gustafsson (1981), Ghil and Malanotte-Rizzoli (1991), Mellor and Ezer (1991) and Daley (1991).) The formalism provides an estimate of the analysis error covariance:

$$\boldsymbol{C}^a = E[(\boldsymbol{w}^a - \boldsymbol{w}^t)(\boldsymbol{w}^a - \boldsymbol{w}^t)^{\mathrm{T}}] = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{C}^f. \tag{1.6}$$

This can be written in the alternate form (see Gelb (1974) or Ghil and Malanotte-Rizzoli (1991) for a derivation):

$$(\boldsymbol{C}^a)^{-1} = (\boldsymbol{C}^f)^{-1} + \boldsymbol{H}^{\mathrm{T}} (\boldsymbol{C}^o)^{-1} \boldsymbol{H},$$

providing the intuitive interpretation that analysis accuracy is the sum of forecast accuracy and observation accuracy. Typically, some fixed functional form is assumed for the structure of the forecast and observational error covariances (Gustafsson 1981). A common approximation for the forecast error covariance is the form:

$$\boldsymbol{C}^f = (\boldsymbol{D}^f)^{\frac{1}{2}} \boldsymbol{C} (\boldsymbol{D}^f)^{\frac{1}{2}},$$

where $\boldsymbol{C}$ is a time-independent spatial correlation matrix and $\boldsymbol{D}^f$ is a diagonal matrix of forecast error variances (Ghil and Malanotte-Rizzoli 1991; De Mey and Benkiran 2002). These forecast error variances may be constant in time, or may be based upon the analysis error variances, $\mathrm{diag}(\boldsymbol{C}^a)$, with some empirical growth rate over time. Of course, if $\boldsymbol{C}^f$ is allowed to vary in time, then both the weight matrix, $\boldsymbol{K}$ (1.5), and the analysis error covariance, $\boldsymbol{C}^a$ (1.6), will also be time-varying. OI has been widely used in meteorological data assimilation. Examples of its use in oceanography may be found in Derber and Rosati (1989), Mellor and Ezer (1991), Dombrowsky and De Mey (1992) and Ezer and Mellor (1994).

Kalman filtering

Optimal interpolation is an approximation to the *Kalman filter*, which derived originally from engineering control theory (Kalman 1960). Instead of the OI

assumption of an *a-priori* structure for the forecast error covariance, $\boldsymbol{C}^f$, the Kalman filter evolves it in time according to the model dynamics. Writing $\boldsymbol{M}$ for the model dynamics (assumed linear here), and introducing explicit time-dependence, then the model evolution from times t_{k-1} to t_k is written:

$$\boldsymbol{w}_k^f = \boldsymbol{M}_{k-1}\boldsymbol{w}_{k-1}^a.$$

This is assumed to capture the evolution of the true ocean with some error:

$$\boldsymbol{w}_k^t = \boldsymbol{M}_{k-1}\boldsymbol{w}_{k-1}^t + \boldsymbol{\varepsilon}_{k-1}.$$

Writing the model error covariance $E[\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}}] = \boldsymbol{Q}_k$, then the Kalman filter provides the following formula for optimally evolving the forecast error covariance:

$$\boldsymbol{C}_k^f = \boldsymbol{M}_{k-1}\boldsymbol{C}_{k-1}^a\boldsymbol{M}_{k-1}^{\mathrm{T}} + \boldsymbol{Q}_{k-1}. \tag{1.7}$$

Thus, the forecast error is the sum of the previous analysis error, as propagated by the model dynamics, and the intrinsic model error. As with time-varying OI, both the weight matrix and the analysis error covariance evolve in time according to (1.5) and (1.6) respectively. The engineering literature refers more suggestively to the weight matrix, $\boldsymbol{K}$, and the model-data misfits, $\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}$, as the *Kalman gain* matrix and the *innovation* vector respectively. In practice, it is not feasible to implement the full Kalman filter (1.3), (1.7), (1.5), and (1.6) for an ocean GCM. If the size of the model state is $\mathcal{O}(N)$ then $\boldsymbol{C}^f$ has storage requirements of $\mathcal{O}(N^2)$ and computing its evolution (1.7) requires $\mathcal{O}(N)$ times more operations than a normal model integration. A moderate-sized model of 10^5 gridpoint variables taking an hour to run would require 10 gigawords storage and a decade for the assimilation! Much effort, therefore, has been put into developing suboptimal approximations to the Kalman filter. OI may be regarded as one such approximation. Other more sophisticated approaches to *order reduction* may be found in Fu and Fukumori (1993), Fukumori and Malanotte-Rizzoli (1995), Oschlies and Willebrand (1996), Faucher et al. (2002) and De Mey (2002), and are reviewed by De Mey (1997). The Kalman filter as described above is limited to linear models $\boldsymbol{M}$. The *extended Kalman filter* (EKF) has been developed for nonlinear models and proceeds by successively linearizing the model about the forecast each timestep. By substituting the linearized model for $\boldsymbol{M}$ in the Kalman filter equations, the evolution of the forecast error covariance, $\boldsymbol{C}^f$, remains correct to first order (Ghil and Malanotte-Rizzoli 1991).

1.6.4 Variational methods

In contrast to the sequential data assimilation methods described above, the variational methods seek to optimise an entire model run by choosing values of the model initialisation or other *control variables* that will minimize the difference between the model trajectory and observations (Ghil and Malanotte-Rizzoli 1991). More specifically, a scalar *cost function* (usually quadratic) is

defined measuring the difference between the model trajectory and the observations; for example,

$$J = \int [(\boldsymbol{w} - \boldsymbol{y}^o)^{\mathrm{T}} \boldsymbol{A} (\boldsymbol{w} - \boldsymbol{y}^o)] d\boldsymbol{x} dt. \tag{1.8}$$

(Note that notation now has been switched so that the model state and observations are continuous vector functions of space and time[7], *e.g.* $\boldsymbol{w} = [T, S, u, v, w]^{\mathrm{T}}(\boldsymbol{x}, t)$.) The weighting functions, $\boldsymbol{A}(\boldsymbol{x}, t)$, reflect the accuracy of observations. In the discrete case, for instance, they may be taken as the inverse of the observational error covariance matrix (Thacker 1988; Marotzke and Wunsch 1993). The variational approach to data assimilation then asks the question: what are the values of the initial conditions, model parameters, or other control variables that will minimze J, subject to $\boldsymbol{w}$ being a solution of the model? Thus, this is a problem of constrained optimisation. The constraint is simply the time-evolution of the model, written here as:

$$\dot{\boldsymbol{w}}(\boldsymbol{x}, t) = \boldsymbol{M}(\boldsymbol{w}(\boldsymbol{x}, t), \boldsymbol{u}(\boldsymbol{x}, t), \boldsymbol{x}, t), \tag{1.9}$$

where the dot notation represents time derivation and $\boldsymbol{u}$ is a vector of control variables (which may include model initialisation, forcing fields, mixing parameters etc.). It is well known that the solution to such problems can be obtained through the use of *Lagrange's undetermined multipliers* (Riley 1974; Le Dimet and Talagrand 1986; Schröter et al. 1993). The method proceeds by forming the *Lagrange function*,

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{u}, \boldsymbol{\lambda}) = J + \int \boldsymbol{\lambda}^{\mathrm{T}} (\boldsymbol{M} - \dot{\boldsymbol{w}}) d\boldsymbol{x} dt,$$

with the *Lagrange multipliers*, $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{x}, t)$, being as yet undetermined functions of space and time. A necessary condition for J to be minimized is that the Lagrange function be stationary, which requires its derivatives all to be zero:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \boldsymbol{0} \tag{1.10}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{u}} = \boldsymbol{0} \tag{1.11}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = \boldsymbol{0}. \tag{1.12}$$

Equation (1.12) retrieves the model equations (1.9), while (1.10) leads to the *adjoint equations* after integration by parts. Equation (1.11), while not generally used directly, must be verified for an optimal solution. (However, if the controls are a function of time, $\boldsymbol{u} = \boldsymbol{u}(t)$, and included in the cost function, then (1.11) solves for them (Speedy et al. 1970).) The adjoint equations can alternatively be found directly by applying a result from the *calculus of variations* (Riley 1974). This states that a functional $\mathcal{F} = \int F(\boldsymbol{y}, \boldsymbol{y}_i, \boldsymbol{y}_{ij}, \ldots, x_i) dx_i$

[7] In the case of observations at discrete times, t_i, then $\boldsymbol{y}^o = \Sigma_i \boldsymbol{y}_i^o \delta(t - t_i)$.

(where $\boldsymbol{y} = \boldsymbol{y}(x_1, \ldots, x_n)$ and $\boldsymbol{y}_i = \partial \boldsymbol{y}/\partial x_i$ etc.) may be minimized by solving the *Euler-Lagrange equations*,

$$\frac{\partial F}{\partial \boldsymbol{y}} = \frac{\partial}{\partial x_i}\left(\frac{\partial F}{\partial \boldsymbol{y}_i}\right) - \frac{\partial^2}{\partial x_i \partial x_j}\left(\frac{\partial F}{\partial \boldsymbol{y}_{ij}}\right) + \ldots - (-1)^n \frac{\partial^n}{\partial x_i \cdots \partial x_{i_n}}\left(\frac{\partial F}{\partial \boldsymbol{y}_{i \ldots i_n}}\right),$$

which in turn provide equivalent differential versions of the stationarity conditions (1.10), (1.11) and (1.12) if $\mathcal{F} = \mathcal{L}$. By way of specific example, consider a simple one-dimensional advection-diffusion model:

$$\frac{\partial w}{\partial t} = c\frac{\partial w}{\partial x} + k\frac{\partial^2 w}{\partial x^2}, \tag{1.13}$$

where c is the advective velocity and k the coefficient of diffusion. Using a quadratic measure for the cost function, the Lagrange function takes the form

$$\mathcal{L}(w, \lambda) = \int \left[(w - y^o)^2 + \lambda\left(\frac{\partial w}{\partial t} - c\frac{\partial w}{\partial x} - k\frac{\partial^2 w}{\partial x^2}\right)\right] dxdt = \int L dxdt,$$

and the Euler-Lagrange equations are

$$\frac{\partial L}{\partial \lambda} = 0 \tag{1.14}$$

$$\frac{\partial L}{\partial w} = \frac{\partial}{\partial t}\left(\frac{\partial L}{\partial w_t}\right) + \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial w_x}\right) - \frac{\partial^2}{\partial x^2}\left(\frac{\partial L}{\partial w_{xx}}\right). \tag{1.15}$$

The first of these retrieves the model (1.13), while the second gives directly the adjoint equation

$$2(w - y^o) = \frac{\partial \lambda}{\partial t} - c\frac{\partial \lambda}{\partial x} + k\frac{\partial^2 \lambda}{\partial x^2}. \tag{1.16}$$

The adjoint equations are partial differential equations (PDEs) in the Lagrange multipliers, or *adjoint variables*, $\boldsymbol{\lambda}(\boldsymbol{x}, t)$. They must be integrated backwards in time for stability (notice the negative diffusion coefficient in (1.16)), and are forced by a non-homogeneous term in model-data misfits (the LHS of (1.16)). The detail of the forcing term depends on the form chosen for the cost function. The boundary conditions emerge in the derivation of the adjoint equations from (1.10), and are homogeneous (Tziperman and Thacker 1989):

$$\begin{aligned} \boldsymbol{\lambda}(\boldsymbol{x}, t) &= 0 \text{ for } \boldsymbol{x} \text{ on the boundaries,} \\ \boldsymbol{\lambda}(\boldsymbol{x}, t_f) &= 0 \text{ at the final time, } t_f. \end{aligned}$$

In principle, the forward model together with the adjoint equations (that is, equations (1.13) and (1.16) for the 1-D advection-diffusion model) is a system of coupled PDEs that jointly can be solved subject to their respective boundary conditions to yield the optimal solution to the variational problem. In all but the simplest problems, however, such a *direct method* is impractical or impossible (Le Dimet and Talagrand 1986; Thacker and Long 1988; Ghil and Malanotte-Rizzoli 1991).

A more tractable approach is a particular iterative technique, which has come to be known as the *adjoint method* for data assimilation. The adjoint method starts by assuming some estimate $\boldsymbol{u}_0$ for the controls. The forward model is then run, accumulating the model-data misfits in the form required for the adjoint model (cf. equation (1.16)). The adjoint model is then integrated backwards in time, forced by the calculated misfits. The values of the adjoint variables so calculated may then be used, it turns out, to calculate the gradient of the cost function with respect to the controls, $\nabla J_{\boldsymbol{u}}$. Thus a gradient-descent algorithm may be used to modify the controls in the direction of the minimum:

$$\boldsymbol{u}' = \boldsymbol{u}_0 - \rho \nabla J_{\boldsymbol{u}},$$

where ρ is a step size suitably chosen for stability (Moore 1991). The whole process is iteratively repeated using the new estimates of the control variables until the cost function J has been minimized adequately. It must be noted that in this iterative approach, the adjoint model is used only to calculate the gradient of the cost function. In principle this could be done also, for example, by successively applying small variations $\delta \boldsymbol{u}$ to the control variables, integrating the forward model and calculating the corresponding variation in the cost function, δJ. In practice, however, it is cumbersome, requiring many integrations of the forward model to calculate a single gradient $\nabla J_{\boldsymbol{u}}$. It was used by Schröter and Oberhuber (as described in Schröter (1989)) to find a global mixed layer model, and increased computation time almost by a factor of 100 over a single forward integration. The relationship between the adjoint variables, $\boldsymbol{\lambda}$, and the gradient of the cost function, $\nabla J_{\boldsymbol{u}}$, depends on the choice of control variables. It can be shown in general (Schröter 1989; Schröter et al. 1993) that

$$\nabla J_{\boldsymbol{u}} = \frac{\partial J}{\partial \boldsymbol{u}} + \int \boldsymbol{\lambda}^{\mathrm{T}} \left[\frac{\partial (\boldsymbol{M} - \dot{\boldsymbol{w}})}{\partial \boldsymbol{u}} \right] d\boldsymbol{x} dt.$$

For the case of control variables which are the model initialisation, $\boldsymbol{u} = \boldsymbol{w}(\boldsymbol{x}, 0)$, the gradient of the cost function is equal to the value of the adjoint variables at initial time (Tziperman and Thacker 1989; Moore 1991; Le Dimet and Talagrand 1986; Schröter 1989),

$$\nabla J_{\boldsymbol{w}(\boldsymbol{x},0)} = \boldsymbol{\lambda}(\boldsymbol{x}, 0).$$

More generally, Hall and Cacuci (1983) have shown that the value of the i-th adjoint variable, λ_i, at time τ gives the response of the cost function with respect to a small perturbation in the model variable w_i made at time τ. Schröter (1989) has shown how to calculate from the adjoint variables the gradients of the cost function with respect to the diffusion coefficient and both a constant or time-varying advection velocity in a 1-D advection-diffusion model. Schröter et al. (1993) also did this with a two-layer quasi-geostrophic model for control variables that included initial streamfunction, Rossby radius, diffusion coefficients, and wind-stress amplitude. Tziperman and Thacker (1989) used diffusion parameters and wind-forcing as control variables in a barotropic vorticity equation model and similarly developed appropriate expressions for $\nabla J_{\boldsymbol{u}}$ in terms of the adjoint variables.

For further discussion of the theory of variational assimilation and the adjoint method see Le Dimet and Talagrand (1986), Talagrand and Courtier (1987) and Schröter et al. (1993). Examples of the adjoint method may be found in Tziperman et al. (1992), Vogeler and Schröter (1995), Lee and Marotzke (1997) and Wenzel et al. (2001), as well as works already referenced.

A disadvantage of variational assimilation *vis-à-vis* sequential methods is that no error analysis emerges from the solution[8]. In reality, an estimate of oceanic fields without some estimate of the error is of limited value. Variational assimilation methods often have been applied with no mention of error estimates, nor even a discussion of the computational difficulties. Thacker (1989) has shown how estimates of model-data error covariances can be made from the inverse of the Hessian matrix, $\boldsymbol{H} = \partial^2 J / \partial u_i \partial u_j$. In addition, the magnitudes and distribution of the eigenvalues of the Hessian determine the convergence rate of the adjoint method. In practice, this matrix is very expensive to compute, requiring as many integrations of the forward and adjoint models as there are control variables (Schröter 1989). Marotzke and Wunsch (1993) used a simplified method to estimate just the diagonal elements of $\boldsymbol{H}$ in their variational estimate of steady-state North Atlantic circulation. Despite the computational burden of their simplified approach, the results of the error analysis were "not wholly satisfactory". Schröter (1989) discussed matrix conditioning in the inversion of the Hessian, and suggested using a truncated set of control variables for the inversion, based on a singular value decomposition. Tziperman and Thacker (1989) examined the *resolution* question, rather than an explicit error analysis. This indicates how well the control variables are resolved by the data. They calculated the Hessian at a coarser resolution than the full model in order to produce the resolution matrix. The analysis showed that their (pseudo) observations of streamfunction were insufficient to resolve the windstresses they used as control variables. Schröter (1994) reviews various approaches to computing the Hessian matrix.

The question of how the adjoint model is produced in practice has not yet been addressed. Until recently, the computer code for the adjoint of a given model had to be produced manually. It has conventionally been accepted that this should be an adjoint implementation of the original (forward) finite-difference model, rather than a finite-difference discretisation of the continuous adjoint equations (Le Dimet and Talagrand 1986; Tziperman and Thacker 1989; Moore 1991; Schröter et al. 1993). Although Sirkes and Tziperman (1997) have recently questioned this in some cases, it remains the method of choice in variational assimilation. Schröter (1989) described a recipe to produce the adjoint code for a finite-difference model. It consists, first, of writing an adjoint subroutine for every subroutine in the forward model. The adjoint model is then run by calling the adjoint subroutines in reverse order. The technical complexity of manually producing adjoint models has been blamed

[8]Variational assimilation as presented here requires the solution to match exactly the model equations. This is the so-called *strong-constraint* formulation. However, a *weak-constraint* formulation also exists, where the model equations are not required to be satisfied exactly, and which allows *a-posteriori* estimates of analysis error. See Chua and Bennett (2001) and references therein for details.

by Sirkes et al. (1996) for the dearth of studies using adjoint primitive equation GCMs. Very recently, so-called "adjoint compilers" have been developed to automatically generate adjoint models (Giering and Kaminski 1998; Marotzke et al. 1999) from the forward code. This is likely to significantly advance the use of variational assimilation.

1.6.5 Initialisation

In the case of sequential assimilation techniques, the possibility occurs of generating dynamical disturbances in the model due to the impulsive changes introduced each analysis time. The linear shallow-water equations admit two classes of wave solutions — the fast inertia-gravity waves and the slower Rossby waves (LeBlond and Mysak 1978). Generally, it is only the slower waves that are of interest, both theoretically and practically (*e.g.* forecasting). The introduction of non-dynamical changes (*i.e.* changes not defined by the model dynamics) into a model may excite unwanted energetic fast-wave responses. For meteorological forecasts, these responses may overwhelm the mesoscale dynamics of interest at periods of hours to days (Daley 1981; Daley 1991); certainly they can make the prediction significantly more noisy. Thus, a tremendous amount of effort has been expended on the so-called *initialisation problem*, whereby the changes at analysis time are modified to eliminate these induced fast-responses. See Daley (1991) and references therein for a review of the meteorological literature. For sequential assimilation in oceanography, there has been little work done. There are several reasons for this. First, quasi-geostrophic models do not admit inertia-gravity waves, and so the problem does not exist for this class of model. Second, primitive-equation models often make the rigid-lid approximation, which also eliminates fast waves. Thus it is only sequential assimilation studies with free-surface primitive-equation models that may be expected to display initialisation effects. Malanotte-Rizzoli et al. (1989) found even in this case that such effects were small and that relatively simple initialisation strategies could be employed to significantly suppress gravity wave noise. Hurlburt (1986) also found a simple initialisation procedure (*i.e.* ensuring the modified fields were geostrophically balanced) was sufficient in a study with a two-layer PE model. Clearly, the requirement for initialisation will depend on the details of the model and assimilation technique, and the purpose for which assimilation is being performed. Discussions of initialisation in ocean data assimilation can be found in Ghil (1989) and Ghil and Malanotte-Rizzoli (1991).

1.6.6 Assimilation in the Southern Ocean

There have been few studies of data assimilation in the Southern Ocean. Vogeler and Schröter (1995) assimilated Geosat SSH data using an adjoint method into an open ocean QG model of the African sector of the Southern Ocean. The control variables were the model initial streamfunction. The assimilation was found to work well for timescales of 20 days or less, but the open boundary conditions were found to influence the results more than the initial

conditions for longer timescales. A later study (Vogeler and Schröter 1999) included the boundary values of streamfunction and vorticity amongst the control variables. While this produced an improvement, there was still a difference between shorter and longer assimilation periods; while an SSH error of less than 5 cm could be achieved with a short assimilation period, a longer term assimilation produced an error of 10-12 cm. Evensen and van Leeuwen (1996) assimilated gridded Geosat altimetry into a two-layer quasigeostrophic model of the Agulhas Current region south of Africa. They employed the Ensemble Kalman Filter, which uses a Monte Carlo method for estimating the forecast error covariance. In essence the method integrates forward an ensemble of states having the sample mean equal to the previous analysis. The covariances between this ensemble of forecasts is then used as an estimate of the forecast error covariance for the Kalman gain matrix in the next analysis. They obtained analyses consistent with the data, and found that the method was able to reproduce eddy shedding which is normally too slow in QG models due to the lack of ageostrophic effects. Seiß et al. (1997) employed a nudging technique to assimilate Geosat altimeter data into a circumpolar eddy-resolving QG model of the Southern Ocean. Convergence to observations was fast in the case of a twin experiment, but less successful with real data. However, the variability of the model was improved with real data. A QG model of the South Atlantic ocean was used by Florenchie and Verron (1998) in a two-year assimilation of TOPEX/POSEIDON and ERS-1 altimeter data using a nudging method. The model was strongly influenced by the assimilated data. A resolution of 1/6° allowed the authors to track individual Agulhas rings, fourteen of which were shed during the assimilation period. Grotov et al. (1998) assimilated hydrographic data from the WOCE S4 section, together with climatological data, into a model using a variational method to reconstruct the large-scale circulation of the Amundsen and Bellingshausen Seas. Stutzer and Krauss (1998) assimilated surface drifter data into a PE model of the South Atlantic to improve surface flow. The subtropical gyre circulation was improved, though little impact on basin-scale integrated quantities was found.

A recent sequential assimilation of TOPEX/POSEIDON and ERS-1 altimetry data into a 1/4° global PE model was performed by Fox et al. (2000a; 2000b). A twin-experiment run achieved significant improvements through assimilation (error reductions of greater than 50% for upper ocean currents after five months of assimilation). Problems were found in the Southern Ocean with temperature and salinity fields being adversely impacted initially. This was thought to be due mainly to the relatively large barotropic component of Southern Ocean variability being incompatible with the assimilation technique, which assumed no change in bottom pressure. An assimilation with real data allowed improvements in SSH prediction over most areas of the ocean, with the exception of the ACC.

1.7 Present study

In this thesis, a regional primitive equation model of the Southern Ocean is implemented and then used for assimilation of altimetry data using an optimal interpolation scheme.

The region chosen is one that is rich in eddy variability, and where topography is thought to play a major role in the ACC's momentum balance. There is also a concentration of in-situ data in the region, including some six repeats of a WOCE hydrographic section, XBT data, and several multi-year current mooring records.

The work presented here includes several novel elements. Few other primitive equation models include an implicit free-surface formulation. Open boundary conditions for such models have not previously been investigated in detail.

This work also represents one of the first dedicated Southern Ocean assimilation studies using either a primitive equation model or TOPEX/POSEIDON altimetry data. Other sequential assimilation studies in the Southern Ocean have generally employed nudging rather than a more sophisticated technique, such as optimal interpolation, as used here. While sea-level variability in the Southern Ocean is known to include a barotropic component, previous Southern Ocean assimilation studies have not explicitly addressed this issue.

The remainder of this thesis is structured as follows: chapter 2 introduces the Hamburg Ocean Primitive Equation (HOPE) model and describes the configuration used here; chapter 3 discusses the modifications made to allow it to operate with oceanic open boundaries; chapter 4 discusses the model's performance in a forty-year spinup run; and chapter 5 reports assimilation results. A summary and conclusions are provided in chapter 6.

CHAPTER 2

Numerical model

2.1 The Hamburg Ocean Primitive Equation (HOPE) model

The Hamburg Ocean Primitive Equation (HOPE) model (Wolff et al. 1997) is a state-of-the-art FORTRAN ocean modelling code developed primarily by Ernst Maier-Reimer at the Max-Planck Institute for Meteorology in Hamburg, Germany. It has recently been used in studies of the El-Niño Southern Oscillation (ENSO) (van Oldenborgh et al. 1999; Venzke et al. 2000a; van Oldenborgh 2000), paleo-climate (Kim et al. 1998; Kim and Crowley 2000), North Pacific decadal variability (Xu et al. 1998; Venzke et al. 2000b), watermasses in the equatorial Indian and Pacific oceans (Rodgers et al. 2000) and the Southern Ocean (Kim and Stössel 1998), and sea-ice (Marsland and Wolff 1998; Stössel et al. 1999). Notable features of the model include an exact representation of model bathymetry at gridpoints, a prognostic sea-level, and momentum advection which conserves second-order quantities (kinetic energy and enstrophy). These features in part motivated the choice of the HOPE model for altimetry assimilation in the Southern Ocean. In particular, topography is known to play a significant role in Southern Ocean dynamics (Treguier and McWilliams 1990; Gille 1997; Witter and Chelton 1998). Therefore its accurate representation is likely to be important in a regional assimilation study. For certain assimilation techniques (direct nudging, for example) and altimetric data, a prognostic sea-level is required. Although direct-nudging was not applied in this study, a prognostic sea-level is likely, nevertheless, to carry some benefits for altimetric assimilation. For example, irrespective of any projection of sea-level anomaly data into the ocean interior, geostrophic adjustment will ensure that the mass transport reflects assimilated sea-level at length scales larger than the barotropic Rossby radius. A prognostic sea-level also allows immediate comparison with altimetric data rather than it having to be diagnosed. Finally, the frontal regions of the Antarctic Circumpolar Current are expected to be highly nonlinear so that accurate momentum advection will be needed to represent adequately the eddy dynamics. In addition, ready access to other users of the HOPE model and the expertise of one of its developers (Dr Jörg-Olaf Wolff) was an important factor guiding its choice for this study.

2.1.1 Model physics

Basic equations

Newton's second law of motion for a material element of a viscous fluid provides the *Navier-Stokes* equations. In a rotating reference frame under gravity these may be written:

$$\frac{D\boldsymbol{u}}{Dt} + 2\boldsymbol{\Omega} \times \boldsymbol{u} = -\frac{1}{\rho}\nabla p - \boldsymbol{g} + \boldsymbol{F}, \tag{2.1}$$

where $\boldsymbol{u} = \{u, v, w\}(\boldsymbol{x}, t)$ is the three-dimensional velocity vector in Cartesian co-ordinates $\boldsymbol{x} = \{x, y, z\}$, $\boldsymbol{\Omega}$ is the angular velocity of the rotating frame, $p = p(\boldsymbol{x}, t)$ is the pressure field, $\rho = \rho(\boldsymbol{x}, t)$ is the fluid density, $\boldsymbol{g}$ is acceleration due to gravity, and $\boldsymbol{F}$ is viscosity (for molecular viscosity this takes the form $\boldsymbol{F} = \nu\nabla^2\boldsymbol{u}$). Notation for the *material derivative* has been used:

$$\frac{D(.)}{Dt} = \frac{\partial(.)}{\partial t} + \boldsymbol{u} \cdot \frac{\partial(.)}{\partial \boldsymbol{x}}.$$

The *primitive equations* of geophysical fluid dynamics simplify the Navier-Stokes equations by making two approximations. The first is the *hydrostatic approximation* (Gill 1982; Pedlosky 1987; Kantha and Clayson 2000). This recognises that vertical velocities in the ocean and atmosphere are much smaller than horizontal velocities, so that the material derivative term is neglected in equation (2.1) for the vertical velocity. Horizontal components of gravity are neglected, $\boldsymbol{g} = g\boldsymbol{k}$. Finally, the vertical component of the *Coriolis term*, $\boldsymbol{\Omega} \times \boldsymbol{u}$, and molecular viscosity are much smaller than other terms and so also are neglected. Thus, equation (2.1) for vertical momentum conservation is replaced by a hydrostatic balance:

$$\frac{\partial p}{\partial z} = -\rho g. \tag{2.2}$$

The second approximation of the primitive equations is the *Boussinesq approximation* (Boussinesq 1903; Spiegel and Veronis 1960), which recognises that density variations are small over typical motions: their effects on horizontal momentum are neglected, but not on buoyancy. It amounts to taking the density to be a constant, $\rho(\boldsymbol{x}, t) = \rho_0$, in equations (2.1) for horizontal components of velocity, but retaining density variations in calculating pressure from the hydrostatic relation (2.2).

In component form, then, the primitive equations are the hydrostatic balance equation (2.2) and the two horizontal momentum equations:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} - fv = -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + F_x \tag{2.3}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} + fu = -\frac{1}{\rho_0}\frac{\partial p}{\partial y} + F_y. \tag{2.4}$$

The *Coriolis parameter* $f = 2|\boldsymbol{\Omega}| \sin\phi$, where ϕ is latitude.

The nonlinear equation of state for seawater depends on pressure and the material thermodynamic quantities *temperature*, T, and *salinity*, S:

$$\rho = \rho(T, S, p). \tag{2.5}$$

Conservation equations may be written for these quantities if temperature is replaced by the conservative property *potential temperature*, θ. This is the temperature a parcel of fluid would have if it were moved adiabatically to some reference pressure (usually 1 bar). The conservation equations for θ and S reflect advection-diffusion physics:

$$\frac{D\theta}{Dt} = -\theta\nabla.\boldsymbol{u} + F_\theta \tag{2.6}$$

$$\frac{DS}{Dt} = -S\nabla.\boldsymbol{u} + F_S, \tag{2.7}$$

where F_θ and F_S are the diffusivities for potential temperature and salinity, respectively (for molecular diffusion these take the form $F_\theta = (\theta k)/(C_p T\rho)\nabla^2 T$ and $F_S = \kappa_D \nabla^2 S$ where k is thermal conductivity, C_p is specific heat at constant pressure, T is temperature, and κ_D is the *diffusivity* of salt in water (Gill 1982; Pedlosky 1987)).

The basic equations are completed with the equation of volume conservation (or *continuity*) for an incompressible fluid:

$$\nabla.\boldsymbol{u} = 0. \tag{2.8}$$

Since the prognostic momentum equation for the vertical velocity, w, is replaced in the primitive equations with the hydrostatic relation, w may instead be calculated diagnostically by integrating the continuity equation vertically through the water column:

$$w(z) = -\int_{-H}^{z} \nabla_h.\boldsymbol{u}_h(z')dz', \tag{2.9}$$

where $z = -H(x,y)$ represents the model bathymetry and the subscript h represents horizontal components.

Sub-grid-scale parametrizations

While the set of equations (2.2), (2.3), (2.4), (2.5), (2.6), (2.7) and (2.8) is complete for a continuous fluid, discretisation of these equations prevents the explicit representation of processes smaller than the grid size. It turns out that the resolved physics satisfies the primitive equations, with additional viscous terms (Pedlosky 1987, §4.2):

$$F_x = (\text{molecular viscosity}) + \frac{1}{\rho}\left(\frac{\partial\tau_{xx}}{\partial x} + \frac{\partial\tau_{xy}}{\partial y} + \frac{\partial\tau_{xz}}{\partial z}\right), \tag{2.10}$$

$$F_y = (\text{molecular viscosity}) + \frac{1}{\rho}\left(\frac{\partial\tau_{yx}}{\partial x} + \frac{\partial\tau_{yy}}{\partial y} + \frac{\partial\tau_{yz}}{\partial z}\right), \tag{2.11}$$

$$F_x = (\text{molecular viscosity}) + \frac{1}{\rho}\left(\frac{\partial\tau_{zx}}{\partial x} + \frac{\partial\tau_{zy}}{\partial y} + \frac{\partial\tau_{zz}}{\partial z}\right). \tag{2.12}$$

The *Reynolds stress tensor*, $\boldsymbol{\tau}$, represents correlations between the unresolved turbulent velocity fluctuations, $\{u', v', w'\}$:

$$\tau_{xx} = -\rho\overline{u'u'},\ \tau_{yy} = -\rho\overline{v'v'},\ \tau_{zz} = -\rho\overline{w'w'},$$

$$\tau_{xy} = \tau_{yx} = -\rho\overline{u'v'},$$
$$\tau_{xz} = \tau_{zx} = -\rho\overline{u'w'},$$
$$\tau_{yz} = \tau_{zy} = -\rho\overline{v'w'},$$

While expressions for these stresses can be written down, they depend on higher-order moments. Expressions for these higher-order moments may also be written, but they in turn depend on even higher-order terms and so on. Thus, there is no simple analytical *closure scheme* for the *turbulent* (or *eddy* or *mixing*) *stresses.* As described earlier (section 1.5.3), their parametrization in terms of resolved quantities represents a very active area of current research in ocean modelling. Typically, the parametrizations of sub-grid-scale processes mimic molecular viscosity and diffusion in form, but are larger by several orders of magnitude and thus the molecular terms are neglected. The HOPE model parametrizes the turbulent stresses separably in the horizontal and vertical,

$$\{F_x, F_y\} = \boldsymbol{F}_H + \boldsymbol{F}_V. \tag{2.13}$$

The horizontal turbulent viscosity optionally includes both harmonic and biharmonic terms, as well as a term dependent on the local rate of strain:

$$\boldsymbol{F}_H = A_H \nabla_h^2 \boldsymbol{u}_h - B_H \nabla_h^4 \boldsymbol{u}_h + \nabla.(\nu_A T^2 \nabla_h \boldsymbol{u}_h), \tag{2.14}$$

where A_H, B_H and ν_A are coefficients of horizontal viscosity and the *strain*[1], $T^2 = (\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y})^2$. The vertical turbulent viscosity is parametrized as:

$$\boldsymbol{F}_V = \frac{\partial}{\partial z}\left(A_V \frac{\partial \boldsymbol{u}_h}{\partial z}\right), \tag{2.15}$$

where A_V is the coefficient of vertical viscosity.

Similarly, the effects of sub-grid-scale turbulence on tracer conservation are parametrized as additional diffusive terms:

$$F_\theta = \frac{\partial}{\partial z}\left(D_V \frac{\partial \theta}{\partial z}\right) + D_H \nabla_h^2 \theta + \nabla.(\nu_D T^2 \nabla_h \theta), \tag{2.16}$$

$$F_S = \frac{\partial}{\partial z}\left(D_V \frac{\partial S}{\partial z}\right) + D_H \nabla_h^2 S + \nabla.(\nu_D T^2 \nabla_h S), \tag{2.17}$$

where D_H and ν_D are coefficients of horizontal diffusivity, and D_V is the coefficient of vertical diffusivity.

The coefficients of both vertical viscosity and diffusivity may optionally be formulated with a term depending on the local Richardson number, Ri (Pacanowski and Philander 1981), and a mixed layer contribution:

$$\frac{\partial A_V}{\partial t} = -\lambda A_V + \lambda\left(\frac{A_{V_0}}{(1 + C_{R_A} Ri)^2} + \delta_{\Delta\rho} W_\rho + A_b\right), \tag{2.18}$$

$$\frac{\partial D_V}{\partial t} = -\lambda D_V + \lambda \max\left[\left(\frac{D_{V_0}}{(1 + C_{R_D} Ri)^2} + \delta_{\Delta\rho} W_\rho\right), D_b\right]. \tag{2.19}$$

[1]Note that T here is different to the temperature, T, used earlier in the expression for molecular diffusion, F_θ.

In this formulation, A_b and D_b are constant background values, and A_{V_0}, D_{V_0}, C_{R_A}, C_{R_D} are constant parameters. The parameter λ defines a "memory" time-scale. The mixed layer contributions W_ρ are included using the switch $\delta_{\Delta\rho}$, which is one in the mixed layer and zero elsewhere. The Richardson number is the ratio of the squares of the buoyancy frequency and the vertical shear:

$$Ri = \frac{\frac{g}{\rho_0}\left(\frac{\partial \rho}{\partial z}\right)}{\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2}.$$

The values used for the various sub-grid-scale parametrizations, and other model parameters, in this thesis are mentioned in section 2.2.3.

Boundary conditions

A no-slip condition is applied to model velocities at all solid boundaries:

$$\boldsymbol{u}(\boldsymbol{x}, t)|_{\boldsymbol{x}\in\delta V} = 0,$$

where δV represents the bathymetric boundary of the modelled ocean volume V. This guarantees zero-flux conditions for tracers:

$$(\boldsymbol{u}\theta).\boldsymbol{n} = 0,$$
$$(\boldsymbol{u}S).\boldsymbol{n} = 0,$$

where $\boldsymbol{n}$ is a unit vector normal to the solid surface. A Newtonian friction law is also applied to the velocities near the bottom:

$$\{F_x, F_y\} = \boldsymbol{F}_V + \boldsymbol{F}_H - \varepsilon \boldsymbol{u}(\boldsymbol{x}, t)|_{|\boldsymbol{x}-\delta V|<\Delta}, \tag{2.20}$$

with ϵ the coefficient of bottom friction.

At the sea surface, wind stress[2], $\boldsymbol{\tau}$, imparts a momentum flux:

$$\rho A_V \frac{\partial \boldsymbol{u}}{\partial z}\bigg|_{z=0} = \boldsymbol{\tau}. \tag{2.21}$$

The wind stress $\boldsymbol{\tau} = \boldsymbol{\tau}(t)$ is a monthly mean value linearly interpolated in time.

While HOPE optionally allows the specification of heat and freshwater fluxes at the surface, the surface boundary condition for tracers used in this thesis was Newtonian relaxation to climatological values:

$$D_V \frac{\partial S}{\partial z}\bigg|_{z=0} = \lambda_S(S_o - S), \tag{2.22}$$

$$D_V \frac{\partial \theta}{\partial z}\bigg|_{z=0} = \lambda_\theta(\theta_o - \theta). \tag{2.23}$$

S_o is an annual mean climatological surface salinity, and $\theta_o = \theta_o(t)$ is a monthly mean climatological sea-surface temperature linearly interpolated in time. λ_S and λ_θ are relaxation strengths.

[2]Note that $\boldsymbol{\tau}$ here, used to represent wind stress, is different to that in equations (2.12) for the viscous stress tensor.

The sea-surface, $z = \eta(x, y, t)$, is a material surface which can be represented by the equation

$$G(x, y, z, t) = \eta(x, y, t) - z = 0.$$

Thus for a particle on the surface, G will remain zero:

$$\frac{DG}{Dt} = \frac{\partial \eta}{\partial t} + u\frac{\partial \eta}{\partial x} + v\frac{\partial \eta}{\partial y} - w = 0.$$

HOPE linearizes this *kinematic boundary condition* to obtain:

$$\frac{\partial \eta}{\partial t} = w|_{z=0} = -\nabla_h . \int_{-H}^{0} \boldsymbol{u}_h dz, \tag{2.24}$$

where the continuity equation (2.8) has been used.

The HOPE model also includes an optional dynamic-thermodynamic sea-ice model (Hibler 1979). This sea-ice model was not used in this study in order to simplify the open boundary modifications and altimetry assimilation interface.

For detailed description of the model numerics, the reader is referred to the HOPE model manual (Wolff et al. 1997). The following two sections provide an outline of the discretisation of the model equations in space and time.

2.1.2 Spatial discretisation

The HOPE model solves the equations described above (section 2.1.1) using a finite-difference procedure. The continuous prognostic variables are approximated by representing them at discrete points on a three-dimensional grid. Spatial derivatives in the continuous equations are replaced by finite-difference counterparts and the solutions are advanced in time using a procedure described below. There are numerous ways of arranging model variables on a regular discrete grid and each will represent the continuous solution with a different degree of accuracy. Five particular arrangements of variables in the horizontal (Figure 2.1) were analysed for their dispersion properties under geostrophic adjustment by Mesinger and Arakawa (1976) and Arakawa and Lamb (1977). The grid arrangements have come to be known as the *Arakawa grids*, A through E. They concluded that the B- and C-grids most accurately represent the inertia-gravity waves associated with one-dimensional geostrophic adjustment. In two dimensions, the C-grid is better if the grid-spacing is smaller than the Rossby radius, but the B-grid is better otherwise. (Recently, Randall (1994) has proposed a Z-grid which performs well at all grid spacings.) Most ocean models today use either the B- or C-grids.

The HOPE model discretises model variables on the Arakawa E-grid in the horizontal, as shown in Figure 2.2. 'Scalar' quantities $(\theta, S, \rho, p, \eta)$ are co-located and offset from 'vector' quantities (u, v). In the vertical, the HOPE model uses a $z-$level discretisation with layers of specified thickness, z_k, as shown in Figure 2.3. Most model variables are defined at mid-layer depths,

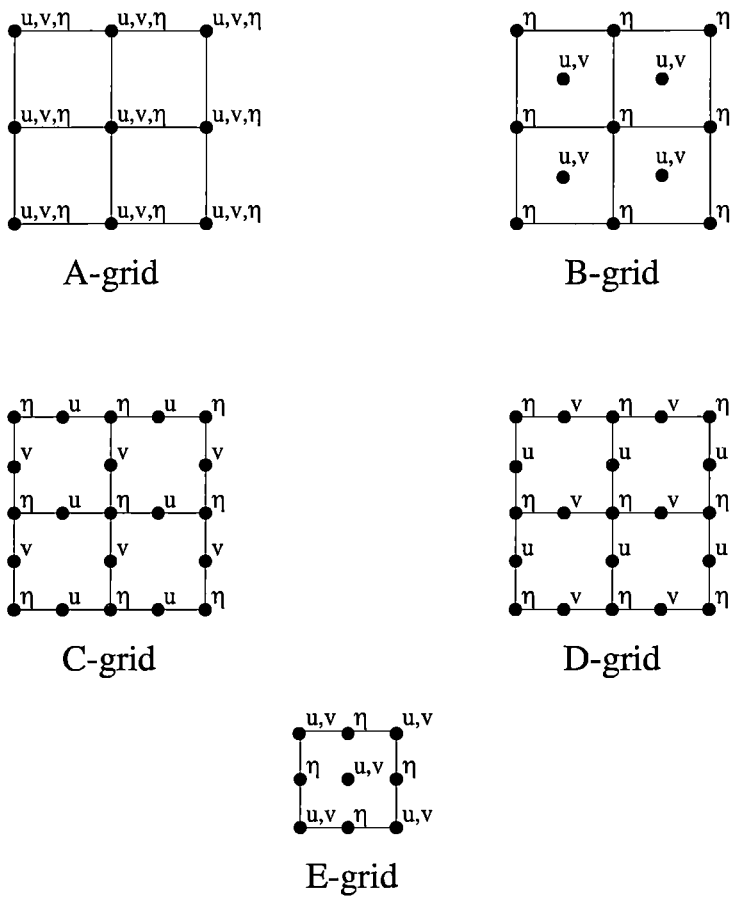


Figure 2.1: The Arakawa finite-difference grids for the shallow-water equations (η is sea-level, (u, v) is horizontal velocity)

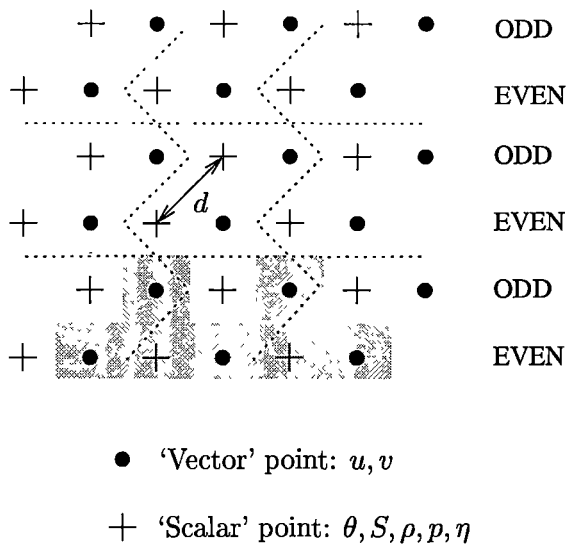


Figure 2.2: The horizontal Arakawa E-grid as used in the HOPE model. The effective resolution is labelled d. The shaded gridpoints represent a land boundary.

but vertical velocities are located at layer interfaces between scalar grid-points. The three velocity components (u, v, w) are to be interpreted as the fluid velocities across the faces of grid-boxes whose average tracer (θ, S) values are given by the enclosed scalar gridpoint. While horizontal diffusivities and viscosities (D_H, A_H, B_H) are constant in space, the vertical mixing parameters (D_V, A_V) may vary with the local Richardson number, and so are discretised in space. They are located at layer interfaces in the vertical, and aligned with the relevant grid-point type (scalar for D_V, vector for A_V) in the horizontal. The local strain rate parameter, T^2, is defined at mid-layer depths in the vertical on scalar grid-points.

The model carries individual grid-box thicknesses for all horizontal grid-points and vertical layers. In this way, the bathymetry is not restricted by the choice of layer thicknesses and may instead be resolved exactly. (Adcroft et al. (1997) have shown that such a scheme provides smoother and more accurate fields than if topography is limited to layer depths.) From the horizontal momentum equations (2.3) and (2.4), it is clear that the evolution of horizontal velocity requires gradients of pressure. This requires that no oceanic velocity point in the model be immediately adjacent to land. The topography in the model is therefore adjusted initially by setting the depth of each scalar grid column to the maximum depth of the four surrounding vector grid columns.

The horizontal grid is Cartesian and specified by the latitude and longitude locations of grid rows and columns. Calculation of spatial gradients is facilitated by storing two-dimensional arrays of inter-grid-point distances. This is necessary with x-derivatives to account for convergence of meridians with latitude. In principle, it provides the opportunity for HOPE to be extended to non-Cartesian orthogonal grid systems. In fact HOPE has recently been reformulated on a non-geographic curvilinear grid in the horizontal (using

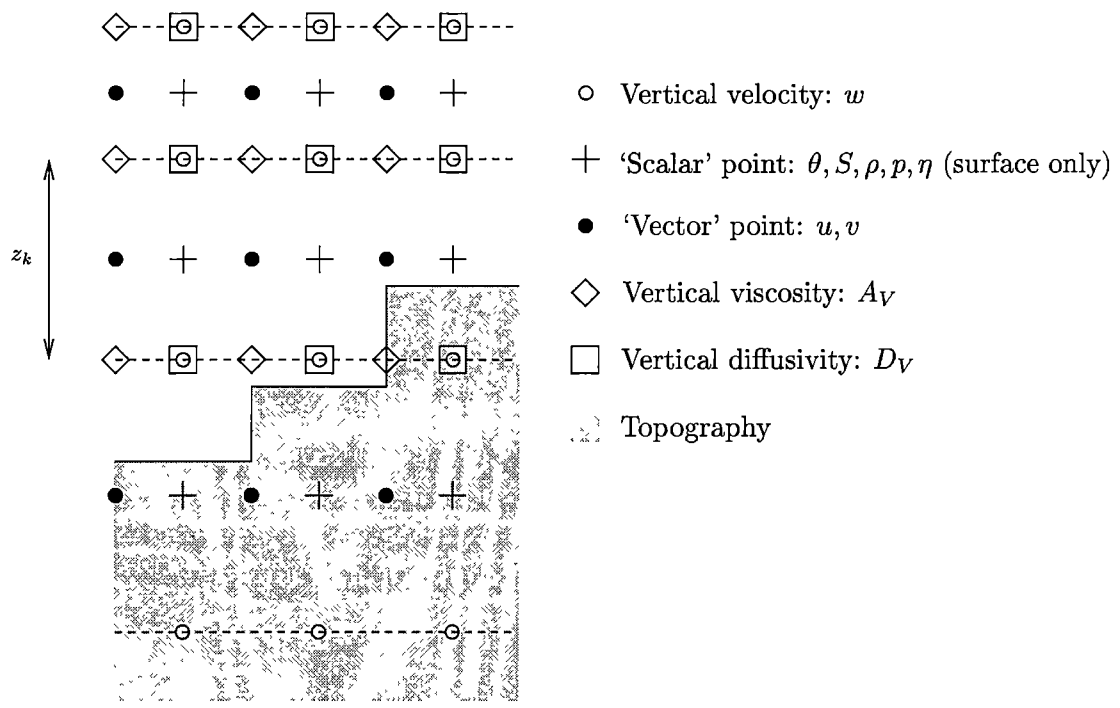


Figure 2.3: The vertical grid used in the HOPE model

an Arakawa C-grid) (MPI 2001).

Successive (latitudinal) grid rows in HOPE are labelled alternately as ODD or EVEN (Figure 2.2), with all model variables carried accordingly in two separate arrays. This internal separation of model variables hints at a more fundamental problem with the Arakawa E-grid: it exhibits a degree of decoupling in its solutions for centred differencing schemes. The problem may be seen by considering a solution in geostrophic balance. Values of u on the EVEN grid, and v on the ODD grid, are determined by centred differences of pressures at ODD scalar grid-points. Similarly, pressures on EVEN scalar gridpoints determine ODD u and EVEN v velocities. Thus there are two independent decoupled systems. A constant sea-level could be added to one of the two (EVEN/ODD) subgrids, for instance, without upsetting either solution but introducing a very severe "chequerboard" pattern into the combined sea-level field. Even with non-equilibrium solutions, it is only the Coriolis, nonlinear, and mixing terms that couple the two solutions (otherwise, time evolution of EVEN u and ODD v points is determined by pressure gradients between EVEN scalar points, while evolution of ODD u and EVEN v is determined by ODD scalar points). Such a decoupling does not occur with other Arakawa grids. The point may further be appreciated by noticing that the E-grid is effectively two overlaid C- (or D-) grids offset diagonally from one another. While severe grid-point decoupling did not often occur in this study, it was found to be prudent to examine both EVEN and ODD arrays of grid point variables to ensure this was the case. With the open boundary investigations described in section 3, for instance, very severe, but stable, chequerboarding sometimes occurred indicating potential problems with coding or algorithms.

2.1.3 Time discretisation

The HOPE model solves the prognostic partial differential equations described above using the method of *operator splitting*. The following description follows that in Press et al. (1992). Also known as the *method of fractional steps*, it may be applied to problems where the differential operator can be written as a linear sum:

$$\frac{\partial \varphi}{\partial t} = \mathcal{L}\varphi, \tag{2.25}$$
$$\mathcal{L}\varphi = \mathcal{L}_1\varphi + \mathcal{L}_2\varphi + \cdots + \mathcal{L}_k\varphi.$$

Suppose a finite-difference algorithm, $\mathcal{F}_i$, is available for each of the k partial differential equations $\frac{\partial \varphi}{\partial t} = \mathcal{L}_i\varphi$ separately:

$$\begin{aligned} \varphi(t^{n+1}) &= \mathcal{F}_1(\varphi(t^n)) \\ \varphi(t^{n+1}) &= \mathcal{F}_2(\varphi(t^n)) \\ &\cdots \\ \varphi(t^{n+1}) &= \mathcal{F}_k(\varphi(t^n)). \end{aligned}$$

Then the original problem (2.25) may be advanced in time from t^n to $t^{n+1} = t^n + \Delta t$ by applying successively the finite-difference operators for each of the components:

$$\begin{aligned} \varphi(t^{n+(1/k)}) &= \mathcal{F}_1(\varphi(t^n)) \\ \varphi(t^{n+(2/k)}) &= \mathcal{F}_2(\varphi(t^{n+(1/k)})) \\ &\cdots \\ \varphi(t^{n+1}) &= \mathcal{F}_k(\varphi(t^{n+(k-1)/k})). \end{aligned}$$

Thus the HOPE model applies operator splitting to the various prognostic equations. Table 2.1 lists, in order of their application, the subroutines used by HOPE to advance the solution each timestep.

Salient features of the differencing schemes used for most of the key processes are now summarised briefly. For further details, the reader is referred to the HOPE model manual (Wolff et al. 1997). Specific parameter values used in this study are given in section 2.2.

Wind forcing (subroutine `OCWIND`)

The surface windstress (equation (2.21)) produces an accelerating force per unit mass (cf. the parametrization of vertical viscosity equation (2.15)):

$$\frac{\partial \boldsymbol{u}}{\partial t} = \frac{1}{\rho}\frac{\partial \boldsymbol{\tau}}{\partial z}. \tag{2.26}$$

In HOPE wind forcing is applied as a body force in the upper layer:

$$\frac{\partial \boldsymbol{u}}{\partial t} = \frac{1}{\rho}\frac{\boldsymbol{\tau}}{\Delta z_1}.$$

Table 2.1: HOPE timestep subroutines

subroutine	*physics solved*	*reference terms/equations*
OCTHER	thermohaline forcing	(2.22), (2.23)
	internal pressure	(2.27)
	convective adjustment[a]	—
	vertical mixing coefficients	(2.18), (2.19)
OCWIND	wind forcing	(2.26)
OCVISC	vertical momentum advection	$w\frac{\partial(u,v)}{\partial dz}$: (2.3), (2.4)
	vertical viscosity	$\frac{\partial}{\partial z}\left(A_V\frac{\partial \boldsymbol{u}_h}{\partial z}\right)$: (2.3), (2.4), (2.15)
	horizontal viscosity	$A_H\nabla^2\boldsymbol{u}_h + \nabla.(\nu_A T^2\nabla_h\boldsymbol{u}_h)$: (2.3), (2.4), (2.14)
	bottom friction	$\varepsilon\boldsymbol{u}$: (2.3), (2.4), (2.20)
OCBIHAE	biharmonic viscosity	$B_H\nabla_h^4\boldsymbol{u}_h$: (2.3), (2.4), (2.14)
OCBIHAR	biharmonic viscosity (rotated operator)	*as above*
OCIADJ7	horizontal momentum advection	$u\frac{\partial(u,v)}{\partial x} + v\frac{\partial(u,v)}{\partial y}$: (2.3), (2.4)
OCMODS	barotropic/baroclinic partitioning	(2.28), (2.29)
OCCLIT	baroclinic solution	(2.34), (2.35), (2.36), (2.37)
OCRHSZ	calculate barotropic system RHS	(2.41)
ZGAUSS	solve barotropic solution	(2.42)
OCVTRO	new barotropic velocities	(2.38), (2.39)
OCVTOT	new total velocities	(2.28), (2.29)
OCADV/ OCADUP	tracer advection	$\boldsymbol{u}\cdot\frac{\partial(\theta,S)}{\partial \boldsymbol{x}}$: (2.6), (2.7)
	horizontal diffusion	$D_H\nabla_h^2(\theta,S) + \nabla.(\nu_D T^2\nabla_h(\theta,S))$: (2.6), (2.7), (2.16), (2.17)
OCVERDI	vertical diffusion	$\frac{\partial}{\partial z}\left(D_V\frac{\partial(\theta,S)}{\partial z}\right)$: (2.6), (2.7), (2.16), (2.17)

[a]Thermohaline forcing of the surface layer may lead to vertically unstable density profiles and a *convective instability*. Physically, convection occurs as a sub-grid-scale process (Marshall and Schott 1999). In the HOPE model, a single vertical sweep downwards through the water column is performed each timestep. If any layer is found to be less dense than the layer above, then θ and S in both are uniformly mixed in proportion to their respective layer thicknesses:

$$\theta(z_{k-1}) = \theta(z_k) = \frac{\Delta z_{k-1}\theta_{k-1} + \Delta z_k\theta_k}{\Delta z_{k-1} + \Delta z_k}$$

$$S(z_{k-1}) = S(z_k) = \frac{\Delta z_{k-1}S_{k-1} + \Delta z_k S_k}{\Delta z_{k-1} + \Delta z_k}$$

Thermohaline forcing (subroutine OCTHER)

As with wind forcing, surface forcing of tracers (equations (2.22), (2.23)) is applied in the surface layer of the model.

Vertical momentum advection and vertical viscosity (subroutine OCVISC)

These are solved together in the HOPE model using an implicit centred-in-space differencing scheme. A tridiagonal system results which is solved by Gaussian elimination and backsubstitution (Press et al. 1992).

Horizontal viscosity (subroutine OCVISC)

Harmonic and strain-dependent horizontal viscosity is solved with an explicit centred-in-space scheme.

Biharmonic horizontal viscosity (subroutines OCBIHAE, OCBIHAR)

A five-gridpoint-wide stencil is applied in an explicit scheme for biharmonic viscosity. Two operators are applied sequentially in these two subroutines: the second rotated horizontally through 45° with respect to the first.

Horizontal momentum advection (subroutine OCIADJ7)

The HOPE model uses a kinetic energy- and enstrophy-conserving formulation of horizontal momentum advection — the finite-difference '$\mathbb{J}_7$ Jacobian'[3] of Arakawa and Lamb (1977). The procedure involves multiple averaging of volume fluxes to obtain fluxes at each of the eight (scalar and vector) gridpoints surrounding the central vector gridpoint. The scheme is explicit in time.

Tracer advection (subroutine OCADUP or OCADV)

HOPE offers the choice of two three-dimensional tracer advection schemes: an explicit upwind scheme or a centred-in-space predictor-corrector scheme. It was expected initially that the upwind scheme may perform better in strong advection regions. A preliminary trial of both in a coarse (1°) model, however, proved the predictor-corrector much better at preserving watermass properties. Figure 2.4 compares salinity sections through 150.2°E for the two schemes three years after initialisation from climatology. Contours of 34.5 psu and 34.72 psu approximately delineate Antarctic Intermediate Water (AAIW) and Circumpolar Deep Water (CDW) respectively. Thus the predictor-corrector scheme was used for this study.

[3]If the independent variables in the two-dimensional advection equation for an incompressible fluid are changed from velocity $\boldsymbol{u}$ to vorticity, $\zeta = \nabla \times \boldsymbol{u}$, and streamfunction, $\boldsymbol{u} = \boldsymbol{k} \times \nabla\psi$, then the advective terms become a Jacobian operator in the resultant *vorticity equation*:

$$\frac{\partial \boldsymbol{u}}{\partial t} = \boldsymbol{u} \cdot \frac{\partial \boldsymbol{u}}{\partial x} \quad \Longrightarrow \quad \frac{\partial \zeta}{\partial t} = J(\zeta, \psi).$$

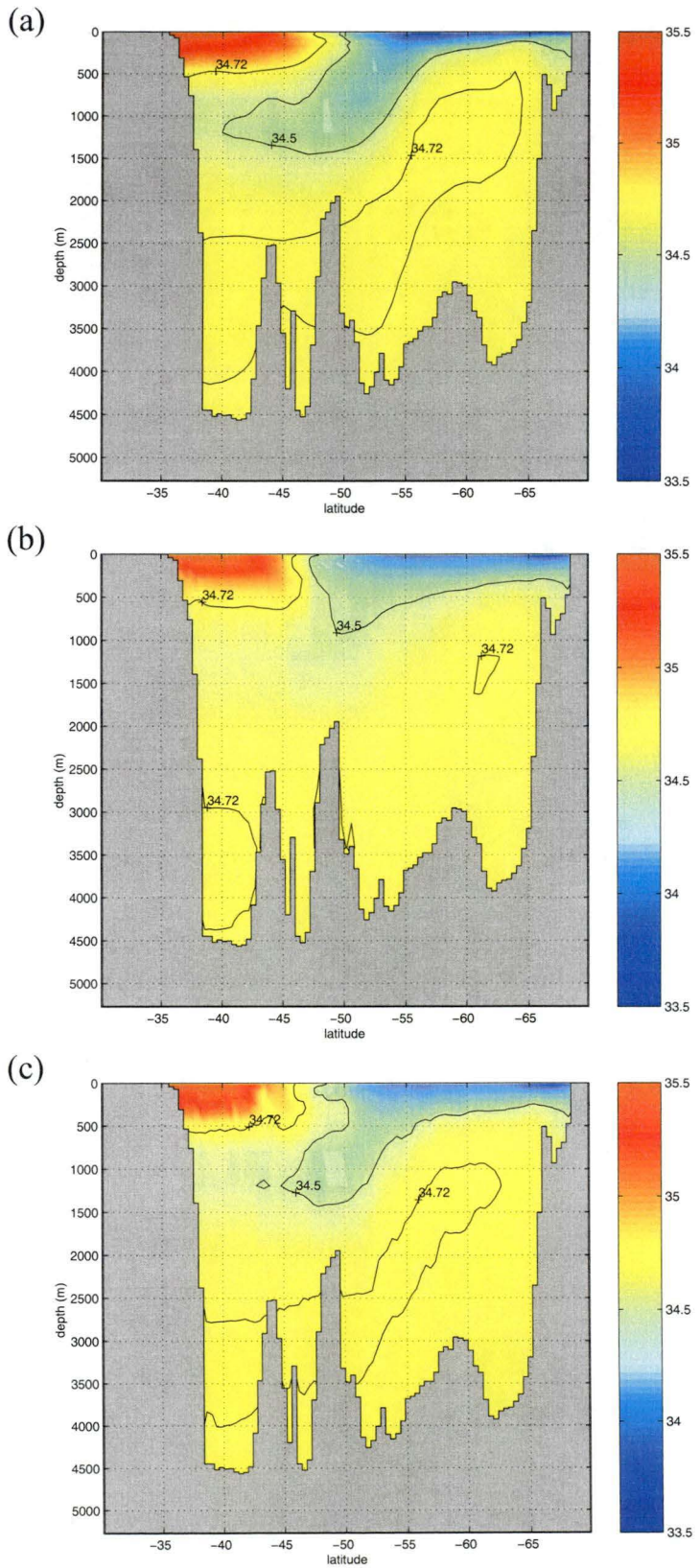


Figure 2.4: Comparison of HOPE's tracer advection schemes for salinity — sections along 150.2°E after 3 years. (a) Climatology, (b) upwind, (c) predictor-corrector.

Horizontal diffusion (subroutine OCADUP or OCADV)

Horizontal diffusion is a centred-in-space explicit scheme.

Vertical diffusion (subroutine OCVERDI)

As with vertical momentum advection and viscosity, vertical diffusion is solved with a centred-in-space implicit scheme.

Internal pressure (subroutine OCTHER)

Integrating the hydrostatic relation (2.2) downwards from the sea surface, the pressure at some depth z may be represented as the sum of two components — a surface (or *external*) pressure p_0, and an *internal pressure*, p':

$$p(z) = p_o + p'(z) = \rho_0 g\eta + \int_z^0 \rho(z')gdz'. \tag{2.27}$$

(In fact, this should include an additional term for the atmospheric pressure, p_a at the sea surface, but atmospheric loading is not included in the HOPE model.) The internal pressure p' is calculated from density using the UNESCO nonlinear equation of state for seawater (Fofonoff and Millard 1983).

Barotropic/baroclinic partitioning (subroutine OCMODS)

The conventional partitioning of horizontal velocities into barotropic, $\boldsymbol{U} = \{U, V\}$, and baroclinic, $\boldsymbol{u}' = \{u', v'\}$, components is made in HOPE by vertically integrating through the water column:

$$\boldsymbol{U} = \int_{-H}^0 \boldsymbol{u}_h dz \tag{2.28}$$

$$\boldsymbol{u}' = \boldsymbol{u}_h - \frac{\boldsymbol{U}}{H} \tag{2.29}$$

Then substituting this partitioning of velocity (2.28), (2.29) and pressure (2.27) into the horizontal momentum equations (2.3), (2.4), and using also the linearized kinematic boundary condition (2.24), prognostic equations are obtained for the barotropic and baroclinic velocities:

$$\frac{\partial U}{\partial t} - fV = -gH\frac{\partial \eta}{\partial x} - \frac{1}{\rho_0}\int_{-H}^0 \frac{\partial p'}{\partial x}dz + G_U \tag{2.30}$$

$$\frac{\partial V}{\partial t} + fU = -gH\frac{\partial \eta}{\partial y} - \frac{1}{\rho_0}\int_{-H}^0 \frac{\partial p'}{\partial y}dz + G_V \tag{2.31}$$

$$\frac{\partial u'}{\partial t} - fv' = -\frac{1}{\rho_0}\left(\frac{\partial p'}{\partial x} - \frac{1}{H}\int_{-H}^0 \frac{\partial p'}{\partial x}dz\right) + G'_U \tag{2.32}$$

$$\frac{\partial v'}{\partial t} + fu' = -\frac{1}{\rho_0}\left(\frac{\partial p'}{\partial y} - \frac{1}{H}\int_{-H}^0 \frac{\partial p'}{\partial y}dz\right) + G'_V. \tag{2.33}$$

$G_{\{U,V\}}$ and $G_{\{U',V'\}}$ are nonlinear and mixing terms for the barotropic and baroclinic velocities respectively. Since these terms are solved separately as described earlier, it only remains to outline HOPE's method for solving the homogeneous barotropic and baroclinic momentum equations.

Baroclinic solution (subroutine OCCLIT)

An implicit differencing scheme is used to solve the baroclinic equations (2.32) and (2.33):

$$\begin{aligned} u'^{n+1} &= u'^{n} + \alpha\Delta t\left[fv'^{n+1} - \frac{1}{\rho_0}\left(\frac{\partial p'^{n+1}}{\partial x} - \frac{1}{H}\int_{-H}^{0}\frac{\partial p'^{n+1}}{\partial x}dz\right)\right] \\ &\quad + (1-\alpha)\Delta t\left[fv'^{n} - \frac{1}{\rho_0}\left(\frac{\partial p'^{n}}{\partial x} - \frac{1}{H}\int_{-H}^{0}\frac{\partial p'^{n}}{\partial x}dz\right)\right] \end{aligned} \tag{2.34}$$

$$\begin{aligned} v'^{n+1} &= v'^{n} + \alpha\Delta t\left[-fu'^{n+1} - \frac{1}{\rho_0}\left(\frac{\partial p'^{n+1}}{\partial y} - \frac{1}{H}\int_{-H}^{0}\frac{\partial p'^{n+1}}{\partial y}dz\right)\right] \\ &\quad + (1-\alpha)\Delta t\left[-fu'^{n} - \frac{1}{\rho_0}\left(\frac{\partial p'^{n}}{\partial y} - \frac{1}{H}\int_{-H}^{0}\frac{\partial p'^{n}}{\partial y}dz\right)\right] \end{aligned} \tag{2.35}$$

The *centering parameter* α determines the degree of 'implicitness' in the solution (a value of one makes the scheme fully implicit, a value of zero is explicit). Thus it is regarded as a stability parameter. Since the pressure gradients are needed at the new time-level $n+1$, an evolution equation for pressure is required. This is obtained in HOPE by first making an assumption that disturbances are small; then density changes may be approximated by

$$\frac{\partial\rho}{\partial t} = -w\frac{\partial\rho}{\partial z}.$$

Combining this with the time derivative of the hydrostatic equation (2.2), the following evolution equation is obtained for internal pressure:

$$\frac{\partial^2 p'}{\partial t\partial z} = gw\frac{\partial\rho}{\partial z}.$$

This is also discretised implicitly in time with another stability parameter β:

$$\left(\frac{\partial p'}{\partial z}\right)^{n+1} = \left(\frac{\partial p'}{\partial z}\right)^{n} + \Delta t g\frac{\partial\rho}{\partial z}\left[\beta w^{n+1} + (1-\beta)w^{n}\right]. \tag{2.36}$$

To complete HOPE's formulation of the baroclinic system, an equation for w at the new time level is required. This is obtained in HOPE by vertically integrating the continuity equation (2.8) for the baroclinic velocities:

$$w^{n+1} = -\int_{-H}^{z}\nabla_h.\boldsymbol{u}'^{n+1}dz'. \tag{2.37}$$

While this is approximate, the contribution to w from the divergence of the barotropic velocities is expected to be small after initial geostrophic adjustment.

To solve this implicit system, HOPE uses an iterative technique. The following steps are repeated until a convergence criterion is satisfied:

1. $\boldsymbol{u}'^{n+1}$ is estimated by solving jointly equations (2.34) and (2.35) (taking $p'^{n+1} = p'^{n}$ for the first iteration)
2. w^{n+1} is estimated from (2.37) using the new estimate of $\boldsymbol{u}'^{n+1}$
3. p'^{n+1} is estimated by integrating up equation (2.36) using the new estimate of w^{n+1}

A maximum limit of 10 iterations is enforced; typically seven were needed. If the limit was reached it was indicative of other problems with the model setup. The baroclinic system accounted for approximately half the CPU time at each time step.

Barotropic solution (subroutines OCRHSZ, ZGAUSS, OCVTRO)

Like the baroclinic system, an implicit differencing scheme is used to solve the barotropic equations (2.30) and (2.31):

$$\begin{aligned} U^{n+1} &= U^n + \alpha\Delta t\left[fV^{n+1} - gH\frac{\partial \eta}{\partial x}^{n+1} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial x}^{n+1} dz\right] \\ &\quad + (1-\alpha)\Delta t\left[fV^n - gH\frac{\partial \eta}{\partial x}^{n} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial x}^{n} dz\right] \end{aligned} \tag{2.38}$$

$$\begin{aligned} V^{n+1} &= V^n + \alpha\Delta t\left[-fU^{n+1} - gH\frac{\partial \eta}{\partial y}^{n+1} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial y}^{n+1} dz\right] \\ &\quad + (1-\alpha)\Delta t\left[-fU^n - gH\frac{\partial \eta}{\partial y}^{n} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial y}^{n} dz\right] \end{aligned} \tag{2.39}$$

The stability parameter α is identical to that used in the baroclinic system. The internal pressure at the new time level, p'^{n+1}, is available as it has already been calculated in the baroclinic system. To complete the barotropic system, an equation for sea-level η at the new time level $n+1$ is needed. This comes from the linearized kinematic boundary condition (2.24) and the definition of the barotropic velocities (2.28):

$$\frac{\partial \eta}{\partial t} = -\nabla_h . \boldsymbol{U}.$$

This is discretised implicitly with the same stability parameter β used for the pressure term in the baroclinic system (2.36):

$$\eta^{n+1} = \eta^n - \beta\Delta t\left(\frac{\partial U}{\partial x}^{n+1} + \frac{\partial V}{\partial y}^{n+1}\right) - (1-\beta)\Delta t\left(\frac{\partial U}{\partial x}^{n} + \frac{\partial V}{\partial y}^{n}\right). \tag{2.40}$$

Thus, the prognostic variables in the barotropic system are the barotropic velocities and sea-level. Since these are two-dimensional (rather than three-dimensional like the prognostic quantities in the baroclinic system), it becomes computationally feasible to solve the system directly rather than with an iterative method. To do this, the barotropic velocities at the new time-level $\{U^{n+1}, V^{n+1}\}$ are first eliminated from the system equations (2.38), (2.39) and

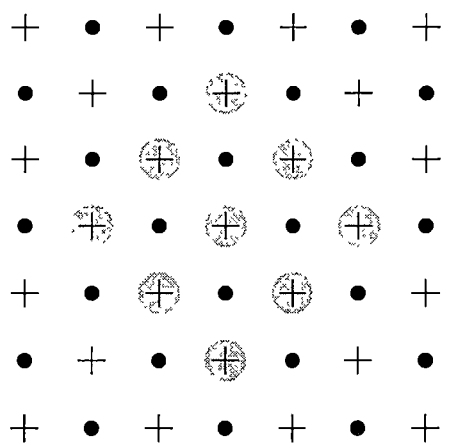


Figure 2.5: Finite-difference stencil used for the barotropic system.

(2.40). This results in an implicit elliptic equation for η^{n+1} which can be written symbolically as:

$$\eta^{n+1} + \Gamma^{n+1} = \Gamma^n, \tag{2.41}$$

where

$$\begin{aligned} \Gamma^{n+1} &= \Gamma^{n+1}(\eta_{xx}, \eta_{xy}, \eta_{yy}) \\ \Gamma^n &= \Gamma^n(\eta, U_x, U_y, V_x, V_y, \eta_{xx}, \eta_{xy}, \eta_{yy}, p'_{xx}, p'_{xy}, p'_{yy}). \end{aligned}$$

The subscripts represent spatial derivatives. The sea-level in Γ^{n+1} is defined at the new time level, while the quantities in Γ^n (with the exception of the already calculated internal pressure p') are defined at the old time level n. Thus, taking the set of sea-level points together, the sea-level equation (2.41) may be written in matrix form:

$$\boldsymbol{A}\underline{\eta}^{n+1} = \underline{\Gamma}^n. \tag{2.42}$$

In HOPE the system matrix $\boldsymbol{A}$ is pre-calculated, triangularised and stored. The stencil of gridpoints shown in Figure 2.5 is used to calculate centred second-order differences in the system matrix. For each element of $\underline{\eta}^{n+1}$, only the eight adjacent surrounding sea-level gridpoints contribute non-zero coefficients in $\boldsymbol{A}$. A systematic ordering of sea-level gridpoints within $\underline{\eta}^{n+1}$ along one of the horizontal grid directions will produce a fixed-width band in the matrix, symmetric about the diagonal, beyond which all elements are zero. In HOPE, only this diagonal band of entries is stored to reduce the computer memory burden. (The triangularisation, with pivots, requires the same storage.)

The steps used by HOPE to solve the barotropic system each timestep may be summarised as follows:

1. calculation of the right-hand-side of the matrix equation (2.42) (subroutine `OCRHSZ`)

2. direct solution of (2.42) for the new sea-level $\underline{\eta}^{n+1}$ using backsubstitution (subroutine `ZGAUSS`)

3. calculation of the new barotropic velocities using equations (2.38) and (2.39)

The combination of the linearized kinematic boundary condition (2.24) and the implicit formulation of the barotropic system enables HOPE to use a barotropic timestep considerably longer than the Courant-Friedrichs-Lewy (CFL) stability criterion would otherwise allow. Indeed there is no separate timestep for the barotropic system; all model physics is solved each time step. This is in contrast, for instance, with both the free-surface Princeton Ocean Model (Blumberg and Mellor 1987) and the explicit free-surface adaptation of Killworth et al. (1991) for the GFDL model where the barotropic system is stepped with many small time increments for each timestep of the complete system. (Dukowicz and Smith (1994) have recently developed an implicit free-surface modification for the GFDL model.) The penalty for this time-stepping advantage is a less accurate representation of shallow-water gravity waves. Distorting their dispersion properties is not expected to cause problems. Rigid lid models, for instance, filter out completely these fast motions. The more relevant barotropic Rossby waves have timescales considerably longer than the model timestep and will not significantly be affected.

New total velocities (subroutine `OCVTOT`)

After the new barotropic and baroclinic velocities have been calculated, they are combined to produce new total horizontal velocities:

$$\boldsymbol{u}_h = \boldsymbol{u}' + \frac{1}{H}\boldsymbol{U}. \tag{2.43}$$

2.2 Southern Ocean model configuration

The configuration of the HOPE model for the limited-area Southern Ocean domain used in this study is now described.

2.2.1 Bathymetry and grid geometry

The model bathymetry is taken from the ETOPO5 data set (NGDC 1988). It is shown in Figure 2.6 with the major topographic features labelled. Depths greater than 6500m are clipped in HOPE; this occurs along the Kermadec Trench northeast of New Zealand. There is no particular reason for this clipping, but it was already built into the version of code the author used. The model domain extends meridionally from 30°S to 70°S and zonally from 110°E to 190°E. This domain was chosen to include eddy-rich locations over the Southeast Indian Ridge and south of the Campbell Plateau. The location of the northern boundary permitted the East Australia Current to enter the domain. All boundaries are open except for the southern boundary where an artificial wall has been extended eastwards from the Antarctic continental landmass, closing off the westernmost part of the entrance to the Ross Sea. The impact on the model of this artificial wall is unlikely to be any worse than

Table 2.2: Model grid geometry

grid parameter	*value(s)*
zonal gridpoints	133
zonal resolution	0.6°(per subgrid)
meridional gridpoints	100
meridional resolution	0.4°(per subgrid)
number of layers	20
layer thicknesses (bottom depths), m	20.7 (20.7), 23.3 (44), 26.5 (70.5), 31 (101.5), 37.3 (138.8), 46.7 (185.5), 61.6 (247.1), 85.9 (333), 277 (610), 375 (985), 416 (1401), 434 (1835), 444 (2279), 451 (2730), 455 (3185), 459 (3644), 461 (4105), 463 (4568), 465 (5033), 466 (5499)

Table 2.3: Model surface forcing fields

field	*dataset/reference*
monthly surface windstress, $\boldsymbol{\tau}(t)$	Hellerman and Rosenstein (1983)
annual mean surface salinity, S_o	Olbers et al. (1992)
monthly mean sea-surface temperature, $\theta_o(t)$	Levitus and Boyer (1994)

that due to the uncertainty of tracer and flow fields in the region, which would be needed to force an open boundary here.

Table 2.2 lists the Southern Ocean model grid parameters. A resolution of 0.6° in longitude and 0.4° in latitude was used for each (EVEN and ODD) subgrid. This produces approximately square grid cells at 50°S, 44 km per side, for an effective resolution ('d' in Figure 2.2) of around 31 km. A fully-implicit discretisation in time ($\alpha = 1$, $\beta = 1$) was used throughtout this thesis for both the baroclinic and barotropic systems in order to allow a long timestep.

The layer thicknesses in the vertical were based on those used in the FRAM model (The FRAM Group 1991; Webb et al. 1991). The uppermost eight layers are identical, with each successive pair of the remaining 24 in FRAM combined into a single layer here, for a total of 20 layers.

2.2.2 Forcing fields

Table 2.3 lists the surface forcing data used in this study.

2.2.3 Model parameters

Table 2.4 lists the model parameters used. Both vertical diffusion and viscosity were constant in the configuration used here — the Richardson-number

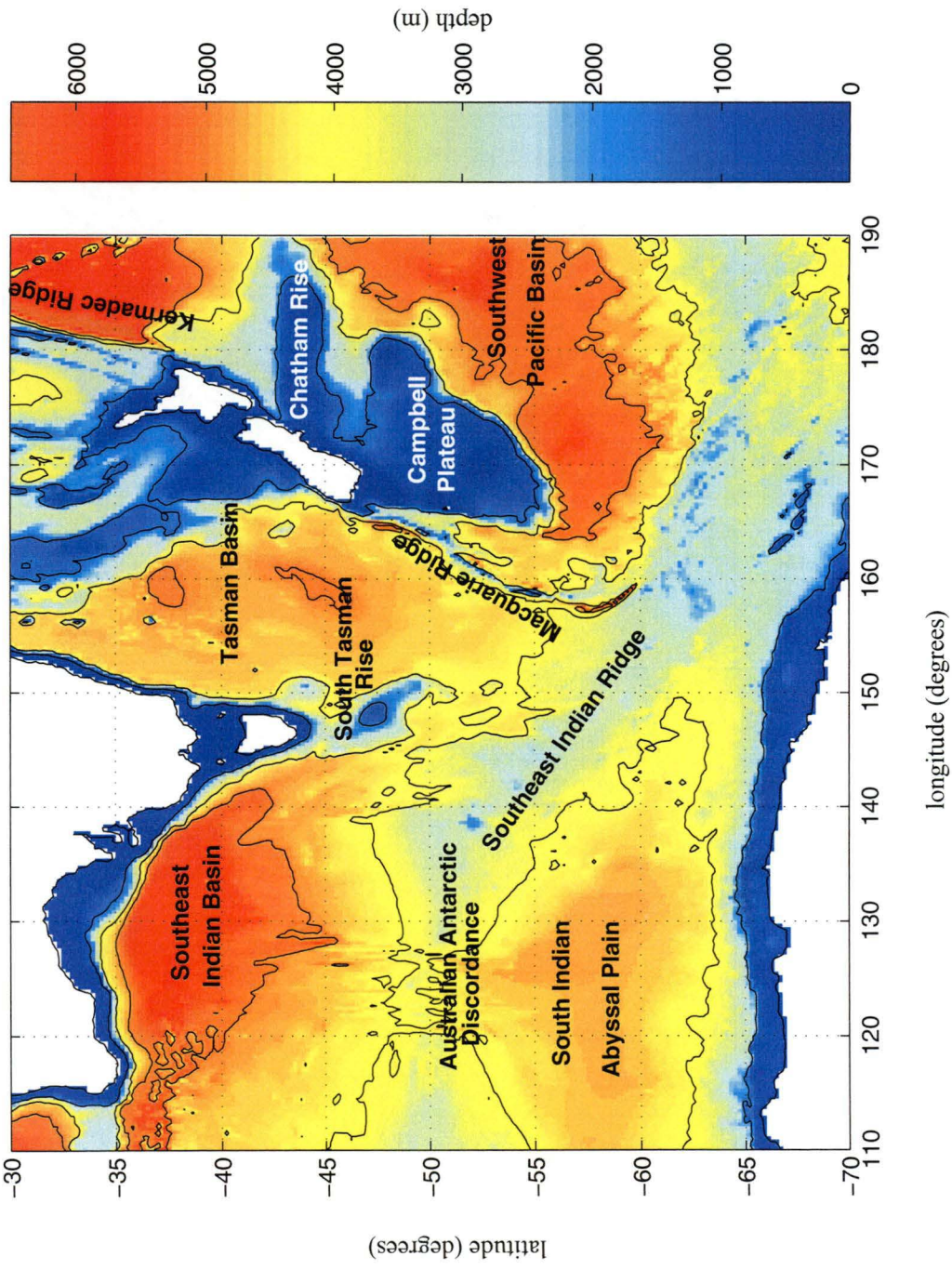


Figure 2.6: Model bathymetry. Solid contours represent 1500 m, 4000 m and 5000 m isobaths.

Table 2.4: Model parameters

parameter	*symbol*	*value*
horizontal viscosity: harmonic	A_H	300 m^2s^{-1}
biharmonic	B_H	0 m^4s^{-1}
strain-dependent	ν_A	0.5 m^2s
vertical viscosity	A_V	1.5 cm^2s^{-1} (constant)
horizontal diffusivity: harmonic	D_H	150 m^2s^{-1}
strain-dependent	ν_D	0 m^2s
vertical diffusivity	D_V	0.7 cm^2s^{-1} (constant)
surface relaxation timescale	$\{\lambda_{S,\theta}\}^{-1}$	30 days
bottom friction	ε	10^{-7} s^{-1}
timestep	Δt	3600 s
implicitness centering parameter	α	1
implicitness centering parameter	β	1

dependent parametrizations were not used[4]. Biharmonic viscosity was not used either, since it preferentially damps smaller scales (Griffies and Hallberg 2000) and these were not found to cause instabilities here.

The numerical values for mixing parameters were chosen after considering those used in a number of other models of similar resolution. For comparison, the values used in FRAM (this study) were $100(200)m^2s^{-1}$ for horizontal viscosity, $1(1.5)cm^2s^{-1}$ for vertical viscosity, $100(150)m^2s^{-1}$ for horizontal diffusion, and $0.5(0.7)cm^2s^{-1}$ for vertical diffusion.

2.2.4 Model initialisation

Model runs starting from an ocean at rest were initialised with climatological values of temperature and salinity taken from the Olbers et al. (1992) climatology.

2.2.5 Computational aspects

The HOPE model has been written to take advantage of vector processing architectures of modern supercomputers. All model runs were performed on the University of Tasmania's CRAY SV1 supercomputing facility. Postprocessing was performed on SUN and SGI desktop servers.

The assimilation runs reported in chapter 5 used around 24 hours of CPU time per model year and required around 50 MW of core memory. With

[4]This scheme was devised by Pacanowski and Philander (1981) specifically for the equatorial ocean where mixing below the thermocline is stronger than in other areas, despite small vertical temperature gradients. This thesis was not concerned with tropical mixing.

a limited amount of parallel optimisation, and in a multi-user environment, one year assimilation runs generally executed in around half a day real time.

CHAPTER 3

Open boundary modifications

In order to resolve the dynamics of interest in the Southern Ocean south of Australia, an eddy-resolving resolution is preferred. The first baroclinic Rossby radius ranges from roughly 30 km at 40°S to 10 km at 60°S (Chelton et al. 1998). The chosen resolution (effectively 31 km at 50°S) was a compromise between this ideal and computational affordability, and is comparable to the resolution of FRAM (35 km × 27.5 km at 50°S). To achieve this resolution, the model had to be limited to a regional domain. Extending the grid zonally around the hemisphere at the same resolution would have more than quadrupled computer memory requirements and execution time; even linearly telescoping a hemispheric grid to a resolution of 4° zonally would have doubled the required resources. However, limiting the model to an open domain required substantial modifications to the HOPE model as it was designed only for closed or zonally periodic domains. The purpose of this chapter is to describe the substantial modifications made to the HOPE model to enable it to be used in an open boundary configuration.

3.1 Theoretical foundations

> "The determination of valid and convenient forms of boundary conditions, particularly at points of outflow, constitutes a major, essentially unresolved, problem in the modeling of many hydrodynamic systems over regional domains." *(Haidvogel et al. 1980)*
>
> "The need to prescribe boundary conditions on open lateral boundaries is the Achilles heel of regional models, whether they are for the ocean, the atmosphere, or coupled." *(Kantha and Clayson 2000)*
>
> "Failures are sometimes ascribed to such mysterious causes as 'nonlinear instability' (a term which is meaningless for lack of a definition and sounds very much like the *Hic sunt leones* label attached to the unexplored land in middle-age maps). In this paper, I will try to show that in mixed initial- and boundary-value problems major problems arise if the boundary conditions are not properly handled." *(Moretti 1969)*

> "If a boundary condition is stable then in most cases it will probably be accurate enough ..." *(Miller and Thorpe 1981)*

The above quotes serve to demonstrate that open boundary ocean modelling is certainly not a mature field. Little attention appears to have been paid to the theoretical foundations of the problem by modelling practitioners. The reasons for this are clear enough. First, the majority of the theoretical contributions have been of a rather abstract and obtuse nature. The ocean modeller, however, requires pragmatic solutions. Second, the link between theoretical well-posedness proofs on the one hand, and their implications for specific model implementations on the other, is often difficult to see. For instance, Kreiss and others (Kreiss 1971; Gustafsson et al. 1972; Gottlieb et al. 1982; Trefethen 1983) have developed sophisticated theory for examining the stability of boundary conditions under specific finite-difference implementations. The purpose of this section is not to attempt to bridge the gap between theory and practice, but to indicate the types of considerations that ought to be borne in mind by the ocean modeller confronting an open boundary problem. As should be expected, appropriate boundary conditions depend heavily on the form of the underlying partial differential equations. The discussion here will focus on theoretical results. Later sections dealing with the implementation of open boundaries in HOPE, will refer back to this discussion as needed. Useful texts elaborating these ideas in greater detail are those of Garabedian (1964), Courant and Friedrichs (1976), Sod (1985), Kreiss and Lorenz (1989) and Gustafsson et al. (1995). In addition, see the paper by Oliger and Sundström (1978). Numerous papers covering both theoretical and practical issues are collated in the proceedings of a 1981 NASA symposium on numerical boundary condition procedures (NASA 1981).

3.1.1 Nature of the problem

Any ocean modeller would be aware that a given set of partial differential equations has, in general, an infinite set of solutions, and that it is only by constraining the set through suitable boundary conditions (including initial conditions for time-dependent problems) that a unique solution may be obtained. Thus the proper formulation of such boundary conditions is fundamentally important. The subject becomes particularly relevant when an artificial computational boundary is imposed in a numerical scheme.

Just as there is considerable flexibility in formulating the finite-difference procedures in the interior, there is, in principle, great freedom in choosing the boundary numerics. Consider, for instance, the linearized one-dimensional shallow water equations:

$$\begin{aligned} \frac{\partial u}{\partial t} + g\frac{\partial \eta}{\partial x} &= 0 \\ \frac{\partial \eta}{\partial t} + \frac{\partial (Hu)}{\partial x} &= 0. \end{aligned} \tag{3.1}$$

Suppose a finite-difference scheme were used to solve these equations. One

such scheme might use centered differences for the spatial derivatives:

$$\left(\frac{\partial\{Hu,\eta\}}{\partial x}\right)_i \approx \frac{\{Hu,\eta\}_{i+1} - \{Hu,\eta\}_{i-1}}{2\Delta x}.$$

At the endpoints of the finite-difference grid, however, such a scheme is not possible. One might instead apply a one-sided difference, *e.g.*

$$\left(\frac{\partial\{Hu,\eta\}}{\partial x}\right)_i \approx \frac{\{Hu,\eta\}_{i+1} - \{Hu,\eta\}_i}{\Delta x}$$

for the left-most boundary. In this manner, one may imagine that no boundary conditions are needed at all — the equations themselves can be discretised and solved for u and η at the boundaries (albeit using a slightly different discretisation to the interior). On the other hand, one could just as easily *prescribe* values of u and η at the boundaries. Indeed this could be done quite arbitrarily, without any reference to the equations. Such "freedom" exists also in HOPE's barotropic system (equations (2.38), (2.39) and (2.40)), where it certainly is possible, for instance, to derive approximations to the spatial derivatives at the boundaries. It is also possible to devise schemes in HOPE that enable one (naïvely) to prescribe boundary values of the prognostic variables U, V and η independently and arbitrarily at each timestep.

Faced with such (apparent) freedom at artificial boundaries, the obvious question is: what *should* one do? Or, more incisively: what do the underlying equations *require* one to do? This is a question of mathematics, not engineering, and one would hope the answer would be largely independent of any specific discretisation, as indeed it will be seen to be. One might use physical intuition to suggest an answer. Surely one ought to be allowed, for instance, to prescribe u at the boundaries for the one-dimensional shallow-water model. Physical intuition, however, is no proof. And intuition becomes progressively less reliable as complexity grows. On physical grounds, it is difficult to decide whether one can likewise prescribe both u *and* η at the boundaries. Remarkably, such considerations have received scant attention in the literature by open ocean modellers. Commenting on a similar phenomenon in the field of compressible fluid dynamics, Moretti (1981) wrote:

> "... many authors have shown a total lack of constructive curiosity, in not inquiring whether numerical methods, which should interpret the physics of a problem, really require more boundary conditions than the physics itself. The answer to the question, of course, should be an unqualified No."

Wang and Halpern (1970) used a coarse mesh hemispheric model to obtain boundary conditions for a limited-area barotropic model, prescribing all variables along the boundaries of the fine-mesh model. Small-scale oscillations in their results are due to overspecification of the boundary conditions (Elvius and Sundström 1973; Camerlengo and O'Brien 1980) and not numerical instabilities as suggested by Shapiro and O'Brien (1970). Wang and Jungclaus (1996) undertook a study of various open boundary conditions using the Princeton

Ocean Model (POM) and stated "(f)or the barotropic flow, OBC's are needed for three variables", which they proceeded to implement independently. This is inappropriate from an appreciation of the relevant theory discussed in the next section. Blumberg and Kantha (1985) developed a radiating boundary condition for tidally forced flows. They claimed (without demonstration) that the boundary condition could be applied to any one of the prognostic variables U, V, or η and that "(t)hese various formulations are equivalent as long as the equations are linearized on the open boundaries". Again, application of the theory of characteristics described below shows this claim to be false. In a recent paper by Beckers (1999), an incorrect application of boundary conditions was found to be the real explanation for anomalous stability behaviour of a finite-difference scheme examined by Deleersnijder and Campin (1993).

Assuming for the moment that one has an answer to the mathematical question of *well-posed* boundary conditions for a given set of partial differential equations, a second issue which arises is the *stability* of their finite-difference implementation. Consider, for example, the simple one-dimensional advection equation:

$$\frac{\partial s}{\partial t} = c\frac{\partial s}{\partial x}, \tag{3.2}$$

where c represents velocity. The general solution is a wave travelling with velocity c in the negative x direction, $s(x,t) = f(x+ct)$. Suppose now the well-known *leap-frog* scheme (Press et al. 1992) is used to solve this on a finite-difference grid:

$$\begin{aligned} \frac{s_j^{n+1} - s_j^{n-1}}{2\Delta t} &= c\frac{s_{j+1}^n - s_{j-1}^n}{2\Delta x} \\ s_j^{n+1} &= s_j^{n-1} + c\frac{\Delta t}{\Delta x}(s_{j+1}^n - s_{j-1}^n), \end{aligned}$$

where the superscripts are a time index, and the subscripts a spatial index. For the left-most boundary, a one-sided difference could be used:

$$s_j^{n+1} = s_j^{n-1} + 2c\frac{\Delta t}{\Delta x}(s_{j+1}^n - s_j^n),$$

for $j = 0$. It turns out, however, that this boundary scheme (with leap-frog in time) is unstable. Figure 3.1 shows the results of applying this model to a triangular wave centered initially at $x = 45$ and with parameters $c = 1$, $\Delta x = 1$, $\Delta t = 1$. The scheme can be stabilised by using *forward Euler*, instead of leap-frog, differencing in time for the boundaries:

$$s_j^{n+1} = s_j^n + c\frac{\Delta t}{\Delta x}(s_{j+1}^n - s_j^n).$$

In summary, both mathematical well-posedness, and numerical stability are required for reasonable boundary conditions. The following two sections address these issues.

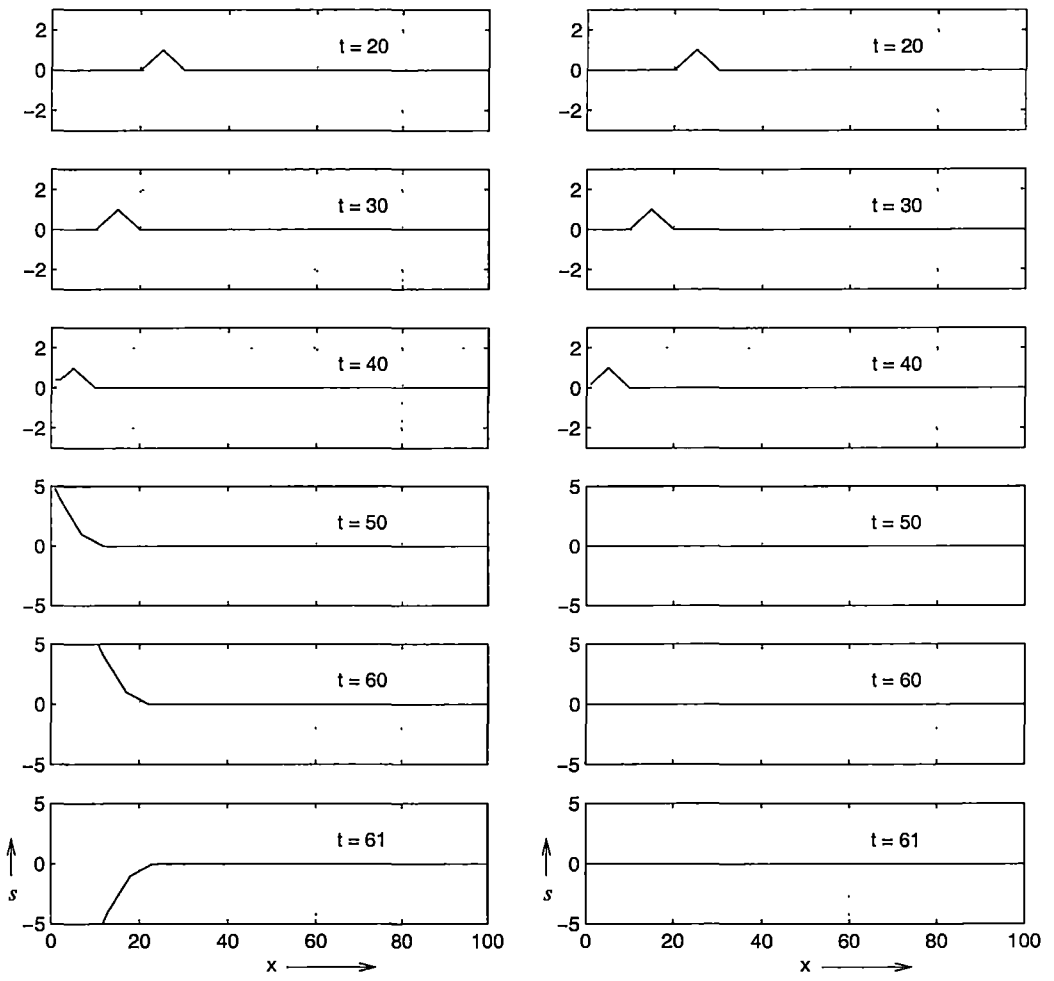


Figure 3.1: Boundary finite-difference schemes: unstable leap-frog scheme (left), and the stable forward Euler scheme (right).

3.1.2 Mathematical well-posedness

As foreshadowed above, the question of mathematical well-posedness for boundary conditions is determined by the partial differential equations themselves. This is discussed in detail by Oliger and Sundström (1978) and Kreiss and Lorenz (1989), and in a rather difficult theoretical paper by Kreiss (1970).

A distinction must first be made between a pure *initial value* or *Cauchy* problem, and an *initial-boundary value* problem. Consider again the advection equation (3.2). If the domain of interest is the real-line $x = (-\infty, \infty)$, then this is a pure initial value (or Cauchy) problem: one is interested in determining $s = s(x,t)$ for all $t > 0$ given initial values $s(x,0) = s_0(x)$. We are interested in determining what form the initial conditions s_0 should take to ensure the problem has a unique, well-behaved solution. If, on the other hand, the domain is limited to, say, $x = [0,1]$ then one is interested in determining the solution subject to the initial conditions and additional boundary conditions $s(0,t) = g_0(t)$ and $s(1,t) = g_1(t)$; the problem is now an initial-boundary value problem. Again, we wish to discover what form should be taken for the boundary conditions g_0 and g_1 to guarantee a well-posed problem.

Well-posed initial and boundary conditions for these respective problems depend on the nature of the equations. A general second-order system of n partial differential equations in r space dimensions may be written in matrix form as:

$$\frac{\partial \boldsymbol{s}}{\partial t} = \sum_{i,j=1}^{r} \frac{\partial}{\partial x_i}\left(\boldsymbol{A}_{ij}\frac{\partial \boldsymbol{s}}{\partial x_j}\right) + \sum_{i=1}^{r} \boldsymbol{B}_i \frac{\partial \boldsymbol{s}}{\partial x_i} + \boldsymbol{C}\boldsymbol{s} + \boldsymbol{f}, \tag{3.3}$$

where $\boldsymbol{s} = (s_1(\boldsymbol{x},t), \ldots, s_n(\boldsymbol{x},t))^{\mathrm{T}}$, $\boldsymbol{x} = (x_1, \ldots, x_r)$, and the real coefficient matrices $\boldsymbol{A}_{ij} = \boldsymbol{A}_{ij}(\boldsymbol{x},t)$, $\boldsymbol{B}_i = \boldsymbol{B}_i(\boldsymbol{x},t)$, $\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{x},t)$, and inhomogeneous forcing function $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{x},t)$ may vary in space and time. If the matrices $\boldsymbol{A}_{ij}$, $\boldsymbol{B}_i$ and $\boldsymbol{C}$ were also dependent on $\boldsymbol{s}$, the system would be nonlinear — we consider here only linear systems, which may be obtained from a nonlinear system by linearizing about a mean state. The above system is sufficiently general to cover the range of physics (suitably linearized) that are modelled in HOPE, with the exception of the biharmonic viscous terms (2.14). Note, however, that these viscous terms were zero in all integrations described in this thesis (see Table 2.4). If all coefficient matrices are diagonal, then the system represents n uncoupled scalar partial differential equations.

The general system (3.3) may be classified as *parabolic* or *hyperbolic* according to the form of the matrices $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_i$. Criteria for well-posed boundary conditions depend on which of these forms the system takes. For present purposes, the system will be said to be parabolic if

$$\boldsymbol{A}_{ij} = \mathrm{diag}(\mu_k),\ \mu_k > 0,\ k = 1, \ldots, n.$$

It will be said to be hyperbolic for $\boldsymbol{A}_{ij} = \boldsymbol{0}$ if $\boldsymbol{B}_i$ is symmetric:

$$\boldsymbol{B}_i^{\mathrm{T}} = \boldsymbol{B}_i.$$

Thus the one-dimensional shallow-water equations (3.1) are seen to be hyper-

bolic (after introducing $\eta' = \sqrt{g/H}\eta$), while the diffusion equation

$$\frac{\partial s}{\partial t} = k^2 \nabla^2 s$$

is parabolic. These definitions are much stricter than required for the well-posedness proofs in Kreiss and Lorenz (1989) that will be cited here, but they are sufficient for this discussion and avoid technical complications arising from greater generality.

Some key results on well-posedness of various systems of equations will now be summarised.

Cauchy problems

The Cauchy problems look for solutions of (3.3) subject to initial conditions

$$\boldsymbol{s}(\boldsymbol{x}, 0) = \boldsymbol{h}(\boldsymbol{x}).$$

If the coefficients, $\boldsymbol{A}_{ij}$, $\boldsymbol{B}_i$, $\boldsymbol{C}$, forcing functions, $\boldsymbol{f}$, and initial conditions, $\boldsymbol{h}$, are periodic in each space dimension, then the Cauchy problems for parabolic and hyperbolic systems are well-posed and have unique solutions (Kreiss and Lorenz 1989, Theorems 6.1.1, 6.2.2).

Initial-boundary value problems

Consider first the parabolic system, but now restricted to the region $0 \leq x_1 \leq 1$, $t \geq 0$. As above, periodicity in the other space dimensions is assumed. Initial conditions are given:

$$\boldsymbol{s}(\boldsymbol{x}, 0) = \boldsymbol{h}(\boldsymbol{x}).$$

Then the resultant initial-boundary value problem is well-posed under the following general inhomogeneous boundary conditions (Kreiss and Lorenz 1989, Theorem 8.1.3):

$$\boldsymbol{L}_{D_i} \boldsymbol{s}(x_i, \boldsymbol{x}_*, t) + \boldsymbol{L}_{N_i} \frac{\partial \boldsymbol{s}}{\partial x_1}(x_i, \boldsymbol{x}_*, t) = \boldsymbol{g}_i(\boldsymbol{x}_*, t), \; x_i = 0, 1, \tag{3.4}$$

where $\boldsymbol{x}_*$ has been used to represent $(x_2, \ldots, x_r)$ and the coefficient vectors, $\boldsymbol{L}_{D_i} = \boldsymbol{L}_{D_i}(\boldsymbol{x}_*, t)$, and $\boldsymbol{L}_{N_i} = \boldsymbol{L}_{N_i}(\boldsymbol{x}_*, t)$ may vary in space and time. Some combination of the variables, $\boldsymbol{s}$, and their derivatives normal to the boundaries, $\partial \boldsymbol{s}/\partial x_1$, must be specified on the boundaries. The special case

$$\boldsymbol{L}_{D_i} \boldsymbol{s} = \boldsymbol{g}_i$$

represents pure *Dirichlet* conditions, while pure *Neumann* conditions are given by

$$\boldsymbol{L}_{N_i} \frac{\partial \boldsymbol{s}}{\partial x_1} = \boldsymbol{g}_i.$$

An additional requirement for smooth solutions is that *compatibility conditions* must be satisfied by the initial and boundary data, $\boldsymbol{h}, \boldsymbol{g}_i$ and the forcing function $\boldsymbol{f}$. For instance, the following relations must hold at initial time:

$$\boldsymbol{L}_{D_i}\boldsymbol{h}(x_i, \boldsymbol{x}_*) + \boldsymbol{L}_{N_i}\frac{\partial \boldsymbol{h}}{\partial x_1}(x_i, \boldsymbol{x}_*) = \boldsymbol{g}_i(\boldsymbol{x}_*, 0).$$

A sufficient condition is that $\boldsymbol{h}, \boldsymbol{g}_i$ and $\boldsymbol{f}$ vanish in some neighbourhood of $(0, \boldsymbol{x}_*)$ and $(1, \boldsymbol{x}_*)$ at $t = 0$.

For the hyperbolic system in the region $0 \leq x_1 \leq 1$, the system can be diagonalised along the dimension x_1, since by definition $\boldsymbol{B}_1$ is symmetric:

$$\boldsymbol{P}^{-1}\boldsymbol{B}_1\boldsymbol{P} = \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

where $\lambda_i = \lambda_i(\boldsymbol{x}_*, t)$ are the (real) eigenvalues of $\boldsymbol{B}_1$. The columns of the transformation matrix $\boldsymbol{P}$ are the eigenvectors of $\boldsymbol{B}_1$. Transforming also the independent variables $\boldsymbol{s}$, one obtains the *characteristic variables*

$$\boldsymbol{s}'(x, t) = \boldsymbol{P}^{-1}\boldsymbol{s} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{s}. \tag{3.5}$$

Thus, the characteristic variables are the original variables transformed by the eigenvectors of $\boldsymbol{B}_1$.

To motivate the significance of the characteristic variables, consider a hyperbolic system in one dimension with no forcing or zeroth order terms:

$$\frac{\partial \boldsymbol{s}}{\partial t} = \boldsymbol{B}\frac{\partial \boldsymbol{s}}{\partial x}.$$

Making the transformation $\boldsymbol{s}' = \boldsymbol{P}^{-1}\boldsymbol{s}$ (with the columns of $\boldsymbol{P}$ the eigenvectors of $\boldsymbol{B}$), the equations can be written in diagonal form (or *characteristic form*):

$$\frac{\partial \boldsymbol{s}'}{\partial t} = \boldsymbol{\Lambda}\frac{\partial \boldsymbol{s}'}{\partial x}, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

where λ_i are the eigenvalues of $\boldsymbol{B}$. Because $\boldsymbol{\Lambda}$ is diagonal, the above equation is an uncoupled system of n partial differential equations:

$$\frac{\partial s_i'}{\partial t} = \lambda_i(x, t)\frac{\partial s_i'}{\partial x}.$$

For the *total derivatives* of the characteristic variables $\boldsymbol{s}'$, we have:

$$\begin{aligned} \frac{d}{dt}s_i'(x(t), t) &= \frac{\partial s_i'}{\partial t} + \frac{dx}{dt}\frac{\partial s_i'}{\partial x} \\ &= 0 \text{ if } \frac{dx}{dt} = -\lambda_i(x, t). \end{aligned}$$

Thus, s_i' is constant along the *characteristic curve* defined by $\frac{dx}{dt} = -\lambda_i$. If λ_i is constant, the solution may be written $s_i' = h_i(x + \lambda_i t)$ — a wave travelling in the negative (positive) x-direction for positive (negative) λ_i. This is illustrated in Figure 3.2 for the region $x \in [0, 1]$. At each boundary, the characteristic variables may be described as *incoming* or *outgoing* according to their direction

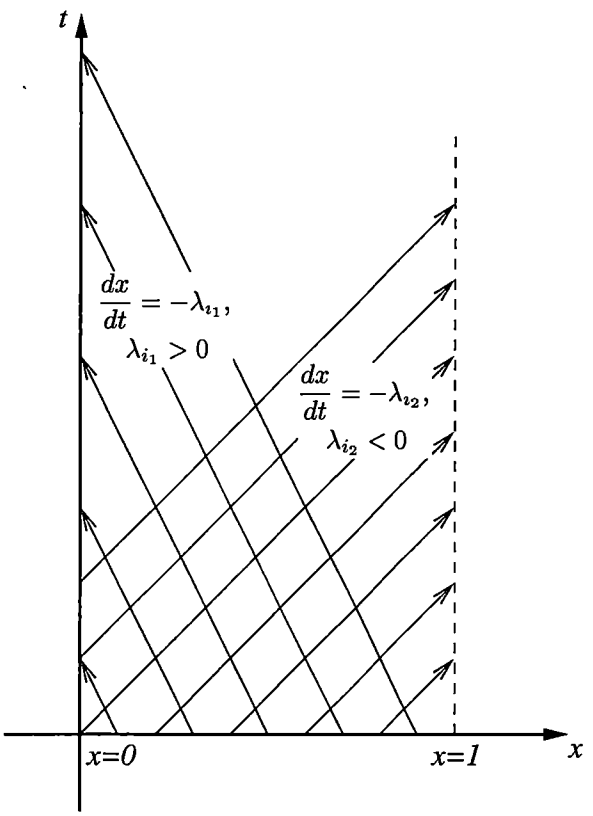


Figure 3.2: Characteristic curves in one dimension. Positive eigenvalue λ_{i_1} corresponds to outgoing characteristics at $x = 0$ and incoming characteristics at $x = 1$; negative eigenvalue λ_{i_2} corresponds to outgoing characteristics at $x = 1$ and incoming characteristics at $x = 0$.

of propagation $\frac{dx}{dt} = -\lambda_i$ relative to the respective boundary. The solution is determined by prescribing values of all characteristic variables initially and those that are incoming at each boundary. Now extend the one-dimensional system by including zeroth-order and forcing terms:

$$\frac{\partial \boldsymbol{s}}{\partial t} = \boldsymbol{B}\frac{\partial \boldsymbol{s}}{\partial x} + \boldsymbol{C}\boldsymbol{s} + \boldsymbol{f}.$$

Then diagonalising as above, we obtain the system

$$\frac{\partial \boldsymbol{s}'}{\partial t} = \boldsymbol{\Lambda}\frac{\partial \boldsymbol{s}'}{\partial x} + \boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{P}\boldsymbol{s}' + \boldsymbol{P}^{-1}\boldsymbol{f},$$

which can be written as a system of *ordinary differential equations*

$$\frac{d\boldsymbol{s}'}{dt} = \boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{P}\boldsymbol{s}' + \boldsymbol{P}^{-1}\boldsymbol{f}$$

along each of the characteristic curves $\frac{dx}{dt} = -\lambda_i$ respectively. Thus, the original system of partial differential equations is transformed to a simpler system of ordinary differential equations along the characteristic curves. This property is exploited in the so-called *method of characteristics* for solving hyperbolic systems (Freeman 1951; Garabedian 1964; O'Brien and Reid 1967; Courant and Friedrichs 1976; Røed and O'Brien 1983).

Having motivated the importance of characteristics, we will now revert to the general hyperbolic system in r dimensions with zeroth order and forcing terms:

$$\frac{\partial \boldsymbol{s}}{\partial t} = \sum_{i=1}^{r} \boldsymbol{B}_i \frac{\partial \boldsymbol{s}}{\partial x_i} + \boldsymbol{C}\boldsymbol{s} + \boldsymbol{f}. \tag{3.6}$$

Suppose that $\boldsymbol{B}_1$ is nonsingular; then diagonalise it as above and partition the characteristic variables according to the signs of the eigenvalues:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_- & 0 \\ 0 & \boldsymbol{\Lambda}_+ \end{pmatrix}, \quad \mathrm{diag}(\boldsymbol{\Lambda}_-) < 0, \ \mathrm{diag}(\boldsymbol{\Lambda}_+) > 0,$$

$$\boldsymbol{s}' = \begin{pmatrix} \boldsymbol{s}'_- \\ \boldsymbol{s}'_+ \end{pmatrix}.$$

As for the parabolic system earlier, initial conditions are given:

$$\boldsymbol{s}(\boldsymbol{x}, 0) = \boldsymbol{h}(\boldsymbol{x}).$$

Assuming as before periodicity and compatibility conditions on initial, boundary and forcing functions, then the hyperbolic initial-boundary value problem is well-posed under the following inhomogeneous boundary conditions (Kreiss

and Lorenz 1989, Theorem 8.2.2)[1]:

$$\begin{aligned} \boldsymbol{s}'_-(0,\boldsymbol{x}_*,t) &= \boldsymbol{Q}_0\boldsymbol{s}'_+(0,\boldsymbol{x}_*,t)+\boldsymbol{g}_0(\boldsymbol{x}_*,t), \\ \boldsymbol{s}'_+(1,\boldsymbol{x}_*,t) &= \boldsymbol{Q}_1\boldsymbol{s}'_-(1,\boldsymbol{x}_*,t)+\boldsymbol{g}_1(\boldsymbol{x}_*,t). \end{aligned} \tag{3.8}$$

At each boundary, the incoming characteristic variables are defined in terms of the outgoing characteristic variables. The problem remains well-posed if $\boldsymbol{B}_1$ is singular due to a single zero eigenvalue (Kreiss and Lorenz 1989, Theorem 8.2.8). Appropriate boundary conditions for a hyperbolic system must provide a number of conditions at each boundary exactly equal to the number of incoming characteristic variables. They must provide values for each of these variables. Values must not be prescribed for either outgoing characteristic variables, or those corresponding to zero eigenvalues. It must be expected that physical accuracy of solutions will be compromised by overprescribing boundary conditions — for example, by applying radiation conditions to all prognostic variables independently at the boundaries.

For the one-dimensional shallow-water equations (3.1) there are two characteristic variables corresponding to a positive and negative eigenvalue of the system matrix. One boundary condition only must be applied at a boundary. The answer to the question posed in section 3.1.1, therefore, is that u and η may not both be prescribed independently at an open boundary.

Characteristic variables have been used to construct nonreflecting and other boundary conditions by Wurtele et al. (1971), Bennet (1976), Hedstrom (1979), Nordström (1995) and Kar and Turco (1995).

An example: the linearized barotropic shallow-water equations

For the sake of definiteness, the above theory is applied here to the example of the nonlinear two-dimensional barotropic shallow-water equations:

$$\begin{aligned} \frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}-fv+g\frac{\partial \eta}{\partial x} &= 0 \\ \frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+fu+g\frac{\partial \eta}{\partial y} &= 0 \\ \frac{\partial \eta}{\partial t}+\frac{\partial[(H+\eta)u]}{\partial x}+\frac{\partial[(H+\eta)v]}{\partial y} &= 0, \end{aligned} \tag{3.9}$$

where (u,v) are the horizontal velocity components, η is sea-level, and H is the water-depth. These equations are first linearized by perturbing about a

[1]The matrices $\boldsymbol{Q}_0=\boldsymbol{Q}_0(\boldsymbol{x}_*,t)$ and $\boldsymbol{Q}_1=\boldsymbol{Q}_1(\boldsymbol{x}_*,t)$ must satisfy an additional technical "smallness" criterion:

$$|\boldsymbol{Q}_0|+|\boldsymbol{Q}_1|\le\kappa, \tag{3.7}$$

where κ depends on $\boldsymbol{B}_1$ in general. The *spectral-norm* of a matrix is defined as

$$|\boldsymbol{Q}|=\max\{|\boldsymbol{Q}\boldsymbol{s}|:|\boldsymbol{s}|=1\}.$$

See Appendix A for an example of how this constraint restricts the allowable boundary conditions.

mean state constant in space and time:

$$\begin{aligned} u(x,y,t) &= \overline{u} + u'(x,y,t) \\ v(x,y,t) &= \overline{v} + v'(x,y,t) \\ \eta(x,y,t) &= \overline{\eta} + \eta'(x,y,t). \end{aligned}$$

Substituting in (3.9) and neglecting terms quadratic in the perturbations, the linearized equations may be written in matrix form as:

$$\begin{aligned} \frac{\partial}{\partial t}\begin{pmatrix} u \\ v \\ \eta \end{pmatrix} = &- \begin{pmatrix} \overline{u} & 0 & g \\ 0 & \overline{u} & 0 \\ H & 0 & \overline{u} \end{pmatrix} \frac{\partial}{\partial x}\begin{pmatrix} u \\ v \\ \eta \end{pmatrix} - \begin{pmatrix} \overline{v} & 0 & 0 \\ 0 & \overline{v} & g \\ 0 & H & \overline{v} \end{pmatrix} \frac{\partial}{\partial y}\begin{pmatrix} u \\ v \\ \eta \end{pmatrix} \\ &- \begin{pmatrix} 0 & -f & 0 \\ f & 0 & 0 \\ \frac{\partial H}{\partial x} & \frac{\partial H}{\partial y} & 0 \end{pmatrix}\begin{pmatrix} u \\ v \\ \eta \end{pmatrix} - \begin{pmatrix} -f\overline{v} \\ f\overline{u} \\ \overline{\boldsymbol{u}}.\nabla_h H \end{pmatrix}, \end{aligned} \tag{3.10}$$

where primes have been dropped for convenience, $\overline{\boldsymbol{u}} = (\overline{u}, \overline{v})$, and ∇_h is the horizontal gradient operator. The mean sea-level, $\overline{\eta}$, has been taken to be zero. Finally, in order to symmetrize the hyperbolic parts, we make the substitution $\eta = \sqrt{(H/g)}\eta^*$. Then (3.10) becomes:

$$\begin{aligned} \frac{\partial}{\partial t}\begin{pmatrix} u \\ v \\ \eta^* \end{pmatrix} = &- \begin{pmatrix} \overline{u} & 0 & c \\ 0 & \overline{u} & 0 \\ c & 0 & \overline{u} \end{pmatrix} \frac{\partial}{\partial x}\begin{pmatrix} u \\ v \\ \eta^* \end{pmatrix} - \begin{pmatrix} \overline{v} & 0 & 0 \\ 0 & \overline{v} & c \\ 0 & c & \overline{v} \end{pmatrix} \frac{\partial}{\partial y}\begin{pmatrix} u \\ v \\ \eta^* \end{pmatrix} \\ &- \begin{pmatrix} 0 & -f & \frac{1}{2}\frac{c}{H}\frac{\partial H}{\partial x} \\ f & 0 & \frac{1}{2}\frac{c}{H}\frac{\partial H}{\partial y} \\ \frac{c}{H}\frac{\partial H}{\partial x} & \frac{c}{H}\frac{\partial H}{\partial y} & \frac{1}{2}H^{-1}(\overline{\boldsymbol{u}}.\nabla_h H) \end{pmatrix}\begin{pmatrix} u \\ v \\ \eta^* \end{pmatrix} \\ &- \begin{pmatrix} -f\overline{v} \\ f\overline{u} \\ \frac{c}{H}(\overline{\boldsymbol{u}}.\nabla_h H) \end{pmatrix} \end{aligned} \tag{3.11}$$

where $c = \sqrt{gH}$ is the shallow-water gravity-wave speed. Equation (3.11) is now in the form of equation (3.6) and well-posed boundary conditions may be formulated. Consider the case of boundaries at $x = a$ and $x = b$ $(a < b)$. To derive the characteristic variables, we must diagonalise the coefficient matrix

$$\boldsymbol{B}_x = -\begin{pmatrix} \overline{u} & 0 & c \\ 0 & \overline{u} & 0 \\ c & 0 & \overline{u} \end{pmatrix}.$$

This has the following eigenvalues and eigenvectors:

$$\begin{aligned} \lambda_1 = -\overline{u} &: \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ \lambda_2 = -\overline{u} - c &: \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \\ \lambda_3 = -\overline{u} + c &: \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \end{aligned} \tag{3.12}$$

The characteristic variables are (see equation (3.5)):

$$\begin{aligned} s_1' &= v \\ s_2' &= u + \eta^* = u + \frac{c}{H}\eta \\ s_3' &= u - \eta^* = u - \frac{c}{H}\eta. \end{aligned}$$

Consider now the left boundary ($x = a$). The incoming characteristic variables are those associated with negative eigenvalues (see Figure 3.2). There are two cases to consider: where the mean flow is into the domain ($\overline{u} > 0$), and where the mean flow is exiting the domain ($\overline{u} < 0$). In the first case, both λ_1 and λ_2 will be negative and λ_3 will be positive (note that we will always have $|\overline{u}| < c$ for oceanic flows[2]). The characteristic variables v and $u + \eta^*$ are both incoming. Two boundary conditions are needed, providing values for these variables. Referring to equations (3.8), the most general forms for well-posed boundary conditions are:

$$\begin{aligned} v &= q_{a_{(1,3)}}(u - \eta^*) + g_{a_1} \\ u + \eta^* &= q_{a_{(2,3)}}(u - \eta^*) + g_{a_2}, \end{aligned} \tag{3.13}$$

where q_a are arbitrary coefficients and g_a allow for prescribed values. Taking $q_{a_{(1,3)}} = 0$ is equivalent to prescribing v independently. Taking $q_{a_{(2,3)}} = -1$ is equivalent to prescribing u, while $q_{a_{(2,3)}} = 1$ allows one to prescribe η^* (or, equivalently, η). Oliger and Sundström (1978) point out, however, that the latter condition is, in fact, ill-posed — the criterion (3.7) is not satisfied. Continuing the analysis for $\overline{u} > 0$, the right boundary ($x = b$) will therefore be an outflow boundary for the mean flow. Well-posed boundary conditions must provide values for characteristic variables corresponding to positive eigenvalues. There is only one of these, λ_3, so only one boundary condition must be provided. The most general boundary condition takes the form

$$u - \eta^* = q_{b_{(3,1)}} v + q_{b_{(3,2)}}(u + \eta^*) + g_{b_3}. \tag{3.14}$$

Appropriate choices of the coefficients allows the condition to prescribe either u or η^* independently. Either of these choices is demonstrated in Appendix A to provide well-posed boundary conditions using a technique known as the *energy method*. We have considered the case $\overline{u} > 0$. The case $\overline{u} < 0$ follows analogously, with the roles of the two boundaries reversed. Summarising, the linearized shallow-water equations require two boundary conditions at an inflow boundary (including one which specifies the tangential velocity component), and one condition at an outflow boundary.

Oliger and Sundström (1978) considered the boundary condition problem for several fluid dynamical systems, including the shallow-water equations and the primitive equations. For the latter, they concluded that no pointwise well-posed boundary conditions could be constructed. Bennett and Kloeden

[2]The case $|\overline{u}| > c$ occurs in gas dynamics where c is the speed of sound and the flow is then *supersonic*. It can also occur for baroclinic modes, where c represents the *reduced gravity* shallow water gravity wave speed.

(1978) reached the same conclusion. The problem is due to the fact that in a hydrostatic baroclinic ocean, the total flow field can be separated into vertical modes, each of which satisfies the *reduced-gravity* shallow-water equations. At a boundary point, for a given mean flow, the flow may be "subcritical" for low-order modes (*i.e.* slower than the reduced-gravity shallow-water wave speed) and "supercritical" for high-order modes. The number of boundary conditions required for well-posedness depends on the signs of the eigenvalues of the system matrix, but these will be different for the two cases (see equations (3.12) — the sign of λ_3 depends on whether the flow is subcritical or supercritical). Unless the system is decomposed into vertical modes and boundary conditions constructed for each mode, some modes may be overprescribed while others may be underprescribed. Boundary conditions which are local and pointwise in the vertical, as implemented in most models, are ill-posed. If such conditions are successfully implemented in a primitive equation model it can only be through an incorrect prescription, with some consequent sacrifice either of the physics or smoothness of the solution at the boundaries. One should not be surprised by difficulties that arise in attaining stability or reasonable solutions. This is discussed by Mahadevan and Archer (1998) and Bennett and Chua (1999).

3.1.3 Numerical stability

Reviewing the previous section, the pure initial value (Cauchy) problem is well-posed for both parabolic and hyperbolic systems. For initial-boundary value problems, parabolic systems are well-posed under Neumann or Dirichlet boundary conditions while hyperbolic systems are well-posed with boundary conditions that prescribe values for incoming characteristic variables at the boundaries. The one-dimensional advection equation (3.2) is a hyperbolic equation already in characteristic form. The variable s is incoming at the right boundary and outgoing at the left boundary. Thus s must be prescribed at the right boundary, but not at the left. The numerical scheme used at the left boundary must be a simple discretisation of the advection equation. As shown in Figure 3.1, however, not all discretisations are numerically stable.

Numerical stability of finite-difference boundary conditions for hyperbolic systems has been examined by Kreiss and collaborators (Kreiss 1971; Gustafsson et al. 1972) and some sophisticated theory (now known as the "GKS" theory, after the authors of the 1972 paper) has been developed. Varah (1971) extended the theory to parabolic systems. A detailed presentation of this theory will not be provided here, but rather a descriptive summary of its general approach and interpretation is given, as well as some results. The theory parallels for the discrete case the *normal mode analysis* developed by Kreiss (1970) for proving well-posedness of boundary conditions in the continuous case. (Note that this is very different from, and should not be confused with, the *normal mode decomposition* of vertical structure that can be done in hydrostatic models.) In practice, the stability of finite-difference implementations of boundary conditions is easy to verify experimentally — failures tend to be explosive.

Typically, a finite-difference method for integrating an initial-boundary value problem consists of a numerical scheme for the interior and one for the boundaries. Implicit in the development of the GKS stability theory is the assumption that the interior difference scheme is stable for the Cauchy problem. This question may be examined with the classical *von Neumann* stability analysis for finite-difference schemes (Forsythe and Wasow 1960; Press et al. 1992). In practice, any useful interior scheme will be Cauchy-stable.

To determine the stability of the initial-boundary value problem, the GKS approach seeks to decide whether the combined (interior and boundary) difference scheme admits any eigensolutions that grow in time. By this method Gustafsson et al. (1972) show, for instance, that any finite-difference extrapolation procedure for estimating boundary values from interior values is unstable with a leap-frog in time scheme used for the interior. Further examples of the application of the GKS theory to examine finite-difference boundary condition stability may be found in Gustafsson and Kreiss (1979), Abarbanel and Murman (1981) and Higdon (1994).

The theory as formulated by Gustafsson et al. (1972) is rather technical and difficult to apply. Trefethen (1983) offered a physical interpretation of the theory which facilitates intuitive application. He demonstrated that the stability criterion could be interpreted in terms of group velocity: if the interior and boundary finite-difference schemes support a set of waves with group velocities entering the domain, then the scheme is unstable. This easier interpretation was used to generalize certain known stability results. The example illustrated in Figure 3.1 was included. Durran et al. (1993) developed nonreflecting boundary conditions and used Trefethen's group velocity interpretation to suggest that their scheme was stable since it required outward-directed group velocities.

Having discussed some theoretical issues relating to open boundary conditions, some specific conditions that have been applied in the literature will now be reviewed before discussing the implementation of open boundaries in the HOPE model.

3.2 Radiation conditions

A commonly used family of open boundary conditions are the so-called "radiation conditions". As discussed later, trials of these were applied to various components of HOPE. Radiation conditions are motivated by the desire for an artificial boundary to be nonreflecting to wave phenomena generated in the interior. The starting point for their development was Sommerfeld's classical condition (Sommerfeld 1949) for solutions to the wave equation

$$\frac{\partial^2 \phi}{\partial t^2} = \nabla^2 \phi. \tag{3.15}$$

Assuming harmonic solutions $\phi = k(\boldsymbol{x})e^{-\imath\omega t}$ then

$$\frac{\partial^2 \phi}{\partial t^2} = -\omega^2 \phi,$$

and the elliptic equation $\nabla^2\phi + \omega^2\phi = 0$ results. The Sommerfeld radiation condition states that the solution to this elliptic equation must satisfy

$$\lim_{r\to\infty}\left|\frac{\partial\phi}{\partial r} - i\omega\phi\right| = 0,\ r = |\boldsymbol{x}|.$$

In one dimension this gives the relation

$$\lim_{x\to\infty}\left|\frac{\partial\phi}{\partial x} + \frac{\partial\phi}{\partial t}\right| = 0.$$

The boundary condition at infinity can be shown to correspond to an outwards flux of energy (Courant and Hilbert 1962).

Israeli and Orszag (1981) motivated the condition on a finite domain as follows. Consider the solution to the one-dimensional wave equation (3.15)

$$\phi(x,t) = F(x-t) + G(x+t),$$

consisting of a right-propagating wave, F, and a left-propagating wave, G. For outwards radiation we require no incoming waves at $|x| > X > 0$, so we must have

$$\begin{aligned} G &= 0,\ x > X, \\ F &= 0,\ x < -X. \end{aligned} \tag{3.16}$$

Then the condition

$$\frac{\partial\phi}{\partial t} \pm \frac{\partial\phi}{\partial x} = 0 \tag{3.17}$$

ensures no incoming waves at $x = \pm X$. The condition (3.17) is regarded as an approximation to the required condition (3.16) on a finite domain. For a wave with phase velocity c, it may be written

$$\frac{\partial\phi}{\partial t} \pm c\frac{\partial\phi}{\partial x} = 0. \tag{3.18}$$

A radiation condition for the barotropic system with $c = \sqrt{gH}$ is also called a *gravity-wave radiation condition* (Chapman 1985; Røed and Cooper 1986; Palma and Matano 1998).

Israeli and Orszag (1981) considered other alternative approximate radiation conditions, including viscous *sponge layers* and damping, as methods to prevent reflected waves from finite boundaries.

Engquist and Majda (1977) considered a hierarchy of higher-order approximations to prevent reflections in two dimensions, taking account of non-normal incidence at the boundary. The first order approximation for an x boundary was precisely (3.17). The second-order approximation, for example, was

$$\frac{\partial^2\phi}{\partial x\partial t} - \frac{\partial^2\phi}{\partial t^2} + \frac{1}{2}\frac{\partial^2\phi}{\partial^2 y}\bigg|_{x=0} = 0.$$

The normal-mode analysis of Kreiss (1970) was used to verify well-posedness of their conditions for the two-dimensional wave equation.

Bennet (1976) attempted to construct radiation conditions for dispersive waves, but ultimately concluded "(i)f the wave fields are dispersive then the outgoing radiation conditions are too complex to implement numerically."

More recently Higdon (1994) had greater success for dispersive waves by using a product of multiple nondispersive conditions:

$$\left[\prod_{j=1}^{m}\left(\frac{\partial}{\partial t}-c_j\frac{\partial}{\partial x}\right)\right]\phi=0.$$

Well-posedness of the above condition was established with Kreiss' (1970) normal mode analysis. The GKS theory (Gustafsson et al. 1972) was used to show stability for a finite-difference implementation of the condition.

Perhaps the most widely used implementations of the radiation condition (3.18) derive from Orlanski's (1976) approximation. He took the approach first of *estimating*, at each timestep, the phase velocity c using points near the boundaries in space and time:

$$c\approx\mp\frac{(\partial\phi/\partial t)}{(\partial\phi/\partial x)}, \tag{3.19}$$

where the upper sign is for a right-boundary, and the lower is for a left-boundary. This estimate of the phase velocity was then used conventionally in radiation conditions (3.18) at the respective boundaries. In the estimate for c, the spatial derivative, $(\partial\phi/\partial x)$ is calculated using one-sided differencing at each boundary. In order to satisfy the CFL stability criterion, c was restricted to the range $0\leq c\leq\Delta x/\Delta t$ where Δx is the grid spacing and Δt the model timestep. Since $c=c(\phi)$, the condition is seen to be nonlinear. Orlanski applied the condition to a collapsing density perturbation in a stratified fluid and obtained good results. Orlanski's discretisation of (3.19) used a leap-frog in time scheme.

A simple variation of Orlanski's method was developed by Camerlengo and O'Brien (1980). These authors distinguished between an "inflow" phase speed ($c<0$) and an "outflow" phase speed ($c>0$) calculated as normal with (3.19). c was then set to 0 for inflow and $\Delta x/\Delta t$ for outflow before being used in the radiation condition (3.18). This is equivalent to no change for inflow, *i.e.* $\phi_b^{n+1}=\phi_b^n$, and pure extrapolation for outflow, *i.e.* $\phi_b^{n+1}=\phi_{b+1}^n$, where the subscripts b and $b+1$ represent a boundary, and next interior, gridpoint respectively, and the superscript n is a time index[3]. In a two-layer model, this simplified condition produced no reflection for a Kelvin wave simulation, but slight reflection for a Rossby wave.

Miller and Thorpe (1981) replaced Orlanski's leap-frog-in-time scheme with one more suited to forward Euler differencing. Their discretisation of the c estimator (3.19) was

$$c=\frac{\phi_{b+1}^n-\phi_{b+1}^{n-1}}{\phi_{b+2}^{n-1}-\phi_{b+1}^{n-1}}, \tag{3.20}$$

[3]Camerlengo and O'Brien's discretisations were slightly different as they used a leap-frog in time differencing scheme.

where the subscript $b+1$ represents the gridpoint one in from the boundary, and $b+2$ the gridpoint two in from the boundary. The radiation condition (3.18) is written

$$\frac{\phi_b^{n+1} - \phi_b^n}{\Delta t} = c\frac{\phi_{b+1}^n - \phi_b^n}{\Delta x}$$

or

$$\phi_b^{n+1} = r\phi_{b+1}^n + (1-r)\phi_b^n,$$

where $r = c\Delta t/\Delta x$ and $0 \leq r \leq 1$. The authors also speculated on improvements to (3.20) including, for instance, the following implicit version

$$c = \frac{\phi_{b+1}^{n+1} - \phi_{b+1}^n}{\phi_{b+2}^n - \phi_{b+1}^n}.$$

Numerous other variations on the discretisations of both the estimator (3.19) and the radiation condition (3.18) were investigated by Tang and Grimshaw (1996). Their most successful version can be derived essentially by averaging over the boundary and next interior gridpoint for $(\partial\phi/\partial t)$ and over two time levels for $(\partial\phi/\partial x)$ in both the estimator and the condition, leading to the final (implicit) condition

$$\phi_b^{n+1} = \phi_{b+1}^n + s(\phi_b^n - \phi_{b+1}^{n+1}),$$

where

$$s = \frac{\phi_{b+1}^n - \phi_{b+2}^{n-1}}{\phi_{b+1}^{n-1} - \phi_{b+2}^n}.$$

The choice of discretisation of Orlanski's condition, (3.19), depends to a large extent on the model being used and the problem under consideration. Röed and Cooper (1986) stressed the importance of constructing the boundary condition according to the problem. Tang and Grimshaw (1996) found that an implicit implementation worked best in their model for coastally trapped waves. As will be seen in section 3.4, an implicit implementation is not possible for the barotropic system in the HOPE model. The literature presents no unambiguously clear "winner"; indeed there is a lack of a theoretical treatment of different discretisations examining reflection properties, for instance, in the same way that Arakawa (1977) examined dispersion properties of various finite-difference grids. Such a treatment would do much to unify the various *ad-hoc* results reported for different numerical implementations.

3.3 Other conditions

While radiation conditions (and their variations) are perhaps the most common open boundary condition applied to limited-area models, there are others which feature in the literature. Two rather unsophisticated approaches are the *clamped condition*, where

$$\left.\frac{\partial\phi}{\partial t}\right|_b = 0$$

and the *gradient condition*, where

$$\left.\frac{\partial \phi}{\partial x}\right|_b = 0.$$

Based on the theory of characteristics, Hedstrom (1979) introduced a nonreflecting condition which leaves the incoming characteristic variables unchanged at the boundary:

$$\left.\frac{\partial s}{\partial t}\right|_b = 0,$$

where s are the incoming characteristics at the boundary b. This forms the basis of the Flather condition ultimately adopted for the barotropic mode in this thesis (section 3.5.4). Røed and Cooper (1987) modified this for the shallow-water equations to obtain the following conditions[4]:

$$\begin{aligned}
\frac{\partial u}{\partial t} &= fv \mp \frac{c}{2}\frac{\partial}{\partial x}(u \pm c\eta) \\
\frac{\partial v}{\partial t} &= -fu - c^2\frac{\partial \eta}{\partial y} \\
\frac{\partial \eta}{\partial t} &= -\frac{\partial v}{\partial y} - \frac{1}{2}\frac{\partial}{\partial x}(u \pm c\eta),
\end{aligned}$$

where the upper sign is for the right-hand boundary and the lower is for the left boundary. This condition was also used by Jensen (1998) and Palma and Matano (1998).

Several authors (Chapman 1985; Røed and Cooper 1986; Røed and Cooper 1987; Jensen 1998; Palma and Matano 1998; Palma and Matano 2000) have undertaken evaluation studies of several open boundary conditions for specific test problems. These papers may serve as reviews of most of the commonly used approaches.

3.4 HOPE modifications: introductory remarks

To the extent possible, this thesis attempts to provide an objective catalogue of results, but it should be emphasized that this is an inaccurate portrayal of the research experience. In fact, a complete solution has not been found, insofar as the complex interactions between boundary conditions for the respective subsystems of HOPE have not fully been characterised. The work was often difficult, with unexpected and frustrating results the norm rather than the exception. Usually, when studying an unstable system, one has some theoretical guidance. However, we have seen already that open boundary conditions for a primitive equation model are ill-posed. The absence of a suitable theoretical framework to guide a rigorous investigation left trial-and-error and *ad-hoc* approaches the most scientific of methods available. The size of the 'parameter space' meant that not all combinations of boundary conditions could

[4]This condition is reviewed in Appendix B together with a discussion of why it appears to be an invalid extension of Hedstrom's (1979) method and the theory of characteristics.

Table 3.1: "Canonical" boundary conditions for the limited-area Southern Ocean assimilation model.

boundary	*barotropic system*	*baroclinic system*	*tracers*
west	Flather	(nothing needed)	upstream advection/relaxation
east	Flather	Orlanski	upstream advection/relaxation
north	Flather	zero-gradient	upstream advection/relaxation

systematically be evaluated; numerous possibilities are available for each of the physical subsystems (and each of the boundaries) and these interact in complex ways. It soon became apparent that an almost infinite succession of variations suggest themselves, each with the promise of fixing some problem or other, but each introducing further complications and confounding an already cloudy picture. Ultimately, the fruit of this labour is a set of boundary conditions that remain stable and reproduce gross aspects of known flow fields.

For reference, the "canonical" set of boundary conditions eventually derived are summarised here in Table 3.1. The following sections of this chapter discuss, in turn, the respective physical subsystems of HOPE — the barotropic, baroclinic, and tracer equations — and variations, for each of these subsystems, on the canonical set. This set eventually will be seen to provide the best overall performance. Of necessity, this thesis cannot describe all of the variations examined. A set illustrating salient points in the development of the canonical conditions is included. A number of early experiments using a zonally connected domain with artificial interpolated topography are not reported here. The results, even using heavy tracer relaxation in the interpolation region, were unrealistic. In addition, numerous additional open boundary experiments for the baroclinic system were performed, but not reported here. These included, for instance, different discretisations of the Orlanski radiation condition. In evaluating the performance of an experiment, both quantitative and qualitative criteria were used. Graphical software was developed to enable the full set of model fields to be examined interactively. Due to space limitations, however, model fields presented here have been restricted to a few illustrative cases (mostly sea-level, surface currents, and stream-function).

In a data assimilation study, one should really evaluate open boundary conditions in the context of assimilation. Modifying model fields may excite internal inertia-gravity waves, for instance, and the performance of the boundary conditions may be different to a free-running model scenario. However, it would be far too large a project to jointly examine both assimilation and boundary condition performance under a suite of different boundary conditions. The approach adopted here was first to derive a set of boundary conditions that performed reasonably in a free-running model, and then simply to use these in the assimilation study.

Apart from the model physics, other incidental changes are needed to adapt the model to open domains. These include, for instance, the input of data fields: topography, wind and surface climatological forcing, etc. Es-

sentially, any of the numerous parts of the HOPE code where periodicity is enforced (via subroutines PERIOxx) needs to be modified. These are not described here for the incidental changes. So long as sensible and self-consistent solutions are devised (*e.g.* ensuring that no oceanic boundary velocity points are adjacent to land), then problems that occur are due to physics and not model configuration.

Although open boundary conditions for the primitive equations are necessarily ill-posed due to the hydrostatic assumption, the approach taken here attempts, at least, to obtain solutions which appear "reasonable" and which remain stable over timescales of interest — typically up to several years for assimilation. The approach is guided by well-posedness considerations for various components of the overall system, as discussed further below.

3.5 The barotropic system

The barotropic system in HOPE (section 2.1.3, page 49) proved to be the most difficult component of the split system to modify for open-boundary integration. There are several reasons for this difficulty.

First, since sea-level is a prognostic variable in HOPE, and since the model is volume conserving, the mean sea-level is susceptible to net imbalances in volume transport across the open boundaries. For the regional Southern Ocean domain used in this study, a net imbalance of just 0.7 Sv (*i.e.* just one half percent of the ACC flow entering the domain) leads to a change in mean sea-level of one metre per year. This may not be important in process studies, but is completely unsatisfactory for altimetry assimilation. The problem of mass conservation for the domain limits the usefulness of unconstrained radiation conditions, for example, as described further below.

A second difficulty lies in the decoupling of the barotropic solutions on the two (EVEN and ODD) subgrids in HOPE's implementation of the Arakawa E-grid, as described earlier. Severe (yet stable) grid-point differences are prone to occur between the two subgrids under open boundary conditions unless the boundary condition itself strongly couples the degenerate systems. In the formulations found to be successful here, such coupling generally relied on averaging between the subgrids along each of the open boundaries.

The third, and most subtle and tedious, of the difficulties results from the implicit numerics used to solve the barotropic system in HOPE. The most obvious limitation this imposes is the requirement for any open boundary condition to be linear and constant-in-time with respect to the prognostic variables. This arises because the (linear) implicit system is solved directly by backsubstitution, as described earlier. The system matrix is calculated and triangularised once at the beginning of an integration. It would be far too expensive, computationally, to recalculate this matrix every timestep. Thus, open boundary modifications to HOPE's barotropic system are constrained to be of the form:

$$U_b^{n+1} = \mathcal{L}_U(U^{n+1}, V^{n+1}, \eta^{n+1}) + \mathcal{G}_U(U^n, V^n, \eta^n) + \mathcal{F}_U^{n+1} \quad (3.21)$$

$$V_b^{n+1} = \mathcal{L}_V(U^{n+1}, V^{n+1}, \eta^{n+1}) + \mathcal{G}_V(U^n, V^n, \eta^n) + \mathcal{F}_V^{n+1} \quad (3.22)$$

$$\eta_b^{n+1} = \mathcal{L}_\eta(U^{n+1}, V^{n+1}, \eta^{n+1}) + \mathcal{G}_\eta(U^n, V^n, \eta^n) + \mathcal{F}_\eta^{n+1}, \qquad (3.23)$$

where U and V are zonal and meridional components of depth-integrated velocity, respectively, and η is sea-level. The subscript b indicates a boundary, and the $\{\mathcal{L}\}$ must be linear and constant-in-time, but the $\{\mathcal{G}\}$ may be nonlinear and time-varying in general. The forcing functions $\{\mathcal{F}\}$ provide for prescribed boundary conditions. The $\{\mathcal{L}\}$ contribute terms to the system matrix $\boldsymbol{A}$, while $\{\mathcal{G}\}$ and $\{\mathcal{F}\}$ contribute to the right-hand-side $\underline{\Gamma}^n$, in equation (2.42). This limitation prevents the consideration, for example, of implicit formulations of Orlanski's classical radiation condition, as considered by Chapman (1985) and Tang and Grimshaw (1996). The tedium of the limitation stems from the fact that the implementation of any new boundary condition requires significant modifications to the model code for the barotropic system. In a solver with numerics that are explicit in time, all that is required is to overwrite the boundary values of prognostic variables with those calculated according to the desired boundary condition. In HOPE, however, each new boundary condition to be trialled requires a reformulation of the system matrix for grid points on and adjacent to the boundaries, as well as modifications to the code for calculating new velocities $\{U^{n+1}, V^{n+1}\}$ from the updated sea-level η^{n+1} (equations (2.38, 2.38)). Further, a boundary condition on velocity, for example, requires often lengthy algebra to determine its influence on sea-level in HOPE's E-grid for incorporation into the system matrix (and subject to the coupling requirement between subgrids mentioned above — see Appendices C and D). Consequently, debugging and verification of implementations was an additional time-consuming and laborious process.

In addition to the requirements already discussed to be satisfied by a suitable open boundary implementation in HOPE (volume conservation, minimal grid-separation, linearity and time-invariance), it is desirable in the case of altimetry assimilation for the boundary condition to be non-reflective to waves generated in the domain interior. For the barotropic system in HOPE, this requires the transmission of shallow-water gravity waves. Unless the sequential assimilation system is re-initialised at each analysis time with balanced barotropic fields, surface gravity waves will be generated. While geostrophic initialisation is straightforward, even slight imbalances will excite such modes. Reflection at the open boundaries will lead to unwanted noise primarily in the sea-level field. Barotropic Rossby waves, which are also admitted by the system, are dispersive and more difficult to handle with an open boundary condition under the above constraints.

In the recent review by Griffies et al. (2000) of ocean modelling, five models other than HOPE are listed as including implicit free surface formulations of the barotropic system: the CANDIE, MIT, MOM, OPA, and POP models. The manuals for CANDIE (Wright et al. 2000) and OPA (OPA 8.1 Ocean General Circulation Model reference manual 1998), however, describe them both as using a rigid-lid approximation for the barotropic mode. Both POP and MOM are derivatives of the GFDL model, and both use the implicit free-surface modification of Dukowicz and Smith (1994) for the barotropic mode. In the case of the POP model, open boundaries are not permitted, while in

the case of MOM, the open boundary option is incompatible with the implicit free-surface formulation. The only other widely available ocean model with an implicit free-surface is the MIT model (Marshall et al. 1997b; Marshall et al. 1997a). While this model supports open boundary conditions, the formulation is fully explicit (equivalent to the $\mathcal{L}$ in equations (3.23) all being zero) (*A. Adcroft, personal communication (2001)*). The present study therefore appears to be one of the first implementations of implicit open boundary conditions in an implicit free-surface ocean GCM.

3.5.1 Well-posedness requirements

The formulation of HOPE's barotropic system was discussed in section 2.1.3 (on page 49). The prognostic equations are reproduced here:

$$\begin{aligned}
\frac{\partial U}{\partial t} - fV &= -gH\frac{\partial \eta}{\partial x} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial x}dz + G_U \\
\frac{\partial V}{\partial t} + fU &= -gH\frac{\partial \eta}{\partial y} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial y}dz + G_V \\
\frac{\partial \eta}{\partial t} &= -\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right).
\end{aligned}$$

After making the substitution $\eta = c^{-1}\eta^*$ where $c = \sqrt{gH}$ is the shallow-water gravity wave speed, the system may be written in matrix form as

$$\begin{aligned}
\frac{\partial}{\partial t}\begin{pmatrix} U \\ V \\ \eta^* \end{pmatrix} = &- \begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ c & 0 & 0 \end{pmatrix}\frac{\partial}{\partial x}\begin{pmatrix} U \\ V \\ \eta^* \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & c \\ 0 & c & 0 \end{pmatrix}\frac{\partial}{\partial y}\begin{pmatrix} U \\ V \\ \eta^* \end{pmatrix} \\
&- \begin{pmatrix} 0 & -f & -\frac{1}{2}\frac{c}{H}\frac{\partial H}{\partial x} \\ f & 0 & -\frac{1}{2}\frac{c}{H}\frac{\partial H}{\partial y} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} U \\ V \\ \eta^* \end{pmatrix} - \begin{pmatrix} \frac{1}{\rho_0}\int\frac{\partial p'}{\partial x}dz \\ \frac{1}{\rho_0}\int\frac{\partial p'}{\partial x}dz \\ 0 \end{pmatrix}. \qquad (3.24)
\end{aligned}$$

This is very similar to the linearized shallow-water equations (3.11) discussed in section 3.1.2 and analysed using the energy method in Appendix A. Differences occur only in the zeroth-order and forcing terms, and the mean velocities about which the equations were linearized (these are zero here). In particular, the eigenvalues of the system (see equations (3.12)) for boundaries parallel to either axis are 0, c and $-c$. Thus one boundary condition is required at each boundary. Well-posed conditions are guaranteed by prescribing values for the incoming characteristic variables (Table 3.2). Alternatively, one may attempt to prescribe either normal velocity (U for eastern and western boundaries, V for northern and southern boundaries) or sea-level, η. For the shallow-water equations, prescribing sea-level at an inflow boundary is ill-posed (Oliger and Sundström 1978). The distinction between inflow and outflow boundaries for the shallow-water equations depends on the sign of the mean velocity about which they are linearized. HOPE's barotropic system does not include momentum advection and therefore need not be linearized, so the distinction between the two types of boundaries does not exist for the purposes of determining

Table 3.2: Characteristic variables to be prescribed in HOPE's barotropic system to guarantee well-posed boundary conditions.

boundary	*characteristic variable*
West	$U + c\eta$
East	$U - c\eta$
North	$V - c\eta$
South	$V + c\eta$

well-posed boundary conditions. On the other hand, momentum advection is performed in another part of HOPE. It might be expected, therefore, that prescribing sea-level may cause problems. While not reproduced here, an attempt to use the energy method to prove well-posedness for prescribing either normal velocity or sea-level in HOPE's barotropic system fails. The use of the energy method to prove well-posedness of boundary conditions is demonstrated in Appendix A, where it is applied to the shallow-water equations. The limitations of the method are also seen there. In cases where the energy method fails, a considerably more complicated technique, such as the *normal mode analysis* of Kreiss (1970), is required. No attempt was made here to apply this theory to HOPE's barotropic system.

Several open boundary conditions trialled for HOPE's barotropic system are now described. Results presented here are for a coarse resolution version of the model (one degree per EVEN/ODD subgrid). Since length-scales for the barotropic system are typically several degrees or more, the conclusions from these experiments are valid for the higher-resolution model used in the assimilation studies, and coarser resolution facilitated rapid evaluation of boundary conditions. All model parameters are the same as for the higher-resolution model (see section 2.2.3). In each case, the boundary condition will be evaluated against its ability to give sensible flow patterns while at the same time conserving mean volume and transmitting surface gravity waves. Significant separation of the solutions on the two subgrids will also be unacceptable.

For the baroclinic system and tracers, the standard boundary conditions of Table 3.1 (discussed later) were used.

3.5.2 Boundary conditions on normal velocity

First, boundary conditions applied to normal velocity at an open boundary are considered. Prescribing normal velocity is well-posed for the linearized shallow-water equations (see section 3.1.2, Appendix A and Oliger and Sundström (1978)). As required by the theory of characteristics, the evolution of the tangential component of barotropic velocity on the boundaries is governed by the relevant momentum equation. The conditions trialled here are listed in Table 3.3. The technical modifications to the model required to implement these boundary conditions is described in Appendix C.

Table 3.3: Boundary conditions applied to normal velocity in HOPE's barotropic system. The upper sign is for the western boundary, the lower for the eastern and northern boundaries.

boundary condition	*equation*
clamped	$\partial U_\perp/\partial t = 0$
gravity-wave	$\partial U_\perp/\partial t = c(\partial U_\perp/\partial x), c = \pm\sqrt{gH}$
zero-gradient	$\partial U_\perp/\partial x = 0$

Clamped condition

For the clamped condition, we require

$$\frac{\partial U_\perp}{\partial t} = 0,$$

where the normal velocity $U_\perp = U$ for the eastern and western boundaries, $U_\perp = V$ for the northern boundary. Equivalently, we have

$$U_{\perp_b}(t) = U_{\perp_0}, \tag{3.25}$$

where $U_{\perp_0}$ is a *prescribed* normal velocity and the subscript b represents a boundary. We allow $U_{\perp_0}$ to slowly vary in time, $U_{\perp_0} = U_{\perp_0}(t)$ (where timescales are much longer than the rapid surface gravity-wave timescales). A range of variations on this condition are discussed, as they also inform the development of the boundary condition eventually adopted (the Flather condition, section 3.5.4) for the barotropic system.

There is considerable freedom to choose the prescribed depth-integrated velocities $U_{\perp_0}$. A common approximation in oceanography is that of a *level of no motion* (Gill 1982, p 216). By referencing geostrophic velocities, with unknown barotropic component, to such a level, the *thermal wind* relation allows absolute velocity profiles to be constructed from density sections. Of course, no uncertainty has been removed by doing this, but the idea of a depth of no motion may have some physical, or at least intuitive, appeal. Linear inverse methods may be used to estimate a level on the boundaries of a closed domain subject to conservation constraints, *e.g.* (Sloyan and Rintoul 2000). Without performing such an inverse study, one simple approximation is to assume a deep reference level, or one at the seafloor in the extreme case. This latter approach is often taken for forming initial estimates in inverse problems (Sloyan and Rintoul 2000; Yaremchuk et al. 2001), and is also used here for the first trial of the clamped boundary condition. While it is certainly not volume conserving for the domain, it allows the numerical implementation of the clamped condition to be tested. Volume conserving modifications are considered in the following pages.

The prescribed normal barotropic velocities are constructed each timestep by *bottom-referencing* the geostrophic flow along each boundary. Setting time derivatives to zero in the prognostic momentum equations (2.30)–(2.33) and requiring $u = 0$ and $v = 0$ at the seafloor $z = -H$ along meridional and

zonal boundaries respectively, we diagnose the bottom-referenced geostrophic velocities from:

$$\begin{aligned} U_0 &= \frac{H}{f}\frac{\partial p'}{\partial y}\bigg|_{z=-H} - \frac{1}{f}\int \frac{\partial p'}{\partial y} dz \\ V_0 &= -\frac{H}{f}\frac{\partial p'}{\partial x}\bigg|_{z=-H} + \frac{1}{f}\int \frac{\partial p'}{\partial x} dz. \end{aligned} \tag{3.26}$$

The model was run for five years with this boundary condition. Figure 3.3(a) shows the resulting total volume transport across each boundary. While the evolution is smooth enough for the first two years, significant problems occur thereafter. Even in the first two years, however, it is clear that the bottom-referenced transport does not conserve volume in the domain. With no constraint on mass divergence over the domain, significant net volume transports can develop. There is a net outflow of around 50 Sv after one year, rising to 600 Sv just before three years. Figure 3.3(b) shows the mean sea-level, which changes as expected, at rates up to 800 m per year.

It is apparent also from Figure 3.3(b) that a divergence in sea-level between the two subgrids evolves over time. The separation is around four metres after two years. This is despite the fact that the implementation of the boundary condition (described in Appendix C) enforces the same volume transport into the respective subgrids along each boundary. The reason for the discrepancy in this case is that the subgrids have a different horizontal surface area, so that a nonzero net inflow or outflow will lead to a differential change in sea-level between the grids. As discussed earlier, such a separation between the subgrids (let alone the absolute change over time) is unacceptable for altimetry assimilation.

At this point, a brief discussion on presentation of model output in this thesis is in order. It was mentioned earlier that the staggered E-grid of HOPE allows a degree of independence between the solutions on the two subgrids. An extreme example of this has just been mentioned. The grid arrangement also complicates the graphical display of model fields. While either of the two (EVEN or ODD) subgrids forms a regular array that is easily rendered, the combined grids for any particular variable have "holes" (see Figure 2.2). These may be filled by averaging the four surrounding values. Such a procedure immediately reveals any separation of the solution between the subgrids (sometimes called "chequerboarding"). Sea-level for the combined grids after two years in the present example is shown in Figure 3.4(a) while the ODD grid alone is shown in Figure 3.4(b). Where severe chequerboarding does not occur, the combined subgrids provide a higher resolution image. Results in this thesis combine the subgrids unless stated otherwise.

Despite the tremendous loss of water from the domain, the flow structure appears quite reasonable initially. Figure 3.5 shows the near-surface currents (layer two, 32 m) after two years. Because the HOPE model has a free-surface rather than a rigid-lid, it is not possible to define a streamfunction for the barotropic velocities. Throughout this thesis, however, an approximate

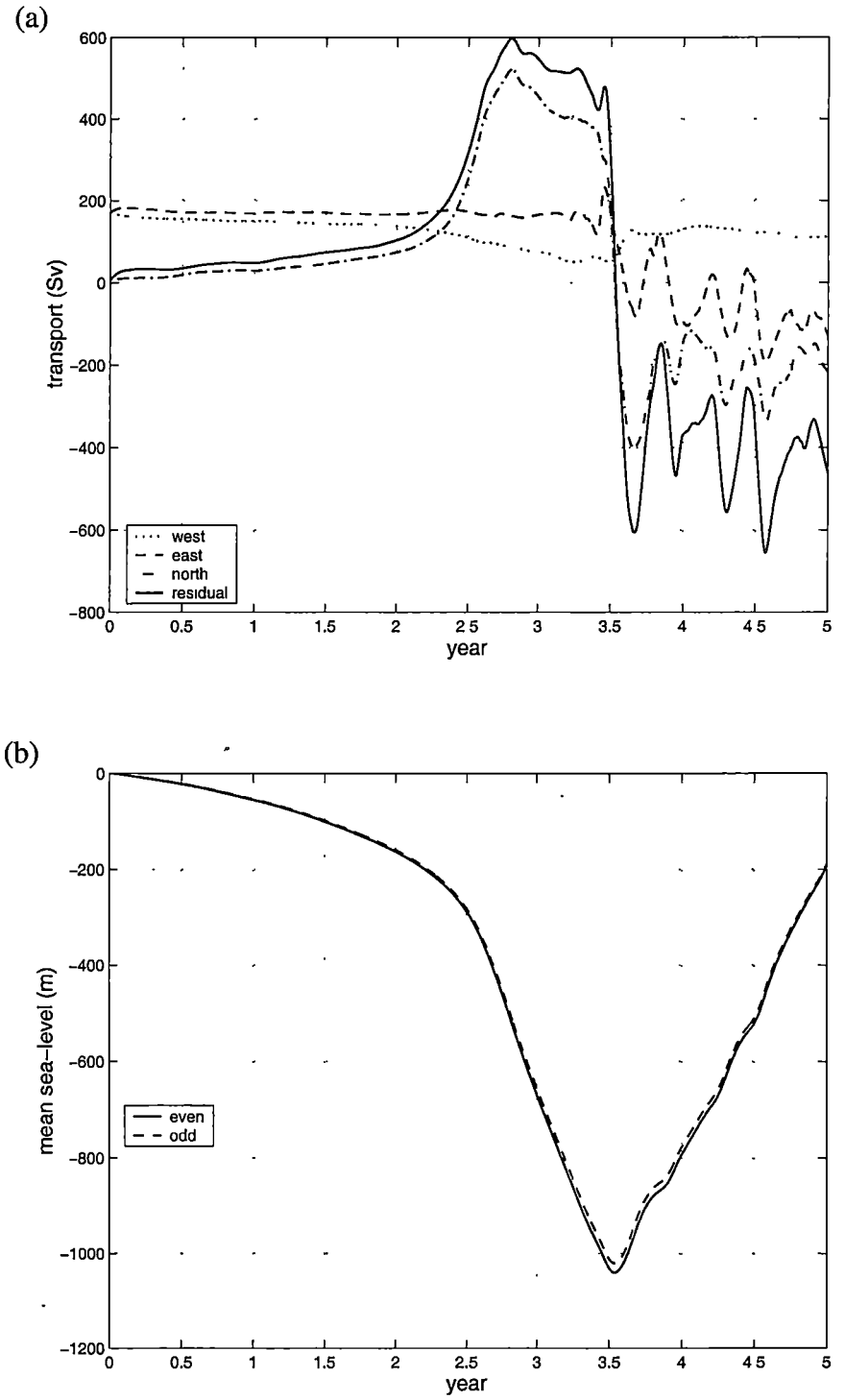


Figure 3.3: (a) Total bottom-referenced geostrophic transport across each boundary for the clamped condition on normal velocity. The residual net transport out of the domain is also shown. (b) Mean sea-level on each subgrid.

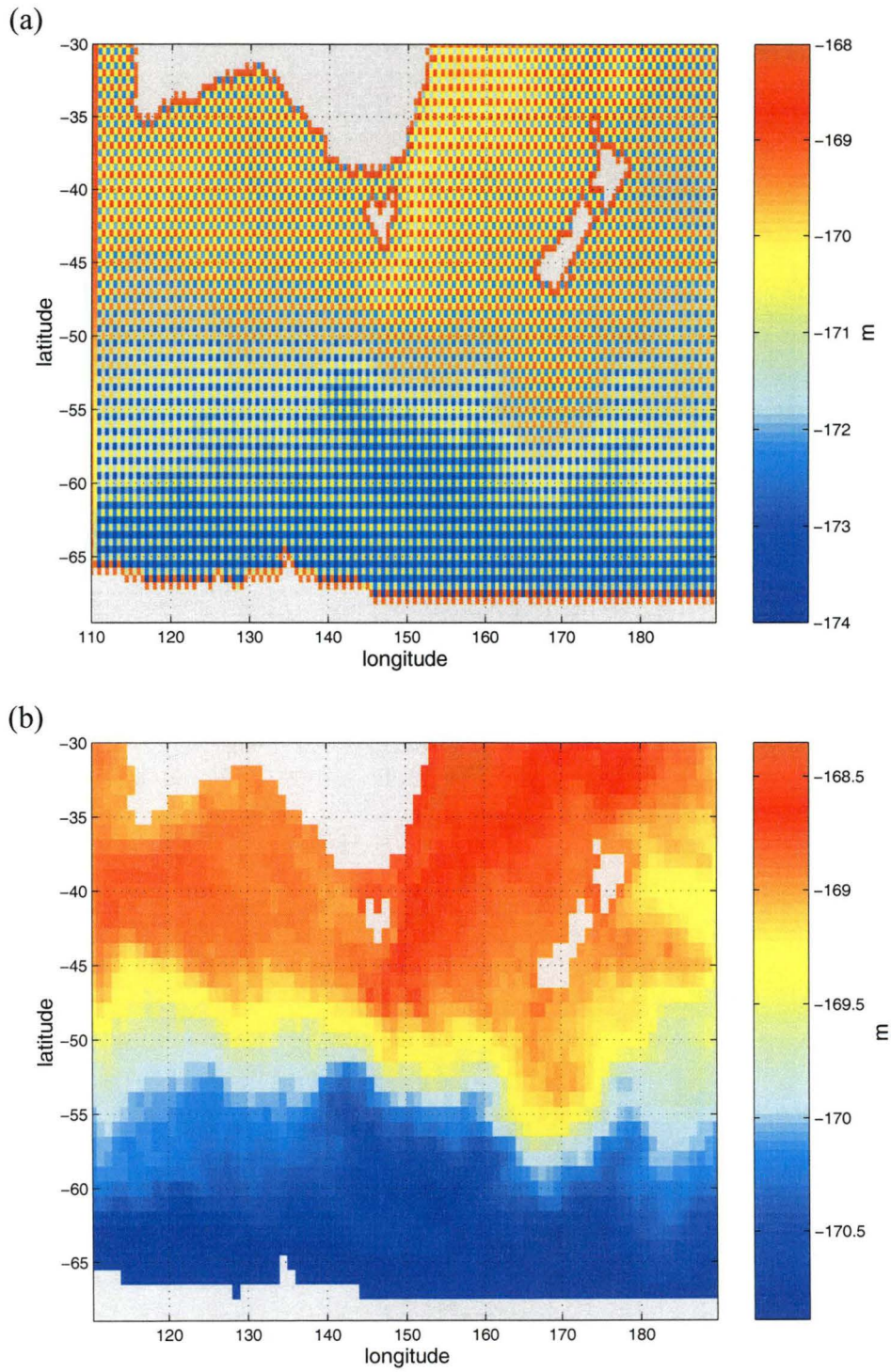


Figure 3.4: Sea-level (m) after two years with the bottom-referenced clamped condition on normal velocity. (a) Combined subgrids. (b) ODD subgrid.

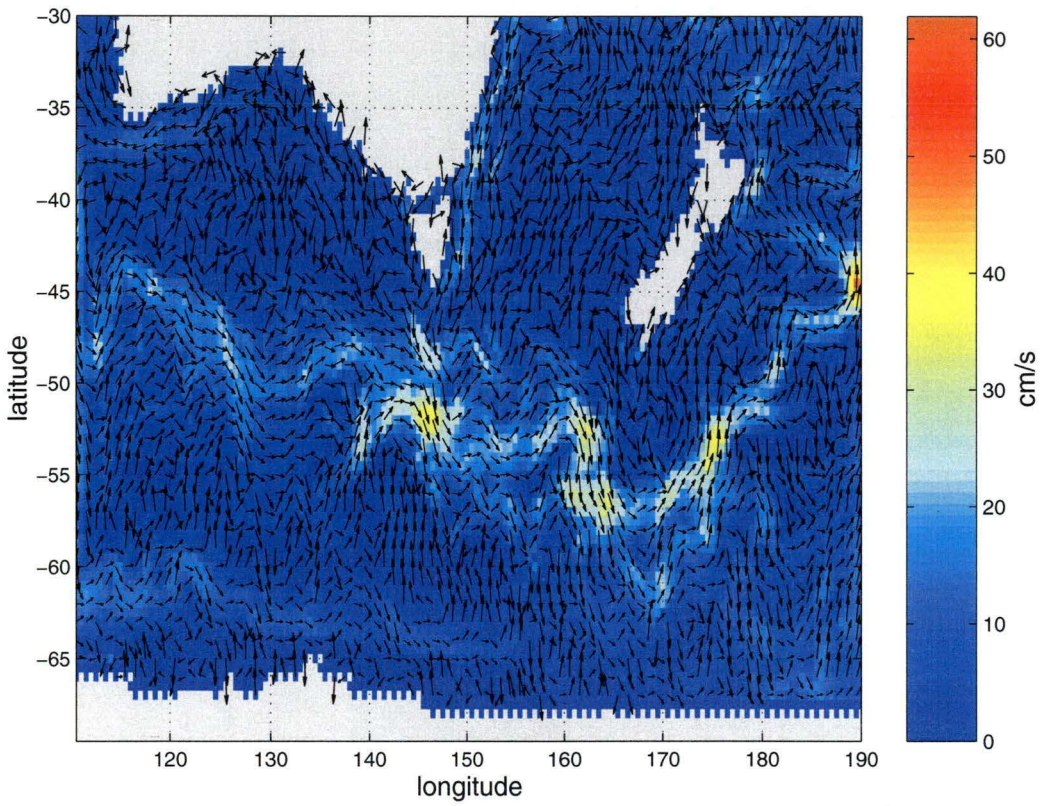


Figure 3.5: Surface currents (32 m depth, layer two) after two years with the bottom-referenced clamped boundary condition on normal velocity.

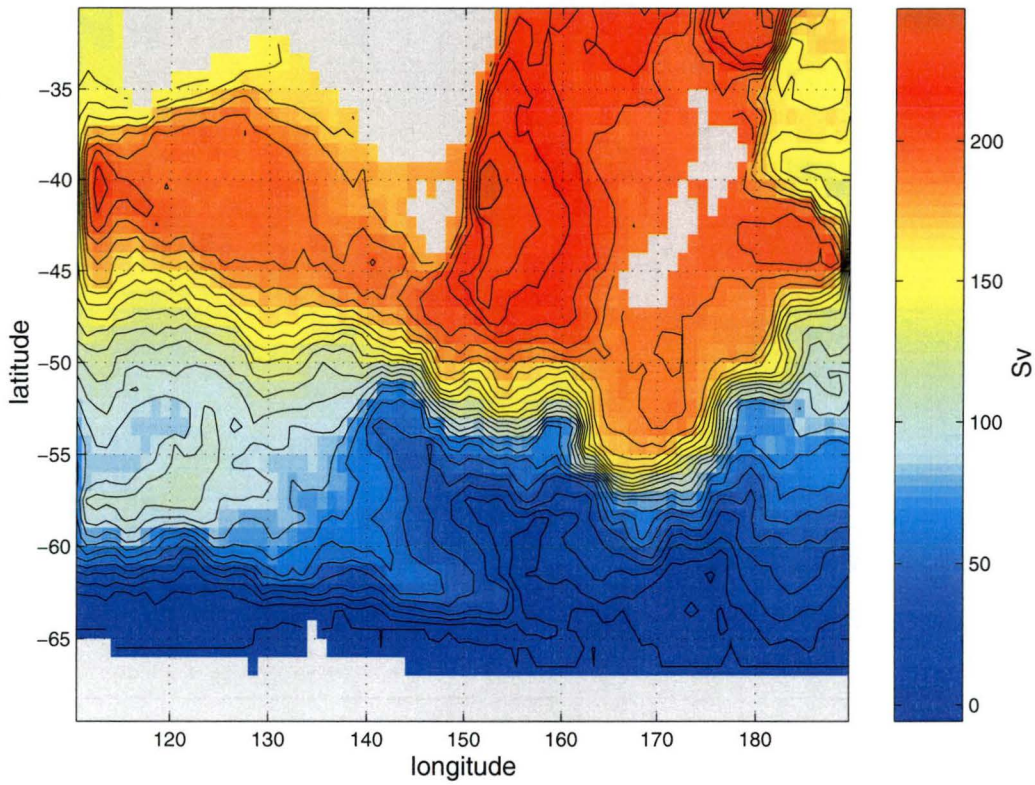


Figure 3.6: Streamfunction (EVEN grid) after two years with the bottom-referenced clamped boundary condition on normal velocity. Increments of 10 Sv are contoured in black.

streamfunction is defined as follows:

$$\psi(x,y) = \int_0^y U(x,y')dy',$$

where the boundary $y = 0$ is taken at the Antarctic continent in the south. This definition is opposite in sign to that conventionally used, but provides more useful values of total eastwards mass transport at a particular location integrated from the Antarctic coast. The circulation is cyclonic around lows of the streamfunction. Figure 3.6 shows this streamfunction after two years. Both the surface and (to a lesser degree) depth-integrated currents show clearly a meandering ACC as well as an East Australia Current.

However, the results at the end of five years are very different and very unrealistic. Figure 3.7 shows the surface currents and streamfunction. Before abandoning the bottom-referenced clamped condition, it is salient to document an example of the complex interaction between open boundary conditions for the various physical subsystems in HOPE. As described in greater detail later in section 3.7, the boundary condition for tracers is essentially upstream advection for outflow and relaxation to climatology for inflow. That is the "canonical" condition eventually used in the assimilation study, and is the condition used here. It also includes along-boundary advection. However, in the present experiment, a completely different result emerges if no along-boundary advection is performed on the northern boundary. The results after five years for

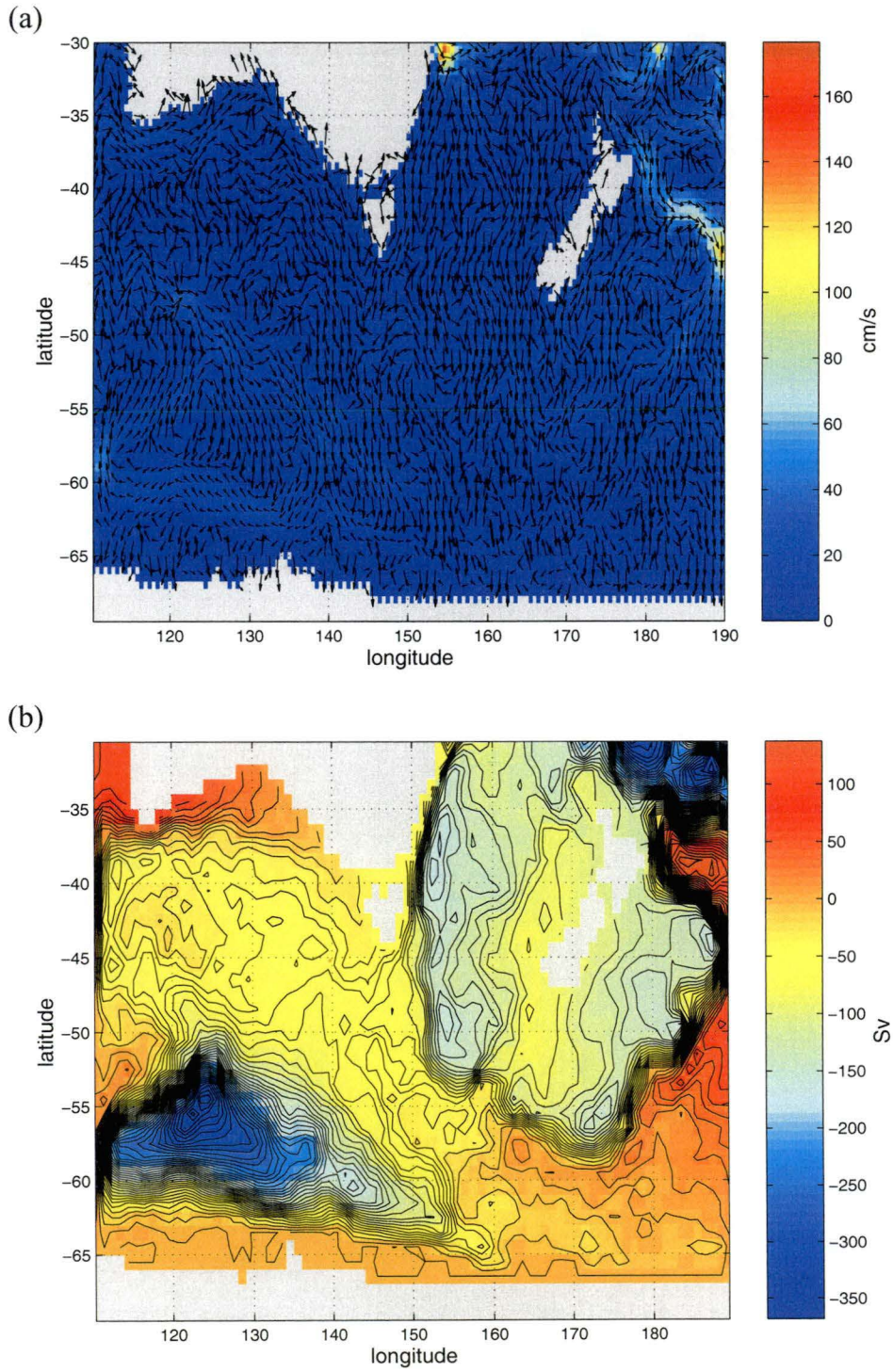


Figure 3.7: (a) Surface currents, and (b) streamfunction (EVEN), after five years for the clamped bottom-referenced condition on normal velocity.

surface currents and streamfunction are shown in Figure 3.8. Not only has the run been stabilised, it has provided even better results: the streamfunction after five years is more realistic than after two years *with* along-boundary advection of tracers on the northern boundary (Figure 3.6). If the bottom-referenced clamped condition were all that was available for the barotropic system, this scheme may be considered acceptable for the tracers. However there is no other good reason *not* to include along-boundary advection (section 3.7), and there are numerous other possibilities for the barotropic system yet to be examined. The results are included here merely to demonstrate just how sensitive are the interactions between boundary conditions for the various physical subsystems of HOPE to rather subtle variations.

A significant problem with this clamped boundary condition is the net volume loss from the domain. Instead of bottom-referencing geostrophic flow fields, it is possible to derive alternate normal transport profiles along the boundaries which do conserve volume. It is best to be guided by knowledge of the flows in the region. For the Antarctic Circumpolar Current, there is some observational data on total volume transport. From current measurements in Drake Passage, Whitworth et al. (1982) reported a transport of around 130 Sv while Nowlin and Klinck (1986) found 134 Sv. More recently, Rintoul and Bullister (1999) found 160 Sv for the geostrophic transport referenced to the deepest common depth from a winter repeat of the WOCE SR3 transect between Tasmania and Antarctica (Yaremchuk et al. 2001). From six occupations of this repeat section, Rintoul and Sokolov (2001) found a mean transport of 147 ± 10 Sv. Combining current meter and hydrographic data in the Subantarctic Frontal region south of Tasmania, Phillips (2000) estimated the mean transport of the Subantarctic Front to be 116 ± 10 Sv. For the northern boundary, there is roughly 15 Sv net inflow comprised as follows:

- 5 Sv of inflow representing the *Leeuwin Current* in a zone roughly 1.5° wide off the coast of Western Australia (Tomczak and Godfrey 1994; Reason and Pearce 1996)

- 30 Sv of inflow representing the *East Australia Current* (EAC) in a zone around 1.5° wide off the coast of New South Wales (Ridgway and Godfrey 1994; Ridgway and Godfrey 1997; Chiswell et al. 1997; Tsimplis et al. 1998)

- 20 Sv of northwards return flow for the EAC consisting of 15 Sv in a zone 2° wide adjacent to the EAC with another 5 Sv in a zone 15° wide further east (Ridgway and Godfrey 1994; Ridgway and Godfrey 1997; Chiswell et al. 1997; Tsimplis et al. 1998)

An alternative to point-wise bottom-referencing of the geostrophic profiles is to enforce realistic integrated volume transports along the boundaries.

There are several ways to do this. One is simply to scale the bottom-referenced transport profiles along each boundary to obtain a zero net outflow. This was implemented here by first calculating, at each timestep, the bottom-referenced profiles along the boundaries as before. These profiles were

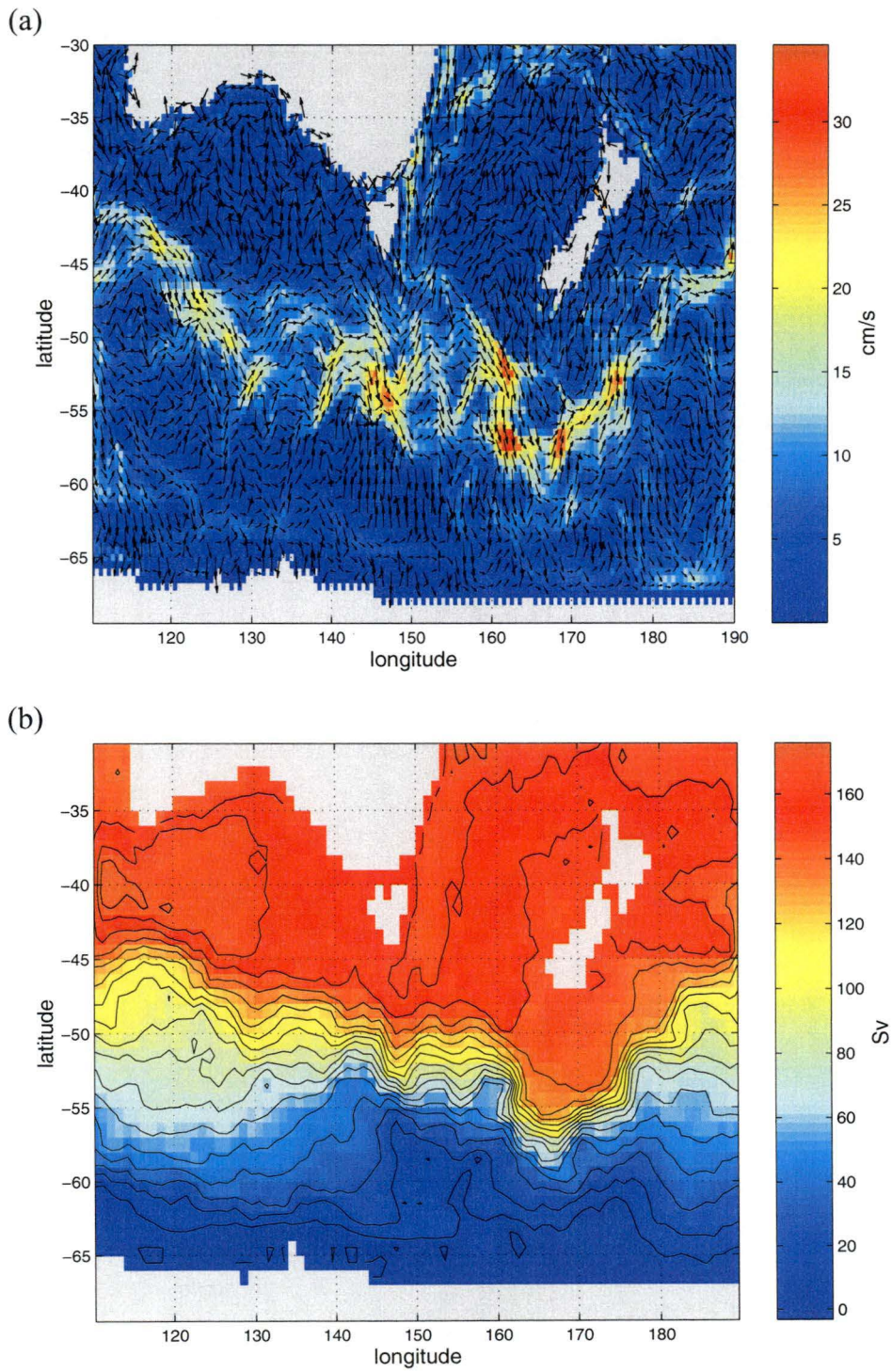


Figure 3.8: Same as Figure 3.7 with no along-boundary advection of tracers on the northern boundary.

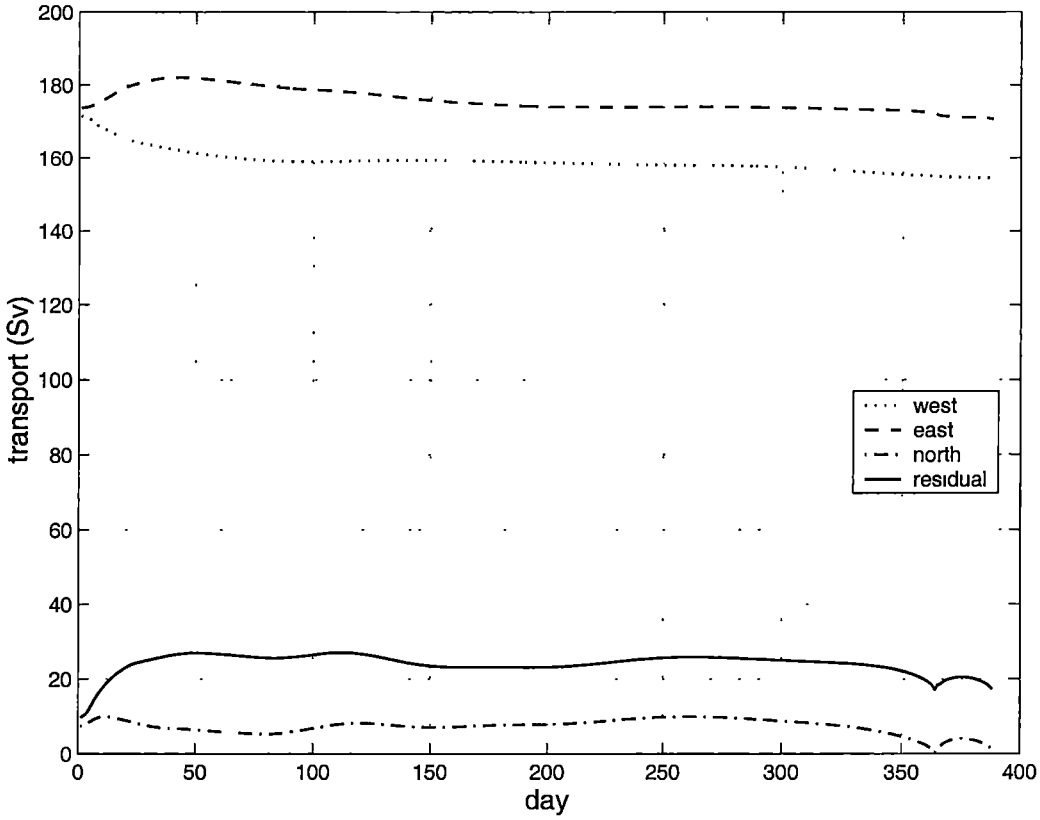


Figure 3.9: Bottom-referenced geostrophic boundary transport prior to uniformly scaling each timestep to 140 Sv (west), 155 Sv (east) and -15 Sv (north).

then uniformly scaled along their respective lengths to obtain 140 Sv inflow at the western boundary, 155 Sv outflow at the eastern boundary and 15 Sv inflow at the northern boundary (corresponding to the transport census noted above). This procedure can be regarded as a very simple inverse approach that constructs a pseudo-reference level by conserving mass transport at each timestep, but not tracer fluxes. A full inverse approach would be considerably more complex, requiring a solution almost every timestep in order to maintain mass balance along the boundaries. This simple approach was earlier used by Treguier et al. (1999; 2000), who noted that Killworth (1992) had found an *equivalent barotropic mode* in FRAM. Gan et al. (1998), Ezer and Mellor (1997), and Barnier et al. (1998) constructed fixed barotropic profiles along the two meridional Southern Ocean boundaries (south of Africa, and at Drake Passage) in their respective Atlantic models. The procedure failed here after approximately 390 days due to the bottom-referenced transport on the northern boundary approaching zero. Of course an attempt to scale this up to 15 Sv inflow results in unbounded scaling coefficients. Figure 3.9 shows the bottom-referenced geostrophic transport along each of the boundaries prior to the scaling procedure. This should be compared with the results for the unscaled case, Figure 3.3(a). The differences on the northern boundary arise through the influence of the altered boundary velocities on the tracer fields. It is interesting to note that the unscaled transports have evolved similarly

over the first year for both the eastern and western boundaries, but the northern boundary is indicating a type of positive feedback between the imposed transport (-15 Sv) and the tracer fields — the bottom-referenced transport is tending towards negative values at the end of the experiment. Apart from the failure after 390 days, the procedure works as expected: mean sea-level remains at zero to within numerical precision. A second serious problem with this approach is that the transport profile imposed along the northern boundary changes sign from its bottom-referenced value. For most of the first year, the bottom-referenced transport across the northern boundary is between 5 and 10 Sv *outflow*. The scaling procedure described here requires that there be 15 Sv *inflow*. This is achieved by multiplying the profile by a negative coefficient at each timestep. Figure 3.10 shows both the bottom-referenced transport profile and its scaled version at 30 days into the run. The initial profile, while representing a net (7.3 Sv) outflow, nevertheless includes around 8.5 Sv of inflow at the location of the EAC. After scaling of the integrated transport, this becomes around 17 Sv outflow for the entire section; the EAC has been forced to flow in the wrong direction.

While still using the simple approach of scaling the boundary transport profiles to obtain zero net outflow, a better method is to separately scale each of the two continuous sections of open boundary in the model: the first along the western and northern boundary up to the coast of Western Australia, and the second along the eastern and northern boundary up to the east coast of Australia. This will prevent both the problem of scaling a zero northern boundary transport up to 15 Sv inflow, and the transformation of the EAC to a strong northwards current. Thus, a five-year run was performed where the bottom-referenced normal velocity profiles along both continuous sections of open boundary were scaled each timestep to 160 Sv (inflow for the west/north boundary section, outflow for the east/north section). Unlike the previous run which failed after 390 days, this run remained stable for the duration of the experiment. The surface and depth-integrated flows at the end of five years are shown in Figure 3.11. While there is a weak EAC at the surface, there is no evidence of it at depth; in fact Figure 3.12 shows clearly a northwards flow at 1618 m depth. As intended, there is no net volume loss or gain from the domain — mean sea-level remains unchanged over the five years. So, while this scaling procedure has stabilised the run and achieved no net volume change, the East Australia Current appears to be poorly represented.

The fourth and final method discussed here for producing the prescribed normal velocities in the clamped condition is to specify an artificial transport profile along the northern boundary. Based on observed transports (mentioned earlier) of the Leeuwin Current, and East Australia Current with its return flow (Ridgway and Godfrey 1994) an artificial profile was constructed, and is shown in Figure 3.13. The clamped boundary condition prescribed this profile along the northern boundary, while scaling the bottom-referenced profiles along the western and eastern boundaries to 140 Sv and 155 Sv respectively. The bottom-referenced transport across the eastern and western boundaries prior to scaling each timestep is shown in Figure 3.14. Compared with Figure 3.3 for raw bottom-referenced profiles, the model transports have been signifi-

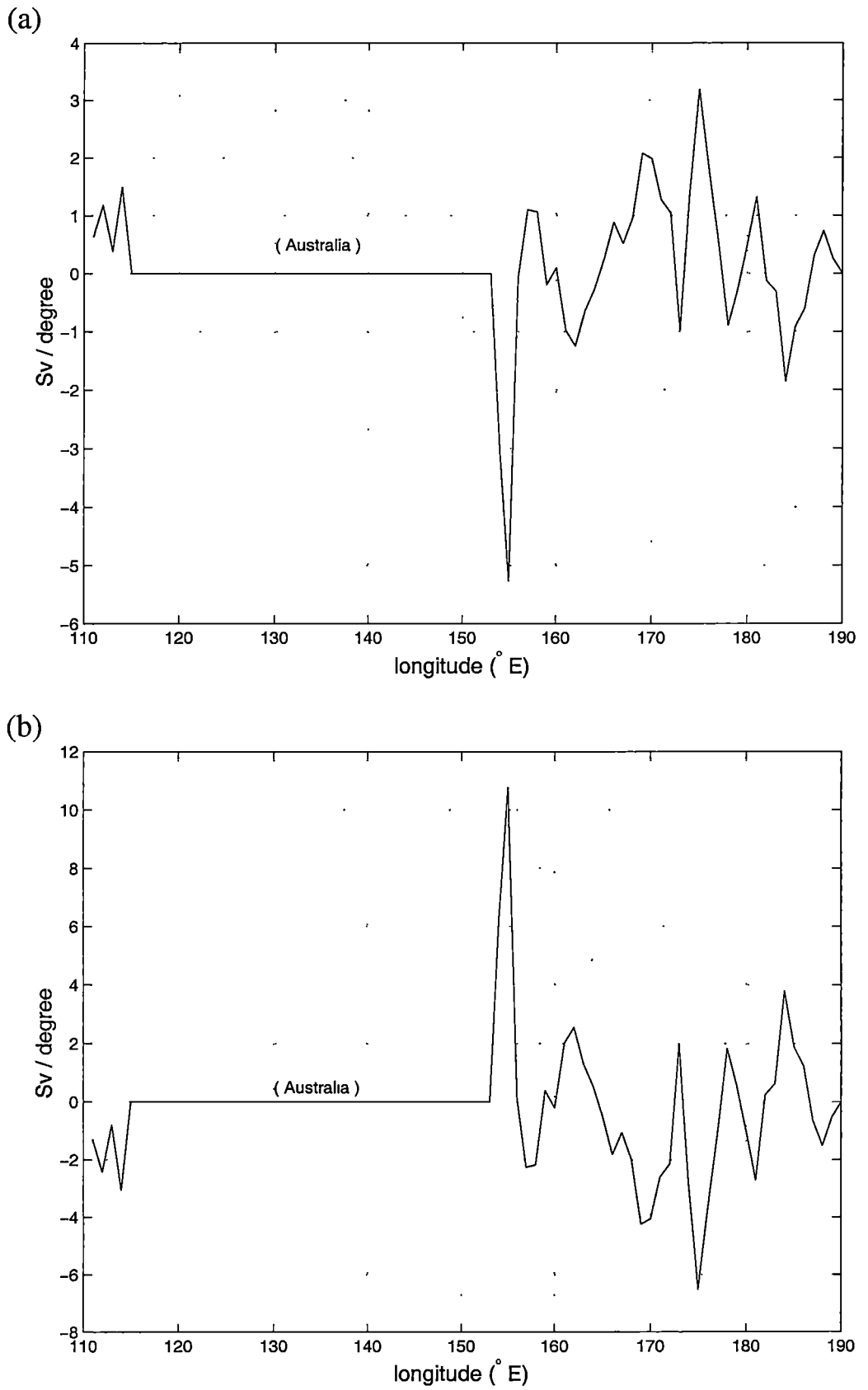


Figure 3.10: (a) Bottom-referenced (7.3 Sv outflow), and (b) scaled (-15 Sv inflow), barotropic velocity profiles across the northern boundary after 30 days.

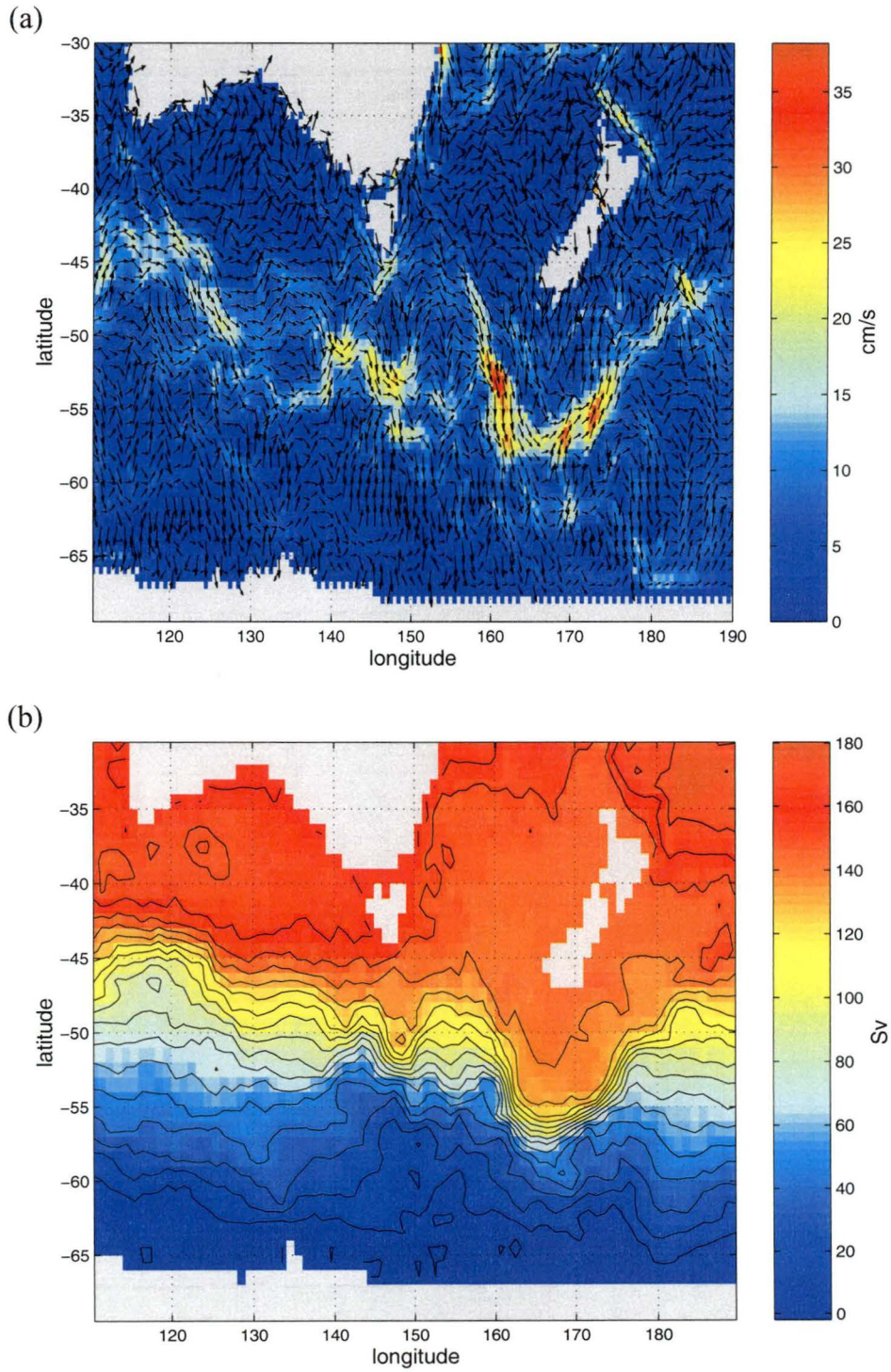


Figure 3.11: (a) Surface currents, and (b) streamfunction (EVEN), after five years for barotropic profiles scaled along each of the two open boundary sections.

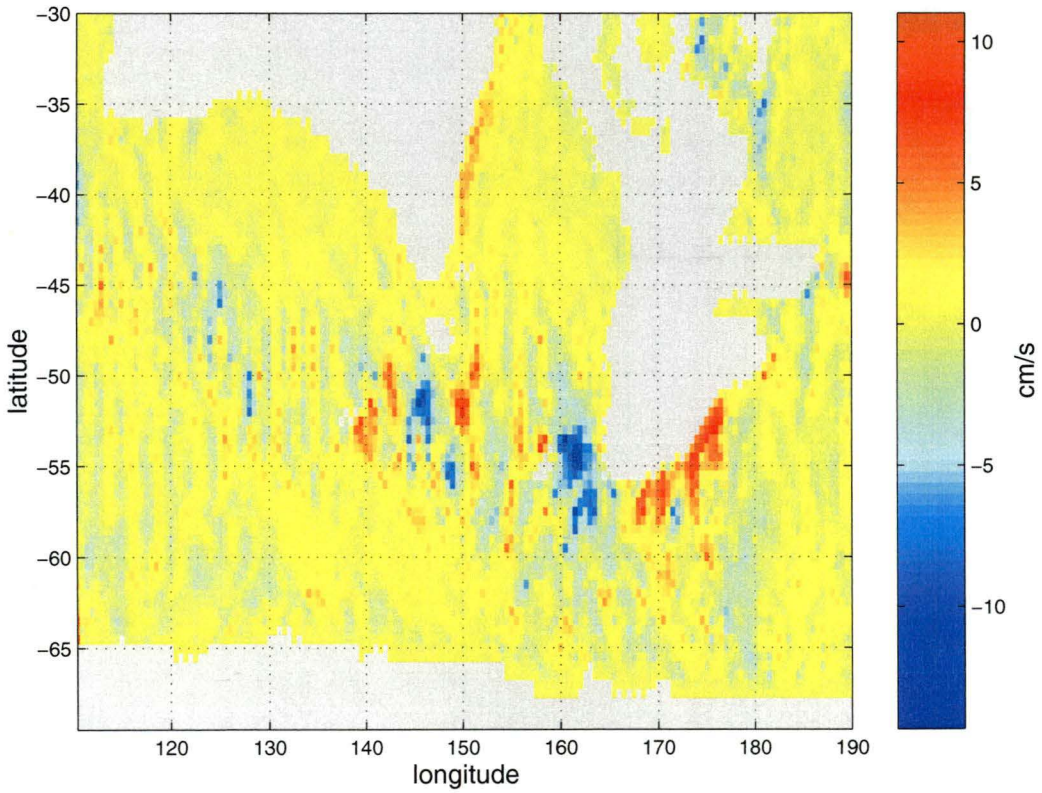


Figure 3.12: Currents at 1618 m (layer 12) after five years for barotropic profiles scaled along each of the two open boundary sections. Note the northwards flow at depth for the EAC.

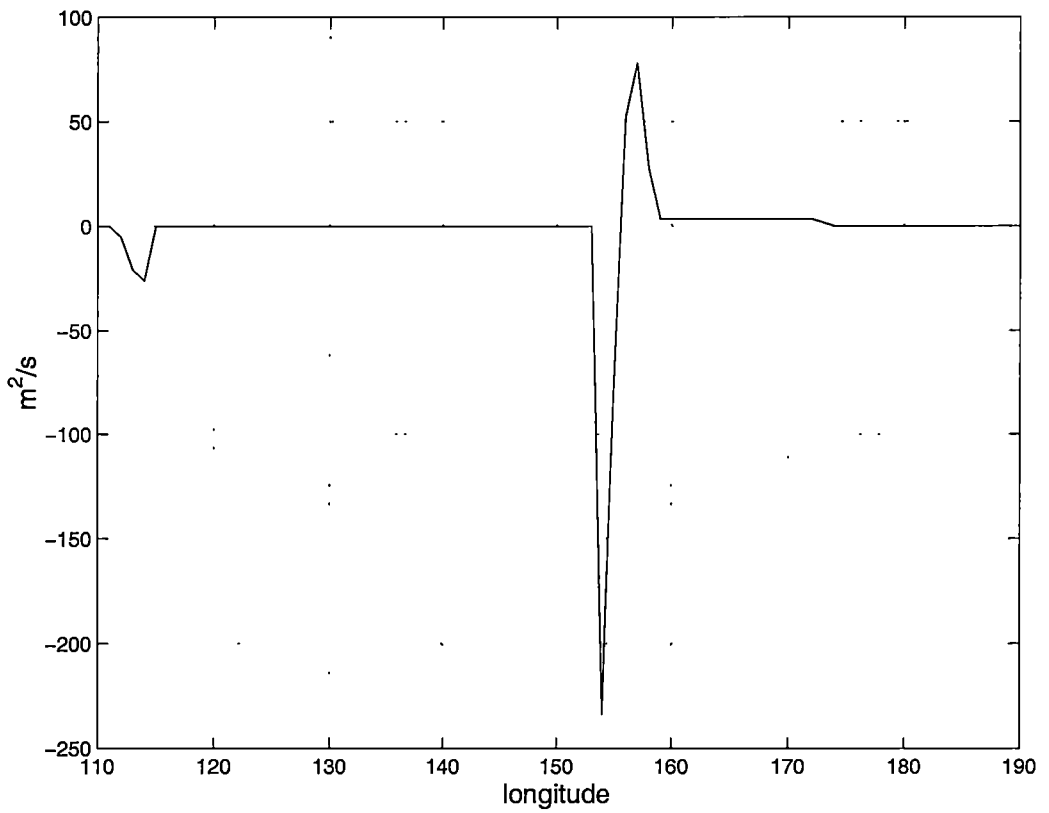


Figure 3.13: Artificial barotropic velocity (V) profile along the northern boundary representing 5 Sv of Leeuwin Current, 30 Sv of East Australia Current, and 20 Sv of EAC return flow.

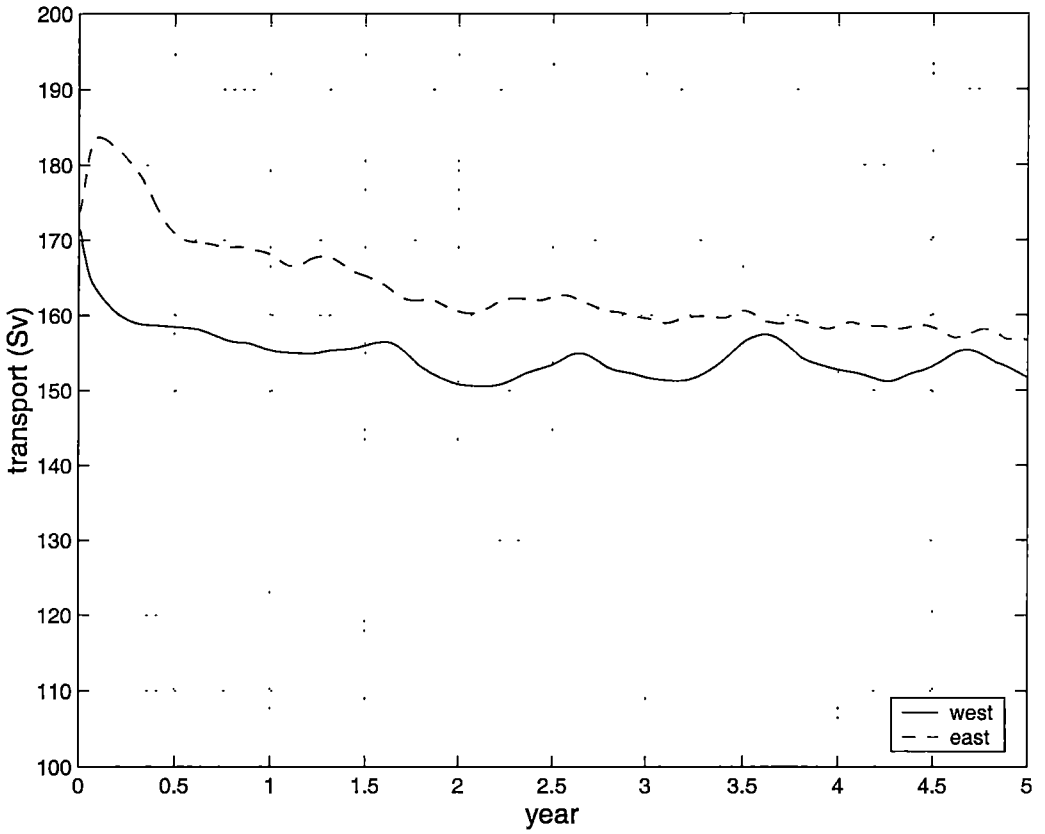


Figure 3.14: Bottom-referenced transport (prior to scaling) integrated along western and eastern boundaries for the clamped condition with an artificial profile along the northern boundary.

cantly stabilised. This is confirmed by examining surface and depth-integrated currents after five years, Figure 3.15. It is interesting that while the depth-integrated EAC extends south along the coast of Australia, its surface signature separates just south of the northern boundary (around 31°S). A clear Tasman Front is seen extending across the Tasman Sea and around the northern tip of New Zealand. This is similar to observations with infrared satellite imagery (Mulhearn 1987). As required, mean sea-level remains at zero within numerical precision. Compared with the results for scaling the continuous west/north and east/north boundary sections (Figures 3.11 and 3.12), the artificial profile along the northern boundary is clearly superior, since it retains an EAC. The streamfunction shown in Figure 3.15 is for the combined subgrids, and shows some separation between the subgrids. Grid separation results from other open boundary condition trials were typically not much better than this. In fact, if the boundary condition for the barotropic system is not implemented (as discussed in Appendix C) by interpolating the transport profile into one subgrid from the other at the boundaries, the results are considerably worse. The separation for sea-level is not as severe (Figure 3.16), which is fortunate since sea-level consistency is more important than streamfunction for assimilation of altimetry data.

Finally, the performance of the clamped boundary condition is evaluated against its ability to transmit disturbances generated in the model interior. A transient gaussian sea-level disturbance was applied in the middle of the model domain after a short 10 day spinup from rest. This allowed sufficient time for initial adjustment to occur. A subsequent 10 day integration was compared with an identical unperturbed run. Both the potential and kinetic energy of the disturbance, integrated over the domain, were calculated during the second ten-day interval. These are defined respectively (Gill 1982) by

$$PE = \int\int \rho g \eta'^2 \, dx \, dy$$

and

$$KE = \int\int \frac{\rho}{H}(U'^2 + V'^2) \, dx \, dy,$$

where primes represent time perturbation quantities. A timeseries of these energies over the second ten day period is shown in Figure 3.17. The total energy of the perturbation settles to around 30% of its initial value after ten days. While there is a slight oscillatory exchange between potential and kinetic energy over the first day or so, the total energy smoothly decreases over the period. Since the clamped boundary condition exactly conserves volume in the domain, the asymptotic value of potential energy is determined by the mean size of the sea-level perturbation (and corresponds to the initial disturbance having flattened out to a uniform level everywhere). Without performing a full energy analysis, the gradual decrease in total energy is probably due, at least partly, to the well-known numerical dissipation of fully implicit time-differencing schemes (Richtmyer and Morton 1967; Wolff et al. 1997). Evidence for numerical dissipation may be garnered by performing an identical experiment using, instead of a fully implicit discretisation ($\alpha = 1$, $\beta = 1$ for

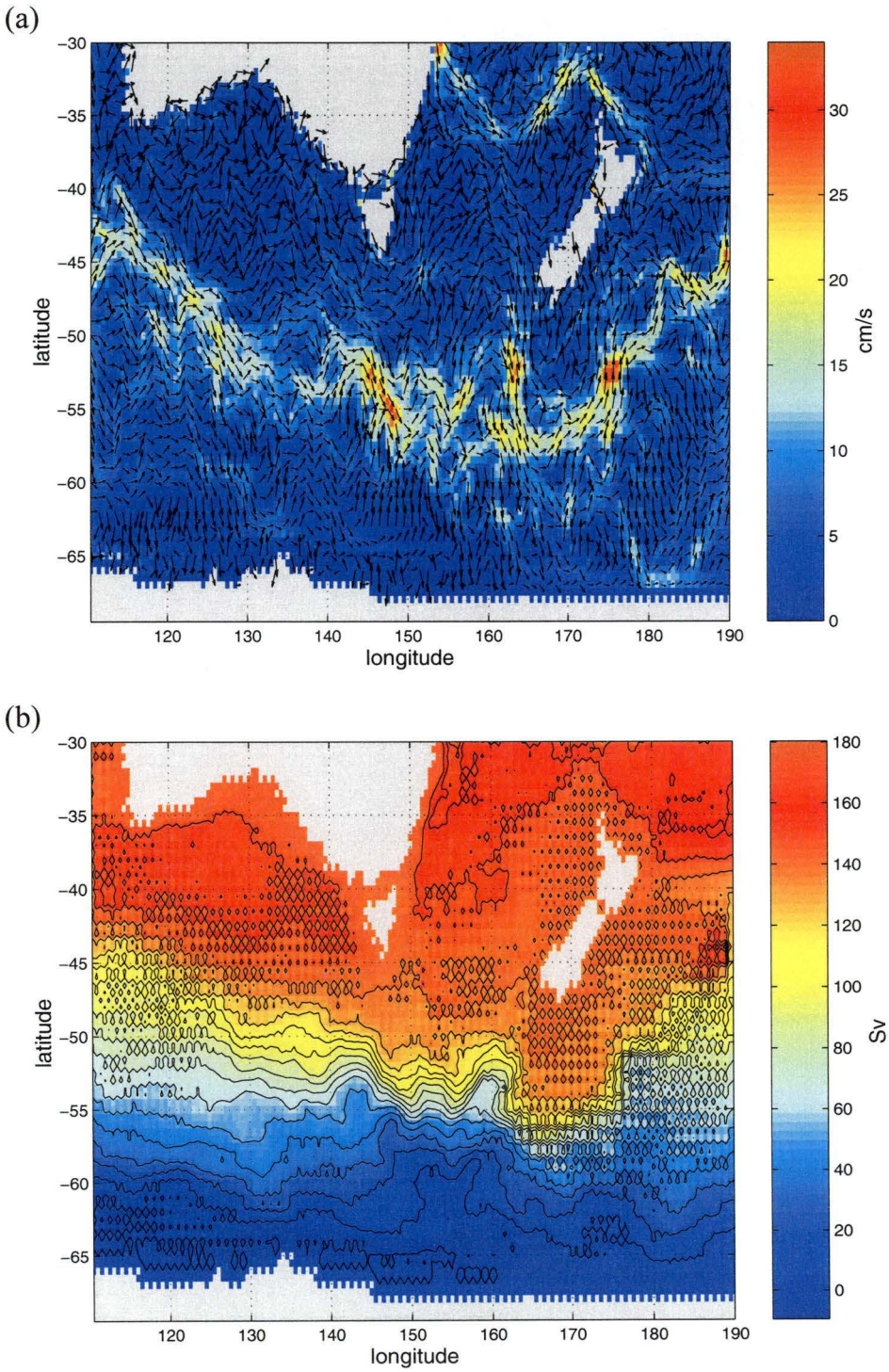


Figure 3.15: (a) Surface currents, and (b) streamfunction, after five years for the scaled bottom-referenced condition on normal velocity with an artificial northern boundary profile.

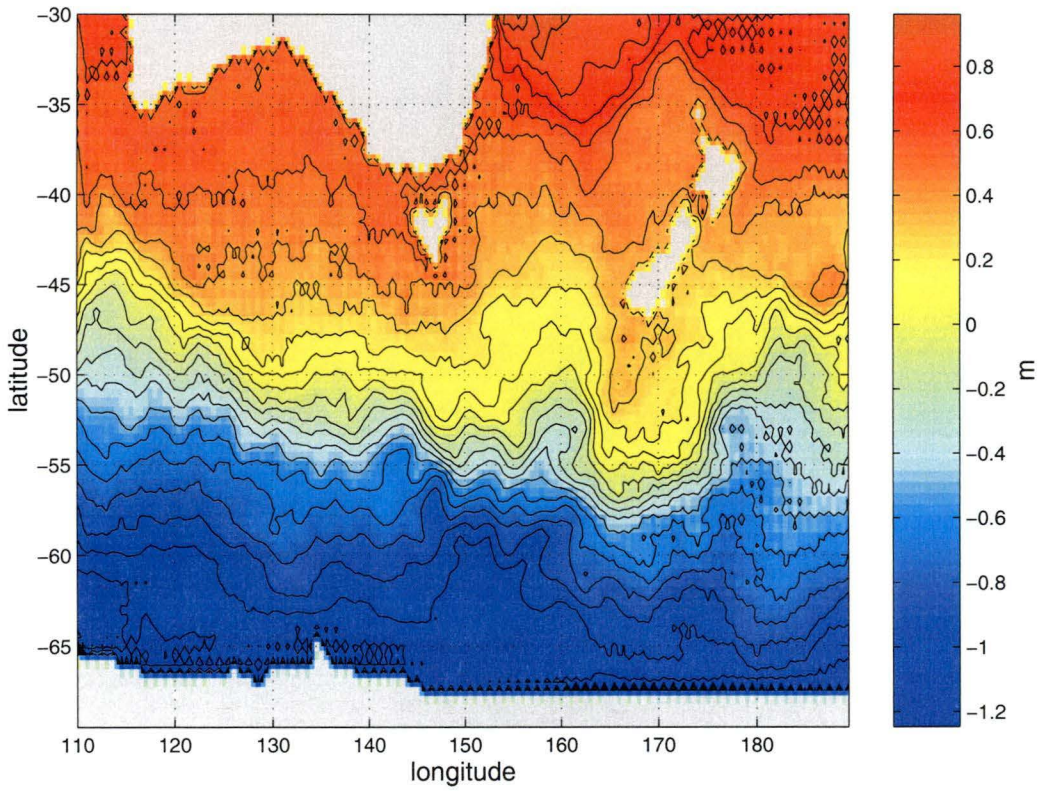


Figure 3.16: Sea-level (m) for combined subgrids after five years for the clamped condition with an artificial profile along the northern boundary. Black contours are at intervals of 10 cm.

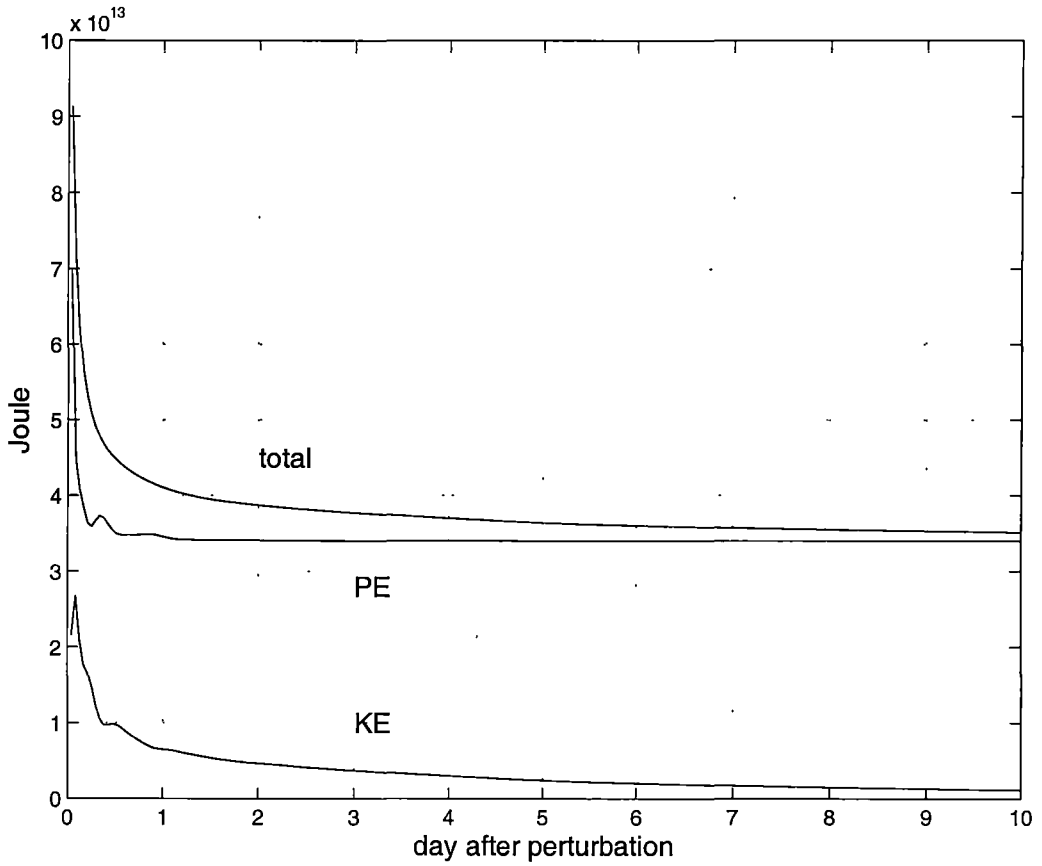


Figure 3.17: Potential (PE), kinetic (KE) and total energies of perturbation over a ten day period for the clamped boundary condition on normal velocity.

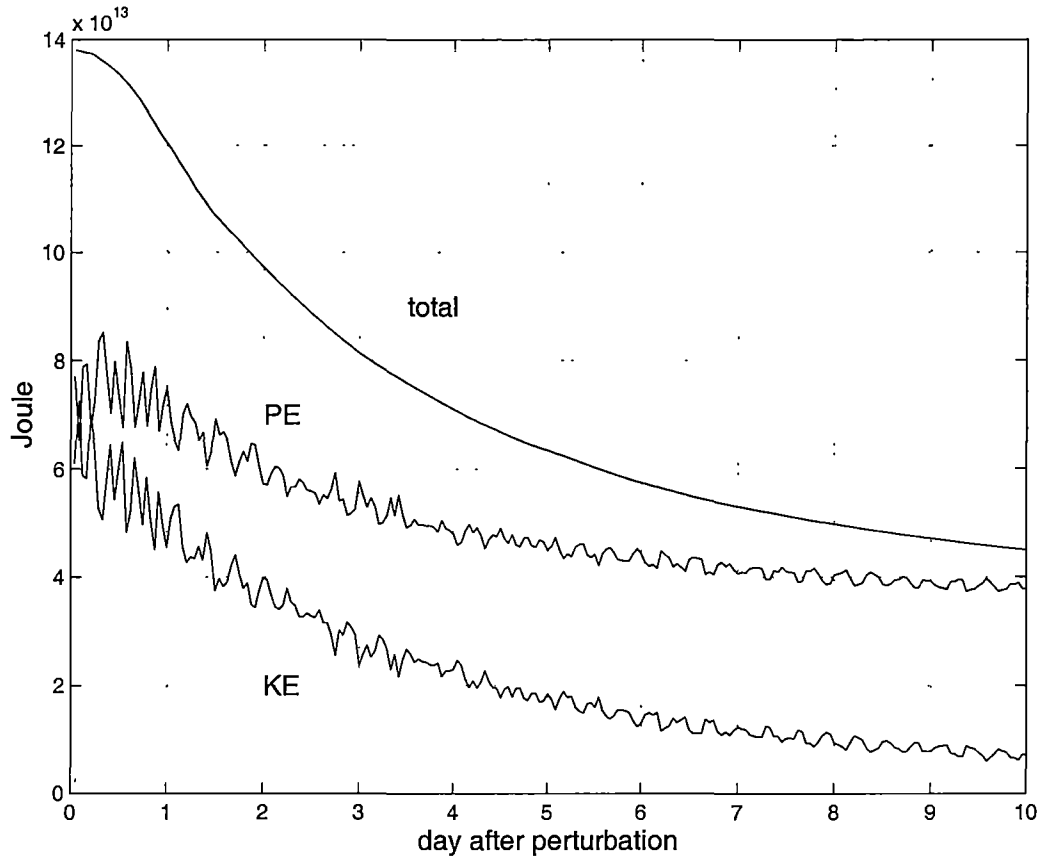


Figure 3.18: Potential (PE), kinetic (KE) and total energies of perturbation over a ten day period for the clamped boundary condition with partly implicit time discretisation.

the barotropic stability parameters, equations (2.38), (2.39), (2.40) in section 2.1.3) a discretisation which is only partly implicit ($\alpha = 0.5$, $\beta = 0.5$). The results for this experiment are shown in Figure 3.18. A much weaker dissipation of energy can indeed be seen, as well as more vigorous exchange between potential and kinetic. Nevertheless, the model again retains around 30% of the initial perturbation energy. Were the numerical scheme not dissipative and bottom-friction in the model set to zero, the particularly strong exchange (with timescales associated with surface gravity-waves) between potential and kinetic energy seen in the first day might be expected to continue undiminished. In an assimilation model, unless perfectly balanced perturbations are applied, the retention of 30% of the residual unbalanced energy in each analysis may contaminate the results.

Summarising the clamped condition on normal velocity, an unchanging mean sea-level can be assured by prescribing volume transports which sum to zero (net) across the boundaries. This can be achieved by scaling bottom-referenced geostrophic velocity profiles along the open boundaries to realistic transport values. A simulated ACC in the model appears reasonably robust to the details of this procedure, but a realistic EAC is harder to maintain. An approach found to be successful was to impose an artificial transport profile

along the northern boundary based on observed total transport values. The clamped condition was reflective to surface gravity waves, retaining around 30% of the energy of a disturbance in the fully implicit discretisation of HOPE's barotropic system used in this study.

Gravity-wave radiation

The *gravity-wave radiation* condition is a radiation condition (section 3.2) using a phase speed fixed at the local surface gravity wave speed, $c = \sqrt{gH}$. In principle, it should perform better than the clamped condition at transmitting disturbances. It is written:

$$\begin{aligned} \frac{\partial U}{\partial t} &= \pm c \frac{\partial U}{\partial x}, && \text{western (+) and eastern (−) boundaries} \\ \frac{\partial V}{\partial t} &= -c \frac{\partial V}{\partial y}, && \text{northern boundary.} \end{aligned} \tag{3.27}$$

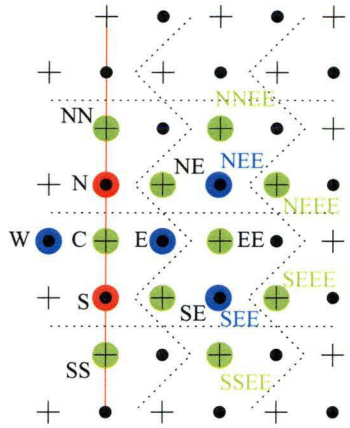
A modified scheme including a temporal relaxation term is also considered later. Since c is constant in (3.27), this condition satisfies the requirements (3.21)–(3.23) for implementation in implicit form in HOPE's barotropic system. It can be discretised in time with an 'implicitness' parameter γ, *e.g.* for the western boundary:

$$\begin{aligned} \frac{U_b^{n+1} - U_b^n}{\Delta t} &= c \frac{\gamma(U_{b+1}^{n+1} - U_b^{n+1}) + (1-\gamma)(U_{b+1}^n - U_b^n)}{\Delta x}, \\ U_b^{n+1} &= pU_{b+1}^{n+1} + qU_b^n + sU_{b+1}^n, \end{aligned} \tag{3.28}$$

where

$$\begin{aligned} p &= \frac{\gamma r}{1+\gamma r}, \\ q &= \frac{1-(1-\gamma)r}{1+\gamma r}, \\ s &= \frac{(1-\gamma)r}{1+\gamma r}, \\ \text{and } r &= c\Delta t/\Delta x. \end{aligned}$$

As described in Appendix C, this was implemented with a procedure that averaged the volume transport along the boundaries between the subgrids. This ensured that at each timestep any net volume change in the domain was the same on each subgrid, in an attempt to minimize sea-level divergence between the grids. It meant that the implicit prognostic equation for ODD sea-level on the boundaries depended on a horizontal grid-point stencil larger than that required in the interior. This is illustrated in Figure 3.19. This extension of the finite-difference stencil has an impact on computational efficiency of the model. As mentioned earlier (page 50), the barotropic system matrix is ordered along one or other of the horizontal grid dimensions. Because of the finite size of the stencil, the matrix has a finite-width of non-zero elements about the diagonal, limited by the stencil size. Only this central band of entries is stored.



Normally, sea–level point C would depend implicitly only on sea–level points NN, NE, EE, SE and SS. However, for the implicit gravity–wave radiation condition, the 'virtual' velocity point W is averaged from points N and S on the boundary. Since these latter depend, via the radiation condition, on points NEE and SEE, the surrounding sea–level points (NNEE, NEEE, SEEE, SSEE, NE, SE) are also involved implicitly in the computation. Points NE and SE were already involved: the inclusion of points NNEE, NEEE, SEEE and SSEE is new with this condition.

Figure 3.19: Extension of the sea-level finite-difference stencil at a boundary for the implicit gravity-wave radiation condition.

By increasing the stencil size, as in this boundary condition, a larger central band needs to be stored. In fact, the size of the matrix needs to be roughly doubled for the ordering in a north-south direction used here[5]. Since the barotropic system matrix accounts for almost one third the model's memory requirements (Wolff et al. 1997, Appendix D) in a standard configuration, this boundary condition carries a penalty on computational resources.

The energy transmission properties of this boundary condition are first evaluated with a perturbation experiment, as performed for the clamped boundary condition. A fully implicit discretisation of the radiation condition is used ($\gamma = 1$ in the discretisation (3.28)). The potential and kinetic energy of the perturbation are shown in Figure 3.20 for ten days after the perturbation was applied. One can immediately see that the transmission of energy from within the domain has been significantly improved in comparison to the clamped condition (Figure 3.17). At around 7.5 days, the total energy of the perturbation has reduced to around 2% of its original value. Also, the energy decay is dominated here by kinetic energy: the potential energy associated with a perturbed sea-level decreases rapidly. Of course, unlike the clamped boundary condition, the gravity-wave condition is not constrained to conserve mean sea-level.

Next, the general performance of the model under this boundary condition must be examined. Figure 3.21 shows the volume transports integrated along each boundary until the model failed at around 38 days. With transports through the western and eastern boundaries of up to 20000 Sv, the gravity-wave radiation condition clearly is completely useless in this form!

Varying the implicitness of the gravity-wave condition did not improve its numerical stability. Indeed a fully explicit scheme ($\gamma = 0$) produced boundary transports exceeding 40000 Sv and failed after only eight timesteps.

[5]This can be appreciated by noticing that the new sea-level points NEEE and SEEE (Figure 3.19) are one whole column further removed from C than the previous most extreme point EE. Thus an additional two columns of grid-points are needed (one at the eastern boundary, one at the west), extending the previous stencil by a factor of two. For an east-west ordering, a similar increase is required due to the extended stencil at the northern (or southern) boundary.

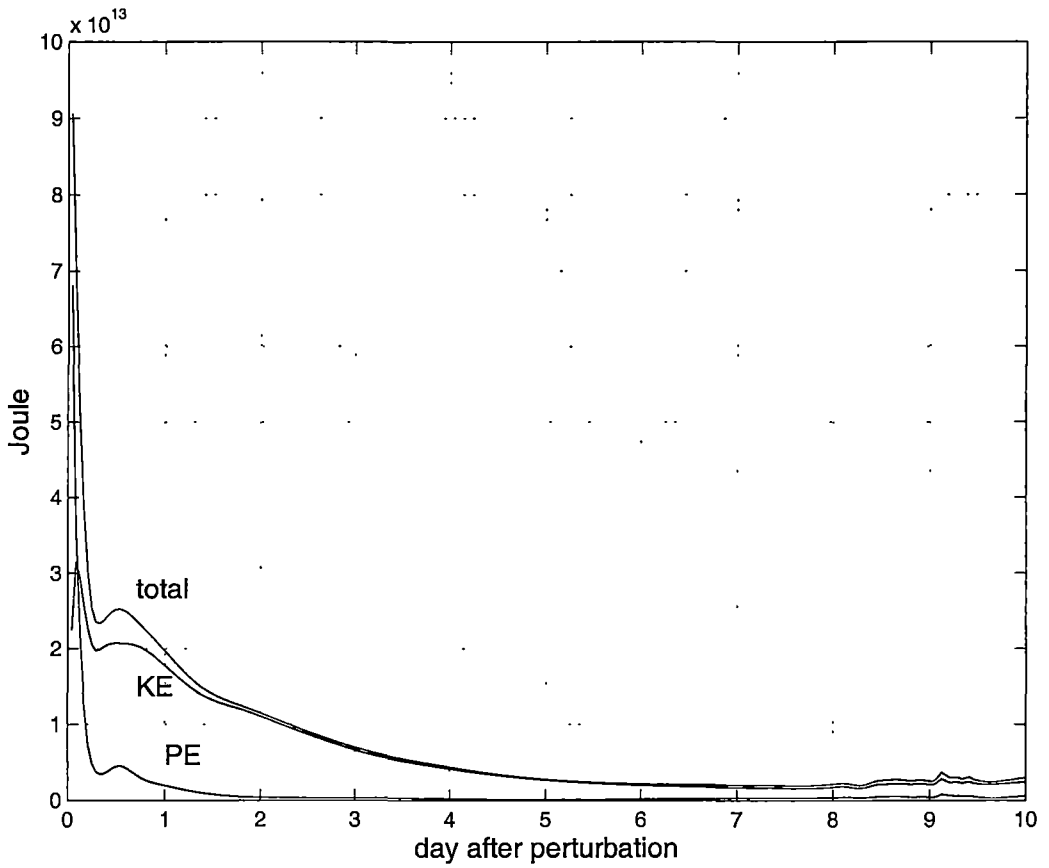


Figure 3.20: Potential (PE), kinetic (KE) and total energies of perturbation over a ten day period for the implicit gravity-wave radiation boundary condition.

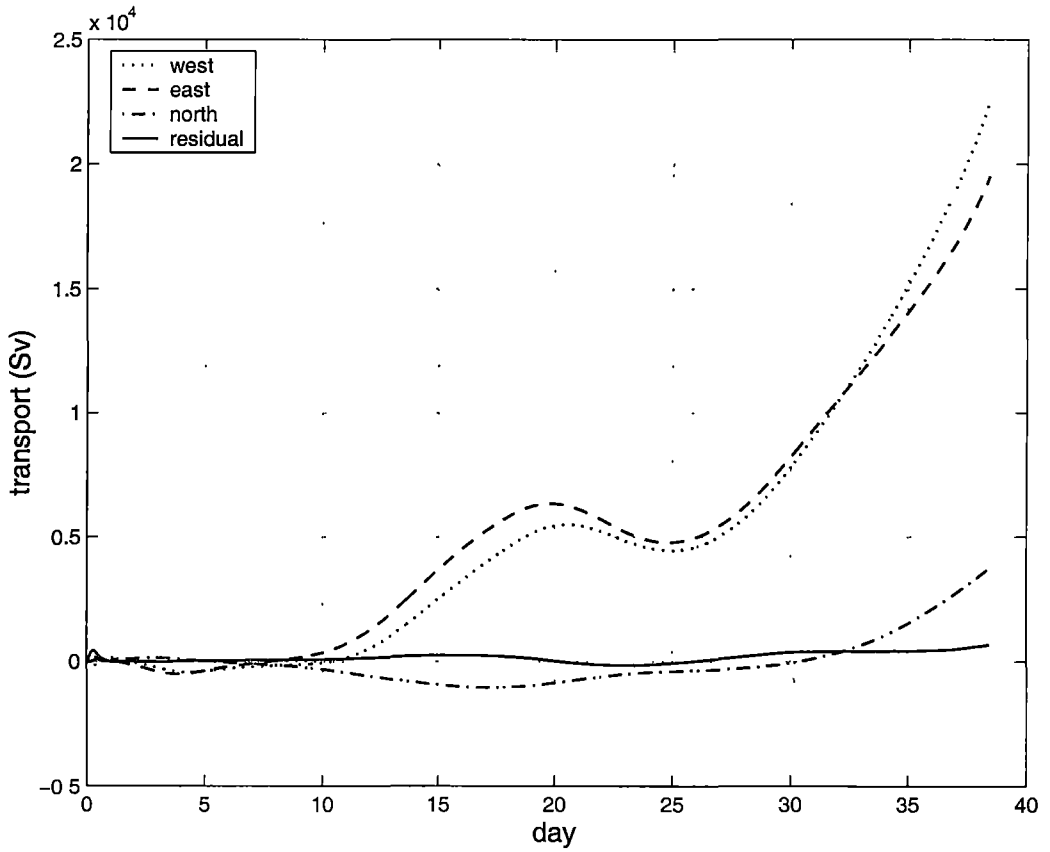


Figure 3.21: Volume transport across each open boundary for the implicit gravity-wave radiation condition.

In an attempt to stabilise the boundary condition, a relaxation term was added to the right-hand side, *e.g.*

$$\frac{\partial U}{\partial t} = c\frac{\partial U}{\partial x} + \frac{1}{\tau}(U_0 - U) \tag{3.29}$$

for the western boundary. Such a scheme was proposed by Blumberg and Kantha (1985) for tidal modelling in order to prescribe a known tidal flow while allowing the transmission of transients through the boundaries. The condition's performance was also reviewed by Chapman (1985) and Palma and Matano (1998). An 'implicitness' parameter, δ, was used for the relaxation term:

$$\frac{U_b^{n+1} - U_b^n}{\Delta t} = c\frac{\gamma(U_{b+1}^{n+1} - U_b^{n+1}) + (1-\gamma)(U_{b+1}^n - U_b^n)}{\Delta x} + \frac{1}{\tau}[U_0 - \delta U_b^{n+1} - (1-\delta)U_b^n].$$

The relaxation transport profile U_0 was set using the scheme found to be successful in the clamped condition: an artificial profile for the northern boundary, and bottom-referenced geostrophic profiles scaled to 140 Sv and 155 Sv total for the western and eastern boundaries respectively.

The relaxation term indeed stabilised the model, allowing runs of five years for various choices of relaxation timescales, τ, and implicitness parameters, γ and δ. As with the un-relaxed scheme described earlier (3.28), explicit discretisations of the gravity-wave part ($\gamma = 0$) were found to fail. For instance, even with a relaxation timescale as short as one hour, the scheme failed after eight timesteps. While the value of the parameter δ, controlling the implicitness of the relaxation part, certainly made a difference to results, no particular value seemed to provide a best choice. For instance, Figure 3.22 shows the streamfunction after five years for a relaxation timescale of one day, and both a fully implicit ($\delta = 1$) and fully explicit ($\delta = 0$) discretisation. Both experiments produce very unrealistic flows. The fully explicit case produces a cyclonic circulation at the western boundary, while the implicit case produces an anticyclonic circulation.

Reasonable circulation patterns were not obtained unless a very short relaxation timescale was used. Even a timescale of six hours produced unsatisfactory results (Figure 3.23 for $\delta = 0$, $\gamma = 0.5$). Reducing the relaxation timescale still further, to one hour, gave acceptable flow patterns, as shown in Figure 3.24.

This result was for values of the implicitness parameters $\delta = 0$ and $\gamma = 0.5$. The flow was marginally better than obtained from a fully implicit discretisation of the gravity-wave part ($\gamma = 1$). On the other hand, the variation in mean sea-level is considerably improved. Figure 3.25 shows the mean sea-level over five years for the half-implicit discretisation ($\gamma = 0.5$) and two fully implicit discretisations ($\gamma = 1, \delta = 0$ and $\gamma = 1, \delta = 1$). Again, there is no real difference for variations of the relaxation implicitness, δ, but the parameter γ, controlling the implicitness of the gravity-wave part, is important. Mean sea-level increases steadily up to 20 m for $\gamma = 1$ (fully implicit), while it oscillates between around -4 m and 8 m for $\gamma = 0.5$. Although the implicitness parameters δ and γ affect the results, it is clear that their effect is less significant than that of the relaxation timescale, τ.

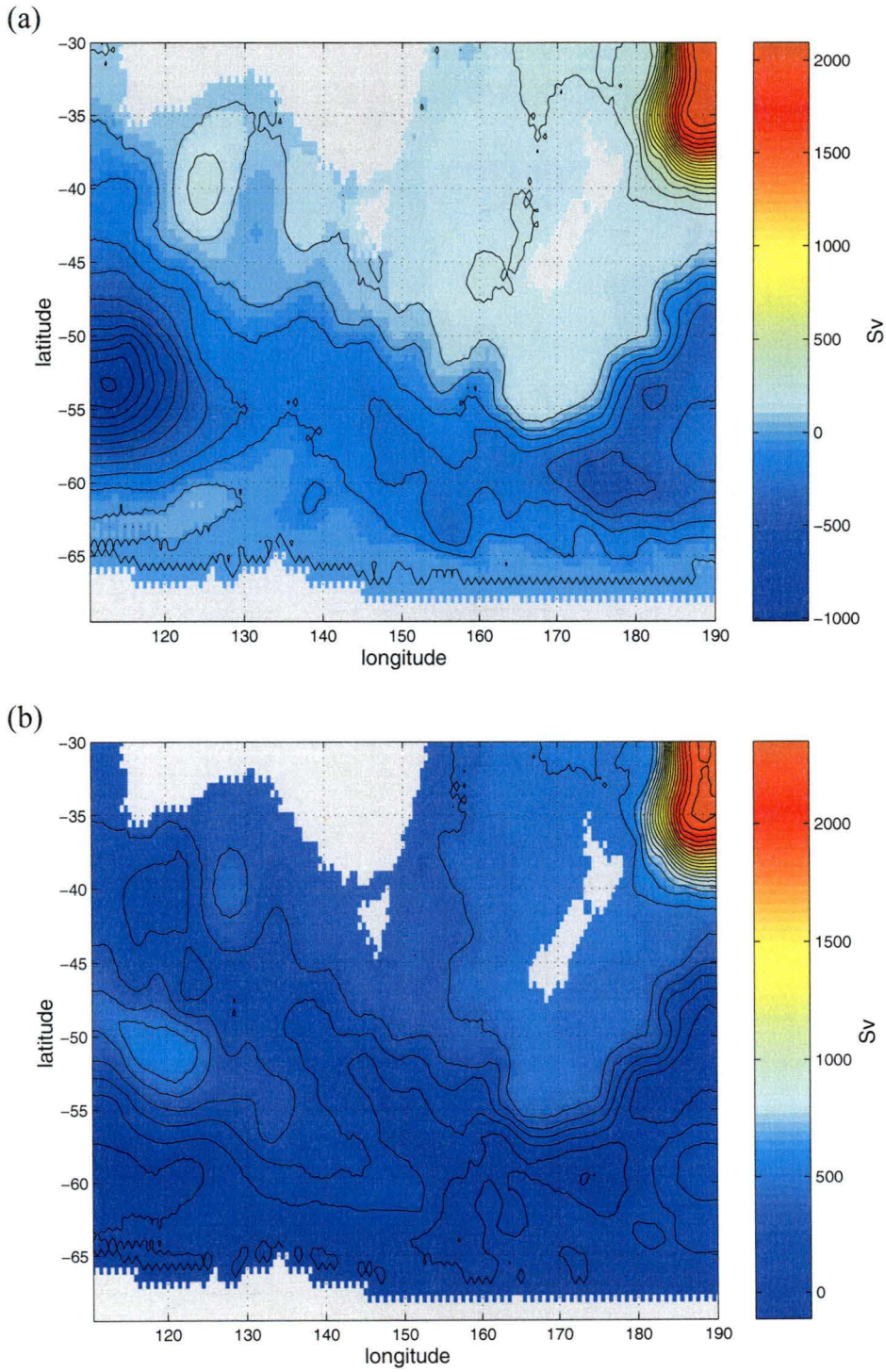


Figure 3.22: Streamfunction after five years for the gravity-wave condition with a one-day relaxation timescale; (a) fully explicit ($\delta = 0$), (b) fully implicit ($\delta = 1$). Contours are at intervals of 100 Sv.

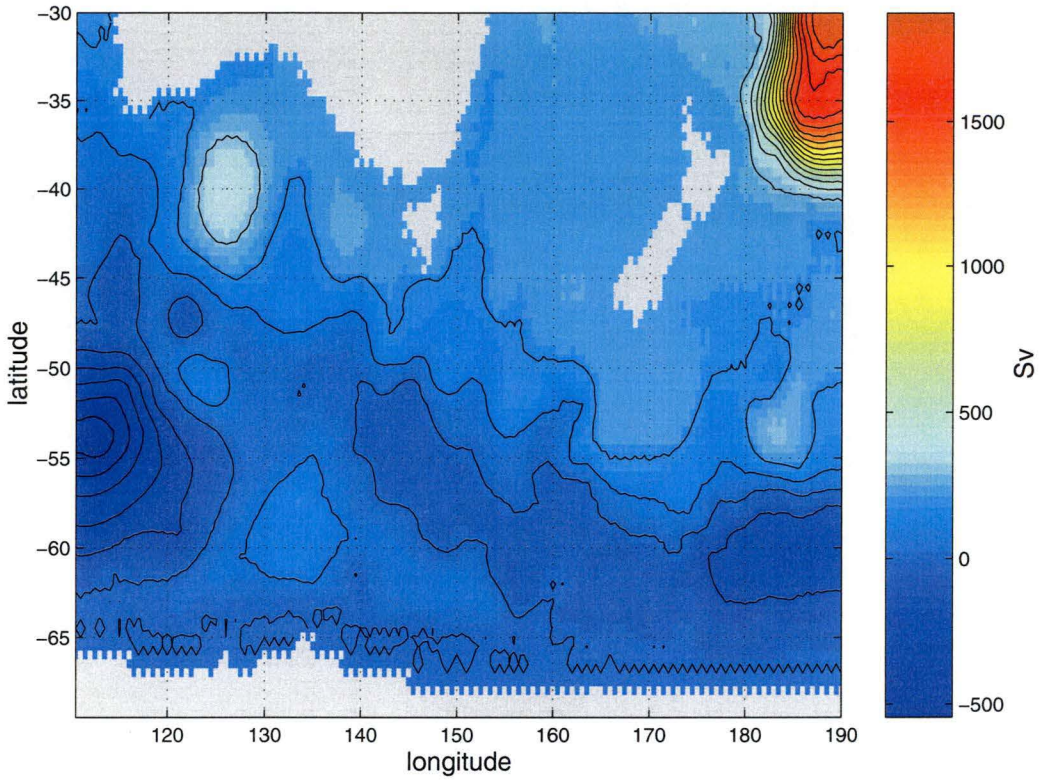


Figure 3.23: Streamfunction (contour interval 100 Sv) after five years for the gravity-wave condition with a six hour relaxation timescale and $\delta = 0$, $\gamma = 0.5$.

Figure 3.26 shows the integrated volume transports across each of the open boundaries for 100 days using a relaxation timescale of one hour and implicitness parameters $\gamma = 0.5$, $\delta = 0$ (the combination providing the best results). The transports across both the eastern and northern boundaries oscillate with a regular period of around 13 days and peak-to-peak amplitude around 100 Sv. They are almost in phase (strongest outflow on the eastern boundary corresponding to strongest inflow on the northern boundary). The western boundary oscillates with a range of around 50 Sv but with more than one timescale apparent. This general pattern continues throughout the five years with slow drifts being responsible for the net change in sea-level (Figure 3.25). It is rather remarkable that despite changes of 100 Sv every 13 days or so, the oscillations of boundary transport are essentially balanced, apart from a much slower drift.

Despite these unrealistic transport oscillations across the boundaries, it remains to be seen whether the imposition of flow relaxation in the gravity-wave scheme has reduced its effectiveness at radiating disturbances. The earlier perturbation experiment (see Figure 3.20) was repeated with the gravity-wave boundary condition including relaxation (with $\tau = 1$ hour and implicitness parameters $\gamma = 0.5$, $\delta = 0$). Figure 3.27 shows the energy evolution over ten days following the perturbation, integrated over the domain. The total energy decreases to around 20% of its initial value after 10 days. This is much worse than the pure gravity-wave condition (without relaxation) where only

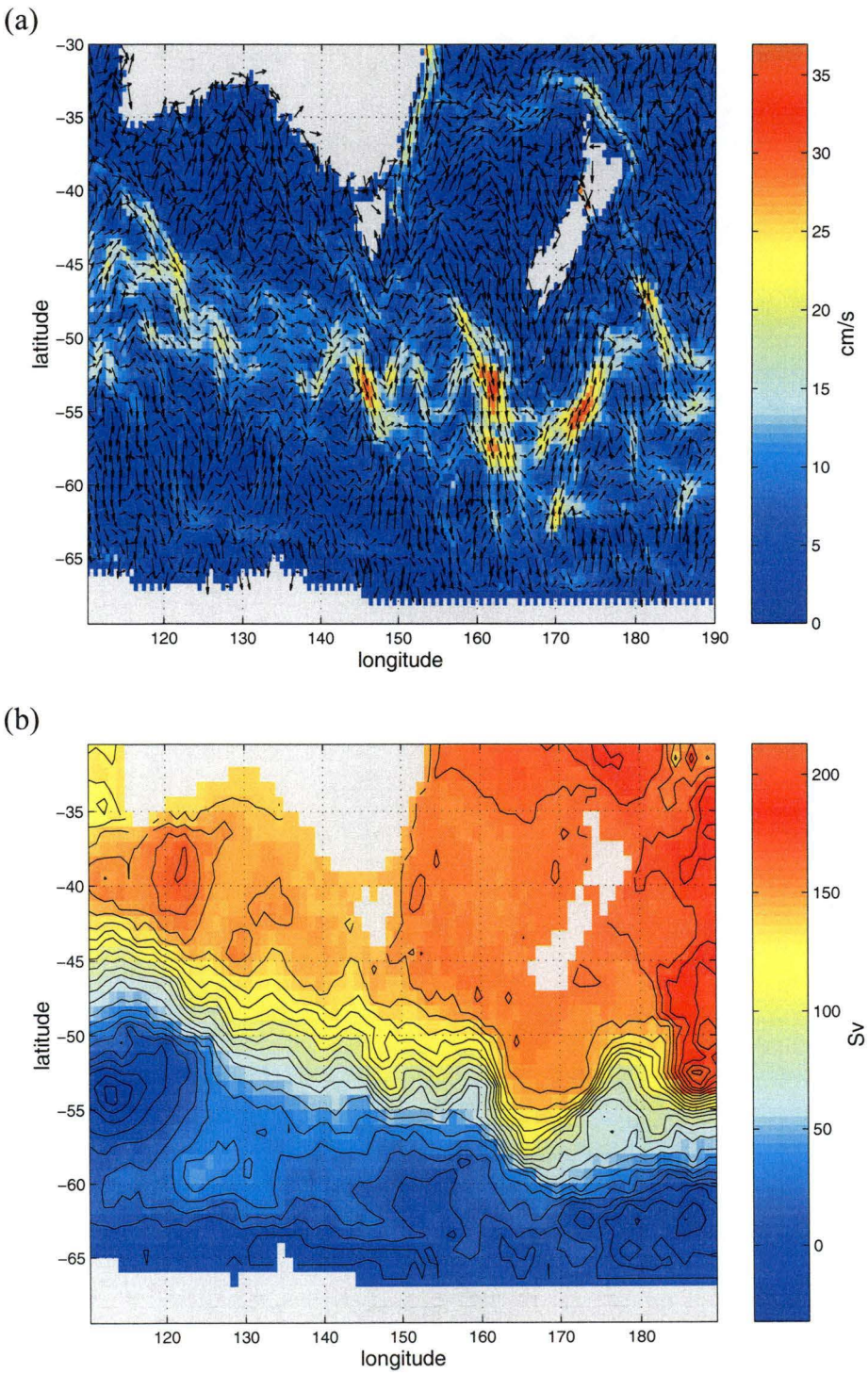


Figure 3.24: (a) Near-surface currents, and (b) streamfunction (EVEN), after five years for the gravity-wave condition with a one hour relaxation timescale.

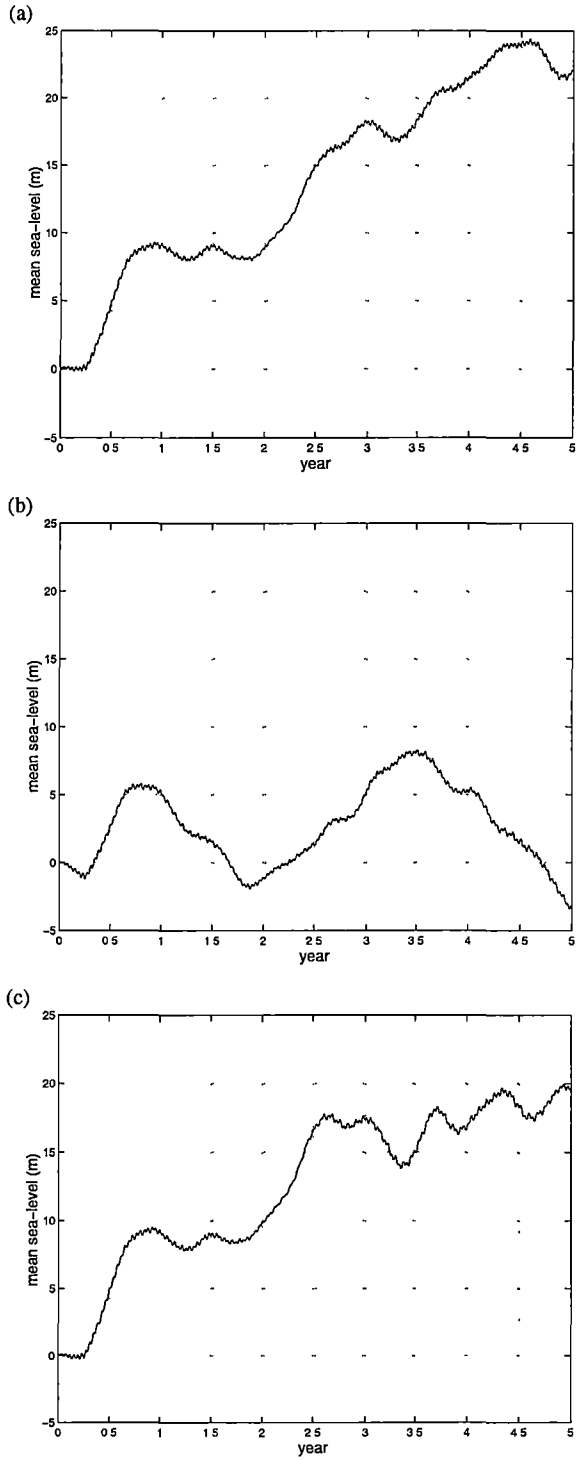


Figure 3.25: Mean sea level for the gravity-wave condition with one hour relaxation and implicitness parameters (a) $\gamma = 1$, $\delta = 0$, (b) $\gamma = 0.5$, $\delta = 0$, and (c) $\gamma = 1$, $\delta = 1$.

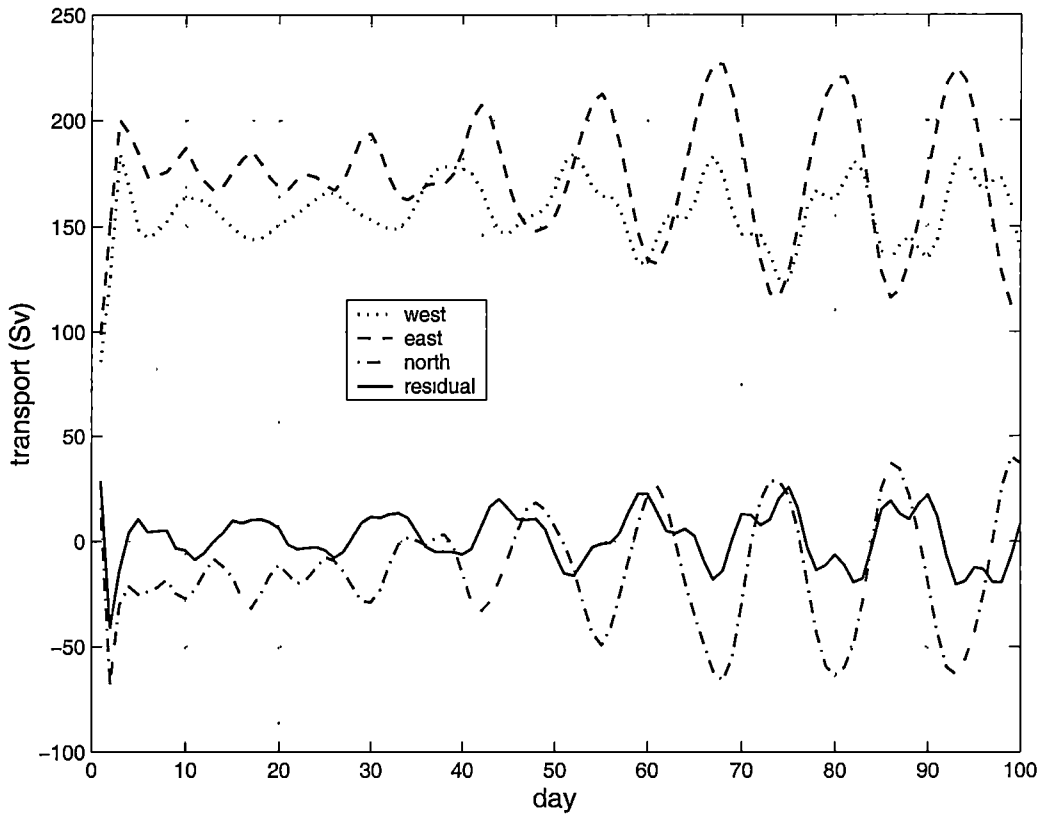


Figure 3.26: Transport across the boundaries, and residual transport, for the gravity-wave condition with a one hour relaxation timescale ($\gamma = 0.5$, $\delta = 0$).

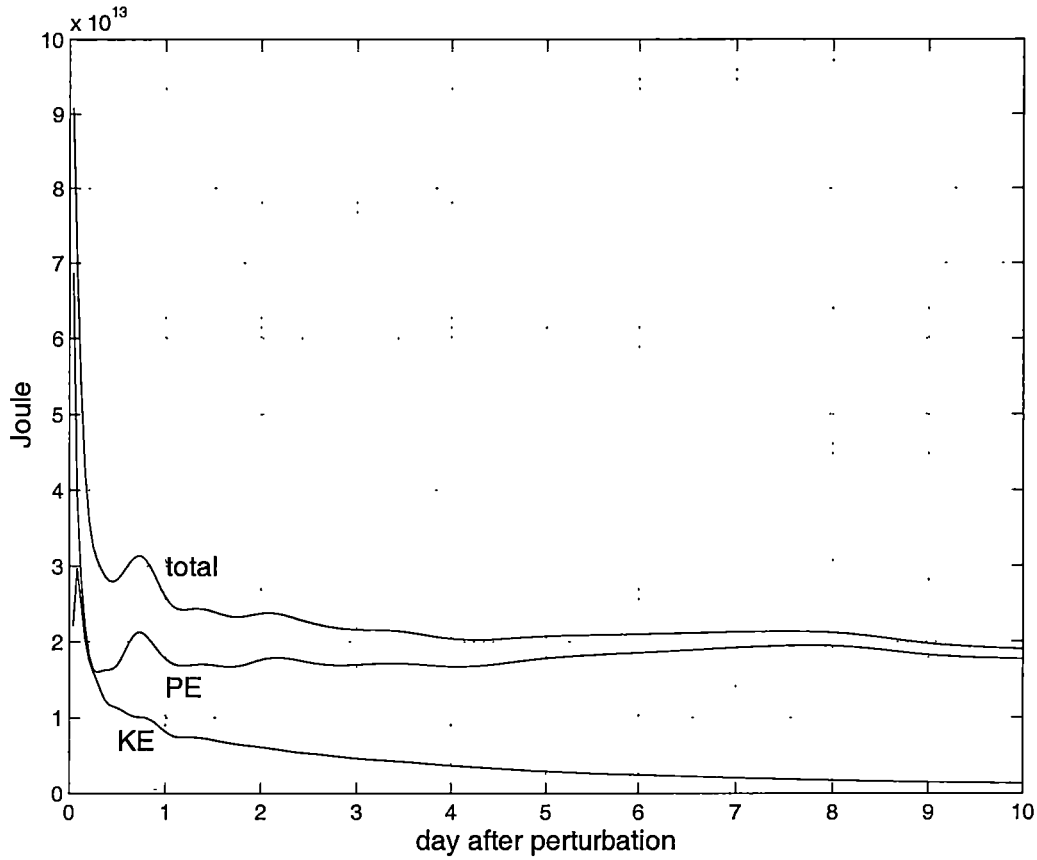


Figure 3.27: Energy transmission for a gaussian sea-level disturbance under the gravity-wave boundary condition with one hour flow relaxation.

2% of the energy was retained. However, it is an improvement on the clamped condition where 30% of the energy was retained (Figure 3.17).

Chapman (1985) found the gravity-wave-with-flow-relaxation scheme to provide similar transmission properties to the pure gravity-wave scheme, but that was likely because of the longer relaxation timescale he used (four hours). However, consistent with the results here, he found both schemes to provide better transmission performance than the clamped condition. Palma and Matano (1998) found only a minor reduction in perturbation energy with an implicit gravity-wave condition.

To summarise the performance of the gravity-wave condition: a formulation which was at least partly implicit in time was needed in order for it to work at all. However, results were poor unless extremely strong relaxation to prescribed boundary flows was also used. With relaxation, this condition transmits around 80% of the energy of a sea-level disturbance applied to the interior. While this is not as high as the 98% transmitted by the pure gravity-wave condition, it is an improvement over the 70% transmitted by the clamped condition. A curious aspect of the condition's performance was the existence of continuous rapid (13 day period) transport oscillations of up to 100 Sv across the boundaries. Despite the large amplitude of these oscillations, the respective boundaries maintained a close phase relationship so that oscillations

of net transport into the domain were considerably smaller. The continuous form of the condition (3.29) does not admit point-wise harmonic solutions, so the oscillations in integrated transport across the boundaries presumably arises through coupling between the boundaries, mediated via the barotropic system dynamics. In other words, the oscillatory behaviour is simply a solution of the barotropic system equations under the boundary conditions (3.29). There is not much more to be said on these oscillations without undertaking more detailed theoretical and numerical investigations.

Zero-gradient

The zero-gradient condition for normal velocity is:

$$\left.\frac{\partial U}{\partial x}\right|_{\{W,E\}} = 0, \qquad \left.\frac{\partial V}{\partial y}\right|_{N} = 0. \tag{3.30}$$

It may be considered a radiation condition (3.18) with infinite phase velocity c. The condition aims to avoid large boundary gradients. It has been used by Chapman (1985) and Jensen (1998) amongst others.

In implicit form, the condition is written (for the western boundary):

$$U_b^{n+1} = \gamma U_{b+1}^{n+1} + (1-\gamma) U_{b+1}^{n},$$

where, as before, γ is an implicitness parameter. Having implemented the implicit gravity-wave radiation condition (3.27), it is straightforward to modify the model code for the implicit zero-gradient condition. This is achieved by setting the coefficients $p = \gamma$, $q = 0$ and $s = (1-\gamma)$ in the discretisation (3.28). Both fully implicit ($\gamma = 1$) and fully explicit ($\gamma = 0$) discretisations were trialled. These gave quite different results, neither of which were satisfactory. The boundary volume transports for both of these are shown in Figure 3.28. Both cases produced transports of several thousand Sverdrups, the implicit case failing after around 55 days and the explicit case after around 740 days.

Despite the poor flow patterns, the condition performed surprisingly well at transmitting a sea-level disturbance, retaining only 1% of the energy of the perturbation in both cases. This is a very different result to that obtained by Chapman (1985), who found a zero-gradient condition even more reflective than a clamped condition. The reason for this discrepancy is not clear, but may be related to the fact that the fully implicit discretisation of the barotropic system used here has significant numerical dissipation. Chapman (1985) used the model of Beardsley and Haidvogel (1981), with an explicit time discretisation for the barotropic system (Platzman 1972). Clearly the behaviour of the open boundary condition is not independent of the numerical scheme used for the model. As mentioned earlier, this study appears to be one of the first examining open boundary conditions in an implicit model.

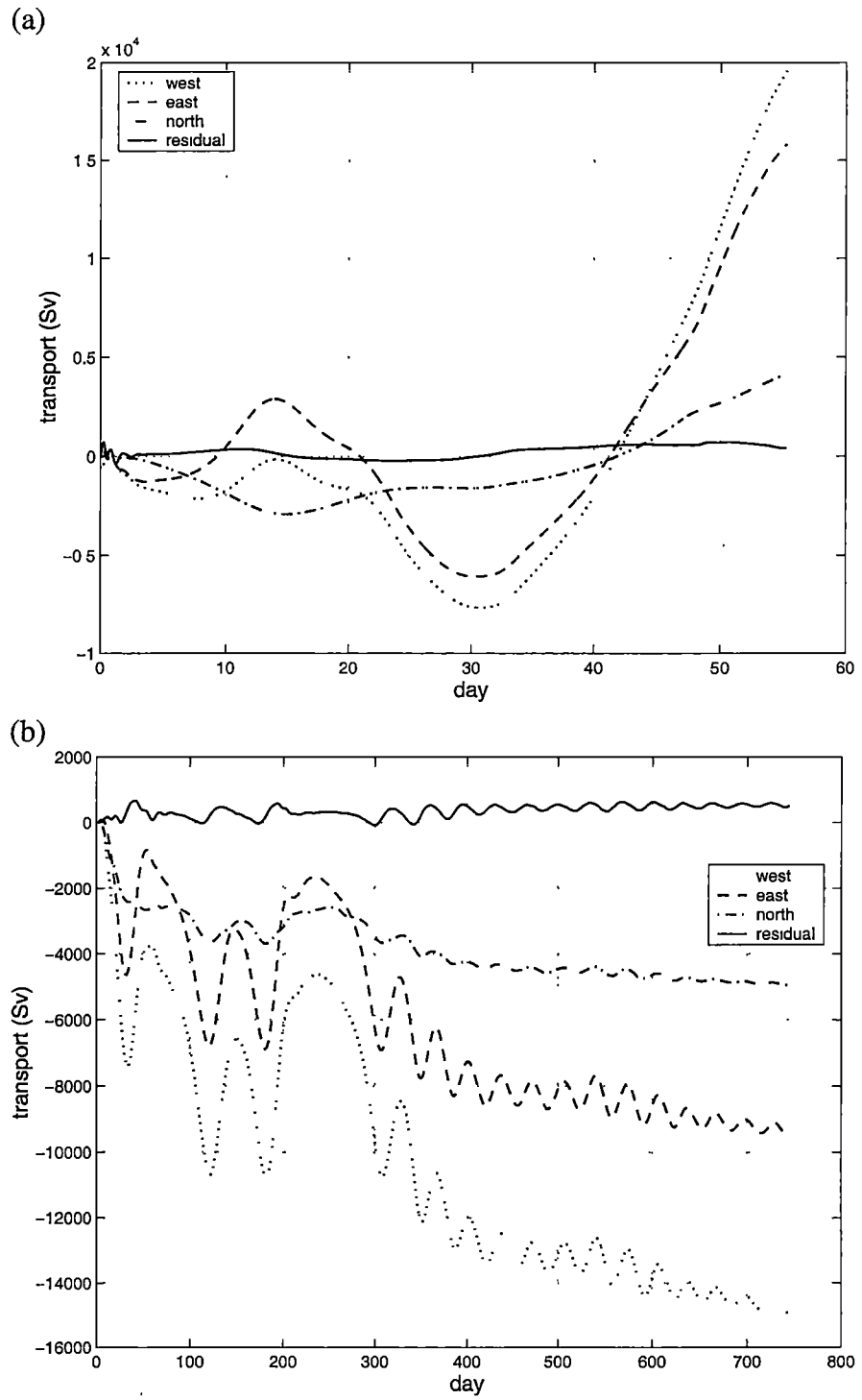


Figure 3.28: Integrated volume transport across open boundaries for (a) implicit and (b) explicit discretisations of the zero-gradient condition on normal velocity.

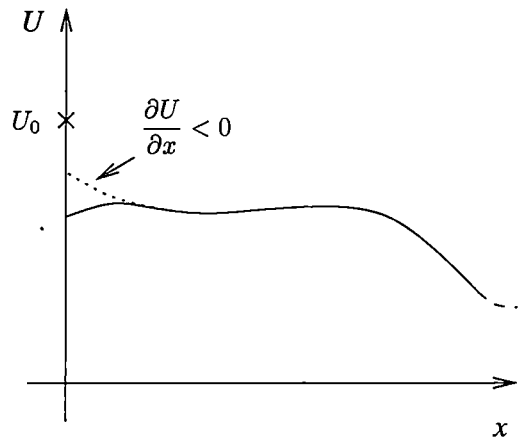


Figure 3.29: Behaviour of zero-gradient condition with "relaxation".

As with the gravity-wave condition, a term can be added to the right-hand side of (3.30) to drive the solution towards prescribed values, *e.g.*

$$\frac{\partial U}{\partial x} = -\epsilon(U_0 - U) \tag{3.31}$$

for the western boundary. There are no time derivatives here, and so this is not formally equivalent to Newtonian relaxation, although the parameter ϵ plays a similar role to a relaxation timescale. The term is motivated by a desire to ensure that the boundary transport U remains close to the reference value U_0, in the context of applying a boundary condition on the normal gradient. Thus if, for instance, the boundary value of U is lower than the reference U_0, the condition (3.31) attempts to increase the value of U on the boundary by ensuring the gradient there is negative. This is illustrated in Figure 3.29. Although the parameter ϵ in (3.31) appears like a spatial decay scale, it is not since the relation only holds on the boundary and not in the interior.

The condition (3.31) was discretised implicitly in time and applied to the model. The reference velocities were again taken to be bottom-referenced geostrophy scaled to 140 Sv and 155 Sv respectively for the western and eastern boundaries, and an artifical profile for the northern boundary. A "relaxation" strength of $\epsilon = 1/(6\Delta x)$ was used (around $(1/430)$ km^{-1}). The boundary transports were indeed stabilised, with the model running for the full five years of the experiment. The resultant boundary transports for the first 200 days, and the mean sea-level for the full five years, are shown in Figure 3.30. An oscillation of transport across the northern and eastern boundaries very similar to the gravity-wave-with-relaxation condition was produced. The surface and depth-integrated flows at the end of five years were very realistic, except for a noticeable lack of a depth-integrated EAC. This was despite the use of the artificial profile on the northern boundary, and different to previous cases where such a profile was used. The net flow into the domain remained roughly constant and lead to a mean sea-level increase of around 30 m, with significant separation between subgrids, after five years. The performance in a perturbation experiment was also very similar to that obtained with the

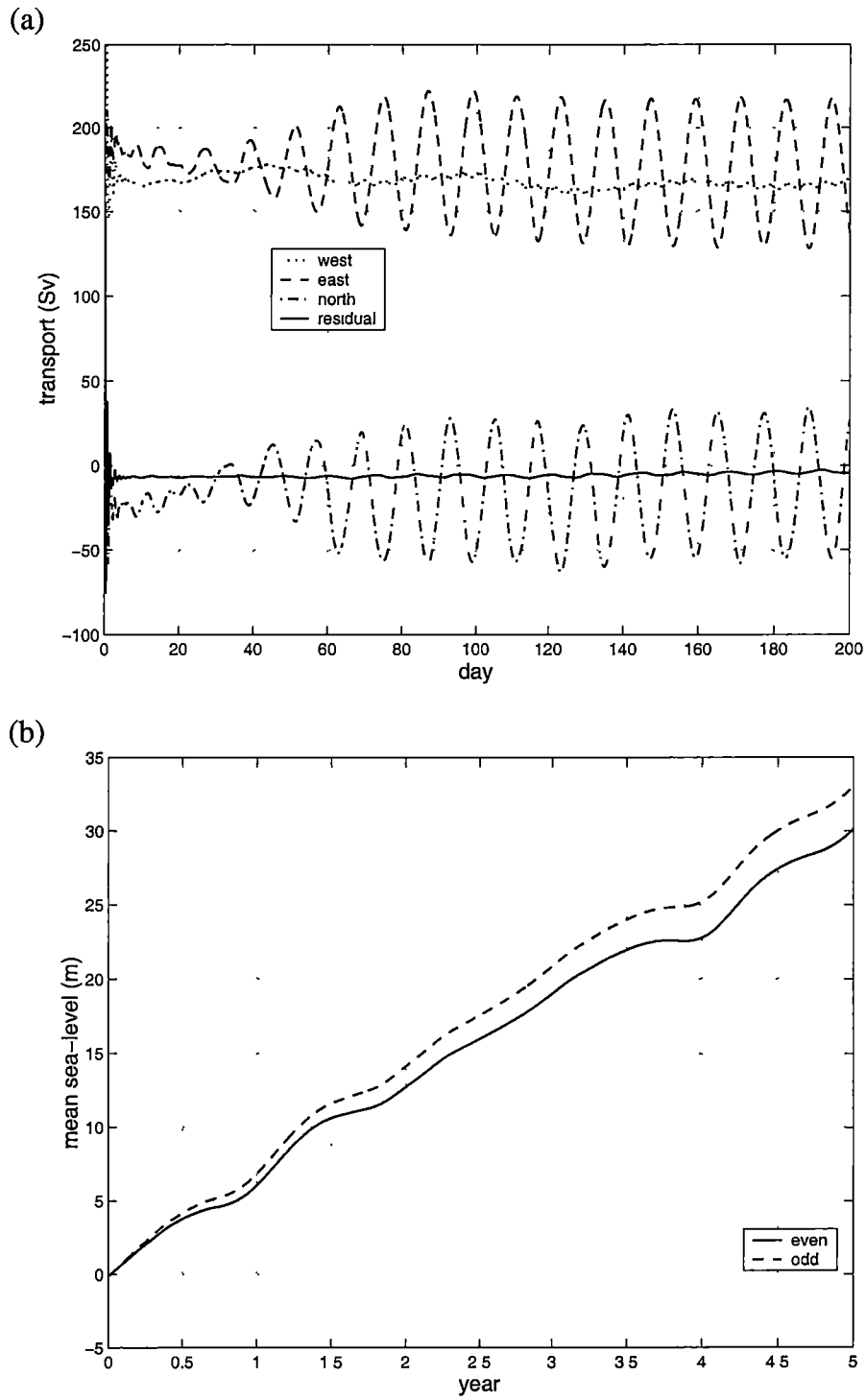


Figure 3.30: (a) Integrated transport across boundaries, and (b) mean sea-level, for the zero-gradient condition with pseudo 'relaxation'.

gravity-wave condition with flow relaxation (Figure 3.27); around 20% of the energy was retained after 10 days.

Summarising the performance of the zero-gradient condition on normal velocity, it was found to fail for the regional Southern Ocean model unless a pseudo-relaxation to prescribed velocity profiles was used. While the very good energy transmission properties were reduced by such a procedure, they were still improved relative to the clamped boundary condition. In many respects, the overall performance of the zero-gradient condition with pseudo-relaxation was similar to the gravity-wave condition with relaxation.

3.5.3 Boundary conditions on sea-level

From the analysis in section 3.5.1, it is clear that for well-posed boundary conditions only one condition must be applied at each open boundary in HOPE's barotropic system. In the previous section, various boundary conditions applied to normal velocity were examined. In this section, we consider conditions applied to sea-level. For the shallow water equations with mean advection, a boundary condition on sea-level is ill-posed (section 3.5.1 and Oliger and Sundström (1978)). In HOPE's barotropic subsystem, there is no momentum advection, but it is not possible to theoretically prove (using the energy method, at least — see section 3.5.1 and Appendix A) well-posedness for prescribing sea-level.

For the sake of brevity, a description of the required technical modifications to HOPE's barotropic system numerics for a boundary condition on sea-level is not provided, but the details follow a similar approach to that described in Appendix C for a condition on velocity. This should not be taken to imply the exercise is a trivial extension of the latter case. On the contrary, as much effort is required to implement an open boundary condition on sea-level as it is on velocity. The details are omitted here precisely because of their excessively cumbersome nature.

Clamped condition

As with the clamped condition on normal velocity (page 80) we require in this case that sea-level remain unchanged on the boundary:

$$\begin{aligned} \frac{\partial \eta}{\partial t} &= 0, \\ \eta_b(t) &= \eta_0. \end{aligned}$$

Although, as with the normal velocity case, we allow the prescribed sea-level η_0 to vary slowly in time.

There is, again, considerable flexibility in choosing the reference sea-level profiles along the boundaries. The approach taken first with the normal velocity case was to assume bottom-referenced geostrophy. We use the same approach here. For zero normal flow at the bottom, the along-boundary gradient of bottom pressure must be balanced by the sea-level gradient:

$$g\frac{\partial \eta_0}{\partial y} = -\left.\frac{\partial p'}{\partial y}\right|_{z=-H} \qquad \text{eastern and western boundaries}$$

$$g\frac{\partial \eta_0}{\partial x} = -\left.\frac{\partial p'}{\partial x}\right|_{z=-H} \qquad \text{northern boundary.} \tag{3.32}$$

(This can also be obtained by setting time derivatives and bottom velocity to zero in the prognostic momentum equations (2.30)–(2.33).) Equations (3.32) fix the sea-level gradient along each boundary, but undetermined constants must also be chosen. Sea-level is required to be continuous at each corner, thus tying together the respective boundaries. In addition, mean sea-level along the entire open boundary (north, west and east) was chosen to be zero. In this manner the reference sea-level η_0 is fully determined.

This condition was stable for a run of five years. The resultant boundary transport and mean sea-level timeseries are shown in Figure 3.31. Forcing the mean sea-level around the boundaries to remain at zero has strongly constrained the mean sea-level over the domain to remain near zero. It is apparent, however, that differences occur between the two subgrids. The numerical implementation of the condition tied the two subgrids together, by averaging 'virtual' sea-level points for the EVEN grid from adjacent sea-level points on the ODD grid around the boundaries, in a scheme analogous to the averaging procedure used for a boundary condition on normal velocity (see Appendix C). The technique was less successful here than for normal velocity boundary conditions. While the mean sea-level was almost the same on the two subgrids at the end of five years, there were localised differences of up to 0.3 m, most noticeably in the Tasman Sea.

The depth-integrated and near-surface flow patterns after five years are shown in Figure 3.32. While the ACC is represented reasonably well, it lies considerably north of its position with the normal velocity boundary conditions (compare, for example, Figure 3.15 for the clamped velocity condition with an artificial transport profile along the northern boundary). Along the east coast of Australia, a very strong northwards flow of over 50 Sv has replaced the southwards flowing EAC. A weak southwards EAC is present only at the surface, where it separates from the coast just south of the northern boundary and continues as a weak Tasman Front across the Tasman Sea. The very strong northwards surface flow around the southeastern corner of Tasmania is very unrealistic. As with sea-level, considerable separation between the subgrids occurs with the depth-integrated velocity — over 50 Sv between adjacent gridpoints in several broad regions within the domain, much greater than the maximum 20 Sv differences shown in Figure 3.15 for the clamped velocity boundary condition.

The clamped sea-level condition performed very well on a sea-level perturbation experiment to examine its reflection characteristics. Figure 3.33 shows the energy of the perturbation for ten days following its application. After ten days, around 1% of the perturbation energy has been retained. This result is surprising given the very high energy retention of the clamped velocity condition (Figure 3.17). It is also different to results obtained by Chapman (1985) with a time-explicit model, who found almost complete reflection for a clamped sea-level boundary condition. To determine to what extent this discrepancy may be related to the numerical dissipation of the implicit dif-

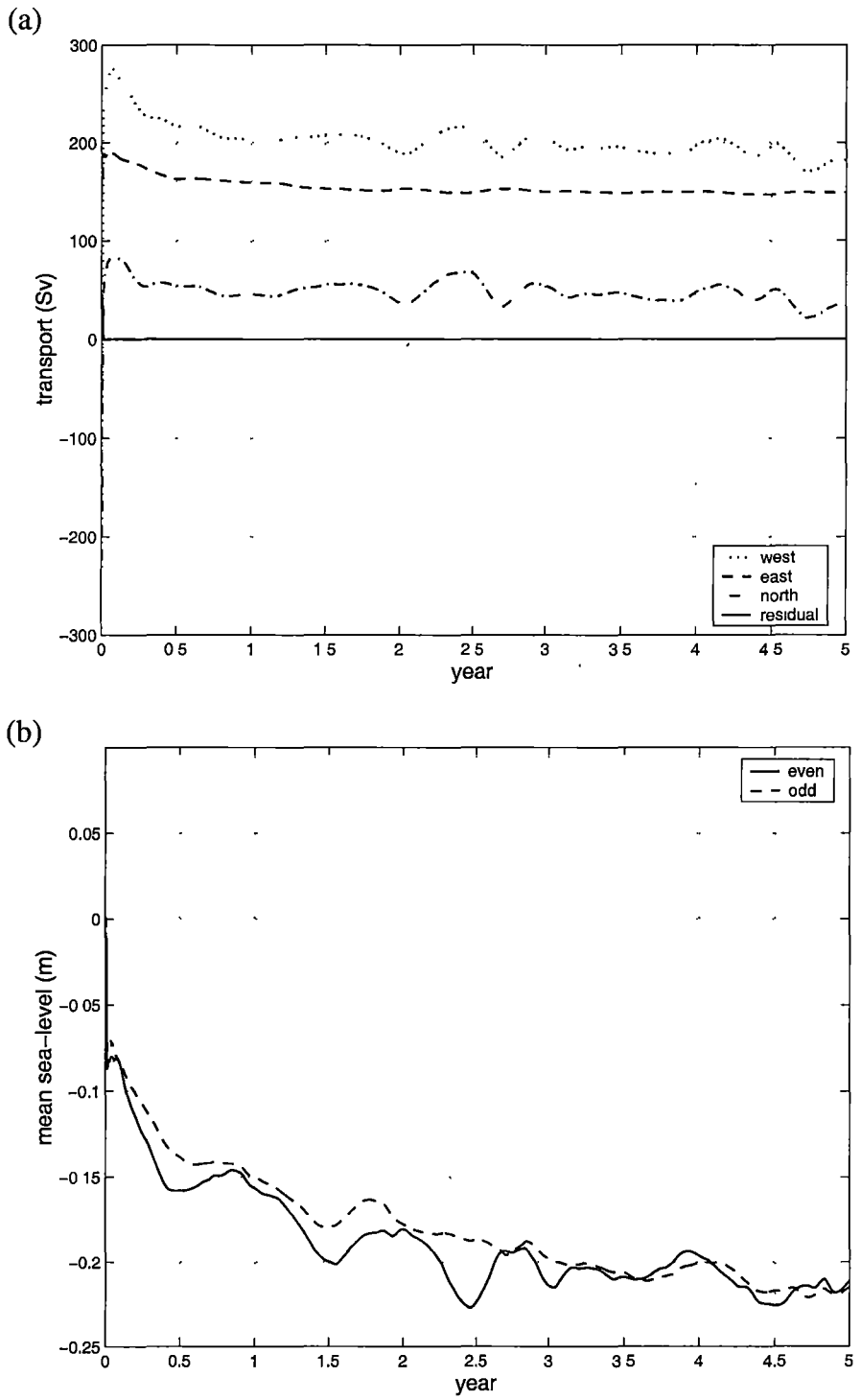


Figure 3.31: (a) Integrated transport across boundaries, and (b) mean sea-level, for the clamped sea-level condition.

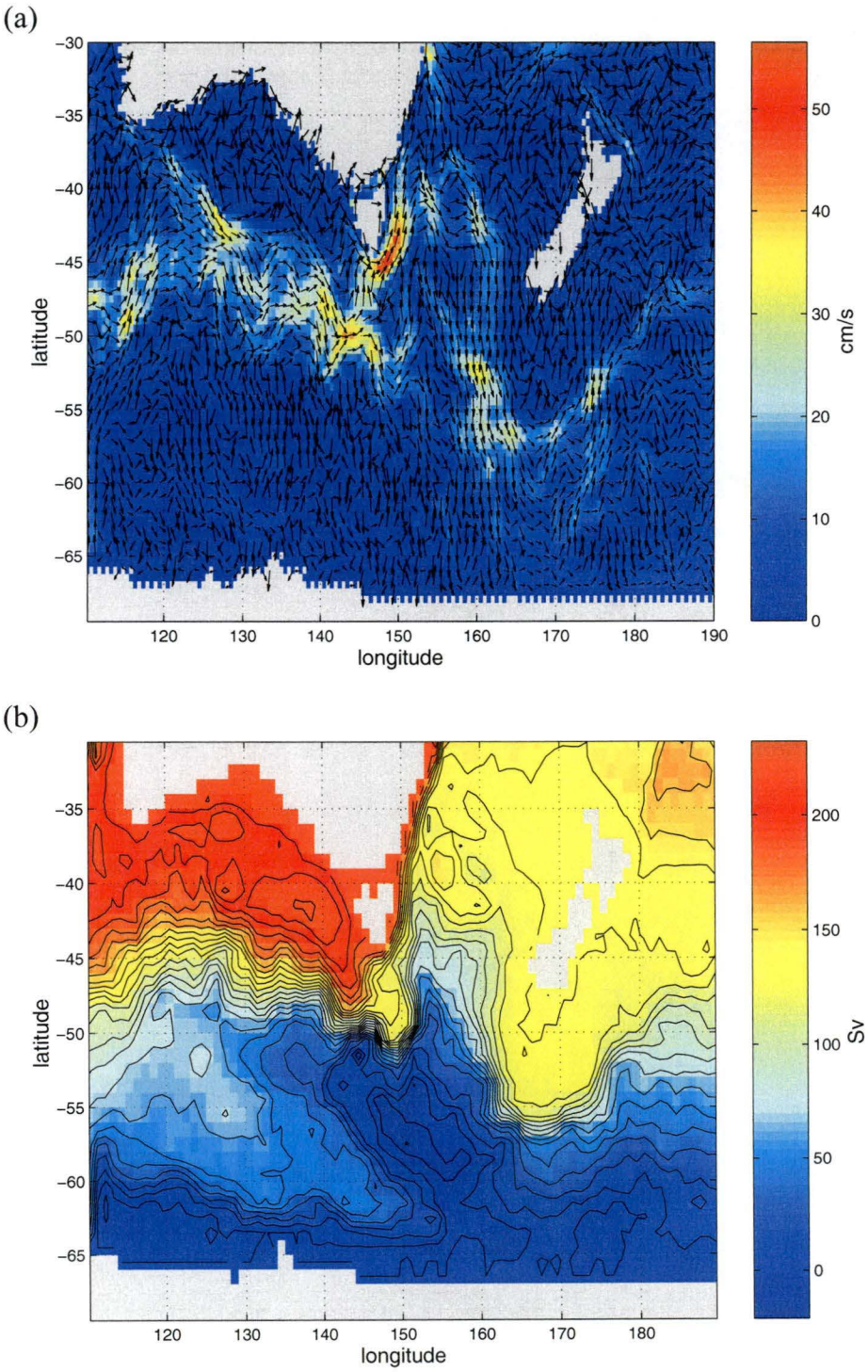


Figure 3.32: (a) Near-surface (32 m, layer two), and (b) depth-integrated flows after five years for the clamped sea-level boundary condition.

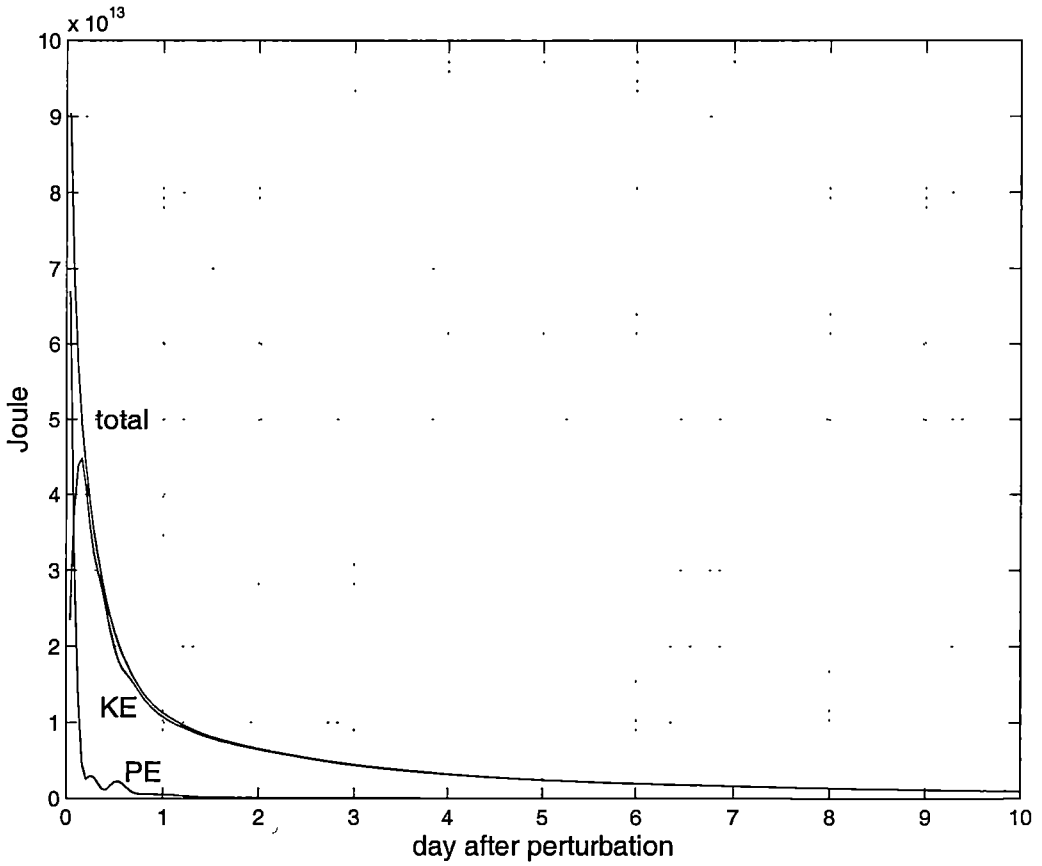


Figure 3.33: Energy transmission for a gaussian sea-level disturbance under the clamped sea-level boundary condition.

ferencing scheme used here for the barotropic system, an attempt was made to perform an identical perturbation experiment using stability parameters $\alpha = 0.5$ and $\beta = 0.5$. (Figure 3.18 shows the results of such an experiment for the clamped velocity condition.) However, the model was unstable under the clamped sea-level condition with this degree of explicitness. The stability parameters had to be raised to around 0.7 in order to maintain stability. With these values, there was only slightly lower energy dissipation than in the fully implicit case. While an extensive study was not performed, it is possible that the ill-posedness of a clamped sea-level condition for the shallow-water equations (Oliger and Sundström 1978) contributes to the instability for numerical schemes that are not fully implicit in time.

A similar range of variations could be trialled for a sea-level boundary condition as were described earlier for the normal-velocity condition (*e.g.* zero-gradient, gravity-wave, relaxation etc.). While some of these were examined, they performed no better than the clamped sea-level condition just described.

3.5.4 Characteristic (Flather) boundary condition

In contrast to the previous two sections which applied boundary conditions on either of the primitive variables (normal velocity or sea-level), this section trials a condition applied to a combination of those primitive variables. From the introductory discussion in section 3.1.2, we have seen that well-posed boundary conditions must be formulated in terms of the characteristic variables, with precisely one boundary condition being required for each incoming characteristic variable at a boundary. Conditions on normal velocity or sea-level can be developed within this framework by writing the incoming characteristic variables in terms of the outgoing characteristic variables (page 68). Well-posedness of such conditions is not necessarily guaranteed. On the other hand, boundary conditions which prescribe the incoming characteristic variables are always well-posed.

Hedstrom (1979) proposed a nonreflecting boundary condition which leaves the incoming characteristic variables unchanged at a boundary. For HOPE's barotropic system, the incoming characteristic variables are (section 3.5.1):

$$\begin{aligned} U \pm c\eta, &\quad \text{western } (+) \text{ and eastern } (-) \text{ boundaries;} \\ V - c\eta, &\quad \text{northern boundary.} \end{aligned}$$

Considering the western boundary, Hedstrom's condition thus becomes

$$\frac{\partial}{\partial t}(U + c\eta) = 0, \tag{3.33}$$

or

$$U(t) + c\eta(t) = U_0 + c\eta_0. \tag{3.34}$$

An identical condition was proposed independently by Flather (1976) for tidal modelling on a continental shelf. The condition has since been used for similar modelling studies by numerous authors, *e.g.* Oey and Chen (1992),

Davies and Lawrence (1994) and Lee et al. (2000). It has also been used in more general contexts by *e.g.* Shulman (1997) and Palma and Matano (1998). Palma and Matano noted it could be derived by combining the gravity-wave radiation condition (3.27) with a one-dimensional continuity equation. For a western boundary we have

$$\frac{\partial U}{\partial t} = c\frac{\partial U}{\partial x}$$

and

$$\frac{\partial \eta}{\partial t} = -\frac{\partial U}{\partial x},$$

from which we can write directly Hedstrom's condition (3.33). With this interpretation, we anticipate that the condition may be useful both for transmitting surface gravity waves as well as conserving volume.

Clearly, the condition requires co-located sea-level and velocity points. This is not the case in HOPE, and so careful thought must be given to its numerical implementation. Appendix D discusses the technical implementation of the condition in HOPE's barotropic system.

As originally proposed by Flather, the condition was intended to drive a coastal model with known tidal forcings at the boundaries. Thus (3.34) becomes

$$U(t) + c\eta(t) = U_0(t) + c\eta_0(t). \tag{3.35}$$

The reference profile U_0 was constructed along the boundary by Flather using an iterative technique. First, given observations of tidal elevation, $\eta_{obs}(t)$, on the boundary, the model was run with $U_0 = 0$ and $\eta_0 = \eta_{obs}$. Successive iterations of model runs substituted the calculated velocity on the boundary, $U(t)$, for the reference velocity, U_0, until convergence was reached, at which point the calculated sea-level on the boundary, $\eta(t)$, closely followed the observed tidal signal $\eta_{obs}(t)$. In Palma and Matano's (1998) recent comparison of the condition with other open boundary conditions, they assumed constant values $U_0 = \eta_0 = 0$ for their model with an initial state of no motion and short duration perturbation experiments.

It is important to note a fundamental difference between Flather's condition (3.35) and Hedstrom's condition (3.34). Flather allowed, indeed required, potentially quite large-amplitude and rapid time variation, while Hedstrom specifically prescribed constant values for the incoming characteristic variables. Thus, while Hedstrom's condition is explicitly designed to be *passive*, *i.e.* non-reflective to wave phenomena, Flather's condition is intended to be *active*, *i.e.* to drive a model with tidal forcing at the boundaries. In fact the only commonality between the two is the fact that they both prescribe values for the incoming characteristic variables. Despite this, recent papers (Shulman 1997; Palma and Matano 1998) have identified conditions on characteristic variables with Flather's condition, even for constant reference values. While the reference values in this thesis are not constant, neither do they vary as rapidly as Flather's. In keeping with the recent trend (but despite the important differences), the boundary condition as implemented in this thesis will be referred to as a 'Flather condition'.

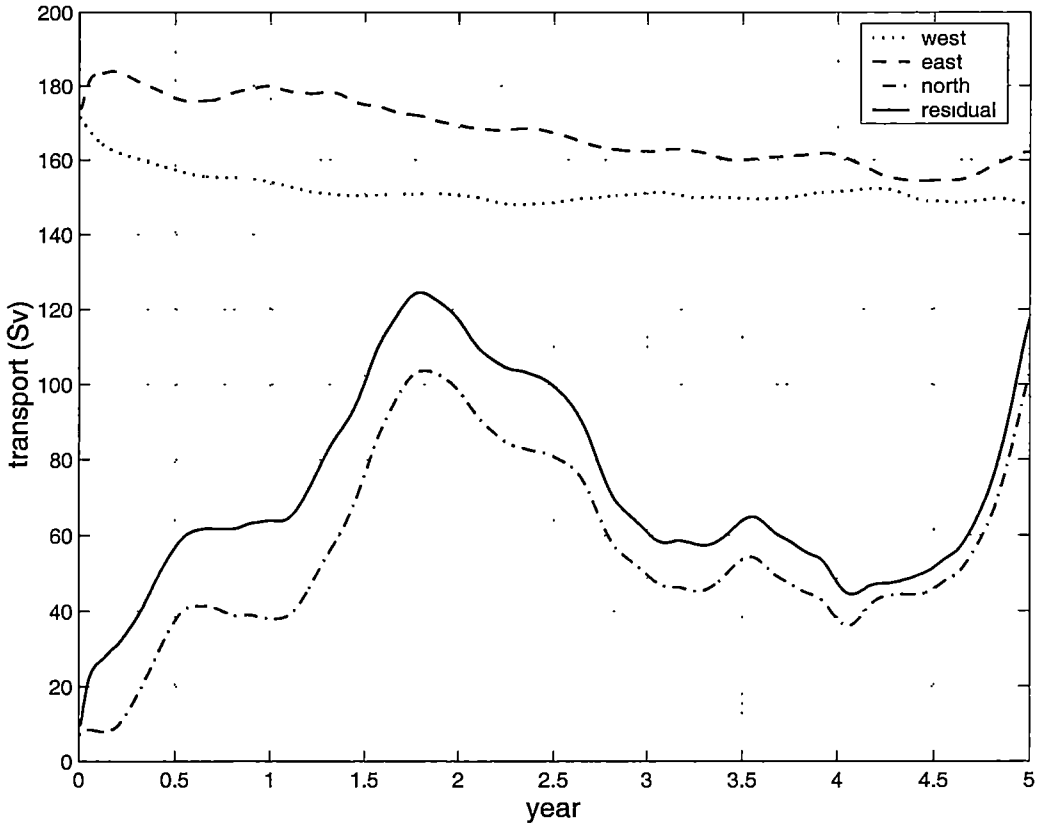


Figure 3.34: Volume transport across boundaries associated with Flather reference velocity profiles U_0, V_0 (derived from bottom-referenced geostrophy).

Reference profiles from bottom-referenced geostrophy

An obvious method for constructing U_0, V_0 and η_0 is to apply the methods of the previous two sections for the clamped conditions on normal velocity and on sea-level: to make an assumption of bottom-referenced geostrophy and calculate fixed U_0, V_0 and η_0 according to equations (3.26) and (3.32).

A trial run using this approach was stable for five years. Figure 3.34 shows the integrated volume transport across each boundary associated with *the reference profiles* U_0 *and* V_0, as well as the residual transport out of the domain that would result *if the modelled velocities on the boundaries followed these reference profiles exactly.* The variation over the five years is due to the evolution of the density fields along the boundaries. Clearly, there is a significant residual outflow from the domain associated with these reference profiles.

We must examine to what extent the modelled velocities on the boundaries (arising from the Flather condition (3.34)) follow these reference profiles. Figure 3.35(a) shows the *actual* volume transports across each boundary, while Figure 3.35(b) shows the residual outflow from the domain. First, it is clear that the *actual* boundary velocities do not track closely the *reference* boundary velocities. The reference transport across the western boundary, for instance, is below 160 Sv for most of the five years of the run, while the actual transport

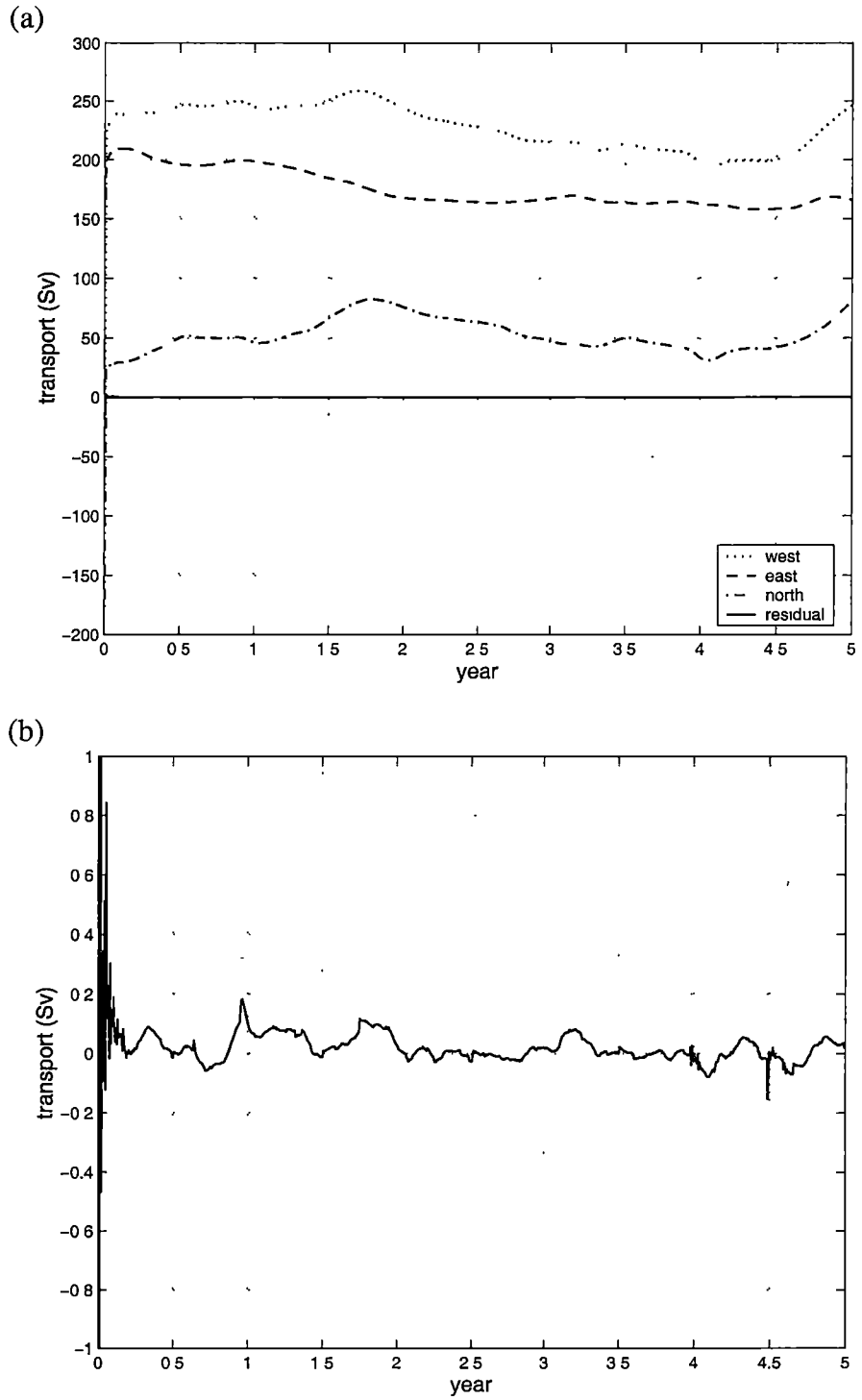


Figure 3.35: (a) Actual volume transport across boundaries, and (b) residual outflow, with the Flather condition imposed using bottom-referenced geostrophy for the reference profiles.

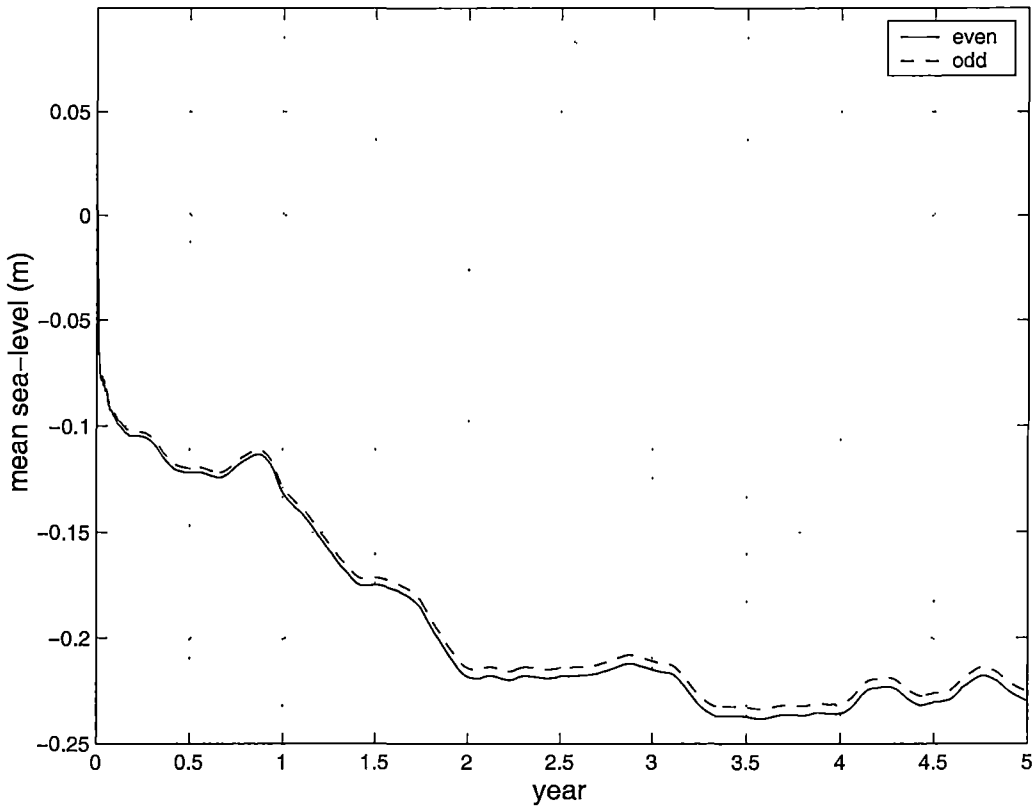


Figure 3.36: Change in mean sea-level with the Flather condition imposed using bottom-referenced geostrophy for the reference profiles.

is greater than 200 Sv. This is seen also in the residual transport out of the domain. While the reference transport profiles have a net loss of well over 50 Sv for most of the run, the actual boundary transports have a residual very close to zero and lower than 0.2 Sv. This is a remarkable feature of the Flather condition. While the respective boundaries are coupled only through the model dynamics in the interior, and not the boundary conditions themselves, the Flather condition nevertheless maintains an almost perfect conservation of mass for the domain. This is despite the actual transport through the various boundaries varying by over 50 Sv during the five year run, Figure 3.35(a). Similar, but less spectacular, examples of emergent coupling between independent boundary conditions were previously seen in both the gravity-wave condition on normal velocity, with flow relaxation (Figure 3.26) and the zero-gradient condition on normal velocity with pseudo-relaxation (Figure 3.30) where, in both cases, the transports across the eastern and northern boundaries oscillated rapidly with amplitudes around 100 Sv, but almost balanced each other. This result was not entirely unexpected for the Flather condition since, as mentioned earlier, Palma and Matano (1998) pointed out that the condition could be derived using conservation of mass. It is, nevertheless, pleasantly surprising in its effectiveness. Figure 3.36 shows the resultant change in mean sea-level associated with the small, but non-zero, net outflow from the domain.

The near-surface and depth-integrated flow fields after five years are shown in Figure 3.37. While the ACC is reproduced reasonably well, the flow in the region of the east coast of Australia is more disappointing. A strong surface and depth-integrated (50 Sv) northwards flow breaks off from the ACC and extends around the southeastern corner of Tasmania, continuing north along the coast of eastern Australia. It joins a weaker EAC before hugging the northern boundary of the model domain. As found in previous sections, a realistic EAC flow is harder to reproduce in the model than the ACC.

Finally, we examine the performance of the Flather condition under a sea-level perturbation experiment. As in previous sections, a gaussian sea-level perturbation is applied to the model after a short (10 day) spinup from rest. The energy of the perturbation is integrated over the domain for another 10 days to determine the ability of the boundary condition to transmit disturbances. Figure 3.38 shows this timeseries. At the end of 10 days, the total energy of the perturbation has reduced to less than 2% of its original value. This is a result essentially as good as any of the results in the previous sections using boundary conditions on either normal velocity or sea-level. It was expected that the Flather condition should be nonreflective to surface gravity-wave disturbances because it can be derived from a combination of the continuity equation and a gravity-wave radiation condition. That its performance is as good as a pure gravity-wave condition on normal velocity is a pleasing result.

The Flather condition with reference profiles calculated from bottom-referenced geostrophy clearly performs very well at conserving volume and transmitting surface gravity waves. It is less satisfactory at reproducing realistic flow patterns and strengths. We now consider two alternative methods for constructing the reference profiles in an attempt to improve this aspect of the condition's performance.

Reference profiles by timestepping η_0

As mentioned in the introduction to the Flather boundary condition, as originally proposed by Flather, the condition was used in an iterative scheme to derive time-varying mutually consistent reference profiles U_0, V_0 and η_0. We saw above that simply bottom-referencing these profiles produces boundary volume transports very different from the reference profiles. Motivated by Flather's original approach to the tidal problem, we trial an alternative method for constructing the profiles. Instead of calculating η_0 from bottom-referenced geostrophy, we prescribe it each timestep according to the model solution on the boundaries:

$$\eta_0^{n+1} = \eta_b^n.$$

We again construct U_0 from bottom-referenced geostrophy, but replacing the profile V_0 on the northern boundary with the artificial profile discussed in section 3.5.2 (see Figure 3.13). Anticipating that the timestepping scheme for η_0 will cause the model to more closely follow the reference profiles, we

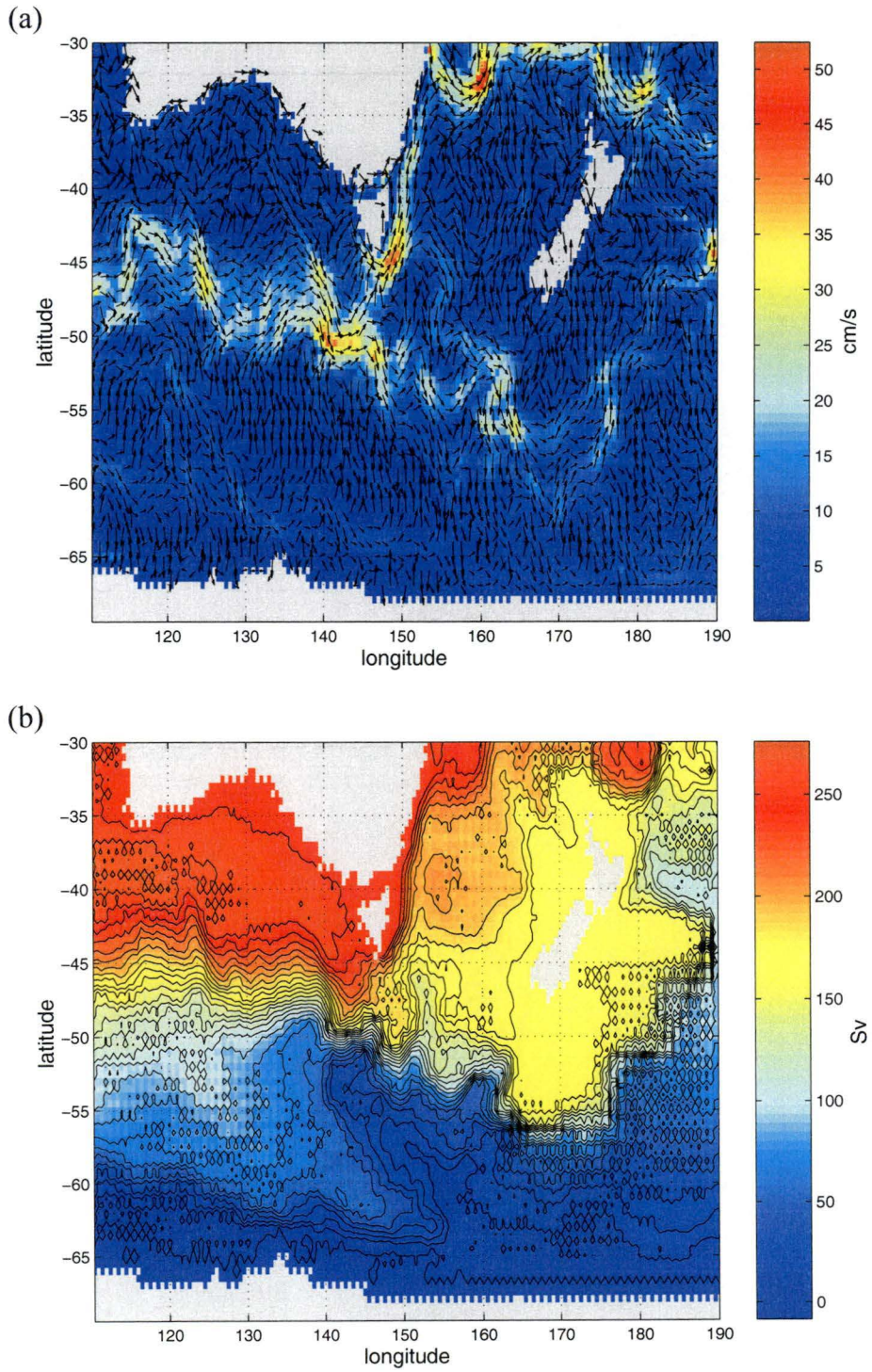


Figure 3.37: (a) Near-surface (layer two, 32 m depth), and (b) depth-integrated flows after five years using the Flather condition with bottom-referenced geostrophy for the reference profiles.

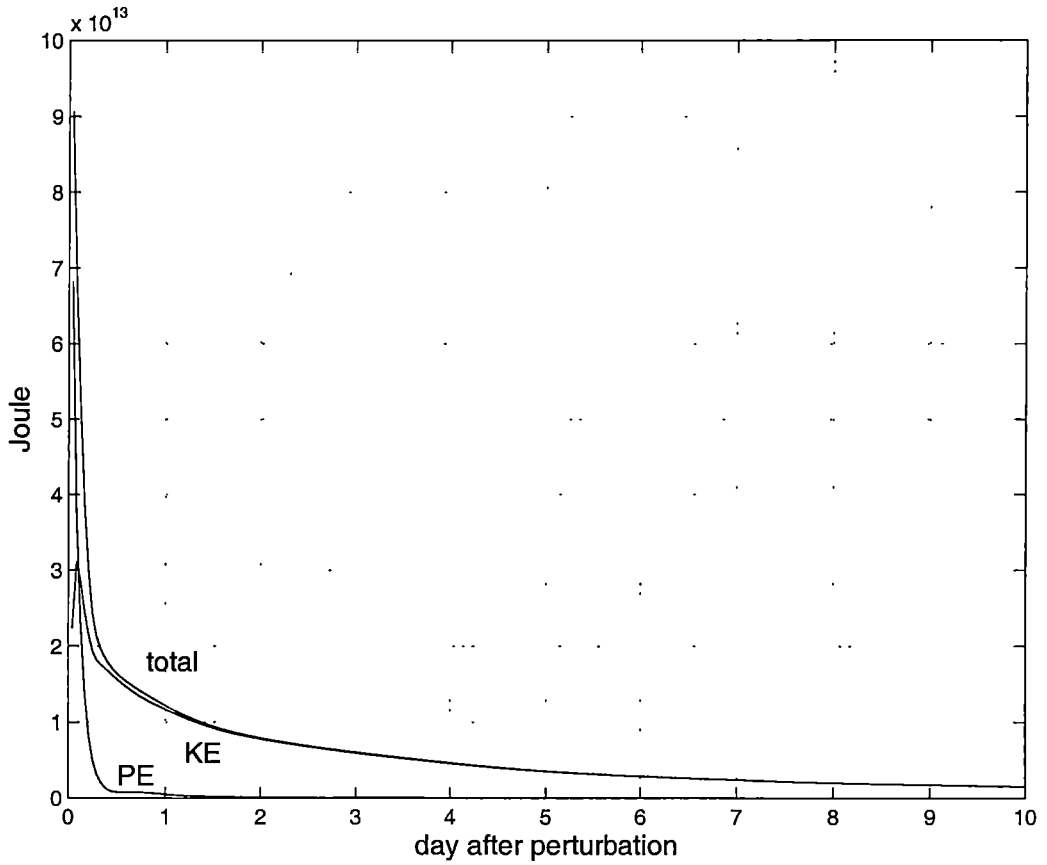


Figure 3.38: Potential (PE), kinetic (KE) and total energy of an initial sea-level perturbation for 10 days with the Flather boundary condition using bottom-referenced geostrophy for the reference profiles.

also scale the bottom-referenced profiles for U_0 on the western and eastern boundaries to 140 Sv and 155 Sv respectively, as done in earlier sections.

The model was stable for the five year duration of the experimental run. The actual transports across the open boundaries are shown in Figure 3.39, together with the mean sea-level. The timestepping method for η_0 has indeed caused the model to track almost perfectly the reference velocity profiles U_0 and V_0. The transport across the western boundary has decreased from over 200 Sv to a value of 140 Sv, consistent now with the reference profile. After the first three months, the residual transport out of the domain stays in the range -0.01 to 0.01 Sv, resulting in the mean sea-level varying by less than 0.5 cm. With the exception of the clamped condition on normal velocity, this is by far the best result yet seen for conserving volume in the domain.

The flow fields after five years for this boundary condition are shown in Figure 3.40. The currents are considerably improved, and almost identical to those obtained with the clamped boundary condition on normal velocity with the same prescribed boundary velocities (Figure 3.15). In particular, the depth-integrated and surface representations of the EAC are now satisfactory.

Since the condition now appears to be performing much the same as the clamped normal-velocity condition, it is necessary to confirm whether it retains the ability to transmit surface gravity-wave disturbances. A sea-level perturbation experiment was again performed, with Figure 3.41 showing the energy timeseries. The boundary condition has become much more reflective, retaining around 20% of the energy of the perturbation after 10 days. This is not as bad as the 30% retained by the clamped normal-velocity condition, but it is much worse than the 2% retained with the bottom-referenced Flather condition.

Thus, the flow fields have been improved with the timestepping method, but the nonreflective properties of the boundary condition have been compromised.

Reference profiles using relaxation

As a tradeoff between reference transport profiles which change only very slowly (the bottom-referenced profiles) and profiles that change rapidly (timestepping η_0), a third variation is trialled where the profiles are relaxed with a timescale τ, *e.g.*

$$\frac{\partial}{\partial t}(U_0 + c\eta_0) = \frac{1}{\tau}[(U_{0_{geost}} + c\eta_b) - (U_0 + c\eta_0)]$$

for the western boundary, where $U_{0_{geost}}$ is the bottom-referenced velocity profile of the last section, and η_b is the model-derived boundary sea-level which was used to timestep η_0 in the last section. The artificial profile, V_0, of Figure 3.13 is used on the northern boundary.

A relaxation timescale of two days was used to trial this variation of the Flather condition. The energy timeseries for a repeated sea-level perturbation experiment is shown in Figure 3.42. The nonreflective properties have clearly been restored. The total energy after ten days is reduced to only 2% of the initial energy of the perturbation.

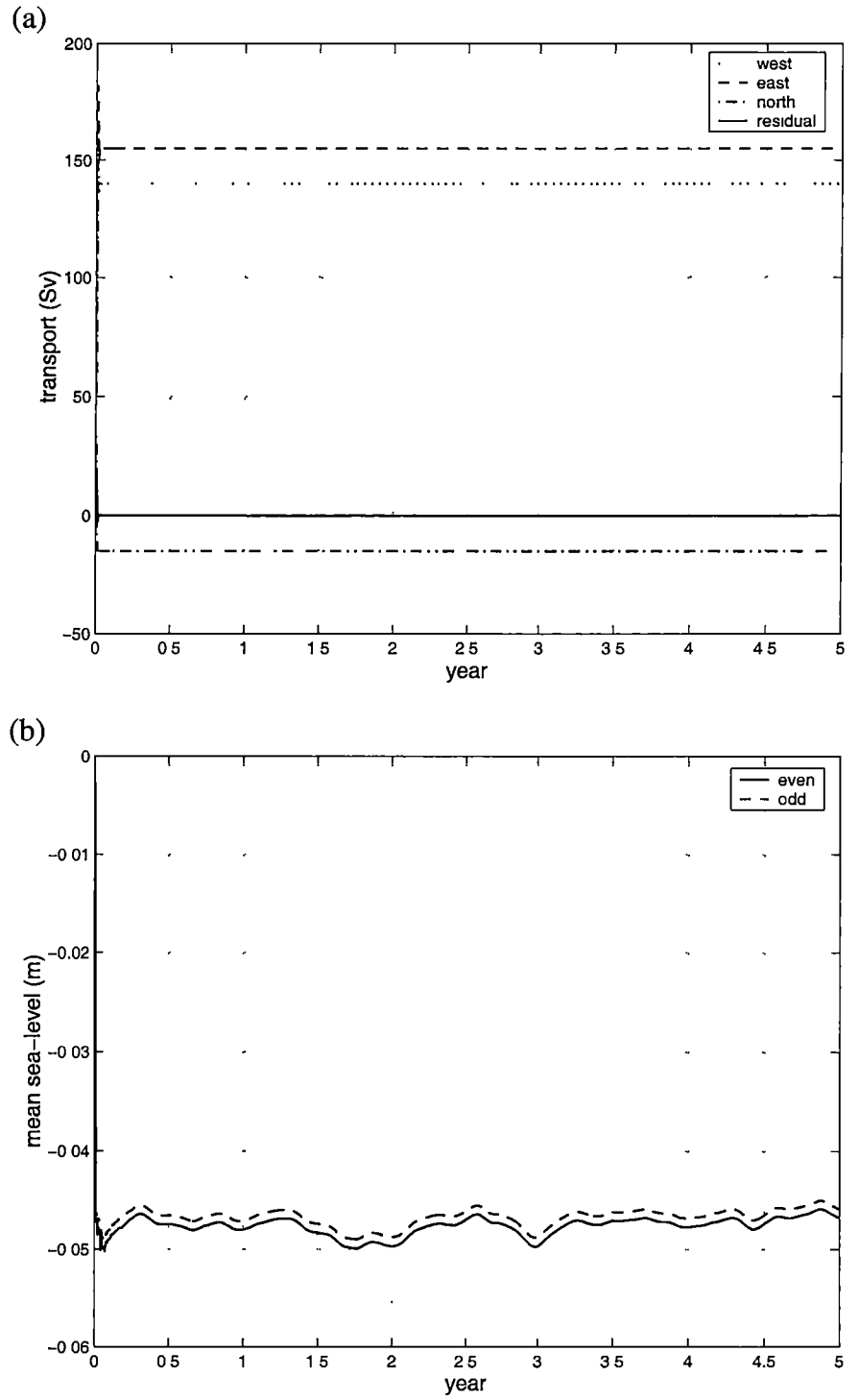


Figure 3.39: (a) Volume transports across boundaries, and (b) mean sea-level, over five years for the Flather boundary condition with a timestepped η_0 reference profile.

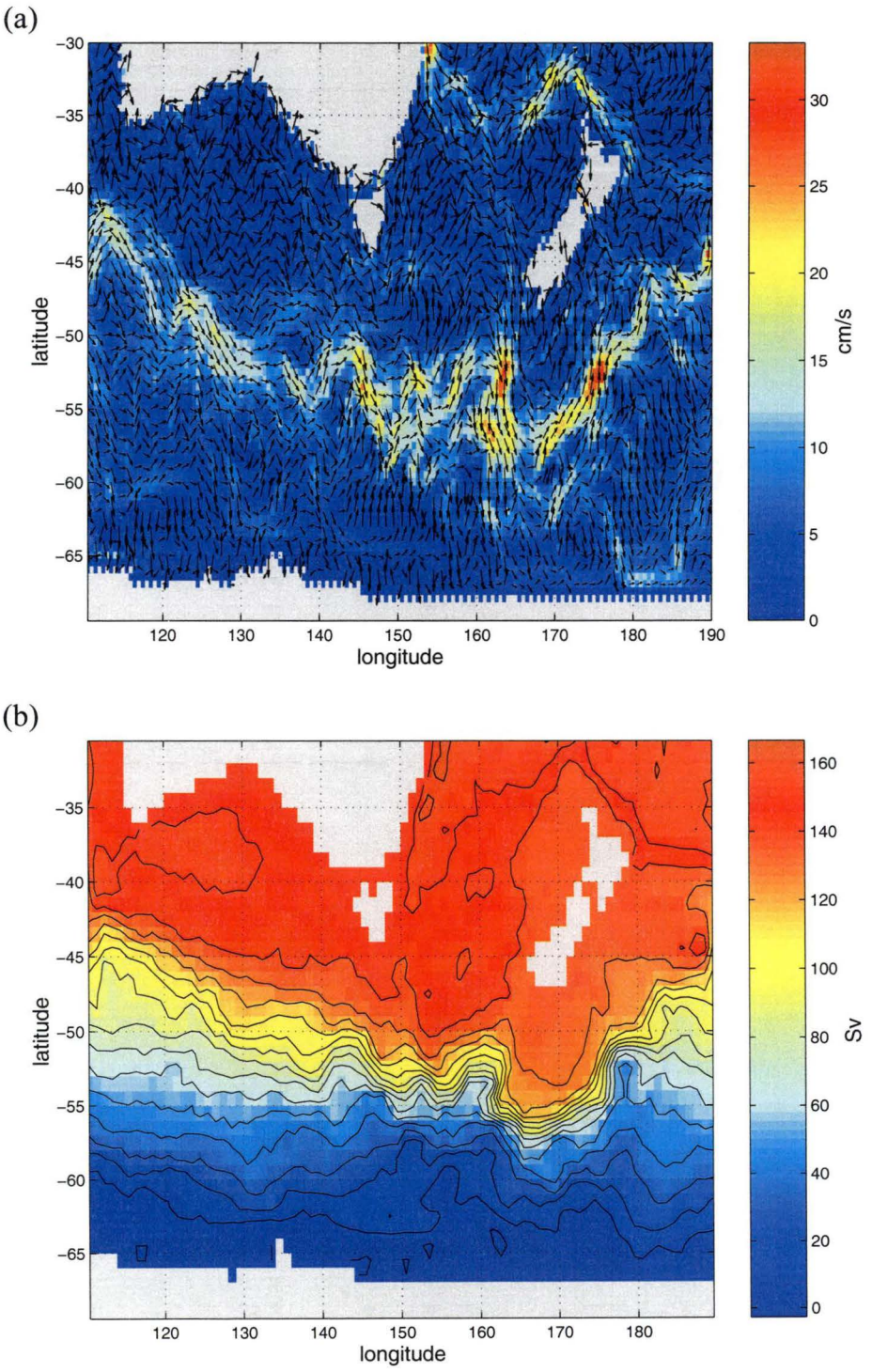


Figure 3.40: (a) Near-surface (layer two, 32 m), and (b) depth-integrated (EVEN) flows after five years for the Flather boundary condition with a timestepped η_0 reference profile.

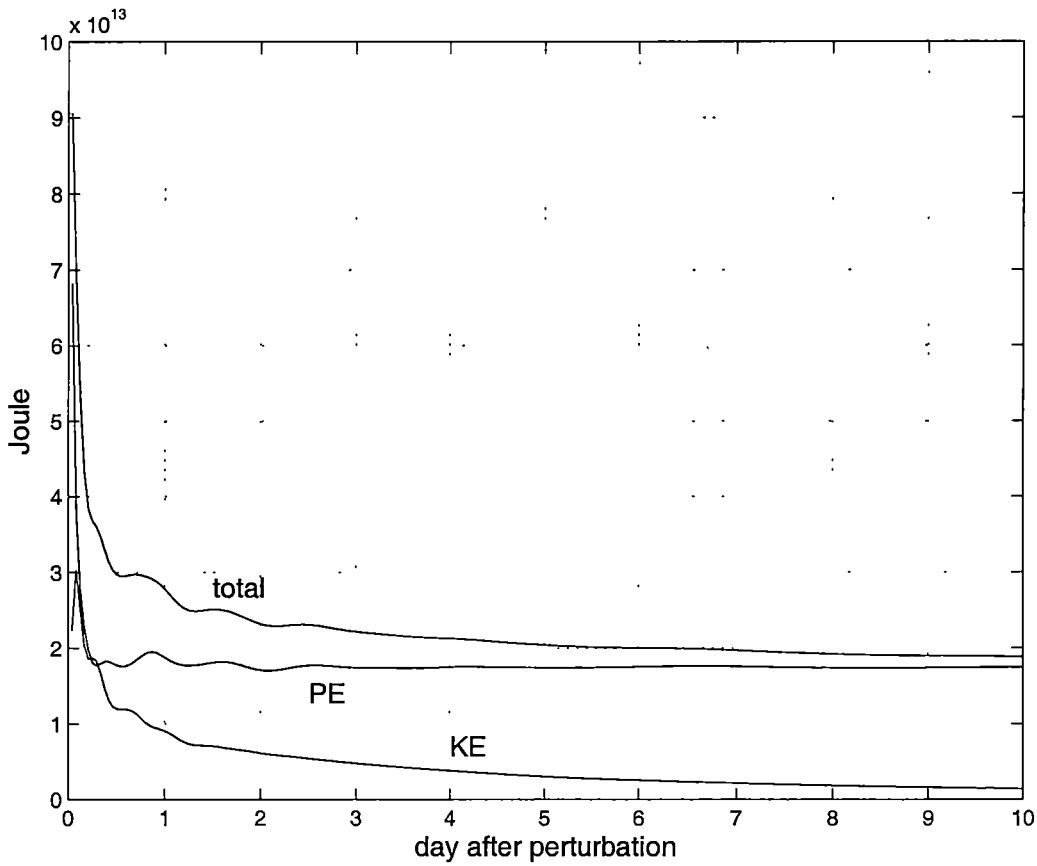


Figure 3.41: Potential (PE), kinetic (KE) and total energy of an initial sea-level perturbation for 10 days with the Flather boundary condition using a timestepped η_0 for the reference profiles.

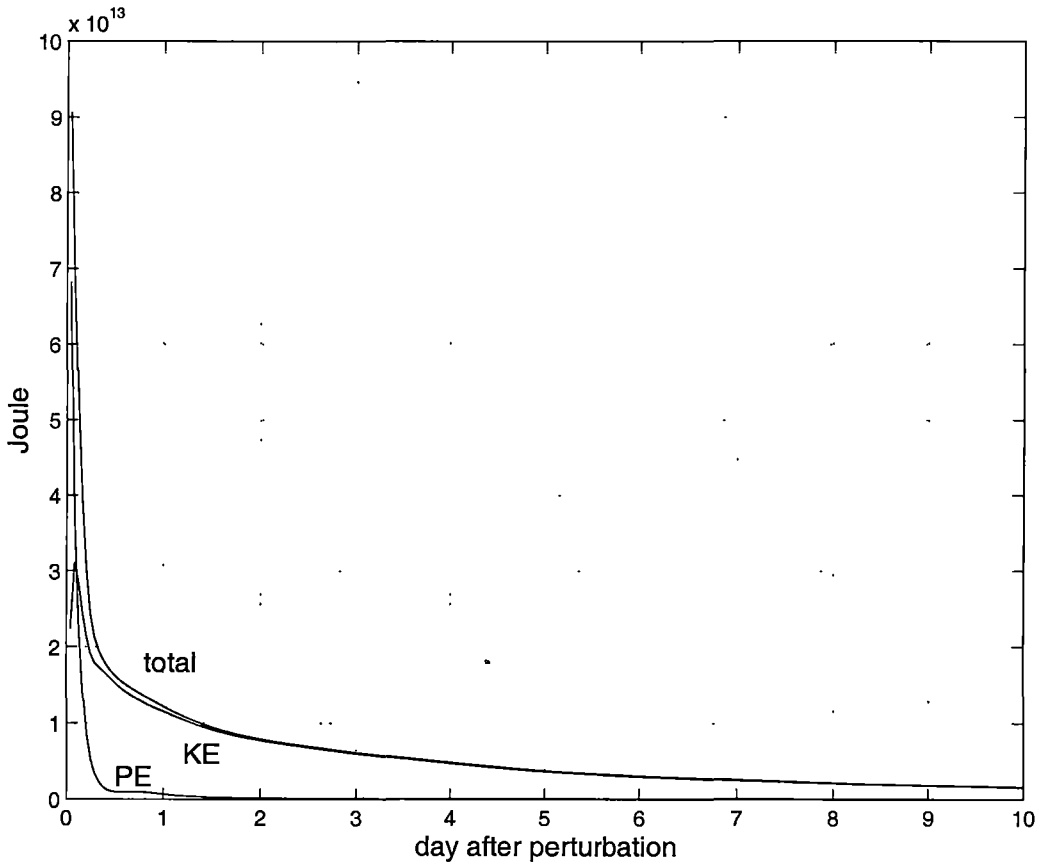


Figure 3.42: Potential (PE), kinetic (KE) and total energy of an initial sea-level perturbation for 10 days with the Flather boundary condition using two day relaxation for the reference profiles.

The flow fields after five years are shown in Figure 3.43. The realism of the flows, particularly in the region of the EAC, has been maintained in comparison to the Flather condition using a timestepped η_0.

The excellent volume conservation property of the timestepped condition has been reduced somewhat. The residual flow out of the domain, rather than varying by ± 0.01 Sv has increased to a range of ± 0.04 Sv, leading to mean sea-level variations over the course of the five year run of up to ± 1 cm, after an initial drop of around 15 cm (Figure 3.44).

Using a longer timescale only lead to decreased realism of flow patterns, and reduced volume conservation. An even shorter relaxation timescale would lead to improved volume conservation but increased reflective properties. A variation in mean sea-level of 1 cm over five years is entirely adequate for the purposes of this project.

3.5.5 Summary

For the barotropic system, only one boundary condition must be applied at each open boundary. It was found to be possible to apply a boundary condition to either normal velocity or sea-level in a stable manner, as well as the characteristic combination. While volume conservation was possible with a boundary condition on normal velocity, significant separation between subgrids occurred with a sea-level condition. Boundary conditions found to be successful at reducing reflections included radiation type conditions and a zero-gradient condition on normal velocity. In order for these to remain stable, however, relaxation to prescribed flows was also required. A clamped sea-level condition was also found to be nonreflective. While the ACC was found to be relatively easy to reproduce in the model, a realistic EAC was much more difficult. In all boundary conditions found to successfully reproduce an EAC, an artificial barotropic velocity profile had to be imposed along the northern boundary. The best performance of all conditions was one based on characteristics. This allowed almost perfect volume conservation, realistic flow patterns, as well as excellent transmission of surface gravity-wave disturbances. While similar to the condition of Flather (1976), and also that of Hedstrom (1979), it was extended here to include a relaxation term on the reference values of the characteristic variables. This allowed an explicit tradeoff between enforcement of the reference values (and therefore realism of flows and volume conservation) and radiation of surface gravity waves out of the domain.

3.6 The baroclinic system

The baroclinic system in HOPE solves the baroclinic momentum equations:

$$\begin{aligned}
\frac{\partial u'}{\partial t} - fv' &= -\frac{1}{\rho_0}\left(\frac{\partial p'}{\partial x} - \frac{1}{H}\int_{-H}^{0}\frac{\partial p'}{\partial x}dz\right) + G'_U \\
\frac{\partial v'}{\partial t} + fu' &= -\frac{1}{\rho_0}\left(\frac{\partial p'}{\partial y} - \frac{1}{H}\int_{-H}^{0}\frac{\partial p'}{\partial y}dz\right) + G'_V.
\end{aligned}$$

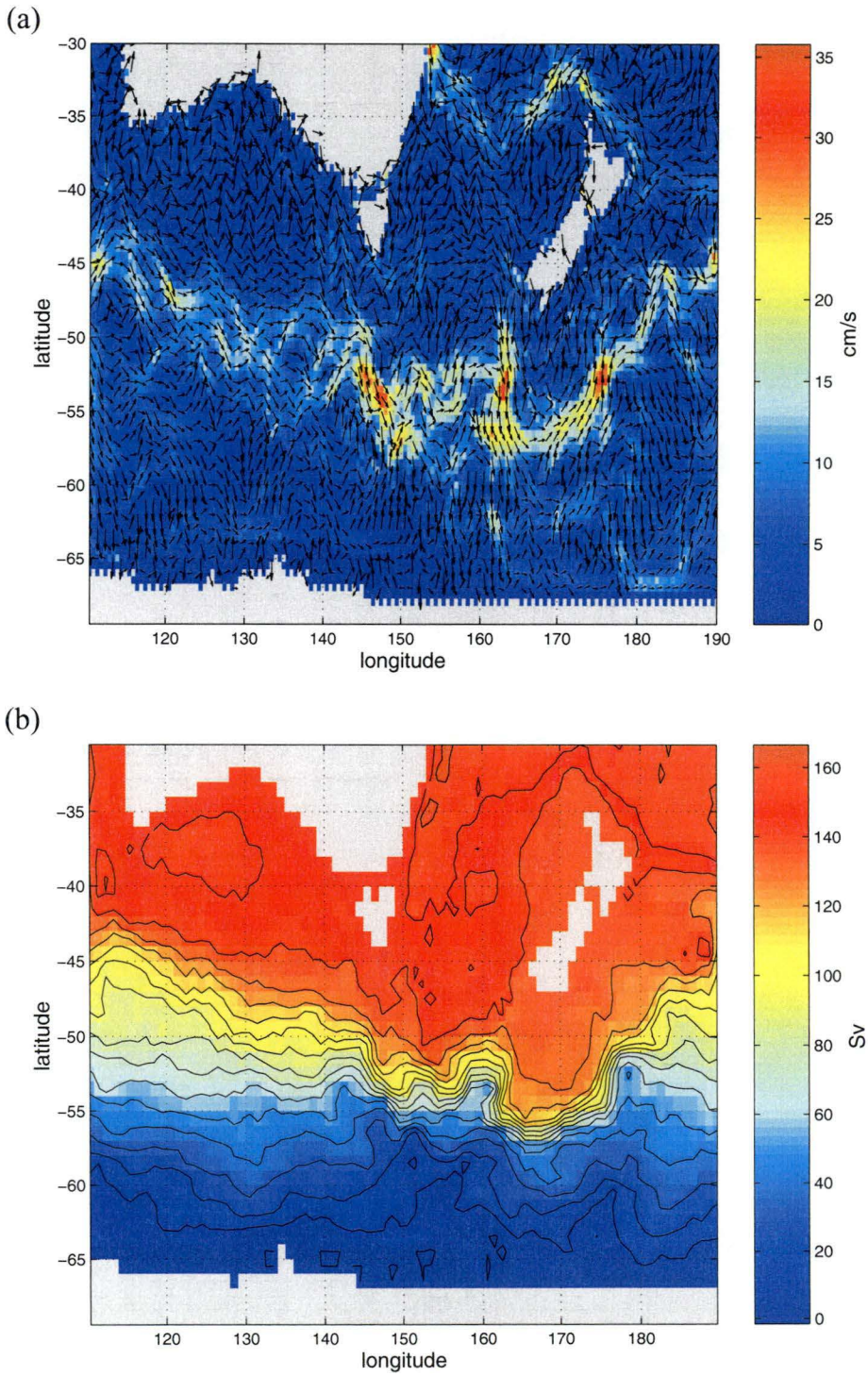


Figure 3.43: (a) Near-surface (layer two, 32 m), and (b) depth-integrated (EVEN) flows after five years for the Flather boundary condition using a two day relaxation timescale for the reference profiles.

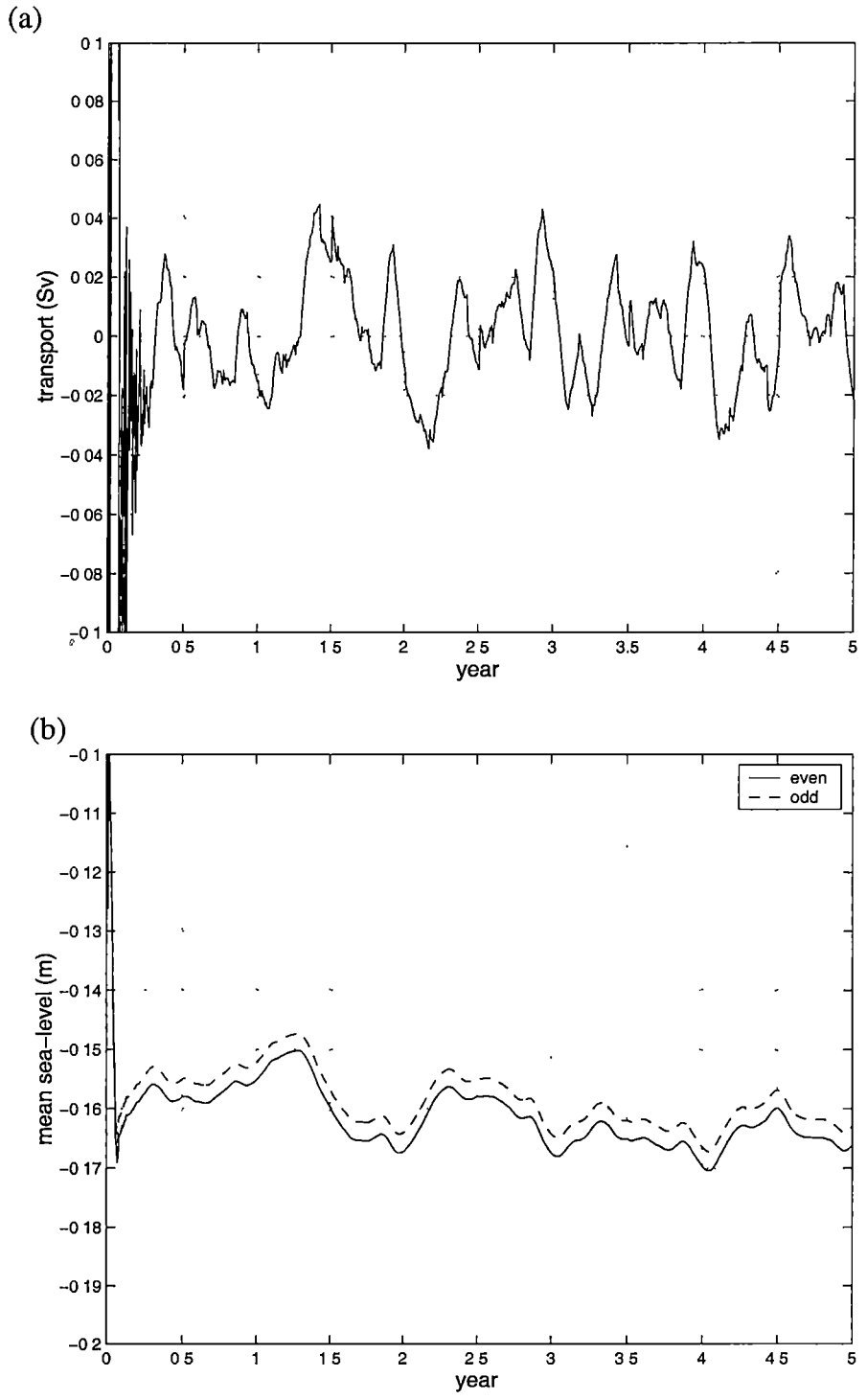


Figure 3.44: (a) Residual outflow from the model domain, and (b) resultant change in mean sea-level, over five years with the Flather boundary condition using a two day relaxation timescale on the reference transport profiles.

The terms G'_U and G'_V represent both momentum advection and viscosity. The hydrostatic relation is combined with the continuity equation to obtain a prognostic equation for pressure:

$$\frac{\partial^2 p'}{\partial t \partial z} = gw\frac{\partial \rho}{\partial z}.$$

Horizontal viscosity is parametrized with harmonic, biharmonic, and strain-dependent terms (see section 2.1.1). The various components of the baroclinic system are solved separately in HOPE's operator splitting procedure.

3.6.1 Well-posedness requirements

As mentioned previously (section 3.1.2), it is not possible to formulate well-posed boundary conditions that are point-wise in the vertical for the baroclinic system due to the hydrostatic assumption. In the absence of topography, baroclinic fields may be decomposed in a series of vertical modes, each of which satisfies reduced-gravity shallow-water equations[6]. Well-posed boundary conditions for these separate modes depend on the flow speed relative to the reduced-gravity shallow-water wave speed. No such decomposition in the vertical was attempted here. The solution of the coupled normal modes for the model each timestep would be completely impractical, requiring to solve an eigenvalue problem the size of the model. For simplicity, an *ad-hoc* approach was preferred, with the aim of finding boundary conditions that remained stable and produced relatively smooth fields near the boundaries. Separate conditions are required for each of the components of the baroclinic system in HOPE's operator splitting scheme.

3.6.2 Prognostic momentum equation

Figure 3.45 shows a limited portion of HOPE's staggered E-grid. Velocity points marked in blue are those for which surrounding pressure points all exist, and which, therefore, may be solved conventionally with the prognostic momentum equations. Similarly, the pressure points marked in green are those for which surrounding (u, v) points all exist, and at which vertical velocities may therefore be calculated (from continuity). The update of pressure on these points may be performed as usual. Thus it is apparent that baroclinic velocities may be calculated prognostically on the western boundary. Open boundary conditions for the baroclinic velocities u, v are required for the northern, eastern and southern boundaries (though the latter is land in the present model).

Radiation condition

Radiation conditions have been widely used for baroclinic velocities in open ocean modelling (Orlanski 1976; Oey and Chen 1992; Jensen 1998; Palma and Matano 2000). Both Gan et al. (1998) and Ezer and Mellor (1997) recently

[6]The presence of topography couples the modes.

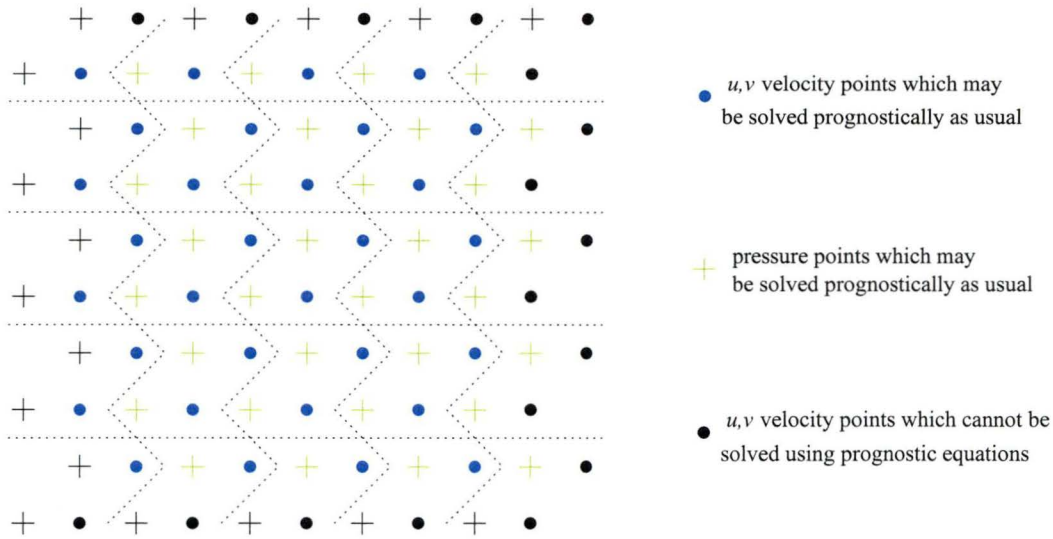


Figure 3.45: HOPE's staggered E-grid showing boundary points that cannot be calculated prognostically.

applied Orlanski-type radiation conditions to the baroclinic velocities at open boundaries in the Southern Ocean of Atlantic models. Barnier et al. (1998) and Treguier et al. (2000) have also applied radiation conditions to baroclinic velocities in the Southern Ocean, and included relaxation to geostrophy.

The version implemented here used Miller and Thorpe's (1981) two time-level modification of Orlanski's leapfrog condition (see section 3.2). Thus, for the unknown baroclinic velocities on the boundaries (those marked in black in Figure 3.45), the following boundary condition was applied:

$$\phi_b^{n+1} = (1-r)\phi_b^n + r\phi_{b+1}^n,$$

where ϕ represents either u or v, the subscript b represents the boundary gridpoint, and $b+1$ means the gridpoint one in from the boundary. The radiation phase speed, r, is estimated as

$$r = \frac{\phi_{b+1}^n - \phi_{b+1}^{n-1}}{\phi_{b+2}^{n-1} - \phi_{b+1}^{n-1}}.$$

In addition, r is clamped to the range $0 \leq r \leq 1$.

Together with the scheme for viscosity on the boundaries described below, this radiation condition proved to be stable. Figure 3.46 shows a partial section of normal velocity v along the northern boundary in the region of the EAC after 30 days[7]. While the radiation condition has certainly given rise to vertical structure, the separation of the EAC into two jets, one at the surface, and one at depth, seems unrealistic. With the radiation condition, this feature persisted throughout a five year run. Given the difficulties of maintaining a reasonable

[7]The full resolution assimilation model was used here instead of the coarse resolution model used for investigating barotropic open boundary conditions in section 3.5. Model parameters were listed in section 2.2.

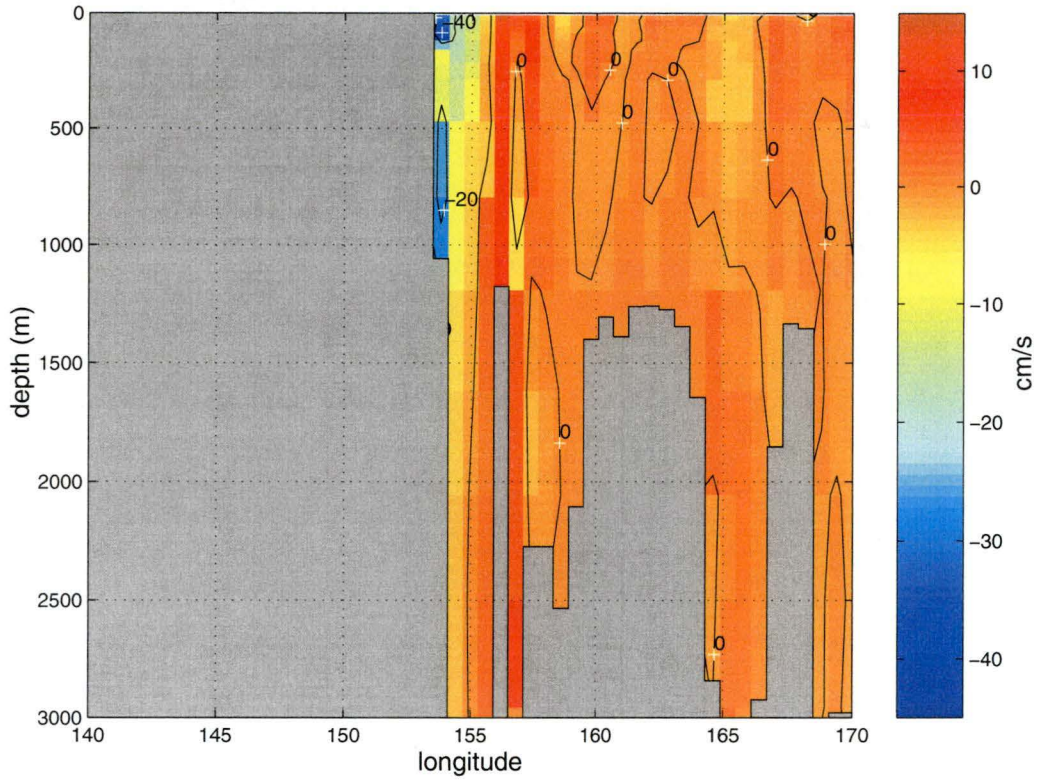


Figure 3.46: EAC baroclinic flow after 30 days using a modified Orlanski radiation condition on the northern boundary.

EAC already noted in the discussion of barotropic boundary conditions, a zero-gradient condition was used here and found to provide more realistic results.

Zero-gradient at northern boundary

The zero-gradient condition for u and v on the northern boundary simply sets the baroclinic velocities equal to those one gridpoint in from the boundary:

$$\{u, v\}_b^{n+1} = \{u, v\}_{b+1}^n.$$

The usual dynamical balance for the surface *Ekman velocity* is between the accelerating wind stress and the Coriolis force. Since the momentum equation, including Coriolis terms, is replaced by the zero-gradient condition on the northern boundary, no wind forcing is performed on the northernmost gridpoints either. If wind forcing is included, then extremely high surface velocities result. Figure 3.47 shows the EAC after 30 days using the zero-gradient condition. The double-jet structure of the EAC has successfully been replaced by a more realistic structure. Since density modifications made through assimilation (described in chapter 5) are typically broader in scale than the first baroclinic Rossby radius, the geostrophic adjustment process will not strongly radiate internal gravity waves (Gill 1982). The dynamic modes excited through assimilation will be predominantly the slower, westwards travelling Rossby

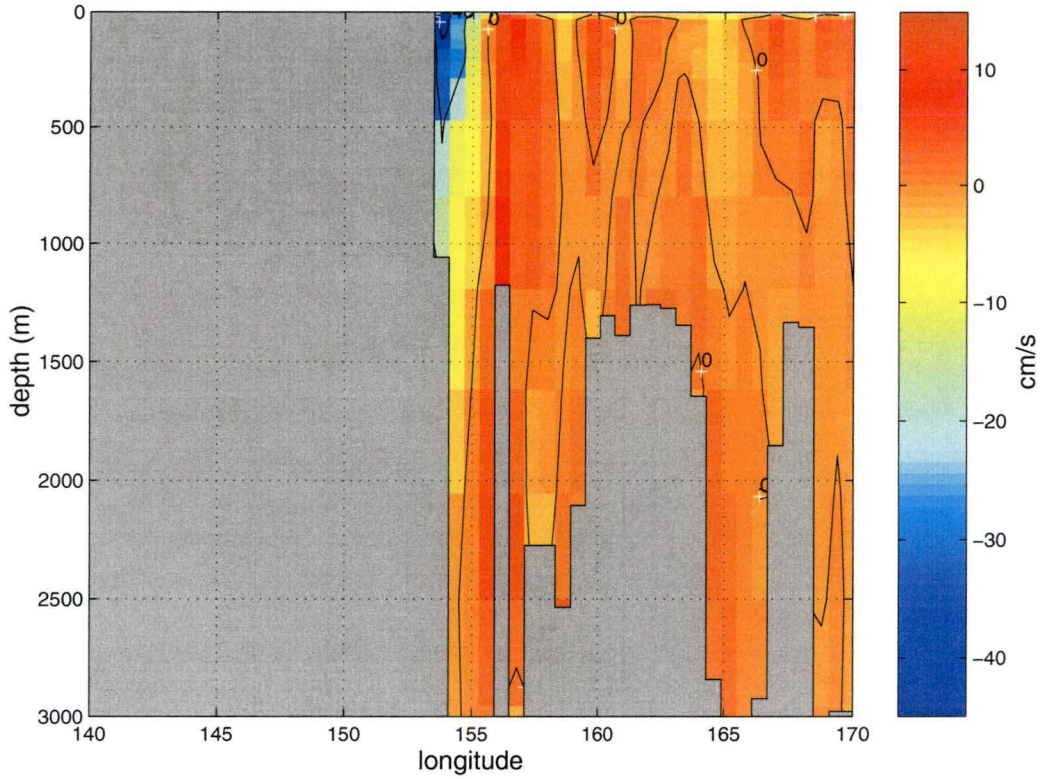


Figure 3.47: EAC baroclinic flow after 30 days using a zero-gradient condition on the northern boundary.

waves. It is therefore not expected that replacing the radiation condition at the northern boundary with a zero-gradient condition will lead to significant reflection of internal gravity waves under altimetry assimilation.

3.6.3 Viscosity

We may consider viscosity as a process separate from the baroclinic momentum equations, consistent with HOPE's operator splitting methodology, for the purposes of developing well-posed boundary conditions. First, vertical viscosity may of course be performed on all lateral boundaries. Horizontal viscosity is parametrized in HOPE with harmonic, biharmonic and strain-dependent terms, although the biharmonic term was not included in the Southern Ocean model constructed here. Both the harmonic and strain-dependent terms are parabolic in form, and so a mathematically well-posed problem results by specifying either Dirichlet or Neumann boundary conditions (section 3.1.2). Neumann conditions were used here, with the normal gradient of baroclinic velocities specified to be zero. Thus in the discretisation of the Laplacian

$$\frac{\partial}{\partial t}\{u, v\} = A_h \nabla^2 \{u, v\}$$

at a boundary, whenever the value of a nonexistent velocity point outside the domain is required, it is taken to be equal to the velocity just inside the

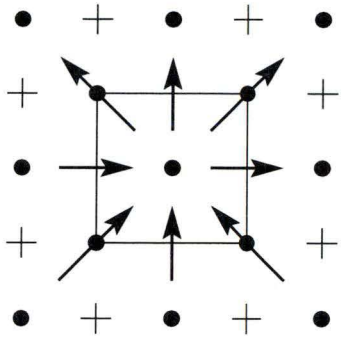


Figure 3.48: Volume transport fluxes are calculated at surrounding gridpoints in the $\mathbb{J}_7$ advection scheme.

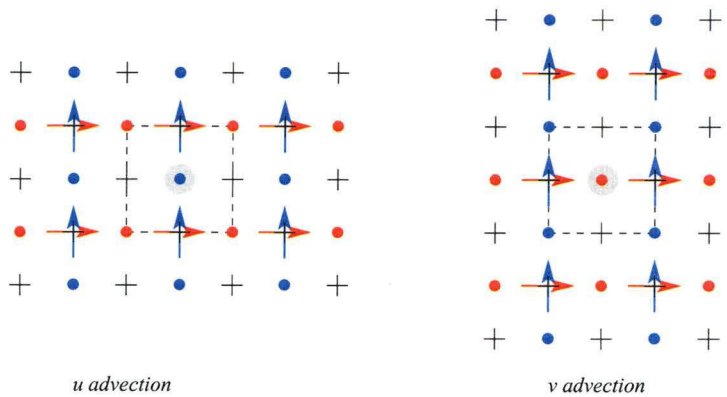


Figure 3.49: Required velocities for calculating momentum advection.

boundary. Compared with other terms in the momentum equations, horizontal mixing is numerically small. Nevertheless it is required to avoid instability. Trial experiments omitting viscosity on the boundaries developed strong and unstable gradients. Similar behaviour was noted by Stevens (1991) who also used a zero normal gradient condition for the nonexistent velocity points outside the computational grid.

3.6.4 Momentum advection

The numerical implementation of momentum advection in HOPE uses the '$\mathbb{J}_7$ Jacobian' of Arakawa and Lamb (1977). This conserves second order quantities (kinetic energy and enstrophy). For advecting momentum at a given velocity point, the formulation calculates volume transport fluxes at each of the eight immediately surrounding (scalar and vector) gridpoints, forming an enclosing cell as shown in Figure 3.48. Omitting the details, in order to calculate these fluxes, an extended stencil of gridpoints is used, as shown in Figure 3.49. For simplicity, it was decided to neglect momentum advection on the boundaries. This approach is common in open ocean modelling (*e.g.* Stevens (1991)).

However, because of the extended stencil, velocity gridpoints one in from the boundary require access to nonexistent velocity points outside the domain in order to calculate the required volume transport fluxes. As with viscosity, a zero-gradient condition was applied, so that the nonexistent velocities were assumed equal to those just inside the boundary.

It should be noted that advection is a hyperbolic process in characteristic form. Therefore, for mathematical well-posedness, the advected quantity should be specified whenever flow is into the domain, with upstream advection used for outflow. The deliberate ignoring of momentum advection here was based on the fact that the small changes that would result are insignificant compared with the boundary conditions used for the prognostic momentum equation described above.

3.6.5 Summary

In summary, the open boundary conditions for the baroclinic system used a modified Orlanski radiation condition for the eastern boundary, together with a zero-gradient condition for the northern boundary. Viscosity was performed on all boundaries, using a Neumann condition to allow calculation of harmonic terms. Momentum advection was ignored on the boundaries.

3.7 Tracers

The physics of tracers in HOPE is governed by advection and diffusion:

$$\frac{\partial \phi}{\partial t} = -\boldsymbol{u}.\nabla\phi + F_\phi,$$

where ϕ represents either salinity or potential temperature. Diffusion, F_ϕ, is parametrized with harmonic and strain-dependent terms.

3.7.1 Advection

As mentioned above, advection is a hyperbolic process where the characteristic variables are equivalent to the advected quantity. They are 'incoming' for flow into the domain. Thus, boundary conditions must be applied to temperature and salinity whenever flow is into the domain. The usual advection must be performed when flow is out of the domain. The boundary conditions implemented here on all open boundaries used a combination of relaxation to climatology for inflow, and upstream advection for outflow. Where a boundary lies along a characteristic curve, no boundary condition should be applied, that is tracers should be advected tangentially along the boundaries as usual. The relaxation timescale for inflow increased with depth, as detailed in Table 3.4. The longer timescales at depth were motivated by the more slowly evolving fields, and smaller velocities in the deeper ocean. It is similar to the increasing relaxation timescales with depth used in the robust diagnostic method of FRAM (The FRAM Group 1991) (180 days for the top 140 m, and 540 days

Table 3.4: Relaxation timescales for temperature and salinity under inflow.

model layer	*relaxation timescale (days)*
1 (20 m)	30
2–8 (20–330 m)	60
9–11 (330–1400 m)	180
12–20 (1400–5500 m)	360

for the deeper levels). Early experiments indicated that if short (*e.g.* 30 day) relaxation was used throughout the water column, then unrealistically strong boundary currents flow along the western boundary.

3.7.2 Diffusion

As with viscosity, well-posed boundary conditions for laplacian diffusion are obtained by applying a Neumann condition of zero-gradient across the boundaries for salinity and potential temperature. Wherever diffusion on a boundary tracer point required nonexistent values outside the domain, these were assumed equal to the first point just inside the boundary.

3.7.3 Summary

The boundary conditions for tracers depended on the local flow direction: for inflow, relaxation to climatology was used, while for outflow, upstream advection was used. For diffusion, a Neumann condition of zero normal gradient was applied.

3.8 Open boundary conditions: a summary

The most sensitive component of HOPE, and the most difficult to implement, were open boundary conditions for the barotropic system. A modified Flather boundary condition, based on the characteristic variables, proved to be the best of all considered. It provided excellent volume conservation as well as the ability to transmit surface gravity waves. Reference profiles of normal volume transport (or barotropic velocity) were needed along the boundaries. These were obtained by bottom-referencing and scaling the geostrophic flow for the western and eastern boundaries, while an artificial profile based on observed flows was used for the northern boundary. For the baroclinic system, a radiation condition was used along the eastern boundary. A similar technique produced unrealistic flows on the northern boundary, and so a zero-gradient condition was applied instead. For the tracers, there is little freedom, and an obvious physically motivated approach of relaxation towards climatology for inflow with upstream advection for outflow was applied.

The examination of boundary conditions here for the barotropic system appears to be one of the first with implicit numerical schemes. Previous studies

of open boundary condition performance (Chapman 1985; Røed and Cooper 1986; Røed and Cooper 1987; Jensen 1998; Palma and Matano 1998) have used numerical schemes explicit in time. While significantly complicating the technical implementation, implicit numerics appear also to change the nature of the results. The findings here that both a clamped sea-level condition and zero-gradient velocity condition are nonreflective are surprising, and at odds with results reported by Chapman (1985), Røed and Cooper (1986), Røed and Cooper (1987) and Jensen (1998). The discrepancy may at least partly be related to the numerical dissipation of implicit schemes. (Interestingly, Israeli and Orszag (1981) proposed sponge layers to dampen reflections at boundaries by introducing artificial numerical dissipation.) Further investigation of the differences between boundary conditions for explicit and implicit numerical schemes would require more detailed energy analyses, as well as comparisons of results across the full range of 'implicitness' (α and β in HOPE covering the full range 0 to 1).

The general performance of the assimilation model under these boundary conditions in a spin-up run is analysed in the following chapter.

Chapter 4

Model performance

Chapter 3 described the implementation of open boundaries in the HOPE model for the limited-area Southern Ocean configuration developed in this thesis. Numerous boundary conditions were trialled by analysing results from short (five year) runs of the model, with a one degree grid used for the barotropic experiments. This chapter provides a more detailed analysis of the model on a 0.6° x 0.4° grid. This is the configuration used for the assimilation experiments in the next chapter.

The analysis serves several purposes. First, the general usefulness of the model for this region needs to be determined. From the boundary condition experiments in chapter 3, we know the model is capable at least of reproducing major current structures (ACC, EAC) for up to five years. The realism of the model in other respects and for a longer integration time is analysed here. Second, variability of subsurface model fields is analysed to guide the implementation of the altimetry assimilation scheme discussed in the next chapter. The assimilation scheme modifies subsurface properties based only on sea-level measurements. While physical arguments strongly constrain the types of subsurface modifications that can be made, an analysis of subsurface variability of the model and its surface expression aids the development of the assimilation scheme. Third, initial conditions for an assimilation of real altimetry data must be determined. The analysis here suggests a sensible starting point. Finally, while an assimilation of real altimetry may be verified against independent (*e.g.* hydrographic) data, it is of interest to know how other aspects of the modelled circulation (*e.g.* heat flux etc.) have been influenced by data assimilation.

A 40 year run of the model forms the basis of the analysis in this chapter. The model parameters were detailed earlier (section 2.2), and the previous chapter described the conditions applied at open lateral boundaries (summarised in Table 3.1). The model was started from rest, initialised with the Olbers (1992) climatology. Monthly climatological wind forcing (Hellerman and Rosenstein 1983) and surface relaxation of tracers to climatological fields (Olbers et al. 1992; Levitus and Boyer 1994) were performed throughout the run.

The analysis of the 40 year run in this chapter is separated by timescale into characteristics which evolve slowly (described as the model climatology)

and more rapid (sub-annual) variations (described as model variability). Complete model fields were stored every 30 days over the course of the run. In addition, a few diagnostic parameters were calculated at every timestep.

4.1 Model climatology

It is usual for general circulation models of both the atmosphere and ocean to exhibit some drift. For example, Saunders et al. (1999) discuss in some detail drift in the OCCAM global model. This simply reflects the fact that models are an imperfect and incomplete representation of the real world. For a model with artificial boundaries, such as that implemented here, the problem may be exacerbated by the artificial constraints imposed on the boundaries as well as the fundamental ill-posedness of the primitive equations under any pointwise open boundary condition. This section examines drift in the model. While not "climate drift" in the conventional sense (timescales are of decades rather than millenia), it represents slow evolution of the model which influences decisions on how and when to attempt data assimilation.

4.1.1 Thermodynamics

Layer averages

Volume averaged potential temperature and salinity were calculated by layer throughout the 40 year run. Figures 4.1 and 4.2 show timeseries of the differences between the model values and the Olbers et al. (1992) climatology from which the model was initialised. In addition, Table 4.1 lists by layer the volume averaged tracer values for both the climatology and the final year of the 40 year run. Changes are most apparent in the uppermost layers where the model has both cooled and freshened with respect to the climatology.

Watermass census

Examining temperature and salinity separately by layer does not indicate how the changes affect the density structure. It is perhaps more useful to analyse the density modifications induced by these changes. Bindoff and Church (1992), Bindoff and McDougall (1994) and Wong et al. (1999) have used hydrographic data to analyse climatological changes of temperature and salinity in density classes, in both the Pacific and Indian Oceans. A less sophisticated analysis is performed here, which counts by density class the total water volume throughout the 40 year run. To avoid complications associated with the choice of a reference level for potential density, the analysis was performed using neutral density, γ^n (McDougall 1987). The density classes chosen are intended to correspond approximately to various recognised watermasses, listed in Table 4.2 together with the density ranges. Figure 4.3 plots the total volume of water in each class over the 40 years of the run. To gain more insight into the changes in watermass properties, Figure 4.4 shows sections of potential temperature and salinity along 160° E from the beginning and end of the 40

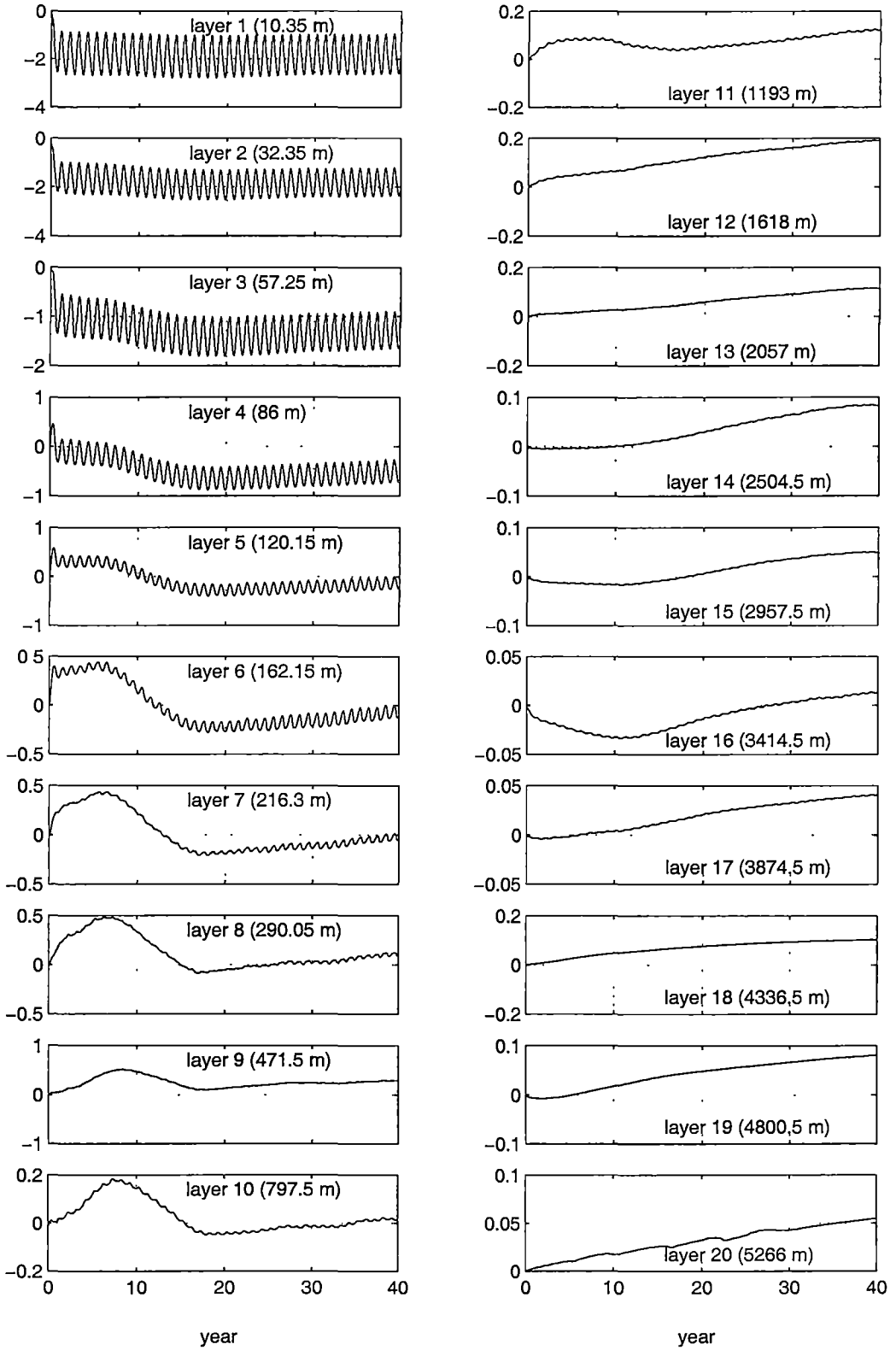


Figure 4.1: Volume averaged potential temperature by layer as differences from climatology (°C).

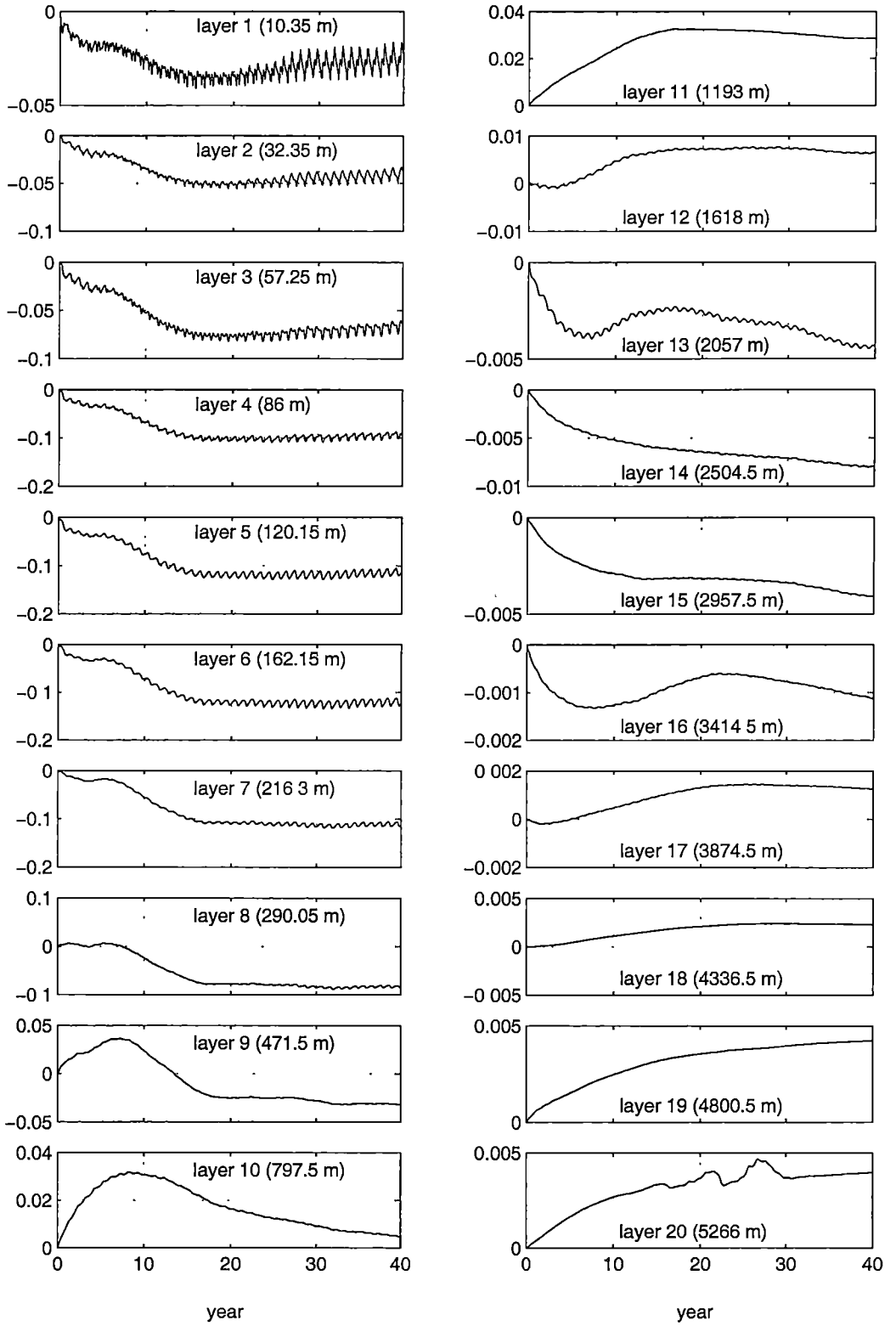


Figure 4.2: Volume averaged salinity by layer as differences from climatology (psu).

Table 4.1: Volume averaged temperature and salinity by layer: climatological values, and difference for mean of year 40.

layer	*mid-depth (m)*	*climatology*		*difference (year 40 mean)*	
		θ (°C)	S (psu)	θ (°C)	S (psu)
1	10	11.36	34.637	-1.84	-0.025
2	32	11.00	34.647	-1.83	-0.040
3	57	10.06	34.653	-1.30	-0.066
4	86	8.97	34.671	-0.52	-0.093
5	120	8.33	34.691	-0.14	-0.113
6	162	8.03	34.703	-0.07	-0.120
7	216	7.75	34.702	-0.03	-0.111
8	290	7.31	34.679	0.11	-0.083
9	472	6.32	34.602	0.29	-0.032
10	798	4.74	34.523	0.02	0.005
11	1193	3.00	34.540	0.12	0.029
12	1618	2.17	34.633	0.19	0.006
13	2057	1.72	34.695	0.12	-0.004
14	2505	1.37	34.717	0.08	-0.008
15	2958	1.06	34.720	0.05	-0.004
16	3415	0.81	34.716	0.01	-0.001
17	3875	0.61	34.711	0.04	0.001
18	4337	0.48	34.708	0.10	0.002
19	4801	0.53	34.707	0.08	0.004
20	5266	0.52	34.706	0.05	0.004

Table 4.2: Watermass classes and their defined range of density used in performing volume census (Schmitz 1995; Bindoff et al. 2000).

watermass	*density range* (γ^n)
Subantarctic Mode Water (SAMW)	26.5–27.2
Antarctic Intermediate Water (AAIW)	27.2–27.5
Circumpolar Deep Water (CDW)	27.7–27.9
Antarctic Bottom Water (AABW)	> 28

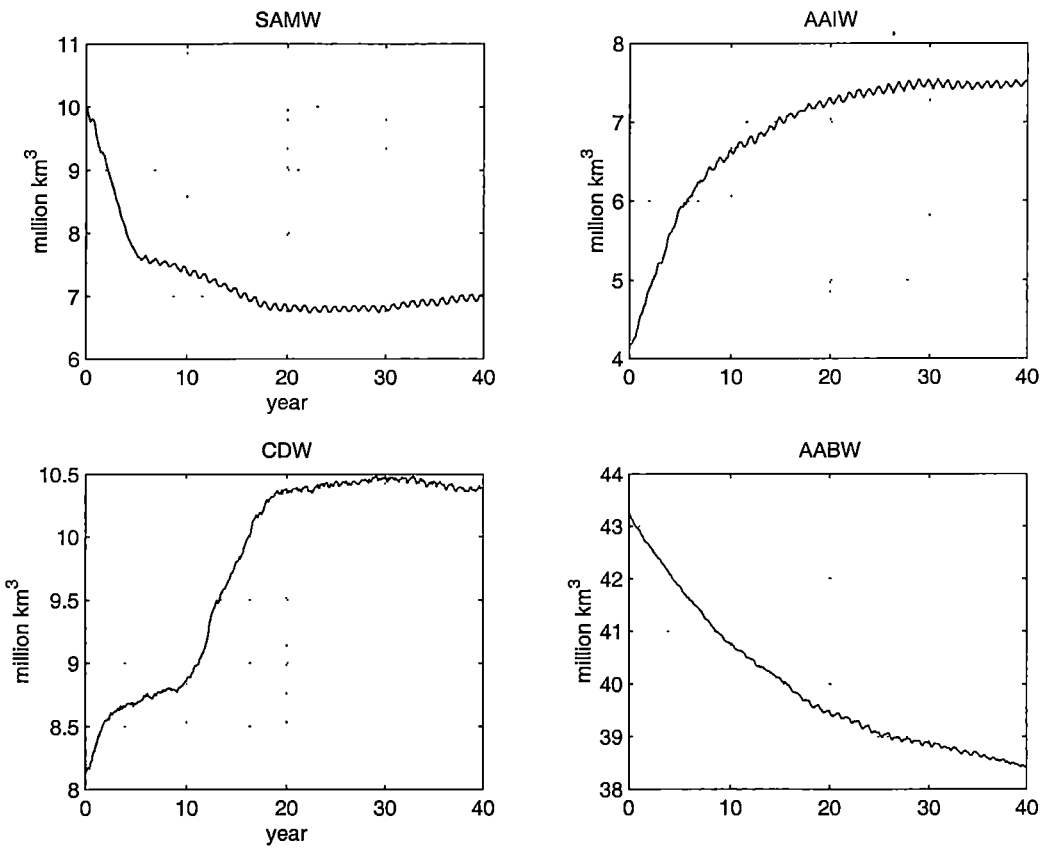


Figure 4.3: Volume census by watermass class over 40 year run of the model (EVEN grid used).

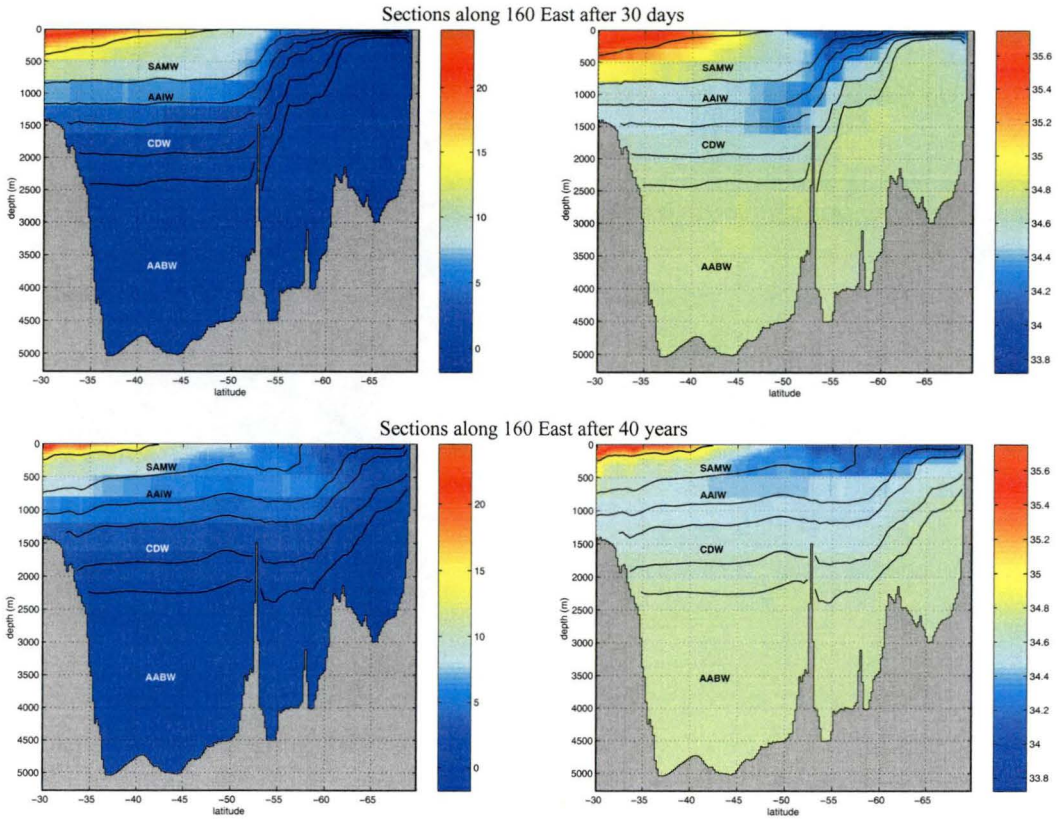


Figure 4.4: Sections of potential temperature (left) and salinity (right) near the beginning (after 30 days), and after 40 years, of the run. Potential density contours delineating watermasses are overlaid.

year run. From these sections, the dominant mode of change appears to be a lifting of isopycnals in the north of the domain and a lowering of isopycnals in the south. From visual inspection, tracer properties are more or less conserved within density layers. The decrease in SAMW volume over time is due to significant lifting of isopycnals north of around 50° S, with outcropping at the surface. The increase in volume of AAIW and CDW almost compensates the loss of AABW and arises through significant lowering of isopycnals south of around 52° S, with stretching south of around 62° S. The layer average tracer differences tend to mask these changes. For example, at around 1200 m depth (layer 11 in the model), there is almost one degree C of cooling in the north, while the waters in the south have warmed by a similar amount. Likewise, salinity increases of up to 0.2 psu occur in the north, with similar decreases in the south. The layer averaged changes of potential temperature and salinity (Table 4.1) are only +0.12° C and +0.029 psu respectively. The sections along 160° E suggest an intuitive explanation for what is occuring: lack of bottom water formation along the Antarctic coast leads to a reduction in AABW over time, with a consequent southwards migration of the ACC and general slumping of isopycnals in the south. In the north of the domain, the loss of warm and salty subtropical waters at the surface leads to a lifting of density surfaces. The following two sections examine the vertical and horizontal circulation in

greater detail.

4.1.2 Meridional circulation

Because the model is not zonally periodic, a streamfunction for the meridional overturning circulation cannot be defined. Zonally integrating the continuity equation gives

$$\int_{x_W}^{x_E} \left(\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) dx = -[u(x_E) - u(x_W)].$$

Vertical divergences are balanced by the zonal transport difference between the eastern and western boundaries; only if this were zero could a meridional overturning transport streamfunction be defined. Instead of a streamfunction, Figure 4.5 shows the zonally integrated meridional volume transport (in Sv) by layer for each successive five year mean of the 40 year run:

$$V_{k_{tot}} = \int_{x_W}^{x_E} \int_{z_{k_{low}}}^{z_{k_{hi}}} \overline{v}(x, y, z)\, dy\, dx,$$

where the overbar represents a five-year mean, and where the $k-$th layer extends in depth from $z_{k_{low}}$ to $z_{k_{hi}}$. Figure 4.6 similarly shows the zonally integrated vertical volume transport:

$$W_{k_{tot}} = \int_{x_W}^{x_E} \int_{z_{k_{low}}}^{z_{k_{hi}}} \overline{w}(x, y, z)\, dy\, dx.$$

The most significant feature of these diagrams is the lack of a strong overturning circulation adjacent to Antarctica associated with deep water formation. In addition there is evidence of upwelling at intermediate depths north of 50° S. These observations are consistent with the intuitive picture formed above from an examination of meridional tracer and density sections. Lack of bottom water formation in the model is not surprising since no sea-ice model is included and the surface relaxation scheme uses climatological fields that are heavily biased towards summer observations.

4.1.3 Horizontal circulation

The open boundary conditions applied to the model were seen in the last chapter to provide reasonable surface and depth-integrated flow patterns for up to five years in a trial run. The circulation is examined here for a run of 40 years. Figures 4.7 and 4.8 show mean surface and depth-integrated currents for each successive period of five years in the 40 year integration. The realistic circulation obtained in the first 10 years of the run progressively deteriorates so that by the final five years, some very unrealistic current patterns are evident. Specific problems include

- the southwards migration of the Antarctic Cirumpolar Current,
- the strong retroflection at 60° S, south of New Zealand,

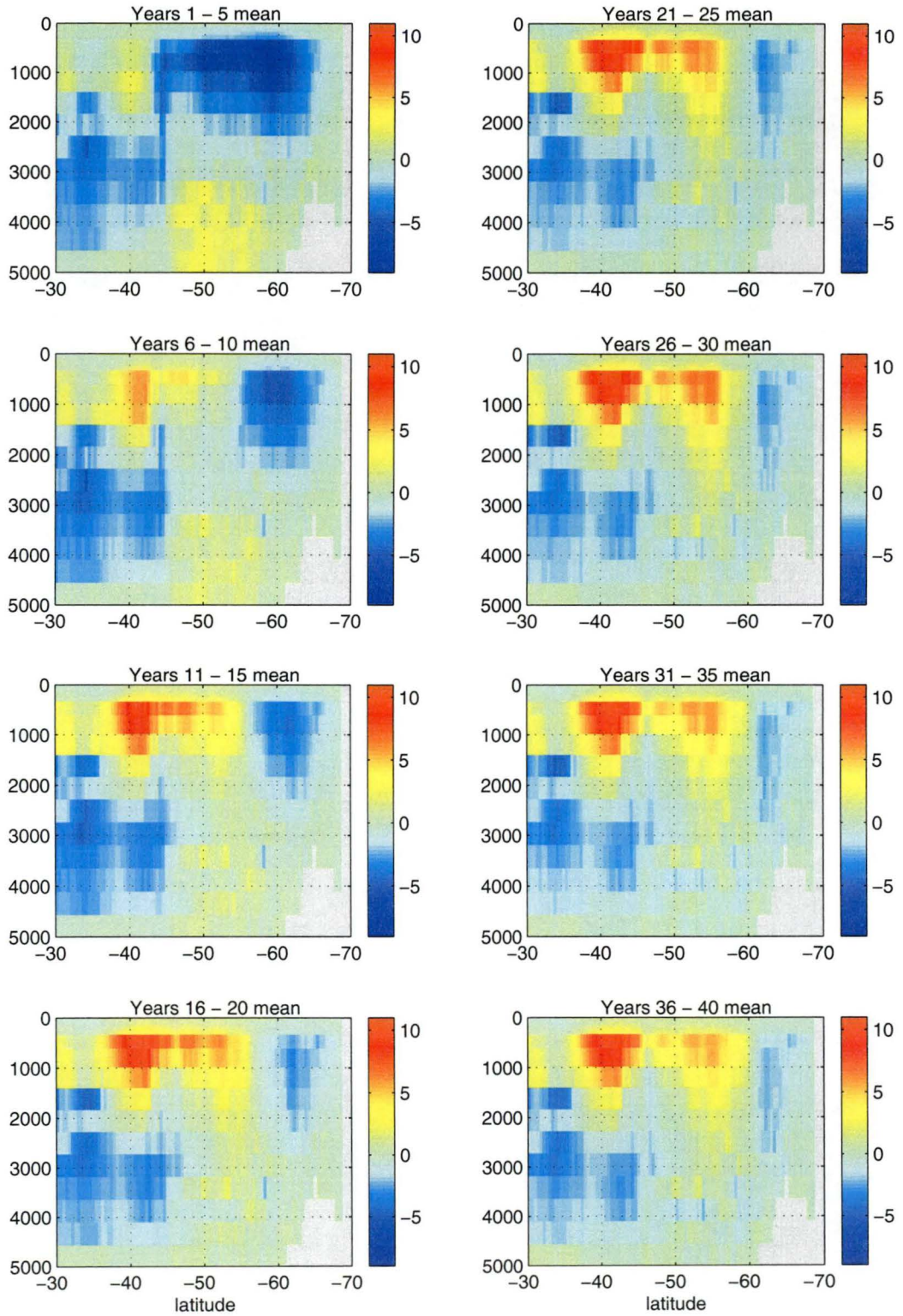


Figure 4.5: Zonally integrated meridional volume transport (Sv) in each layer, by depth (metres) for five-year means of the 40 year run (EVEN grid).

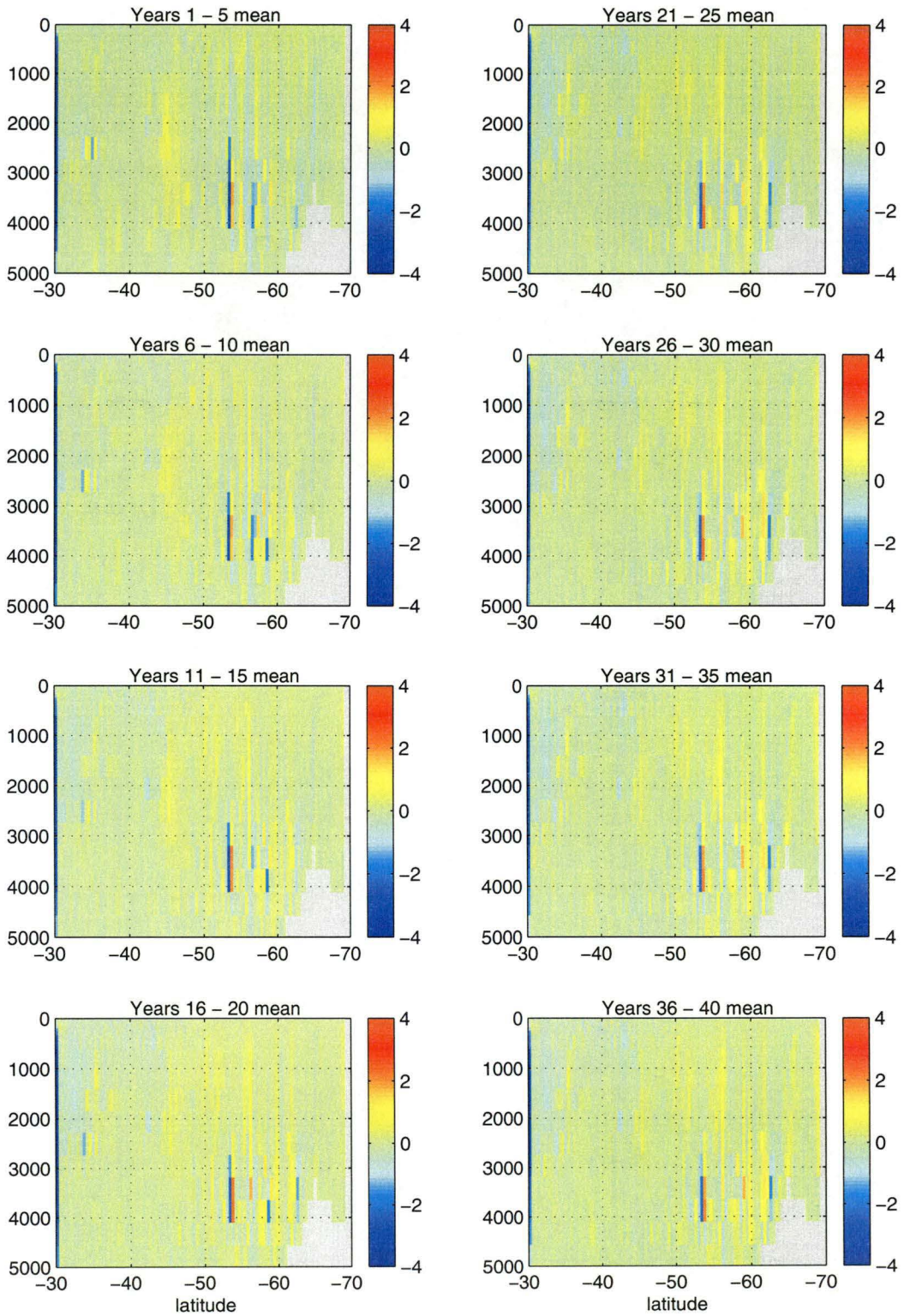


Figure 4.6: Zonally integrated vertical volume transport (Sv) in each layer, by depth (metres) for five-year means of the 40 year run (EVEN grid).

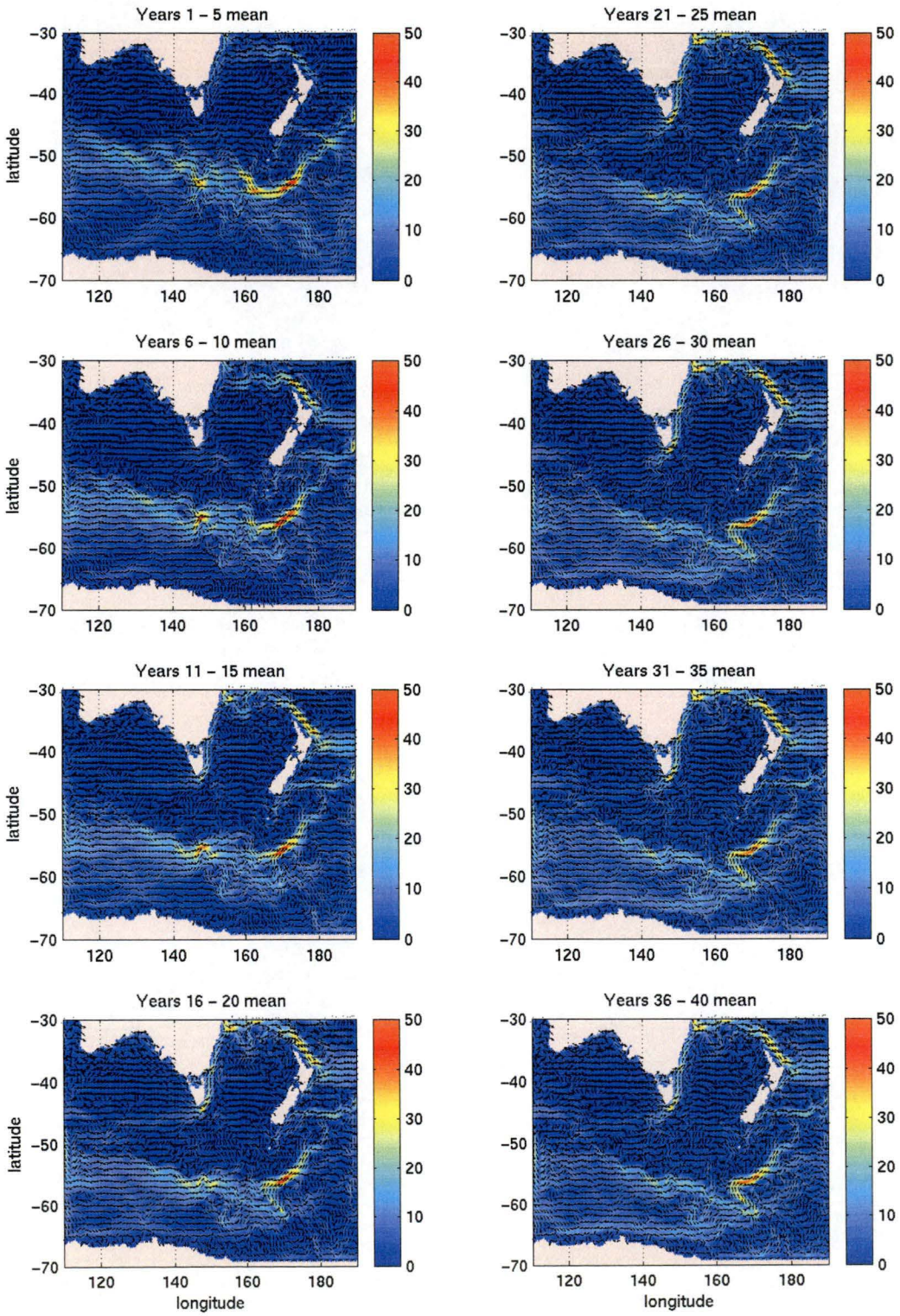


Figure 4.7: Near surface (layer 2, 32 m depth) average currents (cm/s) for each successive five year period of the 40 year run.

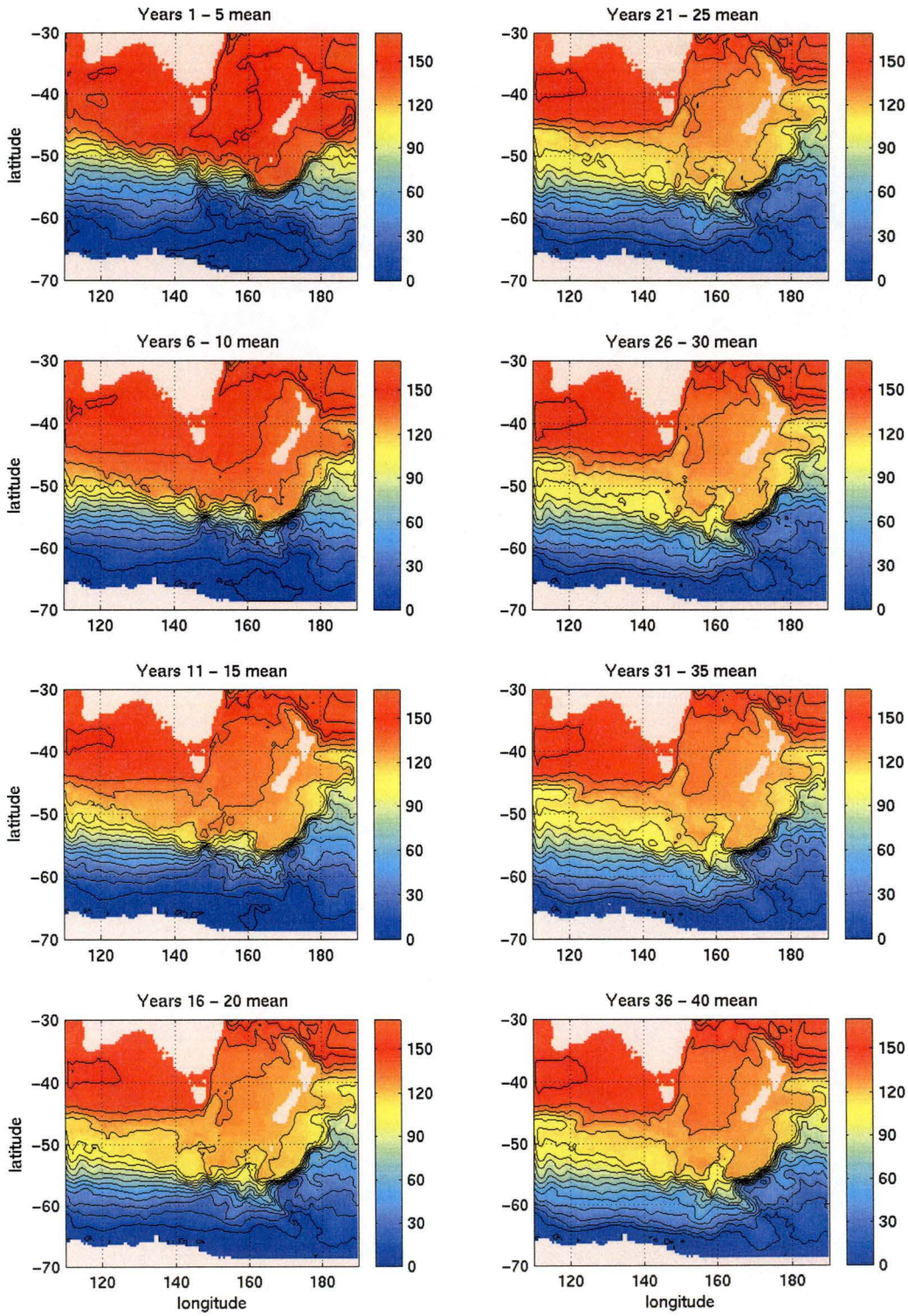


Figure 4.8: Average barotropic streamfunction (Sv) for each successive five year period of the 40 year run.

- the strong northwards flow around the southeastern corner of Tasmania and up along the east coast of Australia,
- an artificial boundary current southwards along the western open boundary, and
- an enhanced eastwards flow along the Antarctic continental shelf.

The southwards migration of the ACC was seen earlier in meridional tracer and density sections, and hypothesized to be at least partly due to the lack of sea-ice around Antarctica.

The peculiar flow south of New Zealand developing from year 15 or so is clearly related to the topography. The flow encounters the Southwest Pacific Basin (see Figure 2.6) at around 160° E and 57° S, and is steered southwards by vortex tube stretching and conservation of potential vorticity. It is not obvious, however, why topographic steering at this location becomes more pronounced later into the 40 year run. While the general southwards migration of the ACC may partly be responsible (thus leading to the current impinging on the Southwest Pacific Basin to a greater extent than the elevated Campbell Plateau), it may also be caused by an enhanced barotropic character to the flow. Recall from Figure 4.4 that the isopycnals tend to flatten out. This implies that the flow is more barotropic and thus more likely to interact with topography.

The unrealistic strong flow around the southeastern corner of Tasmania and on northwards along the east coast of Australia is consistent with the tracer sections seen earlier. At 160° E it was apparent that surface waters in the north of the domain were becoming cooler and fresher – the warm salty waters from the subtropics were not being replaced. The EAC, while persisting in the model after a fashion, appears only to be maintained by the barotropic flow imposed with the artificial boundary condition along the northern boundary. It is less effective by the end of the run at maintaining the surface watermass properties in the Tasman Sea.

An artificial southwards flow along the western boundary was found previously when a strong relaxation to climatology was applied to tracers at the western boundary (section 3.7). While it did not appear in sufficient strength to degrade results in the five year trial runs of open boundary conditions, it is stronger in this longer model integration, and likely to affect the overall circulation. It is not clear to what extent this may contribute to the general southwards migration of the ACC.

An enhanced eastward flow along the Antarctic continental margin is evident after 20 years. This is likely to be due to both the southwards migration of the ACC and a reduced westwards flow associated with bottom-water production.

4.1.4 Meridional heat flux

Introducing the conventional notation (Peixoto and Oort 1992)

$$[f](y,z,t) = \frac{1}{(x_2 - x_1)} \int_{x_1}^{x_2} f(x,y,z,t)\, dx$$

for a zonal average[1], then any field can be decomposed zonally into *zonal mean* and *standing eddy* components:

$$f(x,y,z,t) = [f](y,z,t) + f^*(x,y,z,t).$$

Similarly introducing the notation

$$\overline{f}(x,y,z) = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} f(x,y,z,t)\, dt$$

for a time average, then any field can also be decomposed into *time mean* and *transient* components:

$$f(x,y,z,t) = \overline{f}(x,y,z) + f'(x,y,z,t).$$

Decomposing both temperature and meridional velocity in this manner, we may write

$$\begin{aligned} T &= \overline{T} + T', \\ v &= \overline{v} + v'. \end{aligned}$$

Further decomposing the time-mean components zonally, we have

$$\begin{aligned} T(x,y,z,t) &= [\overline{T}](y,z) + \overline{T}^*(x,y,z) + T'(x,y,z,t), \\ v(x,y,z,t) &= [\overline{v}](y,z) + \overline{v}^*(x,y,z) + v'(x,y,z,t). \end{aligned}$$

The zonally integrated time-mean northwards heat flux at some latitude is given by

$$\mathcal{F}(y) = \int l(y,z)\rho_0 c_p [\overline{vT}]\, dz,$$

where c_p is the specific heat of seawater and $l(y,z)$ is the total oceanic path length at latitude y and depth z. Using the above decompositions, the integrand can be written as the sum of *mean meridional*, *standing eddy* and *transient eddy* components respectively:

$$[\overline{vT}] = [\overline{v}][\overline{T}] + [\overline{v}^*\overline{T}^*] + [\overline{v'T'}].$$

[1]When calculating a zonal average with model data discretised in vertical layers, the definition must be modified to account for the intersection of a layer by topography. The zonal average of a quantity in model layer k is given by

$$[f]_k(y,t) = \frac{1}{(x_2 - x_1)} \int_{x_1}^{x_2} f_k(x,y,t) h_k(x,y)\, dx,$$

where $h_k(x,y)$ is the thickness of the model grid at (x,y) in layer k as a fraction of the nominal layer thickness d_k.

Under this decomposition for the integrand, the heat flux is said also to be decomposed into these three components. The mean meridional and standing eddy components of meridional heat flux are analysed here, with the transient eddy heat flux examined in section 4.2.2.

Figure 4.9 shows both the mean meridional and standing eddy components of meridional heat flux averaged over each successive five-year period throughout the 40 year run. The expected total (mean meridional + standing eddy + transient eddy) oceanic meridional heat flux integrated around the globe was shown earlier (Figure 1.1) and is negative in the Southern Ocean ranging from around 1–2 PW for the latitude band of interest here. It is clear that the mean meridional heat flux in the model is large by comparison and of opposite sign. On the other hand, the standing eddy component is of the correct sign and would integrate to a reasonable value if extended around the globe. The high positive values of mean meridional heat flux obtained here are not too disturbing since the domain is not zonally periodic. Integrated zonally, there can be a net northwards flow if the ACC enters the domain at a higher latitude than it exits for instance. This is precisely the same reason an overturning streamfunction could not be defined earlier. Figure 4.5, showing sections of zonally-integrated meridional transport by layer, suggests that indeed this is the case. Figure 4.10 sums these meridional volume transports vertically to show the total northwards volume transport at each latitude, averaged again over successive five year periods. The large northwards volume transport produces, not surprisingly, a large mean northwards heat flux term $\mathcal{F}_{[\overline{v}][\overline{T}]}$. Of course in a model extending around the globe, there would be zero net northwards transport and this term would be much smaller.

4.1.5 Sea-level

As will be discussed further in the next chapter, assimilation of altimetry data must use an *a-priori* mean sea-level against which the anomalies observed by the satellite are referenced. It is important, therefore, to examine how the modelled sea-level varies with time. It is possible, for example, that assimilation using a fixed reference sea-level may assist with reducing model drift. Figure 4.11 shows sea-level averaged over five-year periods throughout the 40 year run. It is clear that the model drift seen with various other diagnostics above also has a signature in sea-level. The steady drift southwards of the ACC is seen, together with the northwards advection of subpolar waters into the Tasman Sea. The diffusion of circumpolar frontal structure and slumping of isopycnals gives rise to a sea-level that becomes flatter with time.

The Flather open boundary condition for the barotropic system appears to be performing adequately at conserving volume in the domain. Figure 4.12 shows that the mean sea-level changes by around 10 cm over the 40 years of the run. An examination of the differences between the ODD and EVEN subgrids shows that the mean sea-level differs by less than 3 mm throughout the integration (and by less than 1 mm for the final 30 years).

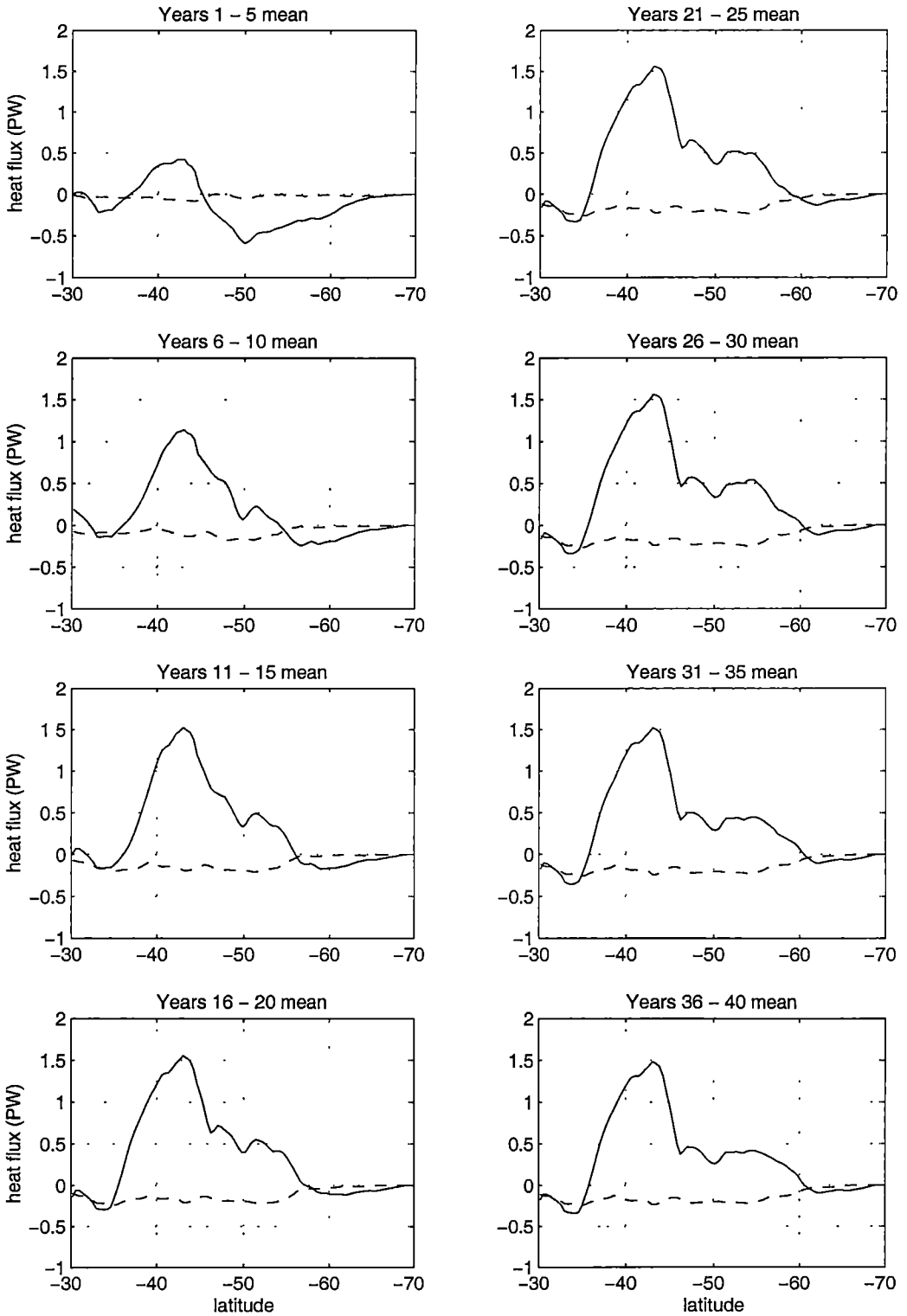


Figure 4.9: Mean meridional, $\mathcal{F}_{[\overline{v}][\overline{T}]}$, (solid line) and standing eddy, $\mathcal{F}_{[\overline{v}^*\overline{T}^*]}$, (dashed line) components of meridional heat flux (PW) by latitude, averaged over successive five-year intervals in the 40 year run.

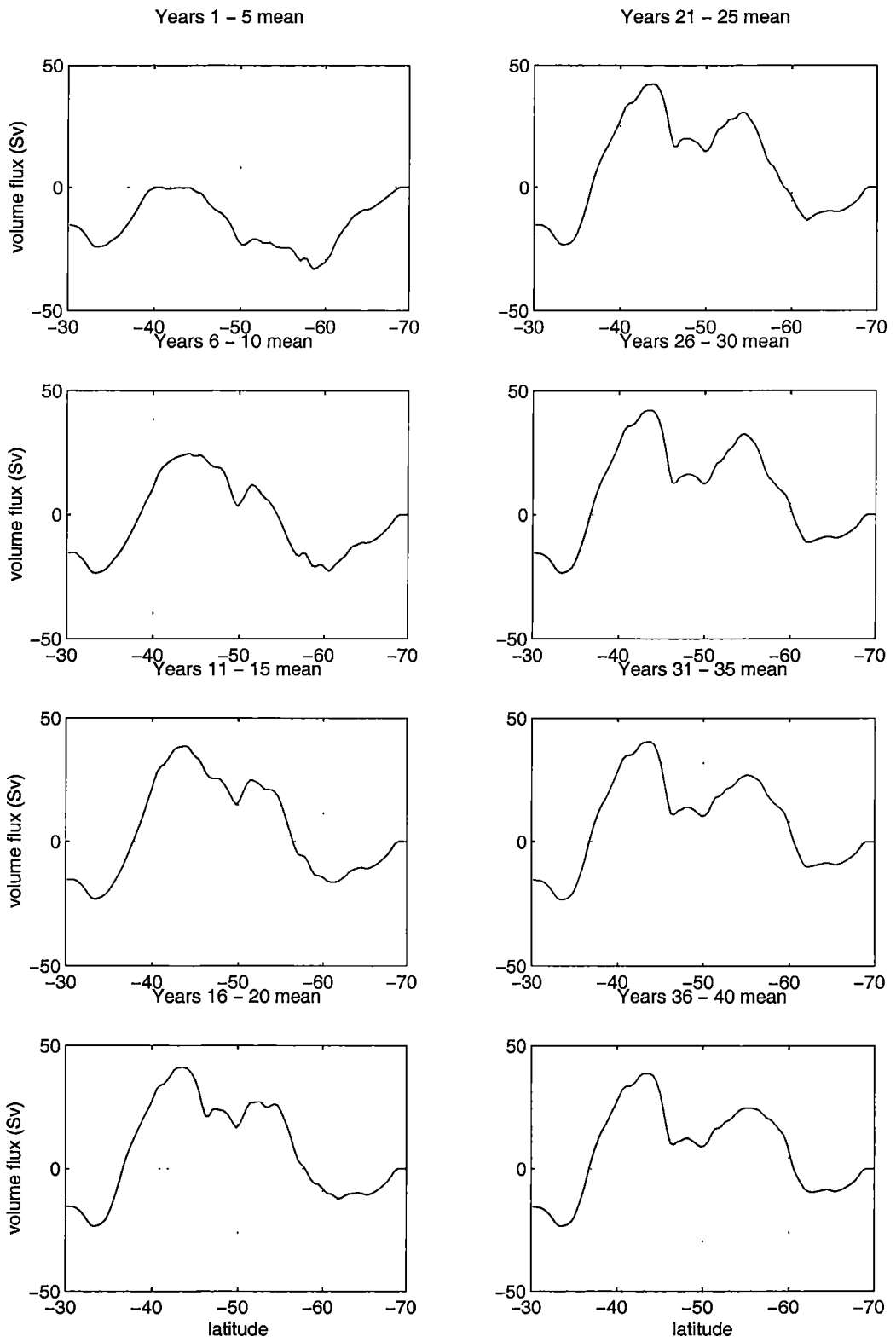


Figure 4.10: Total northwards volume transport by latitude averaged over successive five year periods.

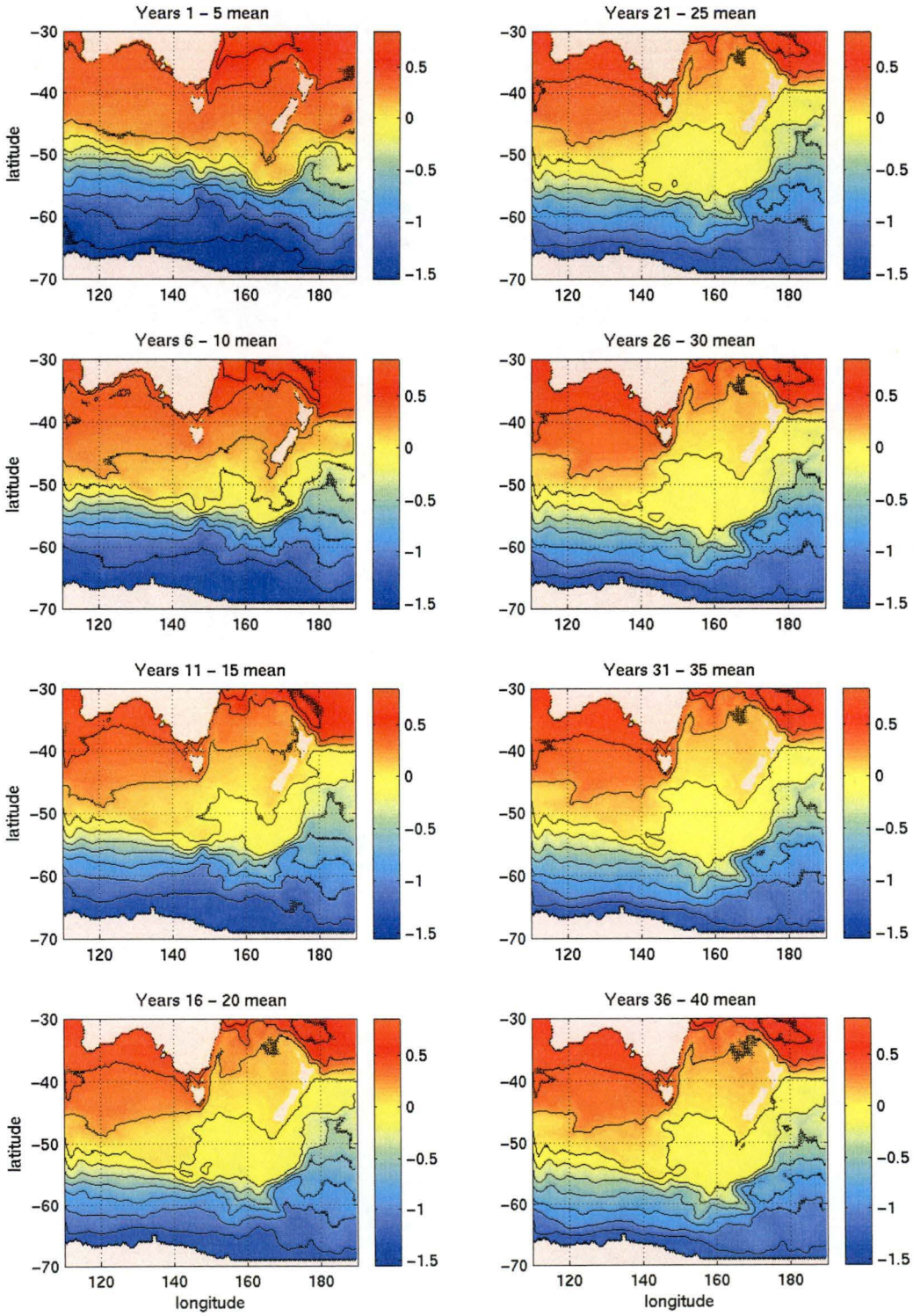


Figure 4.11: Sea-level (metres) averaged over five-year intervals throughout the 40 year run. Black contours are at intervals of 0.2 m.

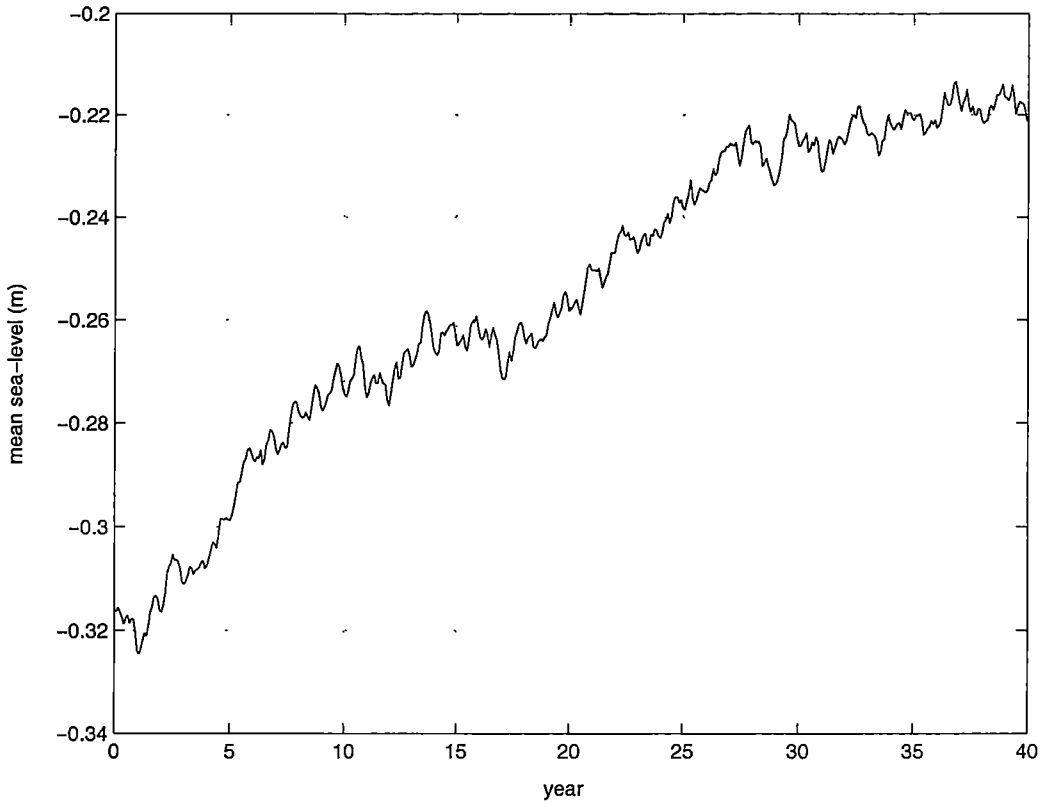


Figure 4.12: Mean sea-level (EVEN) during 40 year run.

4.2 Model variability

The analysis above demonstrates a steady drift of the model away from a state that realistically reproduces the main features of the ocean circulation south of Australia. On the other hand, altimetry assimilation in this thesis is aimed primarily at constraining the evolution of the model on time and space scales characteristic of mesoscale phenomena. The time scales are typically of the order of days to weeks. Assimilation might be expected to perform best when the model naturally displays variability similar to that seen in the real ocean. It is important, therefore, to examine variability in the model. If the variability is unrealistic, it is a crucial test of the assimilation scheme to see whether the variability patterns are thereby improved. In addition, altimetry provides only surface (sea-level) information representing an integrated picture of the subsurface dynamics. Assimilation is unlikely to perform well unless this surface information is used to constrain the model evolution at depth. It is necessary to devise a method for doing this. Analysing both subsurface and sea-level variability in the model provides indications of how this should be performed. It should be noted, of course, that the model itself is in error, due to incorrect forcing, inadequate resolution, etc.. Results from such analyses should not be used quantitatively to design subsurface assimilation schemes. Rather, results that are in qualitative agreement with other sources may suggest suitable approaches. In particular, it will be seen that, consistent with previous studies, the model appears to contain a significant barotropic component to sea-level

variability. This section examines variability in the 40 year run with emphasis on timescales shorter than the slow drift described earlier.

4.2.1 Sea-level

Figure 4.11 above shows how the mean sea-level gradually evolves over the duration of the 40 year run. The standard deviation of sea-level within each of the five year periods is shown in Figure 4.13. It is clear that the patterns of *sea-level variability* in the model also evolve over time, at least for the first 25 years. The most obvious characteristic is a very marked drop in the level of variability. To examine in greater detail the timescales of this variability, the 40 year timeseries of sea-level at each gridpoint was Fourier transformed and the spectral energy separated into three classes: variability on the annual cycle, and both shorter and longer than one year. Figure 4.14 shows the sea-level standard deviation over the entire 40 year period, as well as the variance fractions in these three classes. The sea-level variability in most regions of the domain varies on timescales greater than one year. This is largely associated with the considerable drift of the mean state seen earlier. In addition there is a very strong annual cycle in the sea-level in the Great Australian Bight, particularly along the south coast of Australia. This is most likely due to the strong seasonality in the local winds, which are southerly in summer and westerly in winter (Hellerman and Rosenstein 1983). There are three main locations where sea-level variability is dominated by timescales shorter than one year: around the Campbell Plateau south of New Zealand, across the northeastern flank of New Zealand, and south of Tasmania. These are also regions of strong mean flow for most of the run (see Figure 4.7). The standard deviation of sea-level in the last 15 years of the run shows a similar pattern (Figure 4.13). The picture that emerges is consistent with that formed earlier; there is significant drift in the model initially, sufficient to dominate the total sea-level variation over the 40 years of the run. As the model gradually tends towards an equilibrium mean state, the variability is to a greater extent mesoscale in nature. By the final five years of the run, the strongest sea-level variability occurs as either the annual cycle (in the Great Australian Bight) or at timescales less than one year.

It is necessary to compare this 'equilibrium' variability with that in the real ocean. Figure 4.15 shows the sea-level variability observed by the TOPEX/POSEIDON satellite altimeter. For convenience an interpolated picture is shown, as well as the along-track data. Not surprisingly, the pattern of variability is very different to that obtained with the model after drift diminishes. Variability in the real ocean is associated with the EAC, and with topographic features along the path of the ACC, notably at the Australian Antarctic Discordance, around the southern flank of the Campbell Plateau, and at 150° E where the ACC is steered south by crossing the Southeast Indian Ridge[2]. There is also a considerable difference in the level of variability.

[2]The strong variability in Bass Strait is most likely an inaccuracy due to problems in processing the raw altimeter data over shallow coastal regions, rather than variability of oceanic interest. For instance, different global tidal models are known to have very different

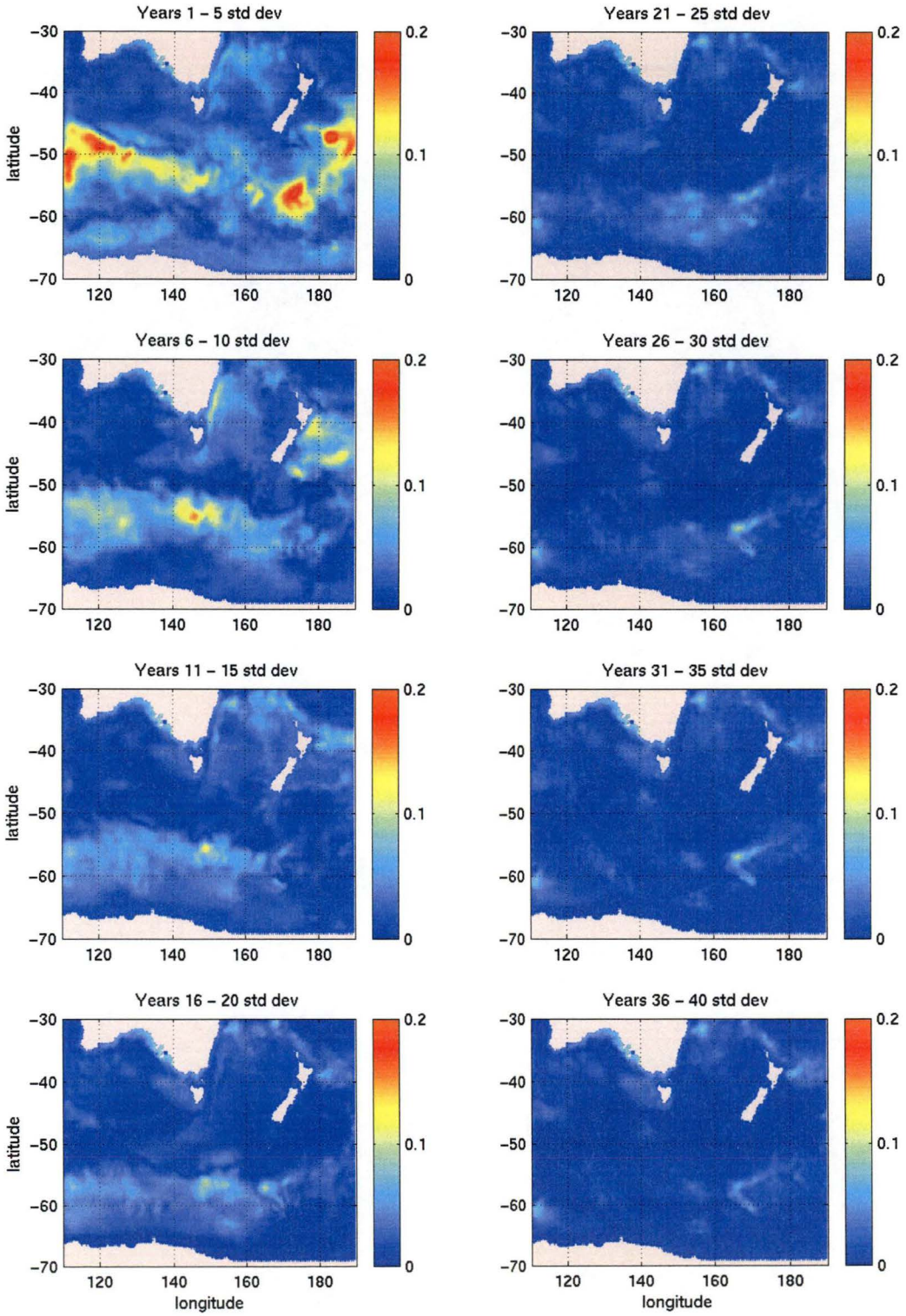


Figure 4.13: Standard deviation of sea-level (metres) within each five year period of the 40 year run.

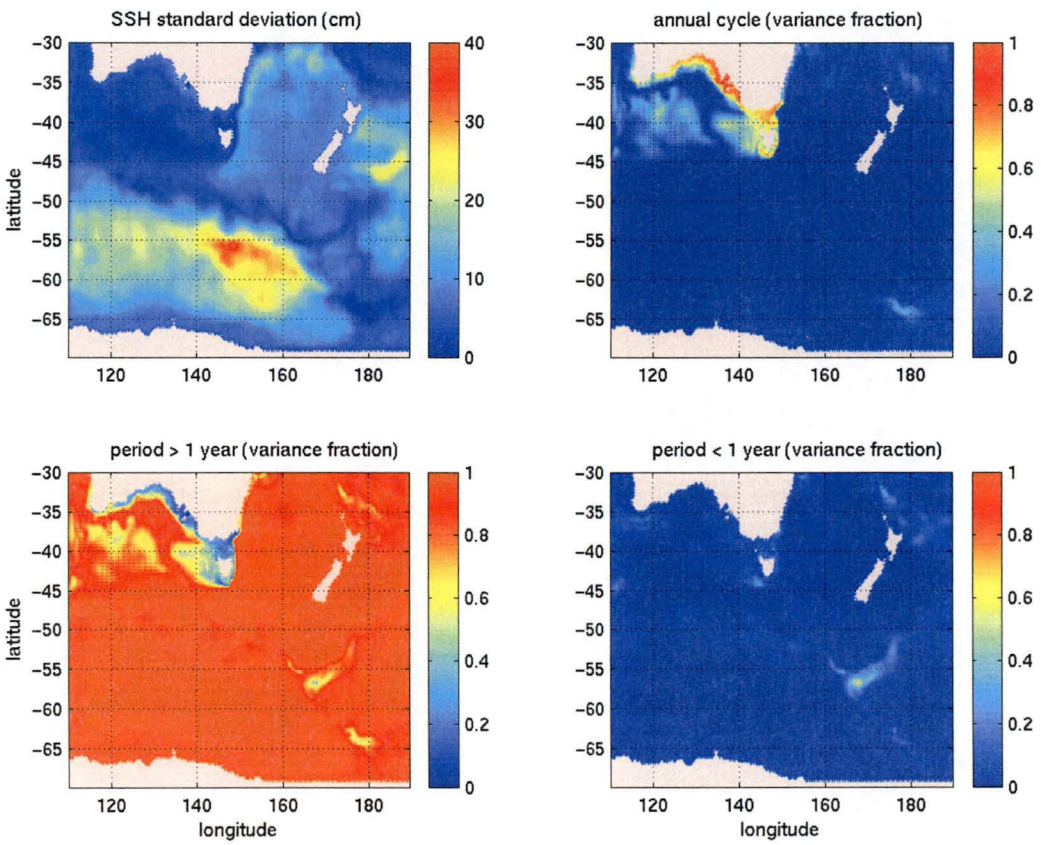


Figure 4.14: Standard deviation of sea-level for entire 40 year run (top left), and fractions of variance at annual periods (top right), periods greater than one year (lower left), and periods less than one year (lower right).

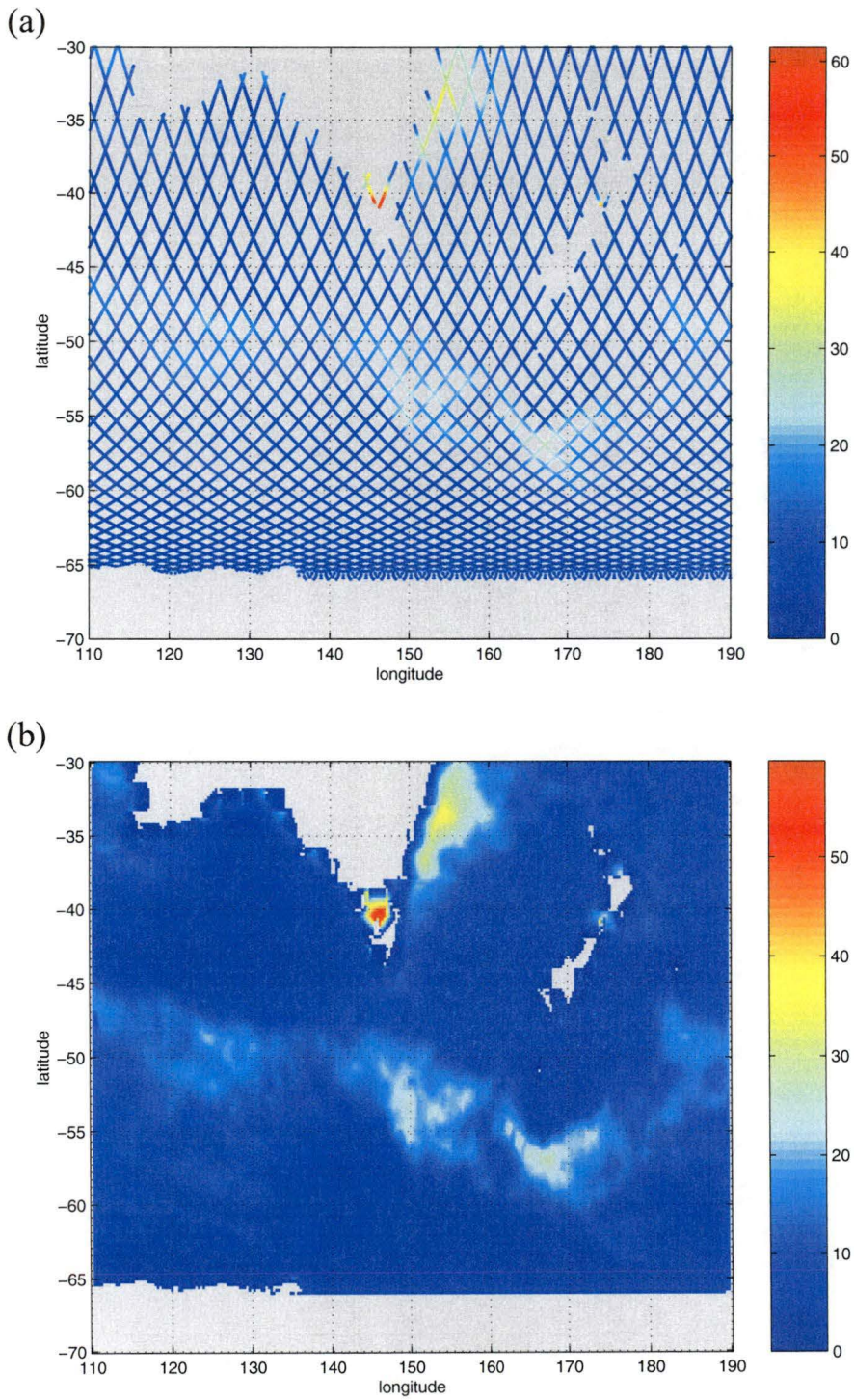


Figure 4.15: Standard deviation of sea-level (cm) observed by the TOPEX/POSEIDON satellite altimeter. (a) Along-track data, (b) interpolated onto model grid.

The three patches of strong variability along the ACC path all have standard deviations of over 20 cm. By comparison, the strongest mesoscale variability obtained with the model in the final 15 years (south of New Zealand, Figure 4.13) has a standard deviation only around 10 cm.

4.2.2 Meridional heat flux

Section 4.1.4 diagnosed both mean meridional and standing eddy components of heat flux over each five year period within the 40 year run. The transient eddy heat flux was defined earlier as

$$\mathcal{F}_{[\overline{v'T'}]}(y) = \int l(y,z)\rho_0 c_p[\overline{v'T'}]\,dz,$$

where c_p is the specific heat of seawater, $l(y,z)$ is the path length at latitude y and depth z, and v' and T' are the transient components of northwards velocity and temperature respectively:

$$\begin{aligned} v'(x,y,z,t) &= v(x,y,z,t) - \overline{v}(x,y,z) \\ T'(x,y,z,t) &= T(x,y,z,t) - \overline{T}(x,y,z). \end{aligned}$$

As before, transient eddy heat flux was diagnosed from the 40 year model run within each five year period. The results are shown in Figure 4.16. Section 1.2 introduced the importance of eddy heat flux to the overall energy budget in the global climate system. From a range of considerations, it is believed that eddy heat flux is the dominant contributor carrying heat poleward at unbounded latitudes in the Southern Ocean (see section 1.2). Estimates range from 0.3 to 0.45 PW. If this were distributed uniformly around the globe, then a total of between 0.07 and 0.1 PW poleward is expected for the longitude band of interest here. Figure 4.16 shows that indeed the transient eddy heat flux is poleward almost everywhere. Very small northwards heat fluxes at some latitudes are negligible by comparison (and not visually evident in the figure). In addition, the values are approximately as expected. While somewhat low perhaps, there are several assumptions behind, and consequently large uncertainty in, the figures quoted above. By comparing Figure 4.16 with Figure 4.13 for sea-level variability, it is clear that latitudes of high transient eddy heat flux correspond to those with significant sea-level variability. In particular, from the discussion in section 4.2.1 this implies that the transient eddy heat fluxes in the final 15 years of the run at least, are likely associated with mesoscale variability and not model drift. Though, as discussed in section 4.2.1, the regions of high variability in the model are different from those observed by satellite.

4.2.3 Subsurface variability

The introduction to this chapter noted one of the reasons for analysing a free-run of the model is to inform the development of the altimetry assimilation

solutions in this region (Lefevre et al. 2000b; Lefevre et al. 2000a). As well, there are known to be radiometer problems that affect the path-length corrections that must be applied to the raw data (White et ál. 1999).

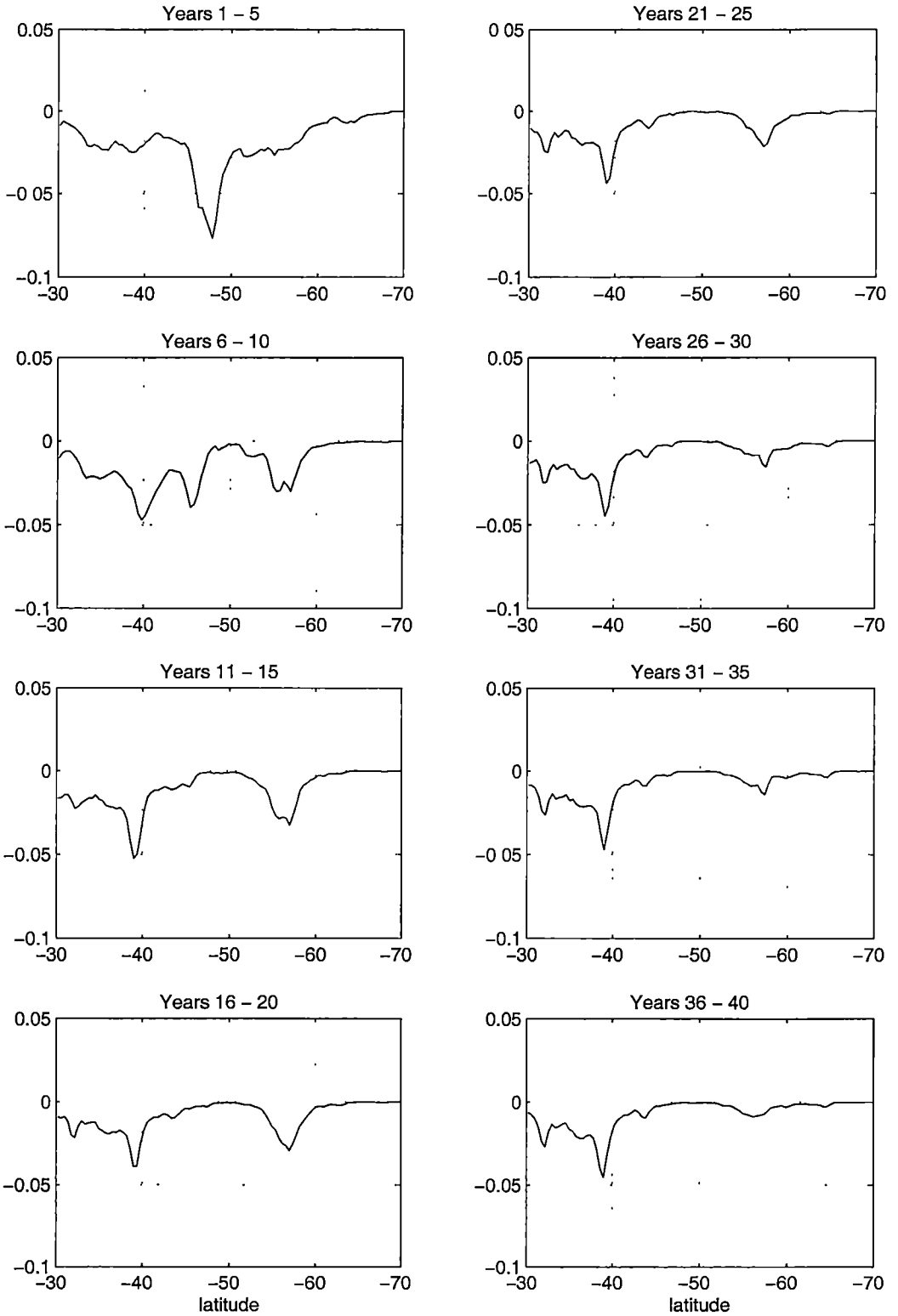


Figure 4.16: Meridional transient eddy heat flux (PW) diagnosed for each five year period of the 40 year run.

scheme developed in the next chapter. Assimilation of altimetric sea-level anomalies requires adjustment of more than just the model sea-level: from geostrophic adjustment theory, it is well known (Gill 1982) that any such perturbations will radiate away with timescales of surface gravity waves, unless on a spatial scale larger than the barotropic Rossby radius (typically several thousand kilometres). In order to impose the altimetric signal on the model, some type of subsurface adjustment is required. This is discussed in more detail in the next chapter, but it is informative here to investigate briefly the relationship between variability of sea-level, and subsurface dynamics, in the 40-year model run.

Several authors have noted the highly barotropic nature of both the time-mean circumpolar current and its variability. Whitworth and Peterson (1985) used bottom pressure gauges in Drake Passage to measure the volume transport through the passage. They found that fluctuations in transport were mostly barotropic, and that these could be as large as 50% of the mean. Pinardi et al. (1995) showed that in a rigid-lid model formulation, the diagnosed surface pressure field (corresponding to sea-level), p_s, could be decomposed into a baroclinic and barotropic component:

$$\nabla.(H\nabla p_s) = -\nabla.\int_{-H}^{0}(z+H)\nabla\rho\,dz + \nabla.(f\nabla\Psi), \tag{4.1}$$

where ρ is density, Ψ is volume transport streamfunction, and H is ocean depth. Under an $f-$plane, flat-bottom approximation this reduces to a corresponding local decomposition directly of sea-level:

$$\eta = \frac{f}{g}\frac{\Psi}{H} - \frac{1}{\rho_0 H}\int_{-H}^{0}(z+H)\rho\,dz. \tag{4.2}$$

Applying the decomposition (4.1) to a run of the GFDL model, they found that the baroclinic component of sea-level dominated everywhere except in the Southern ocean, where the barotropic component represented over 50% of the total. Maes et al. (1999) used the same decomposition to analyse a run of the LODYC model. They found barotropic variability dominated baroclinic in the Southern Ocean. Analysing output from the FRAM model, Killworth (1992) found what he termed an "equivalent barotropic" flow — the horizontal flow throughout a water column was, to a very good approximation, in the same direction, but reduced with depth. Not only the mean, but also the eddy kinetic energy was found to have an equivalent barotropic structure. In a global high resolution assimilation study, Fox et al. (2000a; 2000b) used the assimilation technique of Cooper and Haines (1996), which leaves bottom pressure unmodified. They obtained improved forecasting ability everywhere except the Antarctic Circumpolar Current, due to the significant barotropic variability not being captured by the assimilation scheme.

As discussed in the next chapter, it is possible to make both baroclinic and barotropic changes to the model in order to assimilate sea-level measurements; this is suggested by Pinardi et al.'s decomposition (4.1). The above studies suggest that it may be appropriate to make barotropic changes in the Southern Ocean with greater validity than in other areas of the world ocean.

A simple diagnostic is used here to analyse the relationship between sea-level variability and both baroclinic and barotropic subsurface variability in the 40-year run. If the HOPE model used a rigid-lid approximation, then the diagnostic (4.1) of Pinardi et al. could be used to perform such an analysis. Since it has instead a free surface, an alternative framework is needed. In the geostrophic limit, velocity at the seafloor is determined by horizontal gradient of bottom pressure, p_b. Under the hydrostatic assumption, bottom pressure is simply the sum of surface pressure and internal (baroclinic) pressure:

$$p_b = \rho_0 g\eta + \int_{-H}^{0} \rho g \, dz. \tag{4.3}$$

Rearranging this allows sea-level to be written as the sum of a bottom-pressure ("barotropic") and internal ("baroclinic") pressure term:

$$\eta = \frac{p_b}{\rho_0 g} - \frac{1}{\rho_0} \int_{-H}^{0} \rho \, dz. \tag{4.4}$$

The baroclinic term is equivalent to the conventional bottom-referenced *dynamic height* of the water column (Gill 1982). If the diagnostic is applied to anomalies η', p_b' and $\rho'(z)$, then the baroclinic term is simply the steric effect of a density anomaly in the water column. Writing

$$\eta'_{BT} = \frac{p_b'}{\rho_0 g} \tag{4.5}$$

and

$$\eta'_{BC} = -\frac{1}{\rho_0} \int_{-H}^{0} \rho' \, dz \tag{4.6}$$

respectively for the barotropic and baroclinic components of sea-level anomaly then we have

$$\eta' = \eta'_{BT} + \eta'_{BC}. \tag{4.7}$$

These quantities were analysed through successive five year intervals of the 40-year model run. The anomaly quantities were calculated with respect to the mean for the corresponding five year period. The barotropic (bottom-pressure) term was diagnosed from (4.3), guaranteeing (4.7) to hold exactly at each time.

A statistical check for a systematic relationship was performed by calculating correlation coefficients

$$r_{BT} = \frac{\overline{(\eta'\eta'_{BT})}}{\sigma_{\eta'}\sigma_{\eta'_{BT}}},$$

$$r_{BC} = \frac{\overline{(\eta'\eta'_{BC})}}{\sigma_{\eta'}\sigma_{\eta'_{BC}}},$$

in each of the five-year periods (where $\sigma_{\eta'}$, $\sigma_{\eta'_{BT}}$ and $\sigma_{\eta'_{BC}}$ are standard deviations of sea-level anomaly, and the barotropic and baroclinic components, respectively). With monthly snapshots used for the analysis, each five-year time-series has around 60 samples and correlation coefficients greater in magnitude

than around 0.33 are significant at the 99% confidence level ($p < 0.01$). The correlation coefficients are shown in Figures 4.17 and 4.18 for the barotropic and baroclinic components respectively. Correlations that are not significant at $p < 0.01$ are not shown. The first point to note is that correlations are nearly everywhere significant in each of the five year periods — to a high degree the barotropic and particularly the baroclinic components of sea-level anomaly, defined by (4.5) and (4.6), are linearly related to the total anomaly. Second, there is considerable variation in the regional correlations over the 40 year run. This is associated with the large-scale drift that occurs over the first few decades of the run. By the last 15 years, the correlations in Figures 4.17 and 4.18 appear much more similar in successive five-year periods.

The interpretation of the regions of negative correlation in Figure 4.17 between barotropic (bottom-pressure) component of sea-level anomaly and the total is that there is a systematic deficit between steric anomalies and actual sea-level change: bottom-pressure compensates with the opposite sign. This seems to occur in two distinct regimes during the 40 year run. The first is associated with the dominant mode of variability during the initial 20 years, discussed at length in earlier sections and characterised by large-scale drift of the model with associated watermass changes. This was particularly evident in the Tasman Sea due to the diminishing of the EAC and the migration of sub-polar waters northwards around the southeastern corner of Tasmania. Large steric sea-level effects would be expected but Figure 4.17 demonstrates that this does not occur. After the model reaches a quasi-steady state after the first 20 years of the run, significant negative correlations of bottom-pressure with total sea-level anomaly continue to occur east of Australia. This is precisely where the annual cycle of climatological sea surface temperature is strongest (Levitus and Boyer 1994). This should produce a significant steric sea-level effect. (Gill and Niiler (1973) demonstrate that the steric component of the seasonal sea-level signal is more significant than that due to Ekman pumping.) However, the actual sea-level anomaly is again smaller than the steric component. Nerem et al. (1994) compared the amplitude of the annual sea-level signal in the Levitus climatology (relative to 1000 m) with that observed by the TOPEX/POSEIDON altimeter. Over much of the ocean, the annual cycle observed by altimeter is larger than in the hydrography. However, in the Tasman Sea, the seasonal altimeter signal (up to 6 cm) is smaller than that calculated from the climatology (up to 8 cm). The analysis here is consistent with those results.

Performing a linear regression of barotropic (bottom-pressure) component of sea-level anomaly against the total,

$$\eta'_{BT} = a\eta',$$

then the regression coefficient a is given by

$$a = \frac{\overline{(\eta'\eta'_{BT})}}{\sigma^2_{\eta'}} = r_{BT}\frac{\sigma_{\eta'_{BT}}}{\sigma_{\eta'}}.$$

By construction, (4.7), we also have the regression of baroclinic component

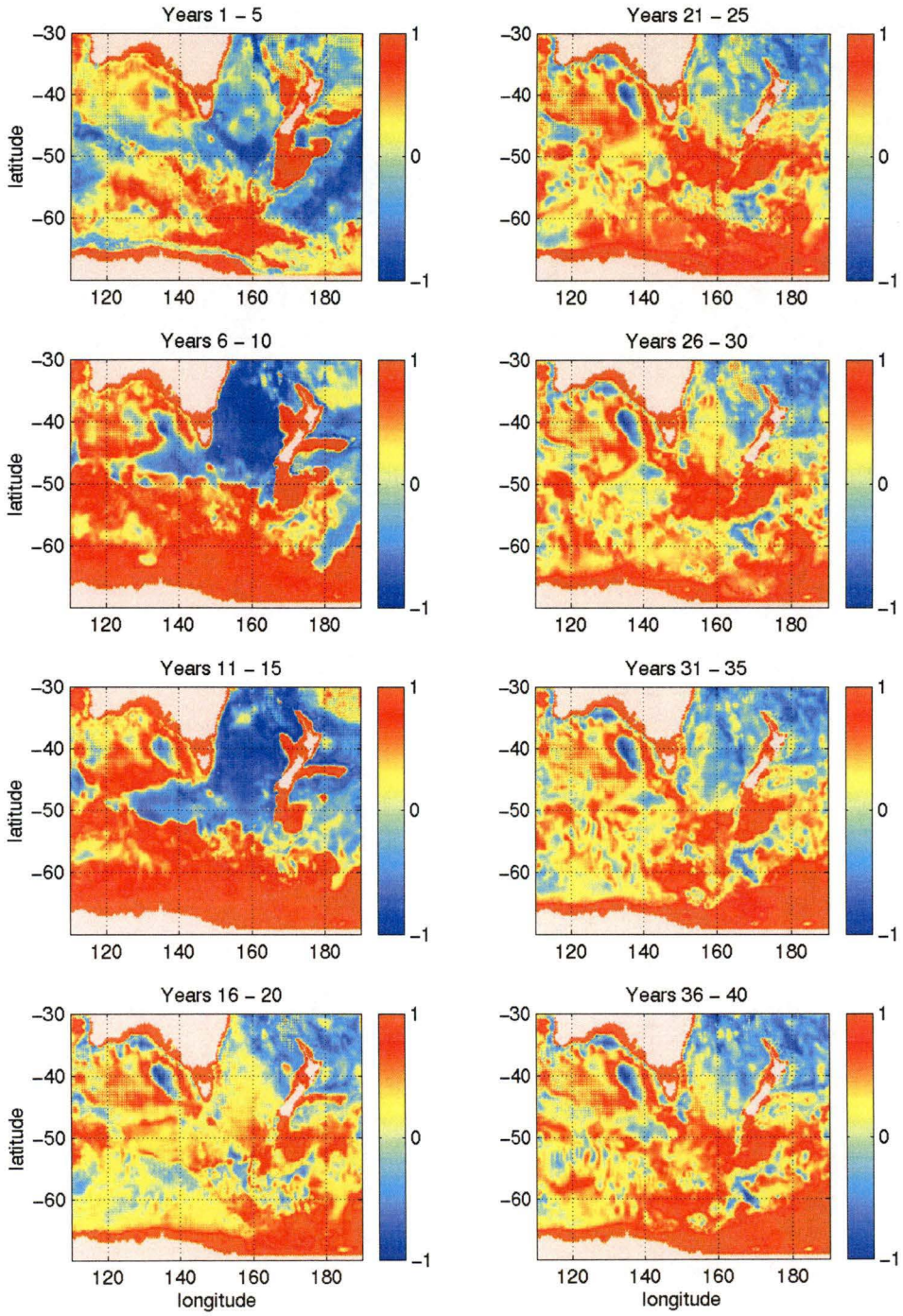


Figure 4.17: Correlation of barotropic component (η'_{BT}) against total sea-level anomaly (η') for each five year period of the 40 year model run.

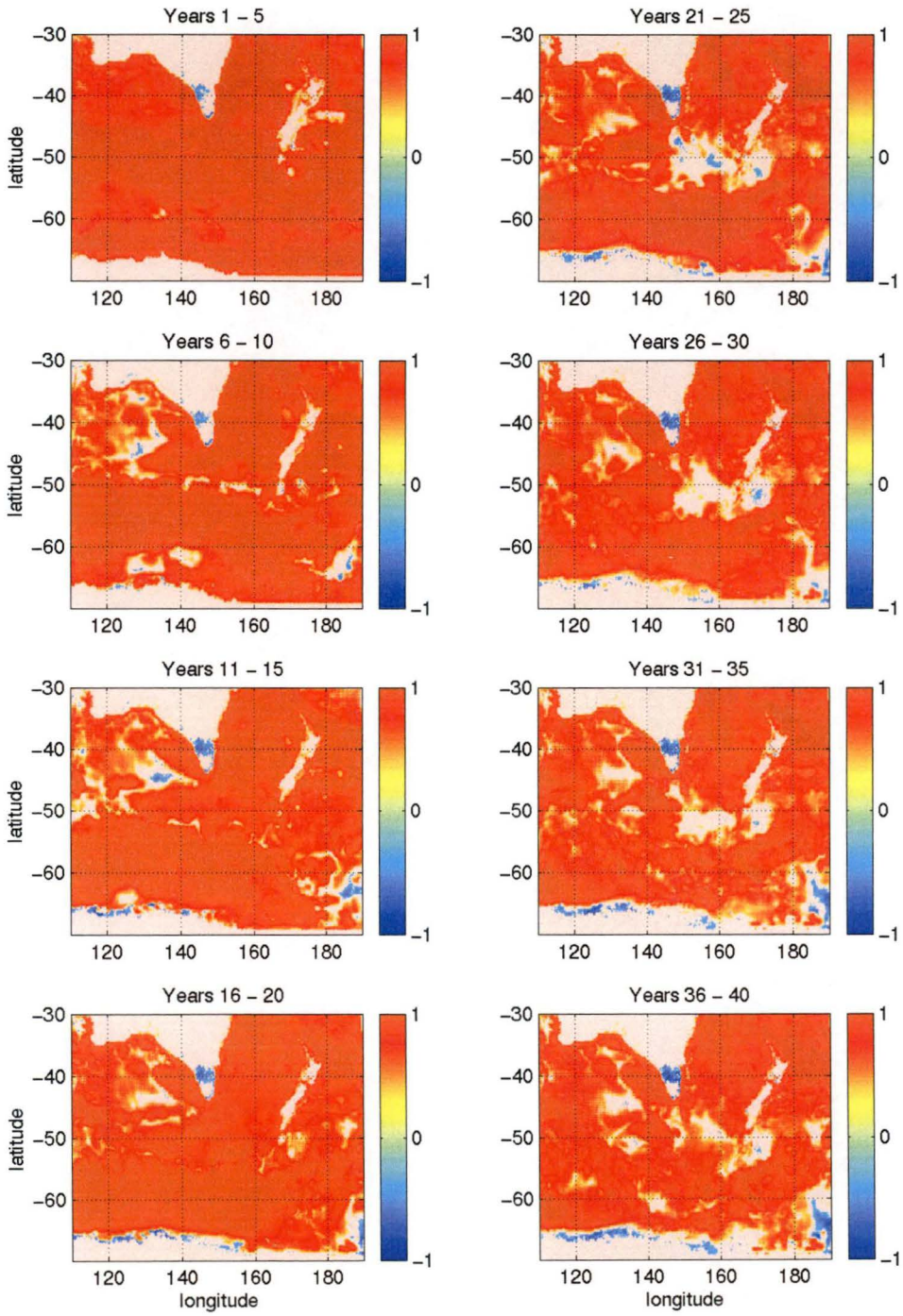


Figure 4.18: Correlation of baroclinic component (η'_{BC}) against total sea-level anomaly (η') for each five year period of the 40 year model run.

against the total

$$\eta'_{BC} = \frac{(\overline{\eta'\eta'_{BC}})}{\sigma^2_{\eta'}}\eta' = (1-a)\eta'.$$

The regression coefficients, a and $(1-a)$, are shown in Figures 4.19 and 4.20 where the respective correlation coefficients are significant at the 99% confidence level ($p < 0.01$). As expected, the steric sea-level anomaly is greater than the actual sea-level anomaly (regression coefficient greater than one in Figure 4.20) in the regions of strong negative correlation of η' and bottom-pressure (Figure 4.17).

The variability timescales were analysed earlier and decomposed in Figure 4.14 into an annual cycle, and periods both longer and shorter than annual. An annual signal of sea-level variability was strongest along the shelf of the Great Australian Bight, and in Bass Strait. From Figure 4.19 it is clear that this variability is almost completely barotropic. While this is the strongest barotropic variability in the model, it is worth noting that there are other regions of significant barotropic variability. The initial drift in the model limits the ability to ascribe to the real ocean these other regions of high barotropic variability, but it is clear that here, as in other studies of Southern Ocean variability, there is a significant barotropic component, although the variability is overwhelmingly baroclinic in nature. The implication for the assimilation scheme developed in the next chapter is that the scheme should attempt to introduce a degree of barotropic variability in the model.

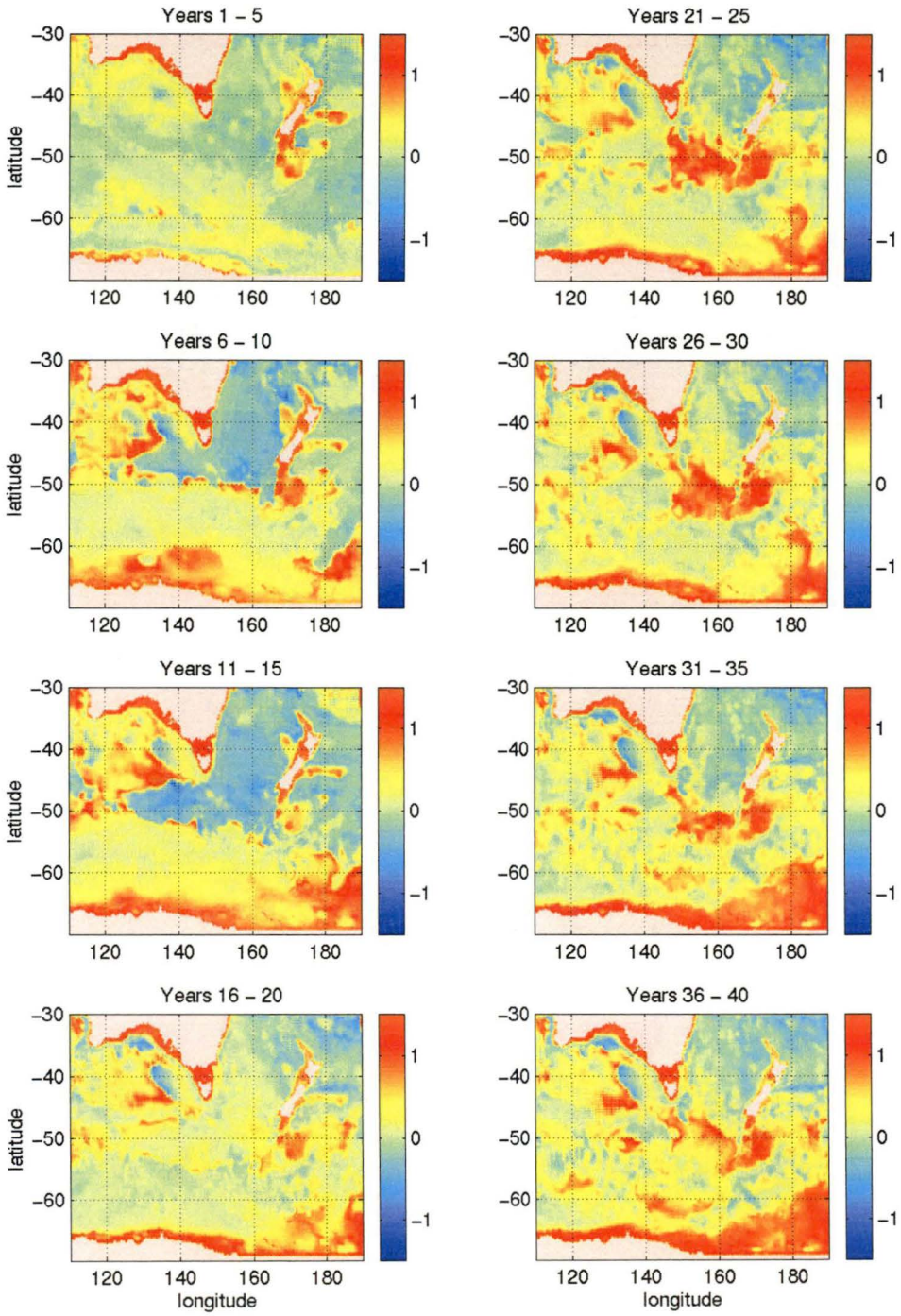


Figure 4.19: Linear regression coefficient, a, of barotropic component (η'_{BT}) of sea-level anomaly against total (η').

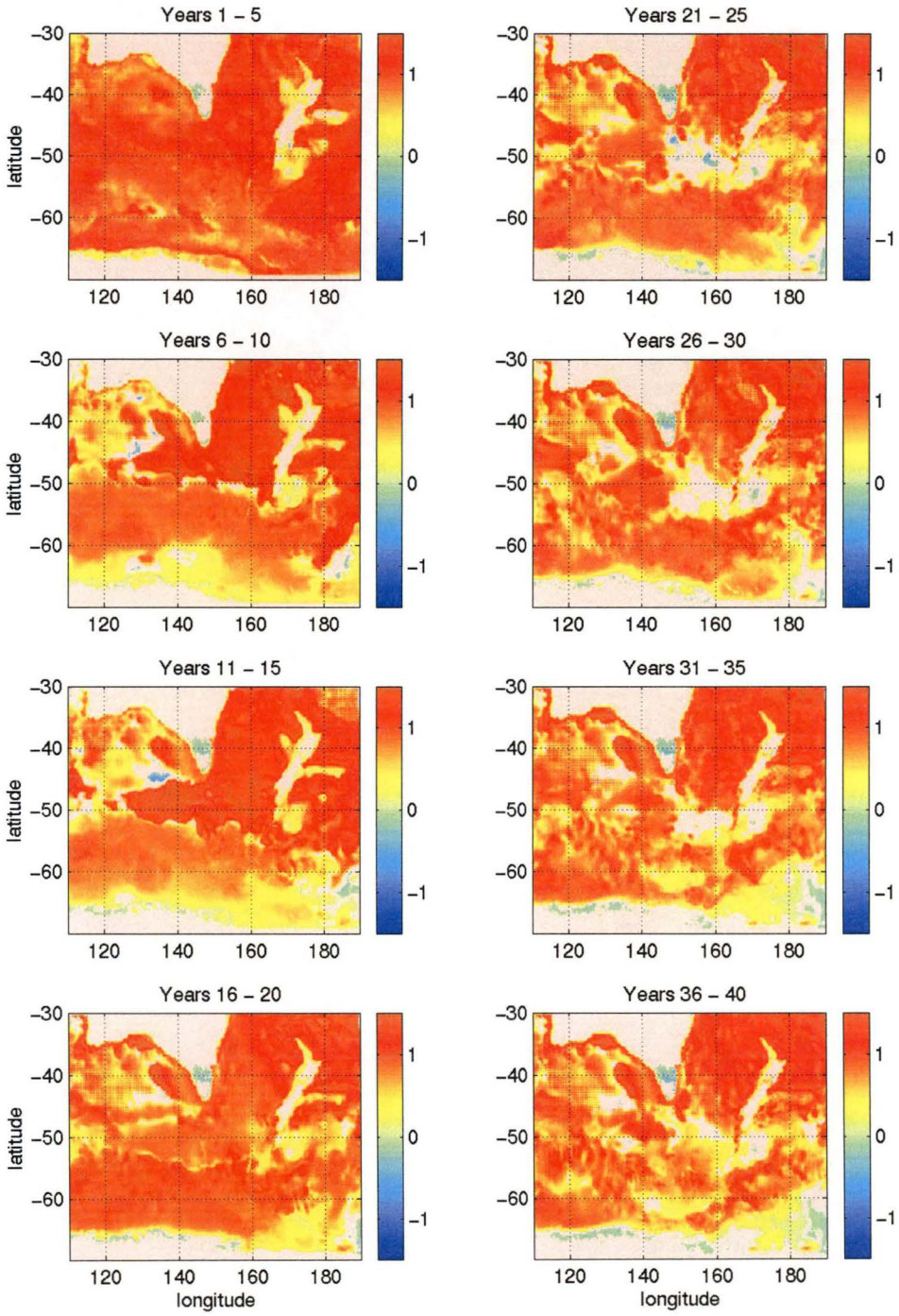


Figure 4.20: Linear regression coefficient, $1-a$, of baroclinic component (η'_{BC}) of sea-level anomaly against total (η'). Grey areas mark locations where the correlation is not significant at the 99% level.

CHAPTER 5

Altimetry assimilation

Section 1.6 introduced data assimilation as the attempt to combine synoptic data with general circulation models to obtain dynamical estimates of the ocean state. A range of approaches to the problem were briefly described and various data sources that have been used for assimilation were mentioned.

The paucity of in-situ observational data in the Southern Ocean was mentioned in section 1.3. Assimilation of satellite altimetry data therefore provides the opportunity to significantly enhance our understanding of Southern Ocean dynamics, and the important role of eddy variability in this region.

This project was intended originally to undertake a thorough investigation of altimetry assimilation in a limited-area Southern Ocean model, with the resulting analyses used to examine in detail the dynamical role of eddies in the region south of Australia. Due to the considerable unforeseen difficulties with implementing open boundaries in the HOPE model, and the time constraints on a project of this type, it was not possible to pursue an assimilation research program as thoroughly as planned. Nevertheless, this chapter discusses preliminary results from assimilation of altimetry data in the limited-area model developed in the previous chapter. Estimates of eddy heatflux and sea-level variability will be seen to improve with assimilation, but a comparison against a WOCE hydrographic section is less satisfactory. There is some evidence that forecast ability is slightly improved over persistence.

The assimilation algorithm is described in the next section, including a discussion of the method for projecting estimated sea-level into the model interior. A brief description of the altimetry data is followed by a summary of assimilation results.

5.1 Reduced order optimal interpolation

Section 1.6.3 discussed sequential data assimilation techniques, including *optimal interpolation* (OI). That discussion is reviewed here for convenience. OI and Kalman filtering is discussed in detail by Gelb (1974), Daley (1991), Fukumori and Malanotte-Rizzoli (1995), De Mey (1997), Bouttier and Courtier (1992) etc. We represent the model as a discrete-time linear system:

$$\boldsymbol{w}_{k+1} = \boldsymbol{M}_k \boldsymbol{w}_k, \tag{5.1}$$

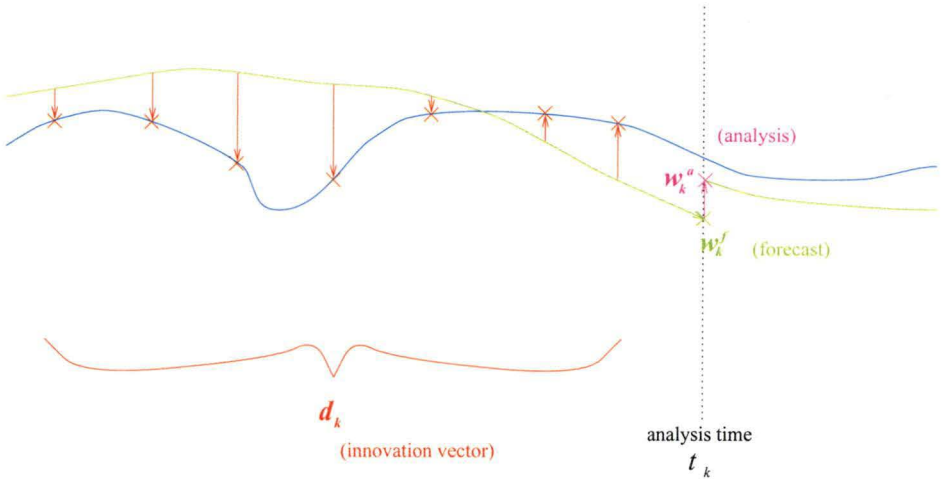


Figure 5.1: Schematic of optimal interpolation, showing analysis, $\boldsymbol{w}_k^a$, derived at analysis time from model forecast, $\boldsymbol{w}_k^f$, and innovation vector, $\boldsymbol{d}_k$. The evolution of the true ocean is represented by the blue trajectory, while the model is shown in green.

where the model state vector $\boldsymbol{w}$ is evolved from time t_k to t_{k+1} by the *state transition matrix* $\boldsymbol{M}_k$. Optimal interpolation forms an estimate, $\boldsymbol{w}_k^a$, of the true ocean state at *analysis times* by correcting a *model forecast*, $\boldsymbol{w}_k^f$, according to the *misfits*, $\boldsymbol{d}_k$, between the model and observations:

$$\boldsymbol{w}_k^a = \boldsymbol{w}_k^f + \boldsymbol{K}_k \boldsymbol{d}_k. \tag{5.2}$$

The vector of misfits between the model and the observations is also referred to as the *innovation vector*, and may be written

$$\boldsymbol{d}_k = \boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}, \tag{5.3}$$

where $\boldsymbol{y}^o$ is a vector of observations, and $\boldsymbol{H}$ is the *observation matrix*, which generates model equivalents to the observations from the model state $\boldsymbol{w}$. It is the data–model misfits that are used in the analysis step, so that, while these may be taken at different times, they are collated into a single vector at analysis time, and explicit time-dependence is omitted on the right-hand side of (5.3). The process is illustrated schematically in Figure 5.1. The *gain matrix*, $\boldsymbol{K}_k$, takes an equivalent form to the optimal *Kalman filter* gain:

$$\boldsymbol{K}_k = \boldsymbol{C}_k^f \boldsymbol{H}^{\mathrm{T}} (\boldsymbol{H}\boldsymbol{C}_k^f \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}^o)^{-1}, \tag{5.4}$$

where $\boldsymbol{C}_k^f$ and $\boldsymbol{C}^o$ are the error covariances (with respect to the true ocean state, $\boldsymbol{w}^t$) of the model forecast and observations respectively,

$$\begin{aligned} \boldsymbol{C}_k^f &= E[(\boldsymbol{w}_k^f - \boldsymbol{w}^t)(\boldsymbol{w}_k^f - \boldsymbol{w}^t)^{\mathrm{T}}] \\ \boldsymbol{C}^o &= E[(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}^t)(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{w}^t)^{\mathrm{T}}]. \end{aligned}$$

The term $\boldsymbol{H}\boldsymbol{C}^f\boldsymbol{H}^{\mathrm{T}}$ in the Kalman gain represents the forecast error covariance calculated between pairs of observation locations. Unlike the Kalman filter,

which optimally evolves the forecast error covariance, OI makes the approximation

$$\boldsymbol{C}_k^f = (\boldsymbol{D}^f)^{\frac{1}{2}} \boldsymbol{C} (\boldsymbol{D}^f)^{\frac{1}{2}},$$

where $\boldsymbol{D}^f$ is a diagonal matrix of *forecast error variances* and $\boldsymbol{C}$ is an *a-priori* correlation matrix with fixed structure. The formalism provides an estimate of the *analysis error covariance*,

$$\boldsymbol{C}_k^a = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{C}_k^f. \tag{5.5}$$

Daley (1991) discusses at length the importance of the forecast error covariance, $\boldsymbol{C}^f$, in optimal interpolation. He states that it is the most important element of the procedure, with its structure governing, to a large degree, the resulting analysis. The forecast error covariance is responsible for "spreading out" observational data in time and space. Daley (1991) also discusses the spectral properties of the OI algorithm. In fact its characterisation is not simple but, roughly speaking, in the case of uncorrelated observation error, the algorithm acts as a low pass filter on the observations; the analysis draws from the observations at large scales and from the forecast at small scales. If the observational error is correlated, then the algorithm gives increased weighting to gradients in the observational data. For more detailed discussion of the filtering and interpolation properties of the algorithm, the reader is referred to the discussions in Daley (1991) and Lorenc (1981), and references therein.

While the OI approximation for the forecast error covariance eliminates the prohibitively expensive calculation required to evolve the matrix in the full Kalman filter, its storage requirements are similarly expensive, except for the smallest problems. For a model state of size $\mathcal{O}(N)$, $\boldsymbol{C}^f$ has storage requirements of $\mathcal{O}(N^2)$. *Reduced order optimal interpolation* attempts to make the problem tractable through projection of the OI equations onto a space of smaller dimension.

5.1.1 Order reduction

Order reduction techniques are discussed by Fukumori and Malanotte-Rizzoli (1995), De Mey (1997) and De Mey (2002). Consider a transformation, $\boldsymbol{T}$, that reconstructs the full model state, $\boldsymbol{w}$, from a state of smaller dimension, $\boldsymbol{w}'$:

$$\boldsymbol{w} = \boldsymbol{T}\boldsymbol{w}' + \boldsymbol{n}, \tag{5.6}$$

where $\boldsymbol{n}$ lies in the null-space of the *pseudo-inverse* $\boldsymbol{T}^*$. By definition $\boldsymbol{T}^*\boldsymbol{T} = \boldsymbol{I}$, but in general $\boldsymbol{T}\boldsymbol{T}^* \neq \boldsymbol{I}$. The corresponding transformation from the full to the reduced state space is given by

$$\boldsymbol{w}' = \boldsymbol{T}^*\boldsymbol{w}.$$

De Mey (1997) reviews various order reductions that have been used for assimilation. These include, for instance, the fit of feature models to the circulation, "vertical projection" transformations that map the full three-dimensional model state to a two-dimensional field, "local inversion" where the analysis uses

only observations in a restricted region local to each model gridpoint, and the use of a coarse grid model.

Substituting (5.6) in the model equation (5.1), we obtain the reduced-space model equation

$$\boldsymbol{w}'_{k+1} = \boldsymbol{M}'_k \boldsymbol{w}'_k + \boldsymbol{n}'_k, \tag{5.7}$$

where $\boldsymbol{M}' = \boldsymbol{T}^* \boldsymbol{M} \boldsymbol{T}$, and $\boldsymbol{n}' = \boldsymbol{T}^* \boldsymbol{M} \boldsymbol{n}$ can be regarded as a new error term produced by restricting the full model dynamics to a space of smaller dimension. If the null space of $\boldsymbol{T}^*$ is *dynamically uncoupled* from the reduced space, then $\boldsymbol{n}' = \boldsymbol{T}^* \boldsymbol{M} \boldsymbol{n} \approx 0$, and (5.7) is equivalent in form to the original model equation (5.1).

The full-state observation equation is written

$$\boldsymbol{y}^o = \boldsymbol{H} \boldsymbol{w} + \boldsymbol{\varepsilon}^o, \tag{5.8}$$

where $\boldsymbol{\varepsilon}^o$ is observation error (and $\boldsymbol{C}^o = E[\boldsymbol{\varepsilon}^o \boldsymbol{\varepsilon}^{o\mathrm{T}}]$). Substituting into this the state-reduction (5.6), we obtain the observation equation in reduced-state form:

$$\boldsymbol{y}^o = \boldsymbol{H} \boldsymbol{T} \boldsymbol{w}' + \boldsymbol{H} \boldsymbol{n} + \boldsymbol{\varepsilon}^o, \tag{5.9}$$

where $\boldsymbol{H} \boldsymbol{n}$ is a new error resulting from observability of the null-space. This may be written in a form equivalent to the full-state observation equation (5.8):

$$\boldsymbol{y}^o = \boldsymbol{H}' \boldsymbol{w}' + \boldsymbol{\varepsilon}'^o, \tag{5.10}$$

where $\boldsymbol{H}' = \boldsymbol{H} \boldsymbol{T}$ and $\boldsymbol{\varepsilon}'^o = \boldsymbol{H} \boldsymbol{n} + \boldsymbol{\varepsilon}^o$.

Thus, the original full-state model and observation equations, (5.1) and (5.8), have been replaced by analogous equations in the reduced state space, (5.7) and (5.10). The OI equations (5.2) and (5.4) also take an analogous form:

$$\boldsymbol{w}'^a_k = \boldsymbol{w}'^f_k + \boldsymbol{K}'_k (\boldsymbol{y}^o - \boldsymbol{H}' \boldsymbol{w}') \tag{5.11}$$

$$\boldsymbol{K}'_k = \boldsymbol{C}'^f_k \boldsymbol{H}'^{\mathrm{T}} (\boldsymbol{H}' \boldsymbol{C}'^f_k \boldsymbol{H}'^{\mathrm{T}} + \boldsymbol{C}'^o)^{-1}. \tag{5.12}$$

The reduced-state forecast error covariance, $\boldsymbol{C}'^f$, now includes contributions arising from any residual dynamical coupling between the null-space of the transformation $\boldsymbol{T}^*$ and the reduced space. That is, it includes errors associated with non-zero $\boldsymbol{n}'$ in equation (5.7). Similarly, the reduced-state observation error covariance, $\boldsymbol{C}'^o$, now includes contributions arising from observability of the null-space (the term $\boldsymbol{H} \boldsymbol{n}$ in equation (5.9)). De Mey (2002) lists desirable properties for a candidate state-reduction transformation. These include

- From equation (5.9), it is clear that the "modes" of the transformation $\boldsymbol{T}$ must be observable. With reference to altimetry data, the reduced state must have a sea-level signature;

- Also from equation (5.9), it is desirable that the observability of the null-space should be marginal so that it is easily captured with the observational error covariance;

- Similarly, the null-space of $\boldsymbol{T}^*$ should be dynamically uncoupled from the reduced space so it can easily be captured by the forecast error covariance.

The two order-reduction techniques used in this thesis are now described.

5.1.2 Horizontal order reduction

If the observations can be separated into groups such that the forecast and observation errors are uncorrelated between groups, then the respective error covariances, $\boldsymbol{H}\boldsymbol{C}^f\boldsymbol{H}^{\mathrm{T}}$ and $\boldsymbol{C}^o$, can be written in block diagonal form. The inverse will also be block diagonal, and the optimal estimate given by (5.2) and (5.4) can be obtained by processing the measurements from independent groups sequentially. While forecast errors, at least, are not strictly independent between any two observation locations, it is a fair approximation that the correlation falls off with increasing distance in space or time.

Considering one row of the analysis equation (5.2), the approximation amounts to using only those misfits in the right-hand side which fall within a localised space-time region around the analysis gridpoint. From the perspective of the formal order-reduction discussion of the last section, the simplification operator $\boldsymbol{T}^*$ simply truncates the model to those gridpoints in some localised horizontal region about each analysis gridpoint in turn.

From a physical perspective, one expects that the analysis at some gridpoint should not be influenced by observations outside the propagation range of the dynamics of interest.

In addition to such a local approximation, observations falling within the local region are usually pruned even further in practice, to further reduce the computational cost.

Regional approximation and *data selection* are characteristic features of OI schemes, and are discussed by Bergman (1979), Lorenc (1981), Daley (1991), Ezer and Mellor (1994), Cohn et al. (1998), and De Mey and Benkiran (2002).

5.1.3 Vertical order reduction

A considerable additional simplification can be made if the three-dimensional model fields can be decomposed in the vertical. The transformation (5.6) then reconstructs the full model state throughout a water column from a single scalar quantity. The forecast error covariance matrix is reduced in size by N^2 where N is the number of model variables in a water column (typically the number of model levels multiplied by the number of prognostic variables). In general such a transformation may not be possible for the raw model variables, but several possibilities arise for anomaly quantities.

For altimetry assimilation, the most obvious vertical state reduction to make is one that maps (two-dimensional) sea-level anomaly directly to anomalies of subsurface fields. Oschlies and Willebrand (1996) used the altimetric sea-level anomaly to calculate surface velocity anomalies from geostrophy. These

were then projected to depth using empirical linear regression of flow at depth with surface currents. Mellor and Ezer (1991) used a similar approach, updating subsurface temperature and salinity according to empirical correlations between these fields and sea-level. In an alternative statistical approach, De Mey and Robinson (1987) and Dombrowsky and De Mey (1992) decomposed the vertical variability of model variables into modes using empirical orthogonal functions (EOFs), each of which has some surface sea-level expression. Altimetric anomalies were then projected to depth by expanding the observed sea-level in the dominant EOF. This technique was extended by Gavart and De Mey (1997) by calculating the EOFs on isopycnals rather than z-levels. By contrast to the various statistical approaches, Haines (1994) and Cooper and Haines (1996) used a dynamical approach which derived subsurface anomalies from sea-level anomalies by applying conservation principles. It is this approach which is used here.

5.2 Vertical projection of sea-level analysis

Geostrophy is assumed at analysis times (either imposed, or through model adjustment), so that we regard the full model state $\boldsymbol{w}$ as comprised of anomalies of sea-level, and temperature and salinity through a water column. We regard the projection of surface altimetry anomalies to depth as corresponding to a rearrangement of water parcels in the vertical. There is, *a-priori*, no good reason to modify in this process material properties that are otherwise conserved. An important dynamical principle in geophysical fluid dynamics is material conservation of potential vorticity (Pedlosky 1987; Rhines 1986). Indeed, this principle leads to the potential vorticity equation which describes the large-scale frictionless motions of the ocean and atmosphere (Pedlosky 1987). Similarly, under adiabatic conditions, potential temperature is materially conserved, as is salinity. The vertical projection should also conserve these quantities.

Potential vorticity, q, is defined (Müller 1995) for a stratified fluid by

$$q(z) = f\frac{\partial \rho_{pot}}{\partial z}, \tag{5.13}$$

where ρ_{pot} here represents potential density. We wish to ensure that this is materially conserved in any vertical rearrangement of water parcels. Suppose a water parcel initially at depth z is moved to a position $z + h(z)$, then the condition can be written mathematically as

$$\frac{\partial \rho_{pot}}{\partial z}[z + h(z)] = \frac{\partial \rho_{pot}}{\partial z}(z)$$

or, using the chain rule of differentiation,

$$\frac{\partial \rho_{pot}}{\partial z}(z)\left(1 + \frac{dh}{dz}\right) = \frac{\partial \rho_{pot}}{\partial z}(z)$$

so that

$$\frac{dh}{dz} = 0$$

and the constraint restricts the allowable motions to vertical translations (lifting or lowering) of the water column which are uniform with depth:

$$\rho_{pot}(z) = \rho_{pot_0}(z - L),$$

where ρ_{pot_0} is the depth profile of potential vorticity before modification and L is the amount of lifting of the material surfaces.

As well as potential vorticity, we wish to materially conserve potential temperature and salinity. Thus depth profiles of these conservative tracers are also translated vertically:

$$\begin{aligned} \theta(z) &= \theta_0(z - L), \\ S(z) &= S_0(z - L), \end{aligned} \tag{5.14}$$

where θ_0 and S_0 are the tracer profiles before modification.

The conservation principles therefore restrict the subsurface modifications to simple uniform vertical translations of the conservative tracer profiles. In the case of lifting, light surface waters are removed and dense water added at the bottom, leading to an increased weight of the water column. If on the other hand the profiles are lowered, light waters are added at the surface and dense bottom water removed, leading to a reduced water column weight. Water in the mixed layer is subject to diabatic forcing, and so the conservation principles used above do not apply. Ideally, direct observations of tracers in the mixed layer is required. No such data were assimilated in this thesis, and the mixed layer was left to adapt on its own in the assimilation experiments performed here.

We must now determine the amount of vertical translation required for an analysed sea-level anomaly. Haines (1994) and Cooper and Haines (1996) made an assumption of no change in bottom pressure in order to close the problem. This induces a primarily baroclinic change. The same approach is used here, but a modification is also considered which includes a more barotropic response.

5.2.1 Purely baroclinic projection

Making an assumption of no change in bottom pressure, Haines (1994) and Cooper and Haines (1996) proposed that the change in surface pressure associated with a sea-level anomaly $\Delta\eta$ should exactly be compensated by the change in weight of the entire water column associated with the lifting or lowering of tracer profiles:

$$\rho(0)g\Delta\eta = -[\rho(-H) - \rho(0)]gL \tag{5.15}$$

or

$$L = \frac{\rho(0)}{\rho(0) - \rho(-H)}\Delta\eta. \tag{5.16}$$

This expression was valid in their model in which the equation of state was a function only of temperature. In the HOPE model, however, a fully nonlinear equation of state is used so that a uniform vertical translation of potential

density does not imply an equivalent uniform translation of *in-situ* density. It is, of course, the *in-situ* density which is required in (5.15) (with L no longer uniform with depth). On the other hand, the assumption of no change in bottom pressure is equivalent to the sea-level anomaly being a purely steric effect. If the changes in potential temperature and salinity at depth z are $\Delta\theta(z)$ and $\Delta S(z)$ respectively, then the change in steric height is given by

$$\Delta\eta = \int_{-H}^{0} (\alpha\Delta\theta - \beta\Delta S)dz,$$

where α and β are the thermal expansion and haline contraction coefficients respectively:

$$\begin{aligned} \alpha &= -\frac{1}{\rho}\frac{\partial\rho}{\partial\theta}\bigg|_{S,p}, \\ \beta &= \frac{1}{\rho}\frac{\partial\rho}{\partial S}\bigg|_{\theta,p}. \end{aligned}$$

For small uniform displacements through the water column, (5.14), then to first order we have

$$\begin{aligned} \Delta\theta(z) &= -L\frac{\partial\theta}{\partial z}, \\ \Delta S(z) &= -L\frac{\partial S}{\partial z} \end{aligned}$$

so the steric sea-level change is given by

$$\Delta\eta = -L\int_{-H}^{0} (\alpha\frac{\partial\theta}{\partial z} - \beta\frac{\partial S}{\partial z})dz.$$

Finally, then, the amount of vertical displacement is given by

$$L = -\frac{\Delta\eta}{\int_{-H}^{0}(\alpha\frac{\partial\theta}{\partial z} - \beta\frac{\partial S}{\partial z})dz}. \tag{5.17}$$

Assembling this into the formalism of section 5.1.1 for the reduced-order optimal interpolation scheme, we therefore have the scalar reduced state w' now being equivalent simply to sea-level anomaly. The full model state $\boldsymbol{w}$ comprises sea-level anomaly, and temperature and salinity anomalies through a water column. The transformation $\boldsymbol{T}$ reconstructing the full state from the reduced state is given by

$$\begin{pmatrix} \Delta\eta \\ \vdots \\ \Delta\theta_k \\ \vdots \\ \Delta S_k \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ -\frac{\frac{\partial\theta}{\partial z}\big|_{z_k}}{\int_{-H}^{0}(\alpha\frac{\partial\theta}{\partial z} - \beta\frac{\partial S}{\partial z})dz} \\ \vdots \\ -\frac{\frac{\partial S}{\partial z}\big|_{z_k}}{\int_{-H}^{0}(\alpha\frac{\partial\theta}{\partial z} - \beta\frac{\partial S}{\partial z})dz} \\ \vdots \end{pmatrix} \Delta\eta, \tag{5.18}$$

where k is the model layer index. The reduced-state observation operator H' directly observes sea-level anomalies and the reduced-state forecast error covariance $\boldsymbol{C}'^f$ is the univariate sea-level forecast error covariance.

The OI equations (5.11) and (5.12) are used to obtain optimal estimates of sea-level anomaly at analysis times, which are then used to update the tracer fields according to (5.18). Velocities may be updated geostrophically or allowed to adjust dynamically to the new three-dimensional density field.

5.2.2 Barotropic projection

The vertical projection scheme described above leaves bottom pressure, and therefore bottom velocities, unchanged initially. The discussion in section 4.2.3, however, noted that there is significant barotropic variability in the Southern Ocean. A simple modification of the above scheme is now described which allows the analysed sea-level anomaly to be projected to depth in a more barotropic manner.

Pinardi et al.'s (1995) geostrophic decomposition of sea-level into a barotropic and baroclinic term was mentioned in section 4.2.3. It holds exactly only for a flat-bottom rigid-lid model on an f-plane and is given by:

$$\eta = \frac{f}{g}\frac{\Psi}{H} - \frac{1}{\rho_0 H}\int_{-H}^{0}(z+H)\rho\, dz.$$

For such a model, this suggests a natural way to impose a sea-level anomaly purely on the barotropic mode — by leaving density unaltered but modifying streamfunction:

$$\Delta\Psi = \frac{gH}{f}\Delta\eta.$$

While the above decomposition does not hold in a free-surface model with topography, an analogous approach is suggested. Applying geostrophy to HOPE's barotropic system (2.30,2.31), provides the following balance (which also holds for anomaly quantities):

$$f(U,V) = \boldsymbol{k} \times \left(gH\boldsymbol{\nabla}\eta + \frac{1}{\rho_0}\int_{-H}^{0} \boldsymbol{\nabla}p'\, dz \right). \tag{5.19}$$

For a purely barotropic response, a sea level anomaly could be imposed by leaving the internal pressure field p' unaltered, but modifying the depth-integrated velocities geostrophically according to:

$$\begin{aligned} \Delta U &= -\frac{gH}{f}\frac{\partial\Delta\eta}{\partial y}, \\ \Delta V &= \frac{gH}{f}\frac{\partial\Delta\eta}{\partial x}. \end{aligned} \tag{5.20}$$

The torque of the bottom velocities acting against topographic slopes induced by such a change would be expected to strongly excite topographic barotropic Rossby waves. In general, one would prefer a modification intermediate between the purely baroclinic and purely barotropic projections. Indeed

this is suggested by the discussion in section 4.2.3 of the subsurface variability in the model.

In principle, we would like to partition the sea-level anomaly into a barotropic part and a baroclinic part, using the vertical projection (5.18) for the baroclinic part and making the geostrophic adjustment (5.20) for the barotropic part. Such a simple partition, however, is not possible. The vertical projection (5.18), while leaving bottom velocity unaltered, nevertheless effects a change in depth-integrated velocity. This change in depth-integrated velocity would not occur only if (for the anomalies) the two terms on the right-hand side of (5.19) exactly compensated. Then geostrophy would ensure that U and V did not evolve (see the barotropic momentum equations (2.30,2.31)). That is, under geostrophy, the depth-integrated velocities would remain unaffected only if the vertically averaged baroclinic pressure gradient balanced the imposed sea-level pressure gradient. We show now that this is not true for the vertical projection (5.18). The vertically integrated pressure anomaly can be written as follows:

$$\begin{aligned}\int_{-H}^{0} \Delta p' \, dz &= \int_{-H}^{0} \left(- \int_{0}^{z} \Delta\rho g \, dz' \right) dz \\ &= \int_{-H}^{0} (z + H) \Delta\rho g \, dz \\ &= H \int_{-H}^{0} \Delta\rho g \, dz + \int_{-H}^{0} z \Delta\rho g \, dz \end{aligned} \tag{5.21}$$

For no change in total bottom pressure, we have

$$\Delta p(-H) = \rho_0 g \Delta\eta + \int_{-H}^{0} \Delta\rho g \, dz = 0$$

which can be substituted in the geostrophic relation (5.21) to show

$$\int_{-H}^{0} \Delta p' \, dz = -\rho_0 g H \Delta\eta + \int_{-H}^{0} z \Delta\rho g \, dz$$

or

$$g H \Delta\eta + \frac{1}{\rho_0} \int_{-H}^{0} \Delta p' \, dz = \frac{1}{\rho_0} \int_{-H}^{0} z \Delta\rho g \, dz. \tag{5.22}$$

The vertical projection scheme (5.18) translates water columns in the vertical. Thus for a stable water column, the sign of $\Delta\rho(z)$ is uniform with depth and the term on the right-hand side of (5.22) is non-zero. Thus even in the flat-bottom case, the right-hand side of (5.19) is non-zero and it is proved that the purely baroclinic vertical projection modifies depth-integrated velocity.

In the presence of topography, this fact leads to the so-called JEBAR (*Joint Effect of Baroclinicity And Relief*) (Huthnance 1984; Mertz and Wright 1992; Slørdal and Weber 1996). The torque of depth-integrated barolinic pressure against sloping topography acts as a forcing term in the barotropic momentum equations (2.30,2.31). The dynamical impact of JEBAR in assimilation schemes does not appear to have received attention in the literature.

Despite the inability to easily partition sea-level anomaly between baroclinic and barotropic parts, we choose, nevertheless, to ignore the effect on the depth-integrated velocities of the vertical projection scheme (5.18) — we assume a zeroth order partition is possible,

$$\Delta\eta = \Delta\eta_{BT} + \Delta\eta_{BC}, \tag{5.23}$$

and use the baroclinic part, $\Delta\eta_{BC}$, to modify tracer profiles according to the vertical projection (5.18) with the remaining barotropic part, $\Delta\eta_{BT}$, used to modify barotropic velocities according to the geostrophic relation (5.20). The degree of partitioning between the two is taken as a free parameter in the assimilation experiments described in section 5.4, with assimilation performance analysed for various choices of the partition. It will be seen that ignoring the residual effect of the vertical projection on the depth-integrated velocities will excite barotropic Rossby waves.

We now investigate the behaviour of the vertical projection scheme for an idealised sea-level anomaly, before examining on hydrographic sections the validity of the conservation principles used in its derivation.

5.2.3 Vertical projection dynamics

To examine the dynamical behaviour of the vertical projection scheme described above, some idealised experiments were performed. A 10 cm amplitude gaussian sea-level anomaly of spatial scale $\mathcal{O}(1000$ km) was applied at each of the two locations ('A' and 'B') shown in Figure 5.2(a). The locations were chosen to examine the effect of topography on the vertical projection. Location 'A' is over very steep topography on the northern flank of the Southeast Indian Ridge and just west of the South Tasman Rise, while location 'B' is over the relatively flat South Indian Abyssal Plain.

The corresponding vertical displacement of water columns required for a completely baroclinic projection, according to (5.17), is shown in Figure 5.2(b). It is clear that to obtain an equivalent steric anomaly, a much greater water column displacement is needed in the south of the domain where stratification is weaker (see, for example, Figure 4.4 for the stratification). The maximum displacements required for the 10 cm disturbances applied here were 56 m at location 'A' and 95 m at location 'B'.

For each location, three model runs were performed, making the projection in turn:

1. 'purely baroclinic', equation (5.18), *i.e.* no change in bottom pressure;
2. 50% barotropic;
3. purely barotropic, equation (5.20).

In each case the model was integrated for around two weeks after the perturbation was applied, and the differences from an identical unperturbed run were examined.

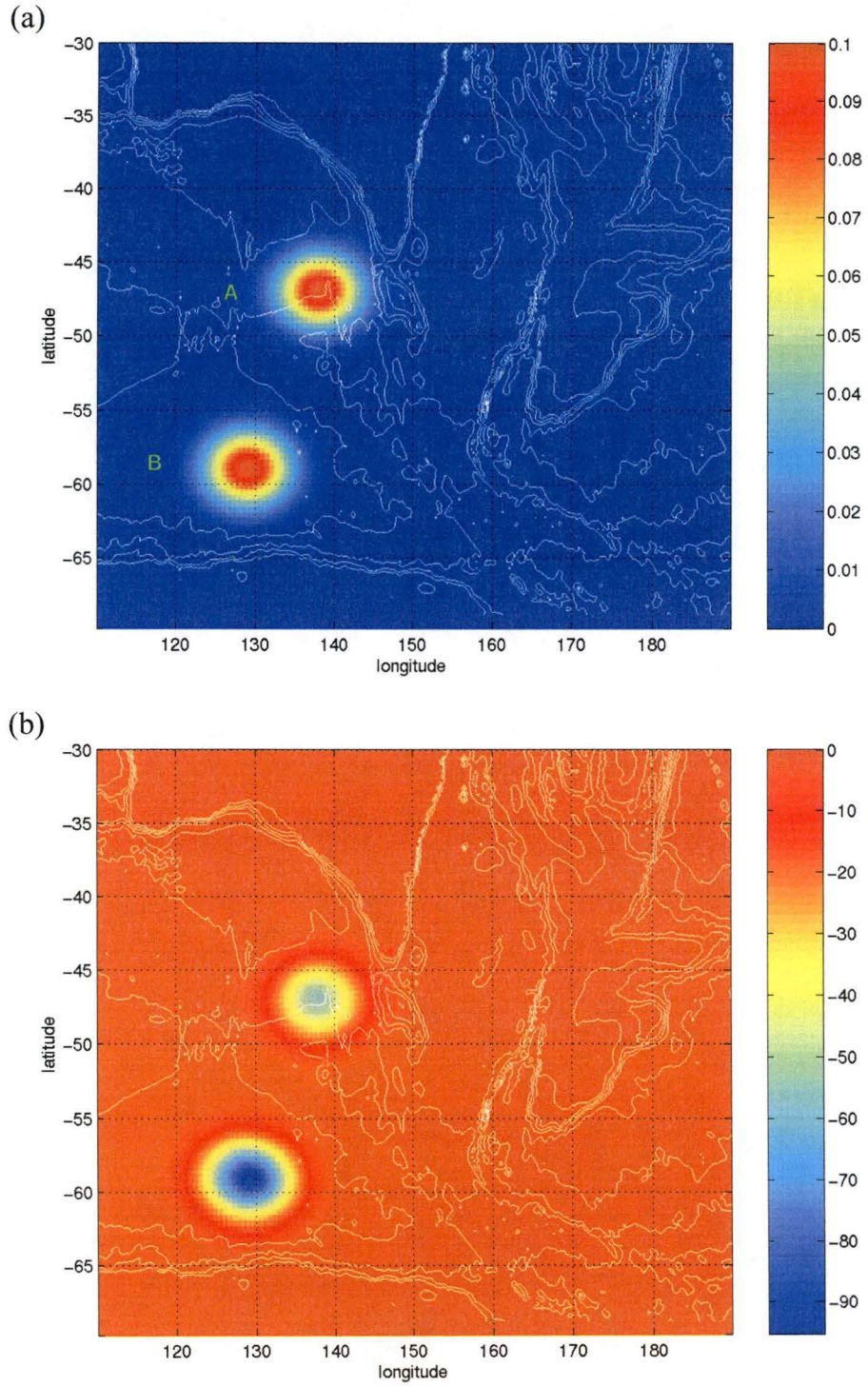


Figure 5.2: (a) Sea-level anomaly (m) and (b) corresponding (pure baroclinic) water column displacement (m) at two locations used to examine behaviour of vertical projection scheme. Topography is overlaid at intervals of 1000 m.

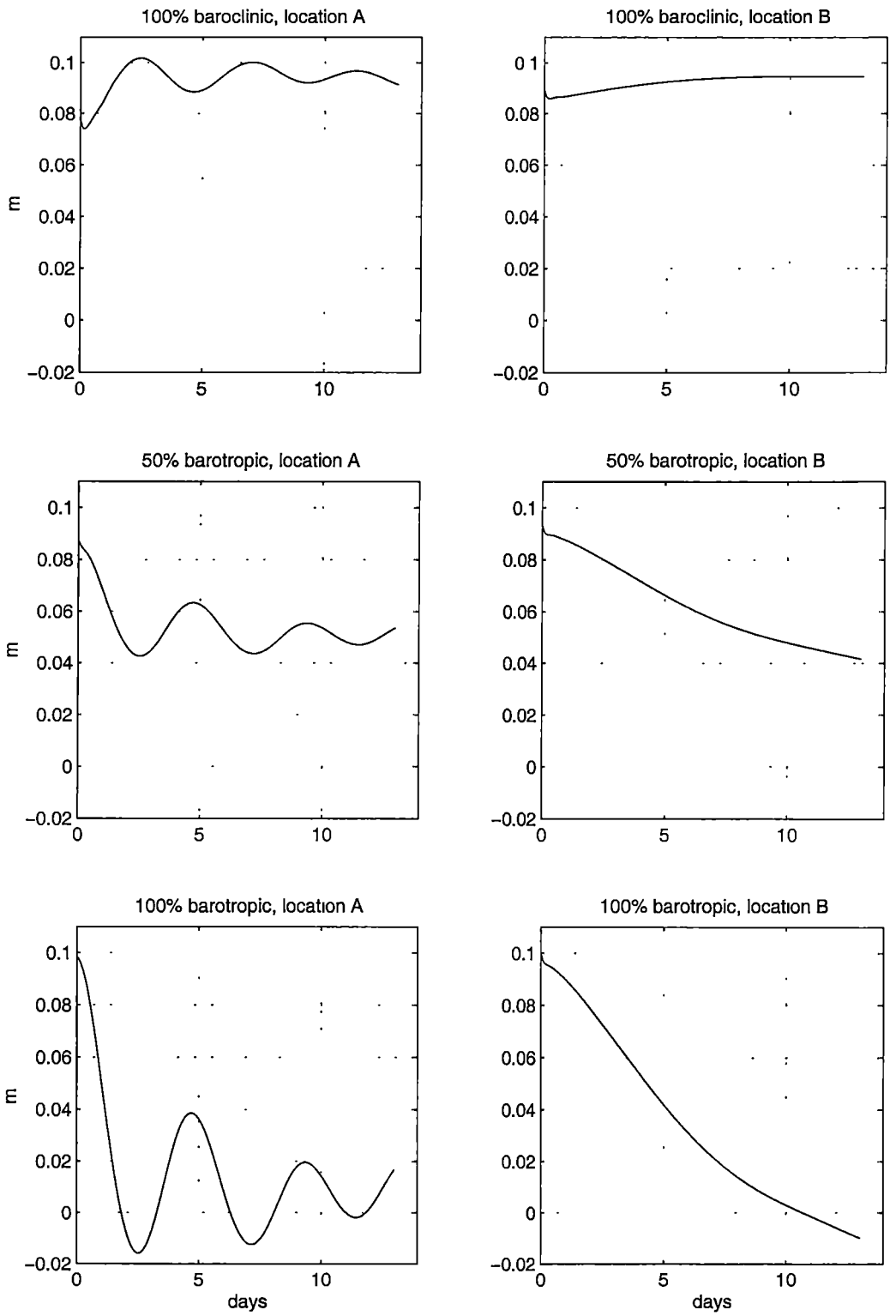


Figure 5.3: Sea-level anomaly at centre of initial perturbation for each location and for the three vertical projection methods: 100% baroclinic, 50% barotropic, and 100% barotropic.

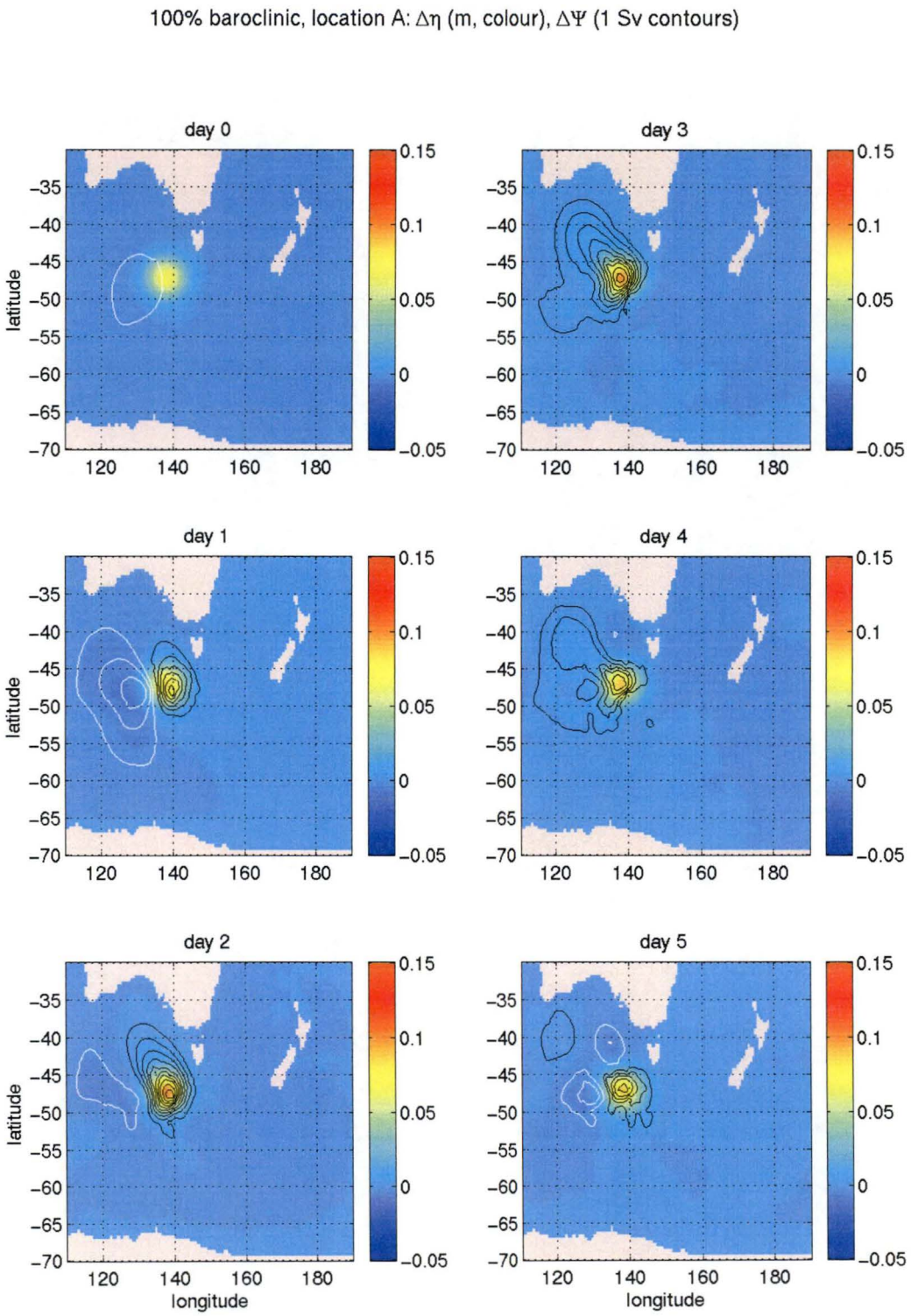


Figure 5.4: Evolution of sea-level and streamfunction anomalies for the perturbation at location 'A', purely baroclinic projection.

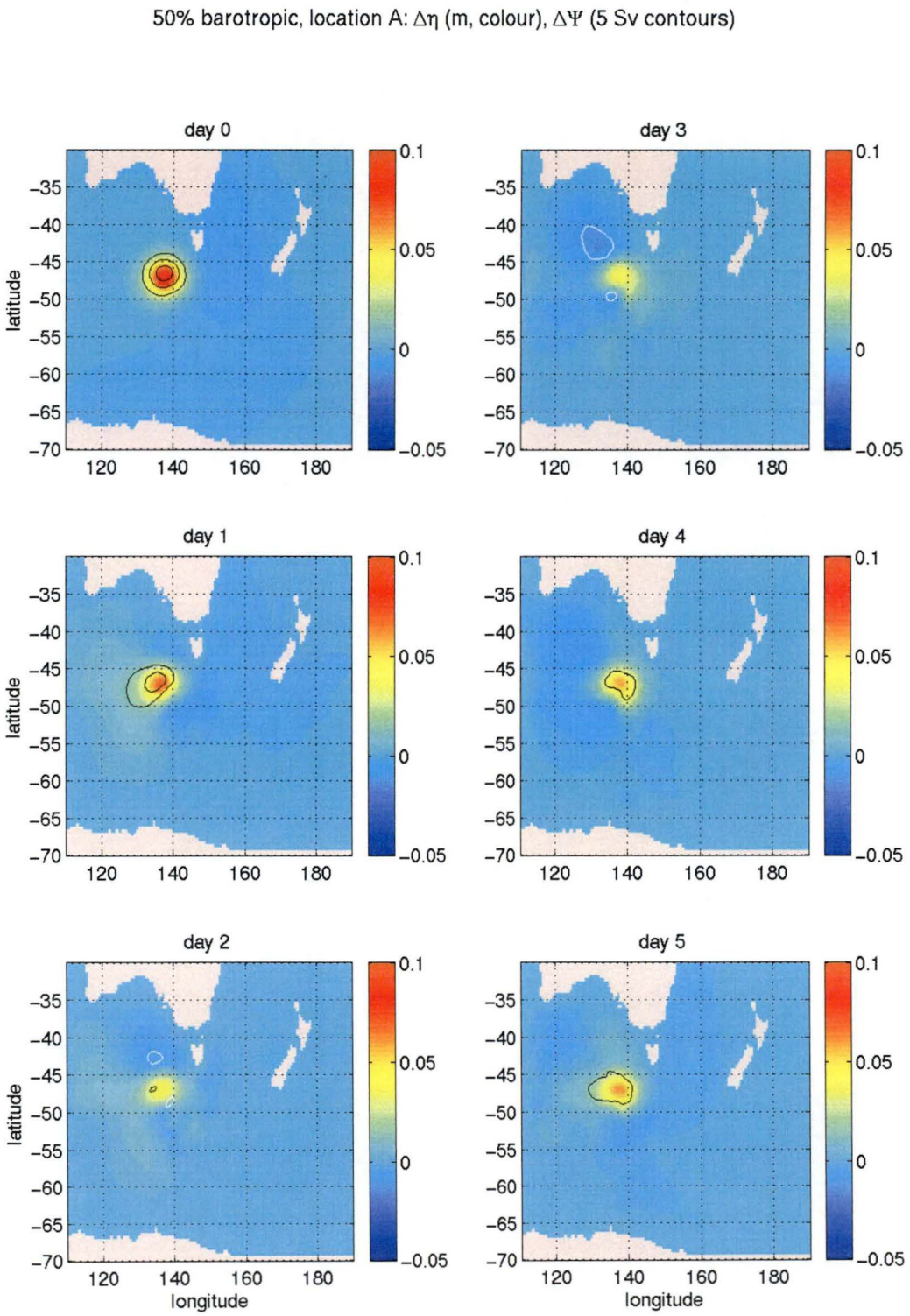


Figure 5.5: Evolution of sea-level and streamfunction anomalies for the perturbation at location 'A', 50% barotropic projection.

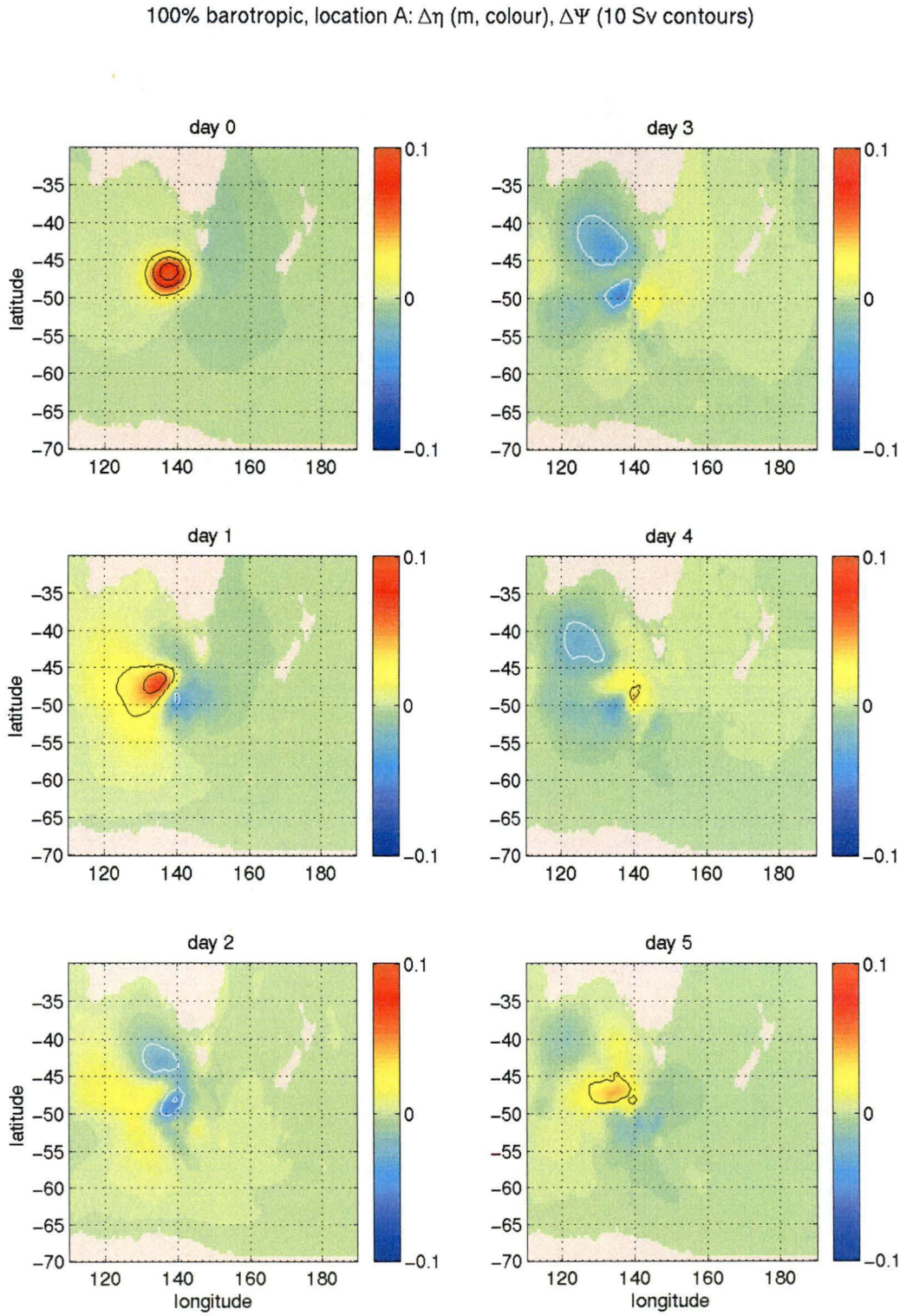


Figure 5.6: Evolution of sea-level and streamfunction anomalies for the perturbation at location 'A', 100% barotropic projection.

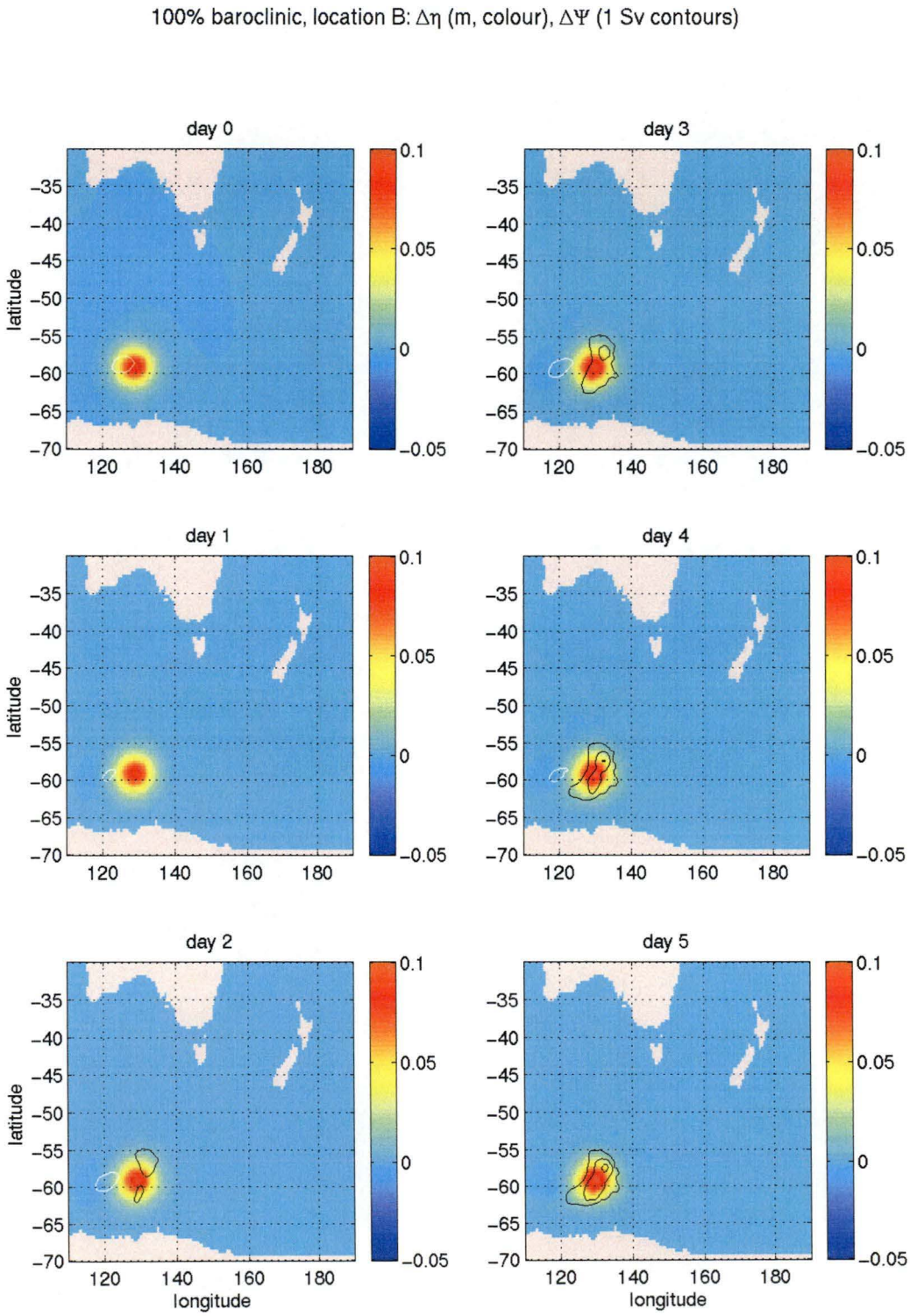


Figure 5.7: Evolution of sea-level and streamfunction anomalies for the perturbation at location 'B', purely baroclinic projection.

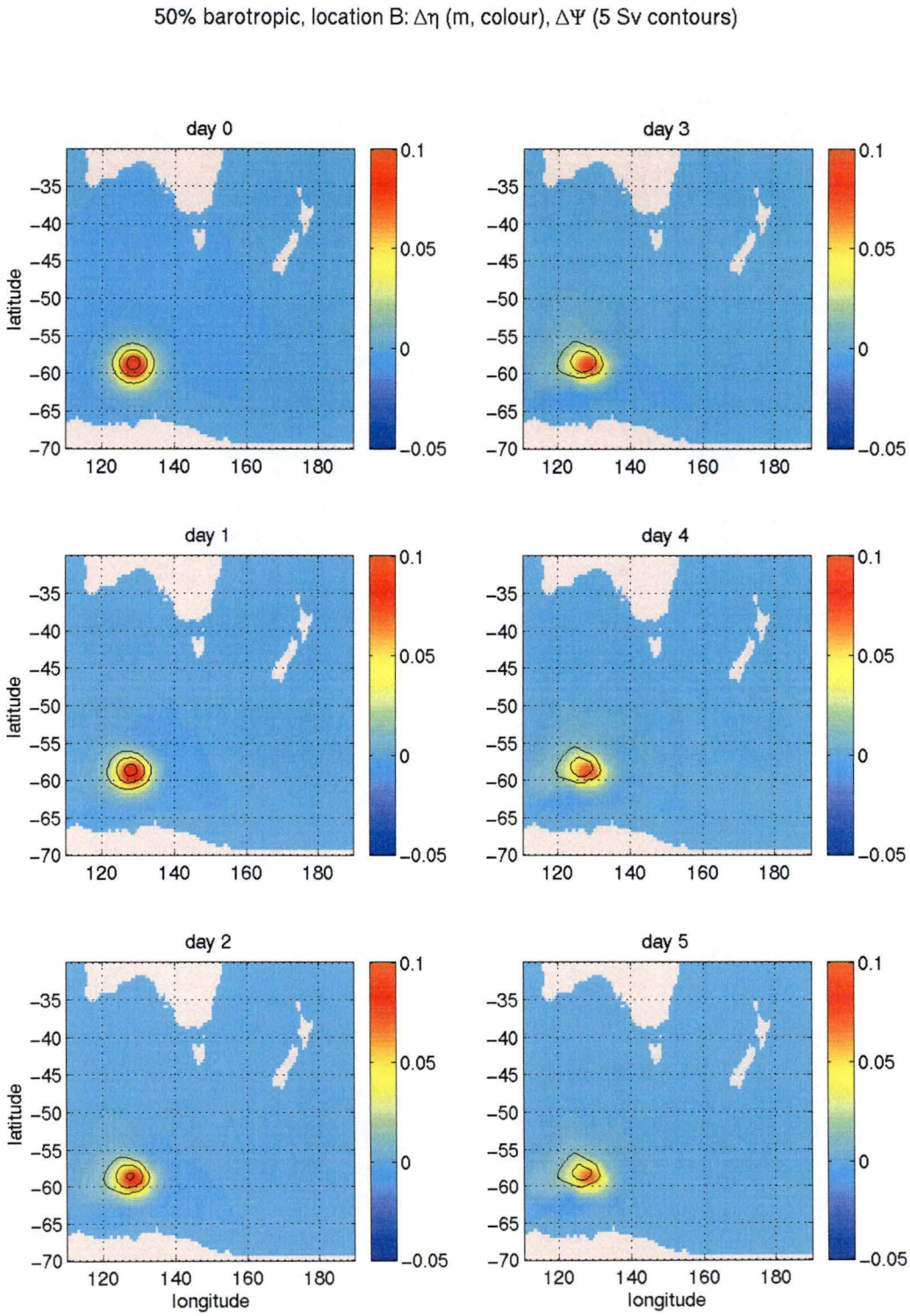


Figure 5.8: Evolution of sea-level and streamfunction anomalies for the perturbation at location 'B', 50% barotropic projection.

Table 5.1: Dispersion relations for waves excited by a density perturbation. In this table, $\boldsymbol{k}$ is the horizontal wavenumber with k_x its zonal component, H_n is the *equivalent depth* for the n-th vertical mode, and $R_e = (gH/f^2)^{1/2}$ is the barotropic Rossby radius. The topographic slope is modeled as an exponential profile with $H_y/H = -1/l$. Reference equation numbers refer to LeBlond and Mysak (1978).

wave type	*dispersion relation*	*reference equation*
surface gravity wave	$c = \sqrt{gH}$ (nondispersive)	17.7
internal gravity wave	$c_n = \sqrt{gH_n}$ (nondispersive)	17.7
planetary Rossby wave (BT)	$\omega = -\frac{\beta k_x}{(\lvert\boldsymbol{k}\rvert^2+1/R_e^2)}$	18.13a
topographic Rossby wave (BT)	$\omega = -\frac{(\beta+f/l)}{k_x}$	20.18
baroclinic Rossby wave	$\omega = -\frac{\beta k_x}{[\lvert\boldsymbol{k}\rvert^2+(n\pi/R_i)^2]}$	18.13b

Table 5.2: Parameters used to calculate timescales of wave processes. Length scales are taken as twice the width at the half-height points of the gaussian perturbation.

parameter	*location A*	*location B*
location	(138°E, 47°S)	(129°E, 59°S)
depth, H	4100 m	4800 m
meridional bottom slope, H_y	2.5×10^{-3}	7.5×10^{-4}
zonal length scale, $L_x = 2\pi/k_x$	1400 km	1000 km
meridional length scale, $L_y = 2\pi/k_y$	980 km	980 km

Figure 5.3 shows the sea-level anomaly at the central perturbation location for each of the experiments, while Figures 5.4–5.9 show daily snapshots of sea-level and streamfunction anomalies for the first five days of each run. Figures 5.10 and 5.11 show meridional sections of zonal velocity, u, through the centre of the purely baroclinic and purely barotropic perturbations respectively at two hours and five days after their application.

There are several timescales evident in these diagrams. In order to interpret the responses, it is necessary to consider the various wave processes excited by the perturbations. These include both internal and external (surface) gravity waves, and Rossby waves. For the barotropic system, the latter include both planetary and topographic Rossby waves. Table 5.1 lists these waves together with their respective dispersion relations. Approximate timescales, T, for these processes can be estimated as $T = 2\pi/|\omega|$. For the (nondispersive) internal and external gravity waves, the length scales are taken as the baroclinic and barotropic Rossby radii, respectively, and the resultant timescale is for geostrophic adjustment. It is equal to the local inertial timescale, f^{-1}. The parameters in Table 5.2 were used to estimate the remaining timescales, listed in Table 5.3.

Guided by these theoretical results, and with reference to the figures, the

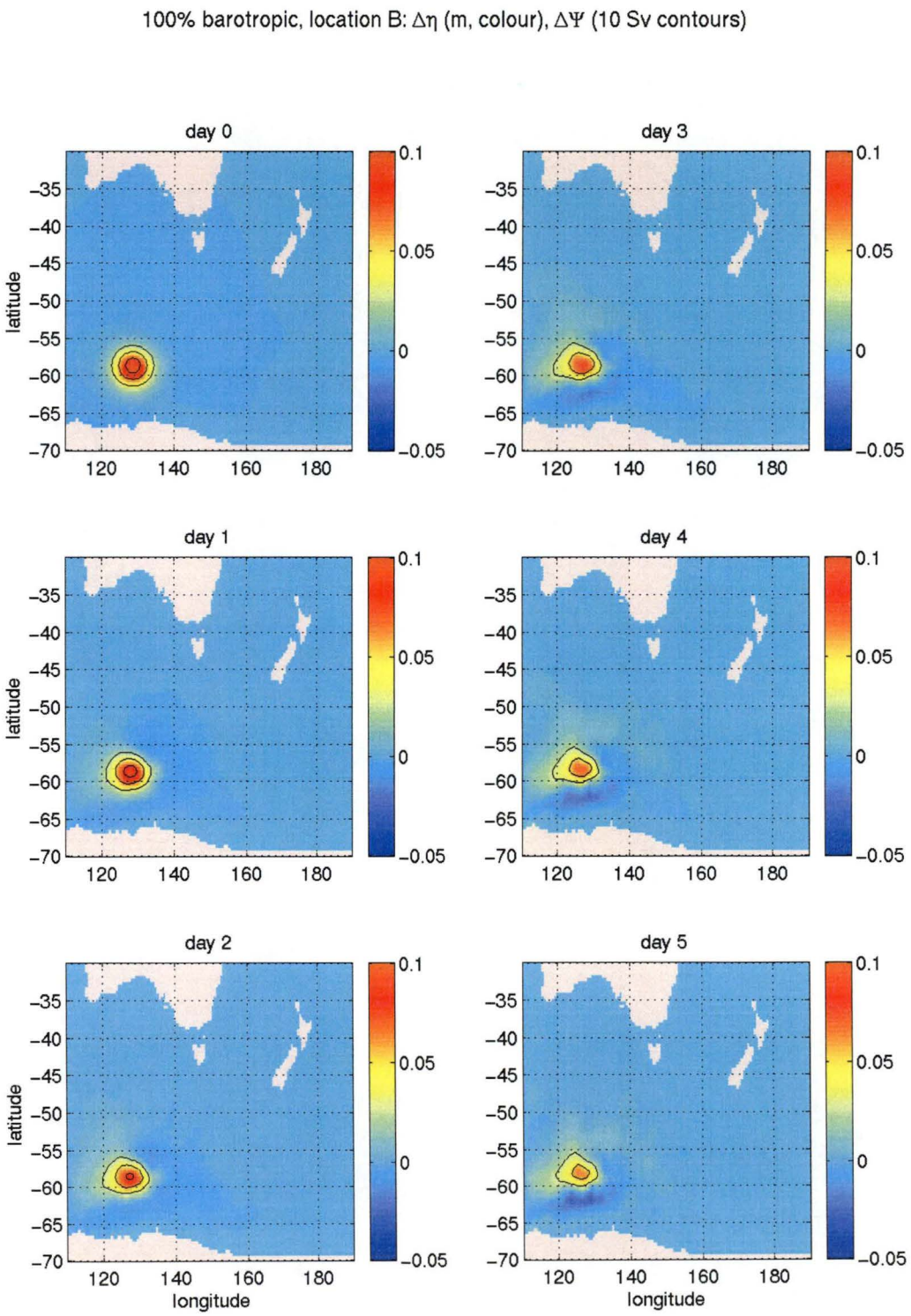


Figure 5.9: Evolution of sea-level and streamfunction anomalies for the perturbation at location 'B', 100% barotropic projection.

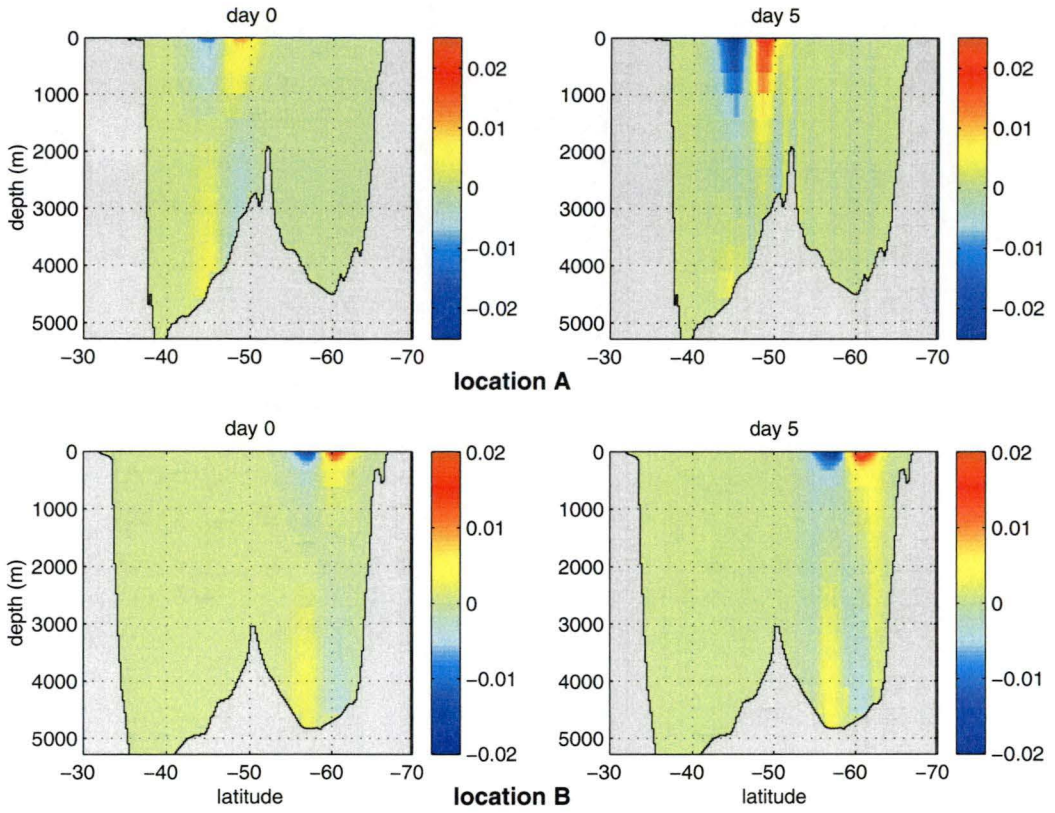


Figure 5.10: Meridional sections of zonal velocity, u, taken through the centre of the purely baroclinic perturbation for locations 'A' (top, 138°E) and 'B' (bottom, 129°E), and after two hours (left) and five days (right).

Table 5.3: Approximate timescales of different wave processes for perturbations at locations 'A' and 'B' with parameters in Table 5.2.

wave type	*location 'A'*	*location 'B'*
surface gravity wave	2.6 hr	2.2 hr
internal gravity wave	2.6 hr	2.2 hr
barotropic Rossby wave (planetary)	64 d	79 d
barotropic Rossby wave (topographic)	4 d	15 d
baroclinic Rossby wave	53 y	185 y

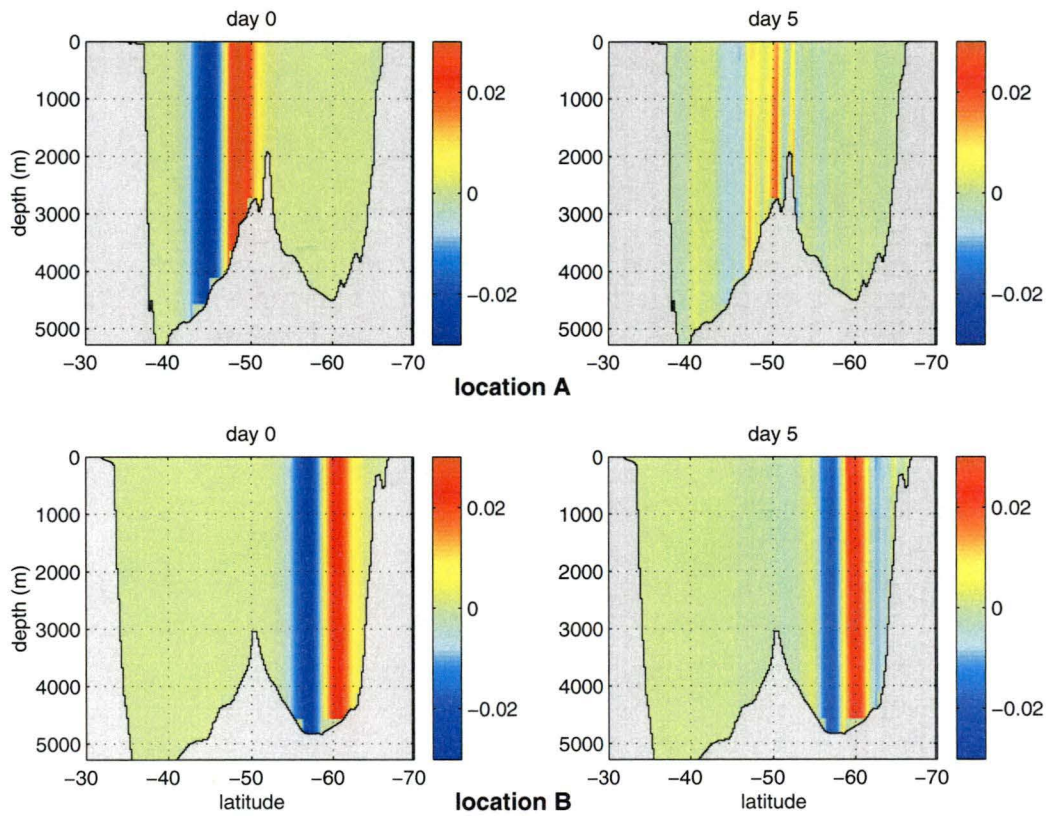


Figure 5.11: Meridional sections of zonal velocity, u, taken through the centre of the purely barotropic perturbation for locations 'A' (top, 138°E) and 'B' (bottom, 129°E), and after two hours (left) and five days (right).

dynamical evolution of the perturbations will now be discussed.

The most rapid process is geostrophic adjustment, completing within a few hours of the initial perturbations (Table 5.3). The spatial scales of the perturbations are much larger than the baroclinic Rossby radii, but smaller than the barotropic Rossby radii. Thus, we expect the (baroclinic) velocities to adjust to the imposed density perturbation for the baroclinic system, and the pressure field (sea-level) to adjust to geostrophy for the barotropic system (Gill 1982, §7.5). The barotropic (depth-integrated) velocities are not expected to evolve substantially during initial adjustment. For the baroclinic perturbation, we expect an initial strong shear to develop in the vertical velocity profile, but for this to integrate to a zero net volume transport ('streamfunction') anomaly. From Figure 5.10 (left panels) we see that this is the case. The meridional section of zonal velocity shows at day zero (actually two hours after the perturbation was applied) strong vertical shear with a change of sign through the water column. The contours of streamfunction anomaly in Figures 5.4 and 5.7 at day zero (again after two hours) show that the anomaly is indeed approximately zero at both locations. As mentioned, we expect sea-level to adjust within the first few hours to provide a geostrophic balance for the barotropic system. More generally, we expect sea-level to adjust continually throughout the runs to provide such a balance on timescales longer than a few hours. The balance in question was discussed earlier in relation to partitioning sea-level anomalies between a baroclinic and barotropic part, and is given by equation (5.19). In that earlier discussion it was noted that for a purely baroclinic perturbation where sea-level attained the desired imposed profile, there would necessarily be a non-zero geostrophic depth-integrated velocity field. Conversely, the sea-level cannot attain the desired profile under geostrophy, despite the steric effect, unless the streamfunction is also modified. In the present purely baroclinic experiments, no initial adjustment of streamfunction is performed and we have just seen that initially it does not evolve. Therefore, the desired amplitude of the sea-level anomaly (10 cm) cannot be attained initially. The deficit corresponds exactly to the non-zero sum of the depth-averaged baroclinic pressure anomaly and the desired surface (sea-level) pressure anomaly. This was quantified in equation (5.22). Examining Figure 5.3, showing the sea-level anomaly at the centres of the perturbations, we see that this is exactly what happens. A small signal associated with (barotropic) geostrophic adjustment is seen in the first few hours in both baroclinic runs, and neither location shows initial sea-level adjustment to the desired 10 cm. At location 'A' the deficit is over two centimetres, while at location 'B' it is almost 1.5 cm. On the other hand, we see that for the 100% barotropic perturbations, sea-level reaches the desired 10 cm initially. This, too, is as expected. Referring again to the geostrophic balance equation (5.19), the vertically-integrated pressure term is now zero, and so the balance is between sea-level and depth-integrated velocity. The latter was chosen, of course, in the barotropic projection experiments exactly to balance the required sea-level anomaly. The analysis here simply confirms that this is the correct thing to do for a purely barotropic projection.

Extending the analysis in time now past the initial adjustment phase,

some very different sea-level responses are seen (Figure 5.3) both between the two locations, and between the projection types (baroclinic *vs* barotropic). Referring to Table 5.3, the next shortest timescale at both locations is that associated with barotropic topographic Rossby waves (four days for location 'A' and 15 days for location 'B'). Sea-level waves at around these periods for the respective locations are seen in all experiments, most strongly in the purely barotropic projection. It was suggested earlier that the large torque of bottom velocity against topographic slope in the barotropic projection would strongly excite barotropic Rossby waves. We see now that this is the case. The westwards propagation of these waves is seen very clearly over the first five days in Figure 5.6 for location 'A'. In the context of the overall analysis, it is useful to bear in mind that the sea-level tracks these waves because it is so tightly bound to maintaining geostrophy in the barotropic system. That these waves are excited by the perturbations is no surprise: a non-zero bottom velocity results from both the barotropic and baroclinic projections. The strength of the excitation depends on the size of the bottom velocities, and the wave speed depends on the slope of the topography. The bottom slope at location 'B' is much smaller than at location 'A', with a consequent reduction in the phase speed of the wave. Figures 5.10 and 5.11 show that in both the baroclinic and barotropic projection schemes, the waves act to reduce the bottom velocity after some days. By examining Figure 5.9 it is clear that most of the reduction in sea-level amplitude at the original location, seen in Figure 5.3, for the barotropic projection is associated with the westwards movement of the whole system. After five days, it has moved westwards by around 4°. This compares with 5.8° expected from the phase speed $c = \omega/k_x$ using the parameters in Table 5.2. The agreement is close, and certainly no worse than approximating the topography as an exponential, implicit in the theory for Table 5.1.

The remaining processes — planetary barotropic, and baroclinic, Rossby waves — have timescales much longer than the relatively short runs considered here, and longer also than the analysis intervals used in the assimilation experiments described later. They have no significant effect on the dynamics of the perturbation over these timescales. In particular, the imposed density anomaly in the baroclinic projection remains fixed apart from advection in the background flow and diffusion.

It is worth making a final comment on the responses for the combined perturbation with 50% of the sea-level used for both barotropic and baroclinic projections. As expected, the responses are intermediate between the purely barotropic and purely baroclinic projections at both locations. Thus, in Figure 5.8, for example, the contours of streamfunction anomaly (up to 10 Sv in this case) are seen clearly to move westwards away from the original perturbation location, which nevertheless maintains a largely stationary sea-level signature due to the baroclinic part of the projection.

Summarising the dynamics of the vertical projection, then, initial adjustment produces a state where baroclinic velocities balance the imposed internal pressure anomaly while sea-level balances both the depth-integrated pressure and streamfunction anomalies. A non-zero imposed streamfunction anomaly is required for sea-level to attain the desired amplitude. Subsequent evolution

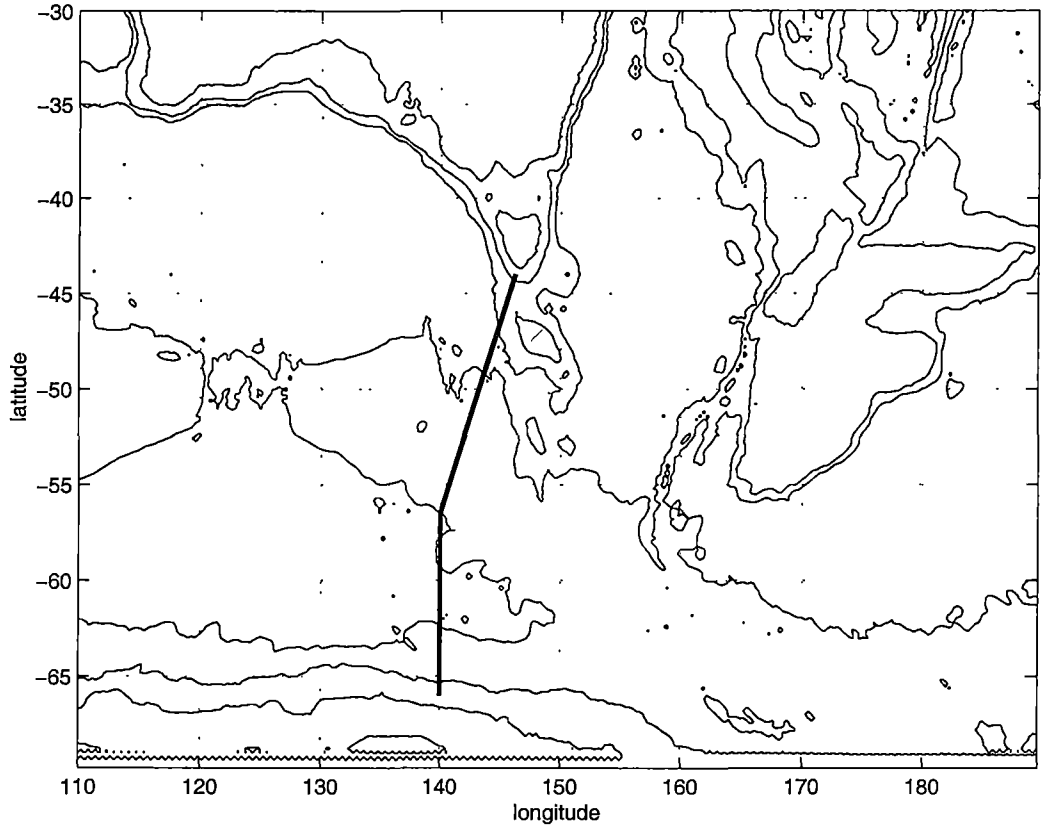


Figure 5.12: Cruise track of WOCE repeat hydrographic section SR3 over topography contours (2000 m interval).

at timescales of interest is dominated by the excitation of topographic Rossby waves.

The validity of the principles used to derive the vertical projection scheme is now examined using repeat hydrographic sections.

5.2.4 Hydrographic validation

The method for projecting analysed sea-level anomalies onto subsurface model fields proposed in section 5.2 was based on conservation of potential vorticity and watermasses on isopycnals. These principles are reasonable if subsurface variability is limited to uniform vertical displacement of isopycnals and tracer profiles through a water column. Such a model of subsurface variability is validated here against hydrographic data from five repeat occupations of the WOCE SR3 section between Tasmania and Antarctica (Figure 5.12). The cruises were undertaken by the Co-operative Research Centre for Antarctica and the Southern Ocean in Hobart, Tasmania on the research vessel *RSV Aurora Australis*. Four of the cruises were made during the successive austral summers 1993–1996, while a fifth was made in the winter of 1995 (Table 5.4).

The hydrographic data are used here to examine the validity of the uniform isopycnal displacement model proposed in section 5.2. Observed density

Table 5.4: Cruise details for repeat hydrographic sections of WOCE SR3 line analysed here.

cruise code	*cruise dates*	*cruise report*
AU9309	11 Mar, 1993 – 3 Apr, 1993	(Rosenberg et al. 1994)
AU9407	2 Jan, 1994 – 16 Jan, 1994	(Rosenberg et al. 1995a)
AU9404	19 Jan, 1995 – 1 Feb, 1995	(Rosenberg et al. 1995b)
AU9501	18 Jul, 1995 – 1 Aug, 1995	(Rosenberg et al. 1997)
AU9601	30 Aug, 1996 – 21 Sep, 1996	(Rosenberg et al. 1997)

anomalies are modelled as a vertical displacement at each location through a water column to examine whether this is indeed uniform with depth. A best-fit uniform displacement is then calculated for each water column, with the inferred dynamic height anomaly compared with the exact anomaly. In addition, the sea-level anomalies observed by satellite altimeter synoptic with the respective cruises are compared with the hydrographic estimates. This provides an indication of the level of barotropic sea-level variability along the sections.

The raw CTD data were calibrated and averaged onto 2 dbar pressure levels, as described in the cruise reports (Rosenberg et al. (1994; 1995a; 1995b; 1997)). To estimate isopycnal displacements, the density anomalies are referenced against mean vertical gradients as described later. While climatological data, such as that of Olbers et al. (1992), could have been used for this purpose, the representation of topography is poor compared with the cruise data. The cruise data itself were consequently averaged to form the mean fields. For the results presented here, then, the individual cast data were linearly interpolated onto a uniform latitudinal spacing of 0.3° and also averaged in the vertical at 20 dbar pressure intervals. Mean fields were formed by averaging the results from the five cruises.

The resulting sections of potential temperature and salinity for each of the cruises are shown in Figures 5.13–5.17, (a) and (b) respectively. Contours of potential density are overlaid. There is significant variability apparent in these sections. Cyclonic features are clearly seen at around 48°S in cruise AU9407 and 46°S in AU9309, for instance; while warm anomalies are evident at around 48°S and 50°S in cruise AU9601.

It seems clear with each of these features, qualitatively at least, that potential density surfaces follow very closely the tracer fields. From visual inspection of the sections it appears that these anomalies correspond to more or less uniform vertical translation. The only place this breaks down is in the upper layers, where surface mixing and buoyancy forcing are active. Water masses and potential density are not expected to be conserved near the surface.

To examine the model more quantitatively, equivalent vertical displacements of isopycnals were calculated pointwise down each water column assum-

ing small displacements from the mean, so that

$$\Delta\rho_{pot}(z) = -L\frac{\partial\overline{\rho}_{pot}}{\partial z}$$

and

$$L = -\frac{\Delta\rho_{pot}}{(\partial\overline{\rho}_{pot}/\partial z)} = -\left(\frac{\alpha\Delta\theta - \beta\Delta S}{\alpha\overline{\theta}_z - \beta\overline{S}_z}\right).$$

This pointwise estimate is shown for each cruise in panel (c) of Figures 5.13–5.17. The calculation was not performed in the upper 300 m. It is immediately apparent that isopycnal displacement is very coherent throughout a water column.

Next, a best-fit uniform displacement of density surfaces below 300 m was calculated for each water column using least-squares. The estimated displacements are shown in Figures 5.13–5.17 (d). Panel (e) in the figures shows, for the density anomalies in each water column, the variance explained with this uniform displacement model. The solid green line in the plots corresponds to the minimum variance explained that is significant at the 99% confidence level ($p < 0.01$). In most locations, over 80% of the observed density variance in a water column can be explained with the uniform displacement model. Typically the fit becomes worse at high latitudes and often over the Southeast Indian Ridge.

The sea-level dynamic height anomaly along each cruise track was calculated and is shown in panel (f) of Figures 5.13–5.17 for both the actual hydrography (red line) and inferred from the best-fit uniform displacement (green line):

$$\Delta\eta = -L\int_{-H}^{-300} \alpha\overline{\theta}_z - \beta\overline{S}_z dz.$$

The correspondence between the two is extremely good. As mentioned in section 5.1.1, De Mey (2002) pointed out that for a reduced-order optimal interpolation scheme, the observability of the null space of the order-reduction transformation should be marginal in order not to affect the observational error covariance. It is clear that this is satisified here; density anomalies that are not uniform with depth make only a minor contribution to the sea-level anomaly compared with the uniform displacement mode.

Finally, the steric sea-level anomalies calculated above are compared with that observed by satellite altimeter (blue lines on panel (f) of the figures) synoptic with the cruises. For this, the gridded altimetry data produced by the French CLS laboratory were used[1]. These products use both TOPEX/POSEIDON and ERS-1/2 data. The orbits are corrected using crossover differences between the two satellites (Le Traon et al. 1995; Le Traon and Ogor 1998), and the resulting track data are gridded using a sub-optimal space and time interpolation method (Le Traon et al. 1998). The spatial correlation function has a zero crossing which decreases with latitude from 250 km at

[1]The altimeter products were produced by the CLS Space Oceanography Division as part of the European Union Environment and Climate project AGORA (ENV4-CT9560113) and DUACS (ENV4-CT96-0357).

14° to 90 km at 60°. The temporal correlation function is a gaussian with an e-folding time of 15 days. Mapped data are available every ten days on a 0.25° grid. These maps were linearly interpolated in space and time onto the cruise track to obtain the corresponding observed sea-level anomaly. No further adjustments, such as removing an average bias, were made. There is again a high correlation between the observed sea-level anomaly and the dynamic height anomaly in each of the sections. However there are also important differences, indicating bottom pressure anomalies and barotropic variability in the ocean flow. Most of the barotropic signal occurs south of around 54°S where the stratification is weaker, but there are also significant differences in more northern locations. The existence of a significant barotropic component of sea-level anomaly is consistent with the results of both previous modelling studies and bottom pressure measurements, described in section 4.2.3. Almost without exception along these sections, the barotropic component of the sea-level anomaly is the same sign as the baroclinic component, suggesting that the simple partitioning scheme for sea-level suggested in section 5.2.2 may be appropriate for altimetry assimilation.

5.3 The SOFA package

A reduced-order optimal interpolation assimilation system was implemented for the HOPE model using the System for Ocean Forecasting and Analysis (SOFA) developed by De Mey (2002). SOFA is a general-purpose FORTRAN code for implementing reduced-order optimal interpolation (ROOI) in a variety of ocean models. It supports 2-d to 3-d univariate projection schemes and calculates assimilation statistics. There are a number of configurable parameters for tuning aspects of the assimilation process.

SOFA was used by Gavart et al. (1999) in an assimilation study of the Azores Current region. It has also been used extensively by the developers in numerous other assimilation projects (including their contribution to the European global assimilation project AGORA (De Mey et al. 1999) and the MERMAIDS and MAST Mediterranean projects (De Mey and Benkiran 1998)). It has also been proposed for the French contribution (De Mey 1998) to the Global Ocean Data Assimilation Experiment (Smith and Lefebvre 1998).

A schematic of the SOFA ROOI algorithm is shown in Figure 5.18. A main assimilation loop iterates over successive analysis periods, T_{int}, during which the model is integrated and (data–model) misfits accumulated. At each analysis time, a loop over each model gridpoint in the horizontal is performed; local misfits are selected and weighted by the appropriate gain matrix to form a sea-level analysis for that gridpoint. The sea-level analysis is then projected onto subsurface model variables using the methods of section 5.2, and the next assimilation cycle begins.

Table 5.5 lists the various statistics calculated by SOFA for each assimilation period. Of particular interest are the following:

- `NMS(MISFIT)`: The normalised mean square misfit provides a measure of the (data–model) misfits relative to the data. This will become less than

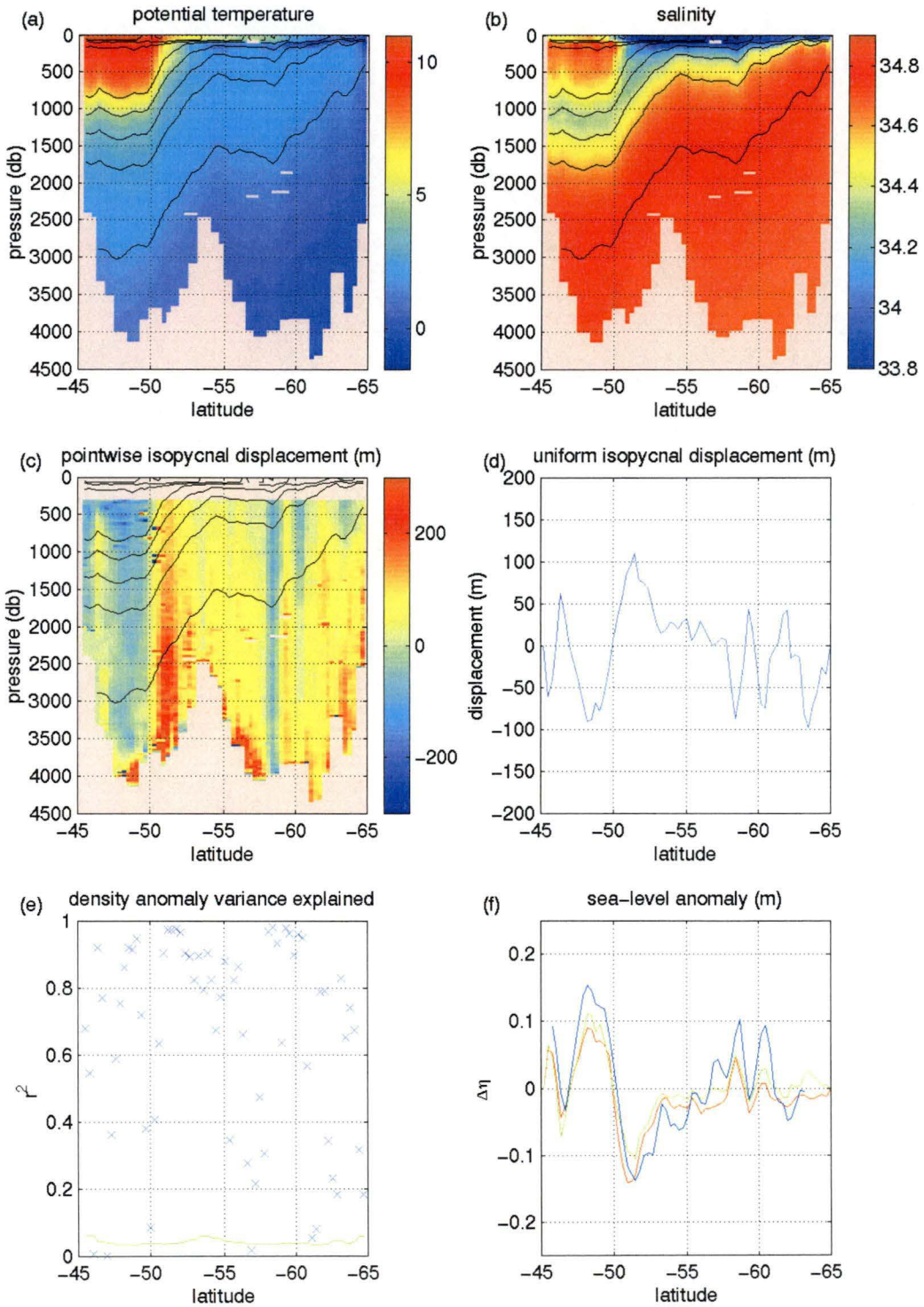


Figure 5.13: *RSV Aurora Australis* SR3 cruise AU9309. Hydrographic sections of (a) potential temperature and (b) salinity; (c) point-wise, and (d) least-squares fit of vertical displacement; (e) density anomaly variance explained with uniform displacement; and (f) sea-level anomaly (red: hydrography, green: estimated with uniform displacement, blue: altimetry)

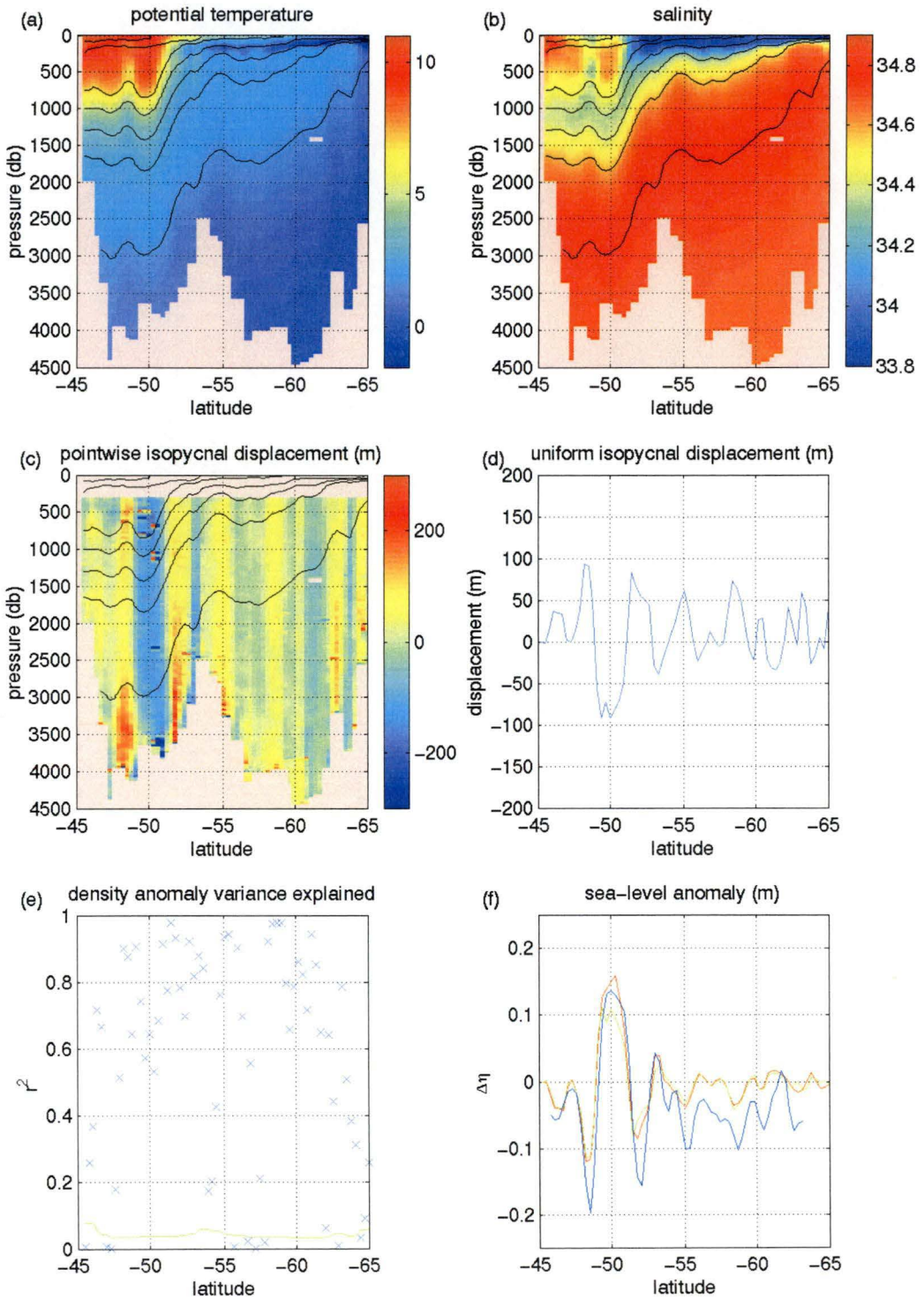


Figure 5.14: *RSV Aurora Australis* SR3 cruise AU9407. Hydrographic sections of (a) potential temperature and (b) salinity; (c) point-wise, and (d) least-squares fit of vertical displacement; (e) density anomaly variance explained with uniform displacement; and (f) sea-level anomaly (red: hydrography, green: estimated with uniform displacement, blue: altimetry)

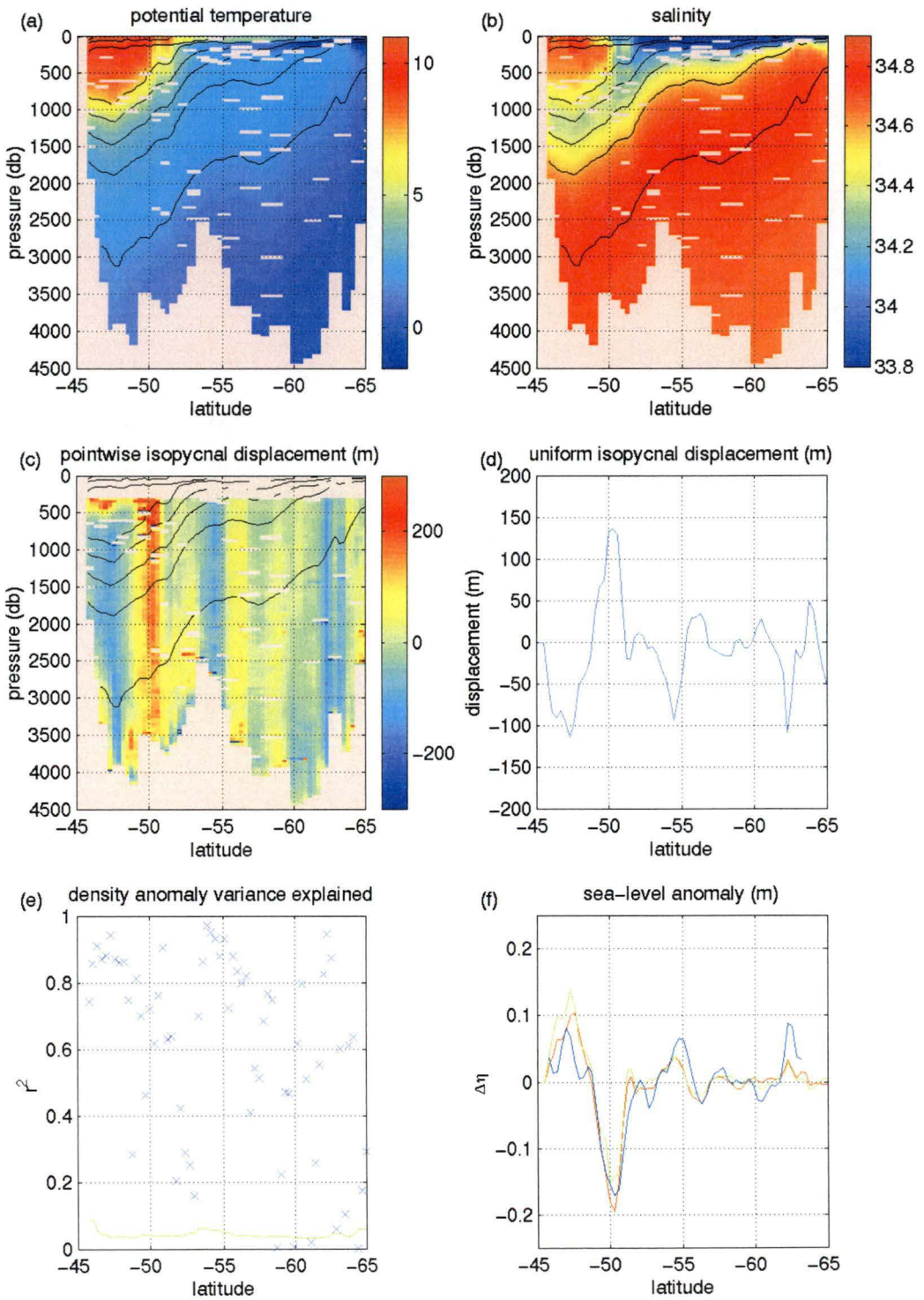


Figure 5.15: *RSV Aurora Australis* SR3 cruise AU9404. Hydrographic sections of (a) potential temperature and (b) salinity; (c) point-wise, and (d) least-squares fit of vertical displacement; (e) density anomaly variance explained with uniform displacement; and (f) sea-level anomaly (red: hydrography, green: estimated with uniform displacement, blue: altimetry)

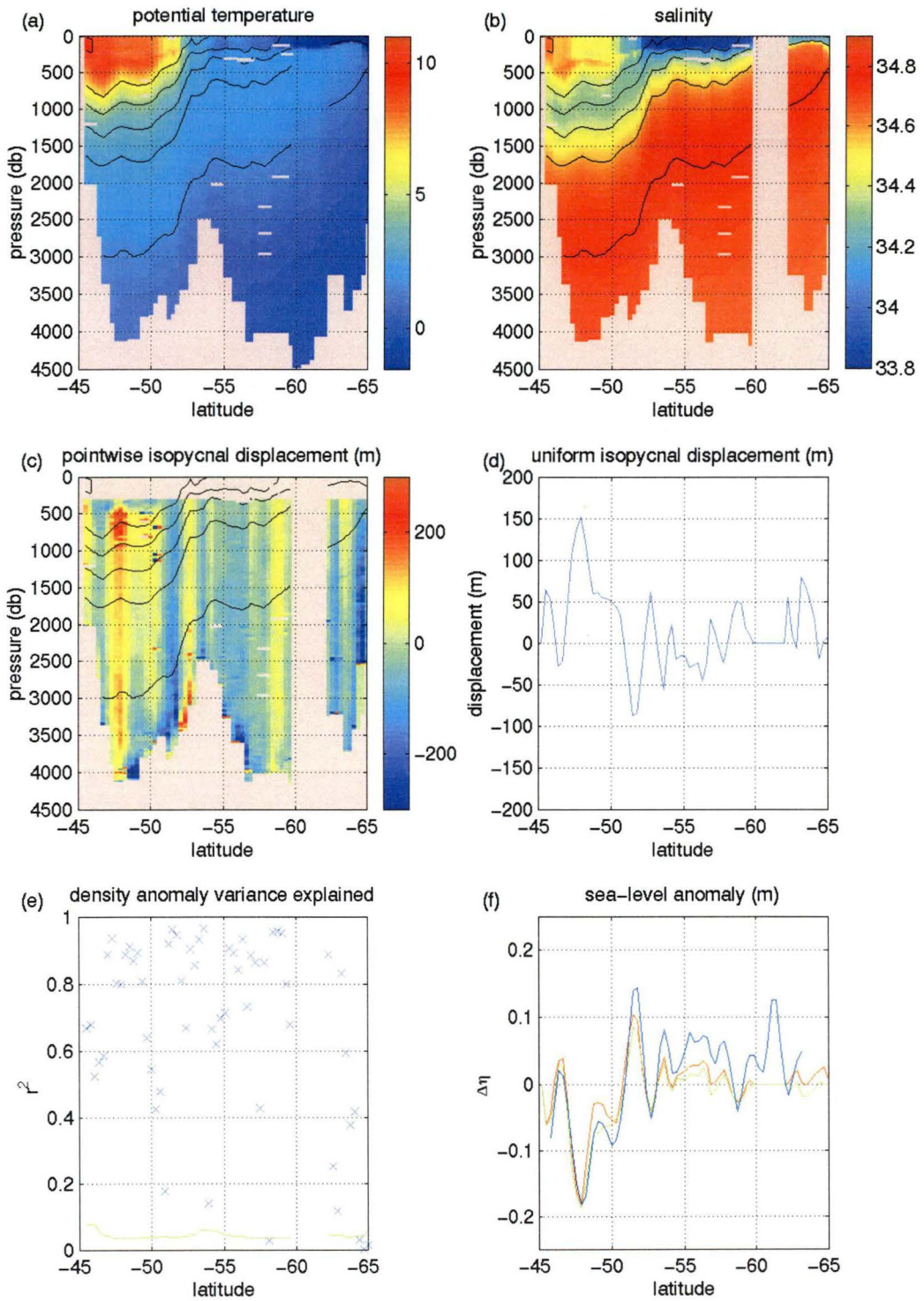


Figure 5.16: *RSV Aurora Australis* SR3 cruise AU9501. Hydrographic sections of (a) potential temperature and (b) salinity; (c) point-wise, and (d) least-squares fit of vertical displacement; (e) density anomaly variance explained with uniform displacement; and (f) sea-level anomaly (red: hydrography, green: estimated with uniform displacement, blue: altimetry)

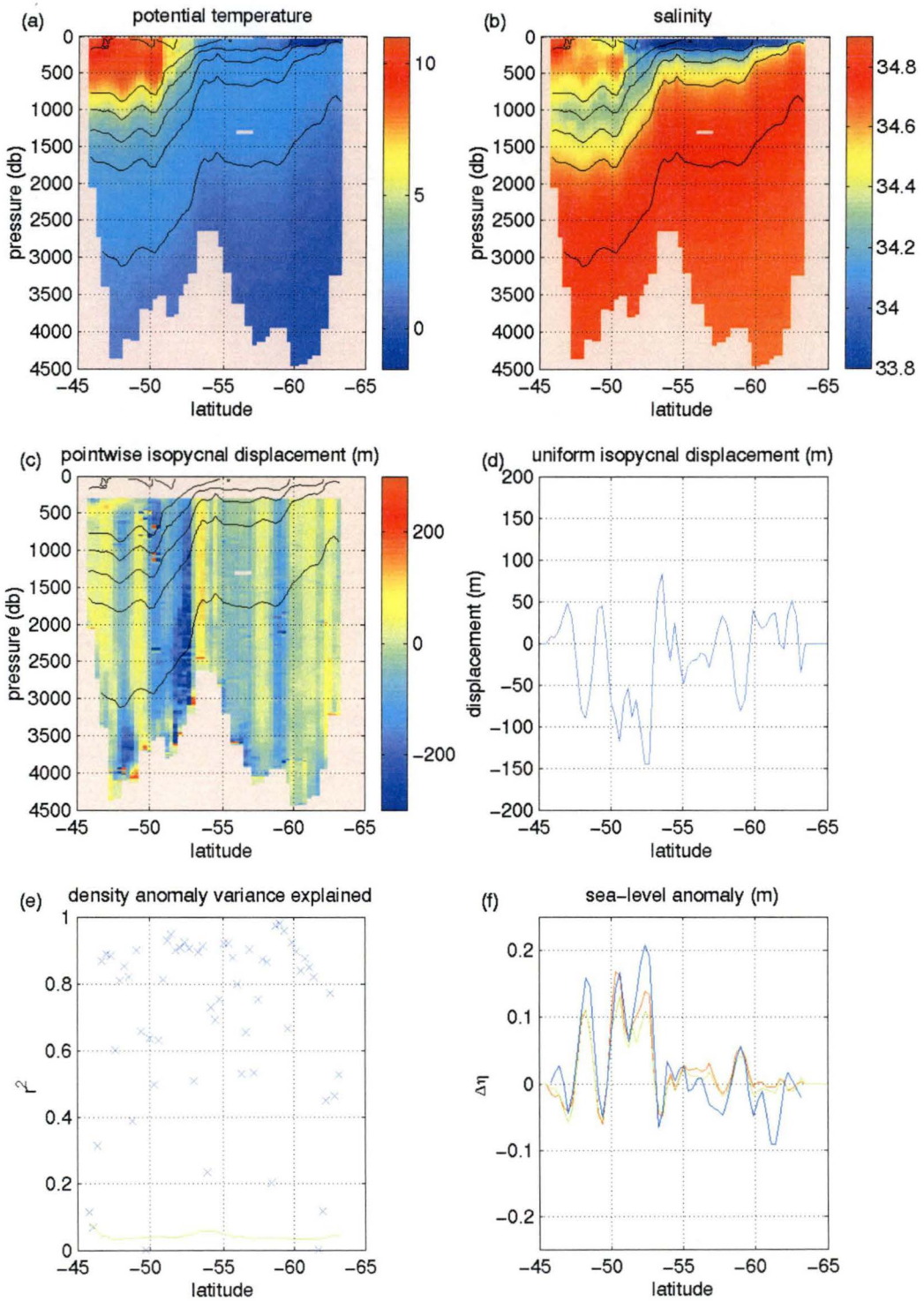


Figure 5.17: *RSV Aurora Australis* SR3 cruise AU9601. Hydrographic sections of (a) potential temperature and (b) salinity; (c) point-wise, and (d) least-squares fit of vertical displacement; (e) density anomaly variance explained with uniform displacement; and (f) sea-level anomaly (red: hydrography, green: estimated with uniform displacement, blue: altimetry)

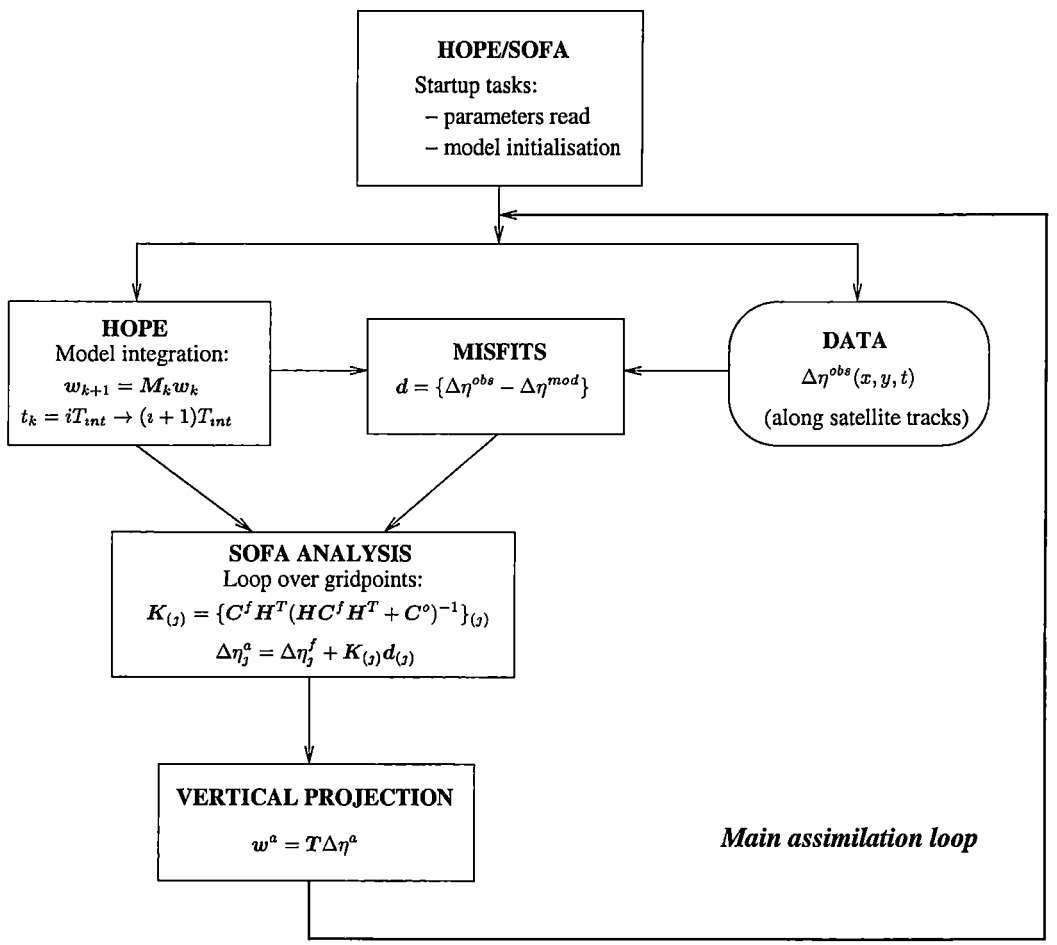


Figure 5.18: Schematic of SOFA ROOI assimilation process.

Table 5.5: Statistics calculated by SOFA for each assimilation cycle, $t = t_0 \rightarrow t_0 + T_{int}$. For the last three statistics, the calculation is performed only over the second half of the assimilation period (from $t = (t_0 + T_{int}/2) \rightarrow (t_0 + T_{int})$).

statistic	*description*	*definition*
`MS(DATA)`	mean square of the observed data	$\overline{(\Delta\eta^{obs})^2}$
`MS(MISFIT)`	mean square misfit	$\overline{(\Delta\eta^{obs} - \Delta\eta^{mod})^2}$
`NMS(MISFIT)`	normalised mean square misfit	`MS(MISFIT)/MS(DATA)`
`AVR(MISFIT)`	average misfit	$\overline{\Delta\eta^{obs} - \Delta\eta^{mod}}$
`MX(ABS(MISFIT))`	maximum absolute misfit	$\max(\lvert\Delta\eta^{obs} - \Delta\eta^{mod}\rvert)$
`MS(PERS.MIS)`	mean square persistence misfit	$\overline{(\Delta\eta^{obs} - \Delta\eta^{mod}_{t_0})^2}$
`FCST/PERS`	ratio of forecast, to persistence, error	`MS(MISFIT)/MS(PERS.MIS)`
`TEND.CORR`	tendency correlation	$corr(\Delta\eta^{mod} - \Delta\eta^{mod}_{t_0}, \Delta\eta^{obs} - \Delta\eta^{mod}_{t_0})$

one if the assimilation process reduces the misfit over time.

- FCST/PERS: The error of persistence is the mean square misfit between the model sea-level at the beginning of the assimilation cycle, and synoptic altimetry through the assimilation cycle. It represents the error in the absence of any knowledge of ocean dynamics. The statistic FCST/PERS is the ratio of the actual model error to the error of persistence. If assimilation improves the model's tracking of the real ocean, then this will become smaller than one over time. It is a measure of assimilation impact on model forecast skill.

- TEND.CORR: The tendency correlation provides a measure of the extent to which the model and observations "go in the same direction". If the model is representing the dynamics of the real ocean, then this will tend towards one.

The version of SOFA used in this study allows the statistics to be calculated over arbitrary subregions of the domain, as well as the whole region.

As well as real altimeter data, SOFA supports the use of 'bogus' observations. These are artificial observations that may be used in certain circumstances to stabilise an assimilation run. They include interpolated values, and observations of zero (altimeter sea-level anomaly), or zero misfit between model and data. The times of the bogus observations are always taken to be analysis time. Bogus observations of zero sea-level misfit were made in the assimilation experiments here at all model locations with water depth shallower than 300 m. This ensured that analyses of sea-level anomalies remained small in shallow coastal regions. Preliminary experiments indicated that very large, unphysical and unstable, vertical displacements of the water column could result under the vertical projection scheme of section 5.2 if the assimilation analyses were not so constrained.

Table 5.6 lists some of the key assimilation parameters that can be configured in SOFA. Variations of some of these form the basis of the suite of assimilation experiments presented in the next section. Some of these parameters are now described in more detail.

Forecast error covariance

As mentioned in section 5.1, reduced-order optimal interpolation replaces the optimal Kalman filter procedure for evolving the forecast error covariance, $\boldsymbol{C}^f$, with a scheme that maintains a fixed correlation structure $\boldsymbol{C}$:

$$\boldsymbol{C}^f = (\boldsymbol{D}^f)^{\frac{1}{2}} \boldsymbol{C} (\boldsymbol{D}^f)^{\frac{1}{2}}.$$

The forecast error variances are captured with the diagonal matrix $\boldsymbol{D}^f$. SOFA provides a choice of the following four functional forms (parameter ICMODEL(1)) for the spatial correlation structure:

1. Delta function: $C(r,t) = \delta(r)\delta(t)$

Table 5.6: Configurable assimilation parameters in SOFA. They are grouped into classes associated with assimilation cycle, data preconditioning, forecast error covariance, observational error covariance, and data selection.

parameter	*description*
TINT	length of assimilation cycle
IFREQ	decimation frequency for altimetry data
ICMODEL(1)	provides a choice of four functional forms for the *a-priori* correlation structure in the forecast error covariance
TAU	controls the error propagation scheme for the forecast error covariance
FNOISE	forecast/model error variance
RCX1/2(1)	zonal and meridional length scales for forecast error covariance
RCT(1)	*e*-folding time for forecast error covariance
ICMODEL(2)	provides a choice of four functional forms for the *a-priori* correlation structure in the observational error covariance
ENOISE	observational error variance
RCX1/2(2)	zonal and meridional length scales for observational error covariance
RCT(2)	*e*-folding time for observational error covariance
RIX1/2	local misfit selection radii
RIT	local misfit selection window in time
NSEL	maximum number of data points selected for analysis at each model gridpoint
RATIO	"pruning" parameter used to eliminate observations that fall within the local space/time window but which are particularly close together

2. Non negative lobe: $C(r,t) = [1 + ar + \frac{1}{3}(ar)^2]e^{-ar}e^{-t^2}$, where $a = 2.103803$.

3. Small negative lobe: $C(r,t) = [1 + br + \frac{1}{6}(br)^2 - \frac{1}{6}(br)^3]e^{-br}e^{-t^2}$, where $b = 3.336912$. The zero-crossing occurs at $r = 1$.

4. Large negative lobe: $C(r,t) = [1 + ar - \frac{1}{3}(ar)^3]e^{-ar}e^{-t^2}$, where $b = 2.103803$. The zero-crossing occurs at $r = 1$.

The parameters r and t are distances in space and time, normalised by the length scales `RCX1/2(1)` and timescale `RCT(1)`.

A simplified error propagation scheme is provided in SOFA for the forecast error variances:

$$D^f(\boldsymbol{x}) = \tau C^a(\boldsymbol{x}) + Q(\boldsymbol{x}),$$

where C^a is the error variance of the previous analysis (calculated according to equation (5.5)), Q is specified with the SOFA parameter `FNOISE` to be a fixed fraction of the observational data variance in the corresponding assimilation period, and τ (SOFA parameter `TAU`) determines the temporal propagation of the previous analysis error. Comparing this equation with the Kalman filter formula for optimally evolving the forecast error covariance (equation (1.7)), Q can be regarded as an analogue to the model error variance[2] in the case $\tau \neq 0$, while τ itself replaces the model dynamics in propagating the previous analysis error. If $\tau = 0$, then Q specifies the forecast error variance to be a constant fraction of the altimetry variance in each assimilation cycle.

Observational error covariance

SOFA allows the observational error covariance to be specified in a similar manner to the forecast error covariance, with a fixed correlation structure and specified variance:

$$\boldsymbol{C}^o = (\boldsymbol{D}^o)^{\frac{1}{2}}\boldsymbol{C}(\boldsymbol{D}^o)^{\frac{1}{2}}.$$

The correlation function used in $\boldsymbol{C}$ can be specified (with parameter `ICMODEL(2)`) as any of the four forms described above for the forecast error covariance while the observational error variance, $D^o(\boldsymbol{x}) = D^o$, is spatially uniform and specified (in parameter `ENOISE`) as a fixed fraction of the altimetry data variance in each assimilation period. The length scales for the error correlation are provided in parameters `RCX1/2(2)` while the e-folding time is specified with the parameter `RCT(2)`.

Data selection

Section 5.1.2 discussed horizontal order reduction as a technique for reducing the size of OI methods. Two approaches are generally used to achieve this. First, only data points within some localised region in time and space are used to form the optimal analysis at any given gridpoint. Second, the selected data

[2]In SOFA, this is not specified absolutely, but rather as a fixed fraction of the altimetry variance in each assimilation period.

points are pruned even further by eliminating redundant data that otherwise only increase the computational burden.

In SOFA, data are selected within a specified space/time window around each model gridpoint at analysis time. The spatial selection radii are specified with the parameters `RIX1/2` for the zonal and meridional directions respectively, while the time window around each analysis is specified with the parameter `RIT`. While SOFA may be configured to operate in a filter mode that uses future observations (with respect to analysis time), the assimilation experiments performed here only used past data to perform the analysis.

The algorithm used by SOFA for further reducing the selected data attempts to eliminate redundancy. The selection procedure iterates over the data points, preferentially accepting those which are closest to the analysis gridpoint. However, additionally, with each new data point that is accepted, the algorithm rejects any remaining data points that are closer to it than 1/`RATIO` times their distance to the analysis gridpoint. The procedure terminates when a maximum number (`NSEL`) of data points have been selected. Thus, the set of selected data is both minimally redundant, and close to the analysis gridpoint in time and space.

5.4 Assimilation experiments

A suite of data assimilation experiments was performed, varying different elements of the optimal interpolation algorithm as well as the vertical projection scheme. This section presents results from these experiments. Each experiment was over the one year period from June 1, 1993 through July 4, 1994. This encompassed the AU9407 cruise of *RSV Aurora Australis* along the WOCE SR3 section through the model domain (see Figure 5.14), and also overlapped the period during which an array of current meters were deployed in the region (Phillips and Rintoul 2000). As well as the various assimilation runs, an unassimilated run over the same period was performed to provide a basis for comparison.

The altimeter data used here were from the joint NASA-CNES mission TOPEX/POSEIDON (Zieger et al. 1991). This satellite was launched in August, 1992, and has been providing measurements of sea-level anomaly continuously since October, 1992. The orbit is at an inclination of 66° with a 10-day repeat cycle. Figure 4.15 showed the ground tracks over the model region.

The data used here were processed from the raw Geophysical Data Records (Benada 1993) by the Commonwealth Scientific and Industrial Research Organisation's Division of Marine Research in Hobart, Tasmania. The standard path-length (dry troposphere, wet troposphere, ionosphere) and sea-surface (sea-state bias, tidal, inverse barometer) corrections (Benada 1997) were made (*Neil White, CSIRO, personal communication (2000)*). Altimetry observations were available every 7 km along-track. To avoid any potential problematic interactions between altimetry assimilation and the open boundary conditions, data within 10° of the eastern and western boundaries, and within 5° of the northern boundary, were not used to form sea-level analyses.

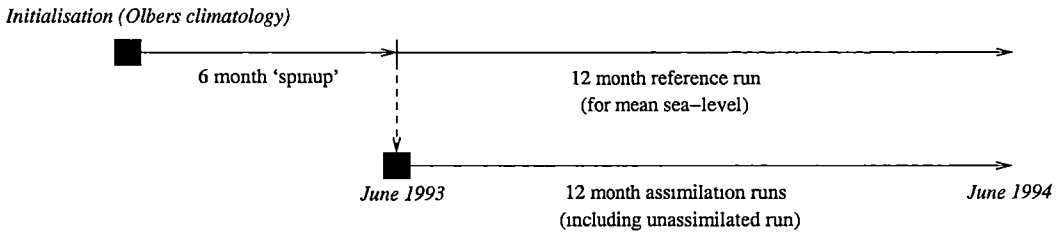


Figure 5.19: Schematic of timetable used for assimilation experiments. A six month spinup was followed by a one year reference run. The assimilation runs started from the same point, with model-equivalent sea-level anomalies calculated against the mean sea-surface height from the reference run.

Given the substantial drift of the model over the first 20 years (chapter 4), the assimilation was not performed using a fully spun-up model. The model was instead run for six months after initialisation from climatology (section 2.2). The model state at the end of this six-month period provided the initialisation for the one year assimilation experiments. A reference run also followed on for a further year. The reference sea-level for calculating the model-equivalent sea-level anomalies in the assimilation experiments was taken as the mean over this one year reference run. This assimilation 'timetable' is shown in Figure 5.19.

To examine the performance of the assimilation experiments, various diagnostics were considered, including some of the SOFA statistics described earlier, comparisons between analysed sea-level anomaly and the gridded fields produced by CLS (used in section 5.2.4, (Le Traon et al. 1998)), comparisons against the WOCE hydrographic section AU9407, and comparisons of model forecasts against persistence. As well, the impact of assimilation on meridional heatfluxes and sea-level variability was examined.

Before discussing the assimilation results, some results from the unassimilated run are first described.

5.4.1 Unassimilated run

The unassimilated run was performed exactly analagously to the respective assimilation runs, except that no sea-level updates were performed. Various diagnostis from the run were calculated to deduce the influence of data assimilation in later runs.

Assimilation statistics

The assimilation statistics described earlier (Table 5.5), calculated over the whole model domain, are plotted by month in Figure 5.20 for the unassimilated run. A nominal assimilation period of seven days was assumed (as used in the assimilation run in the next section); this was the period over which the statistics were calculated during the run.

The initial spike in RMS misfit between model and data is associated with adjustment of the model due to a discontinuity in the forcing regime.

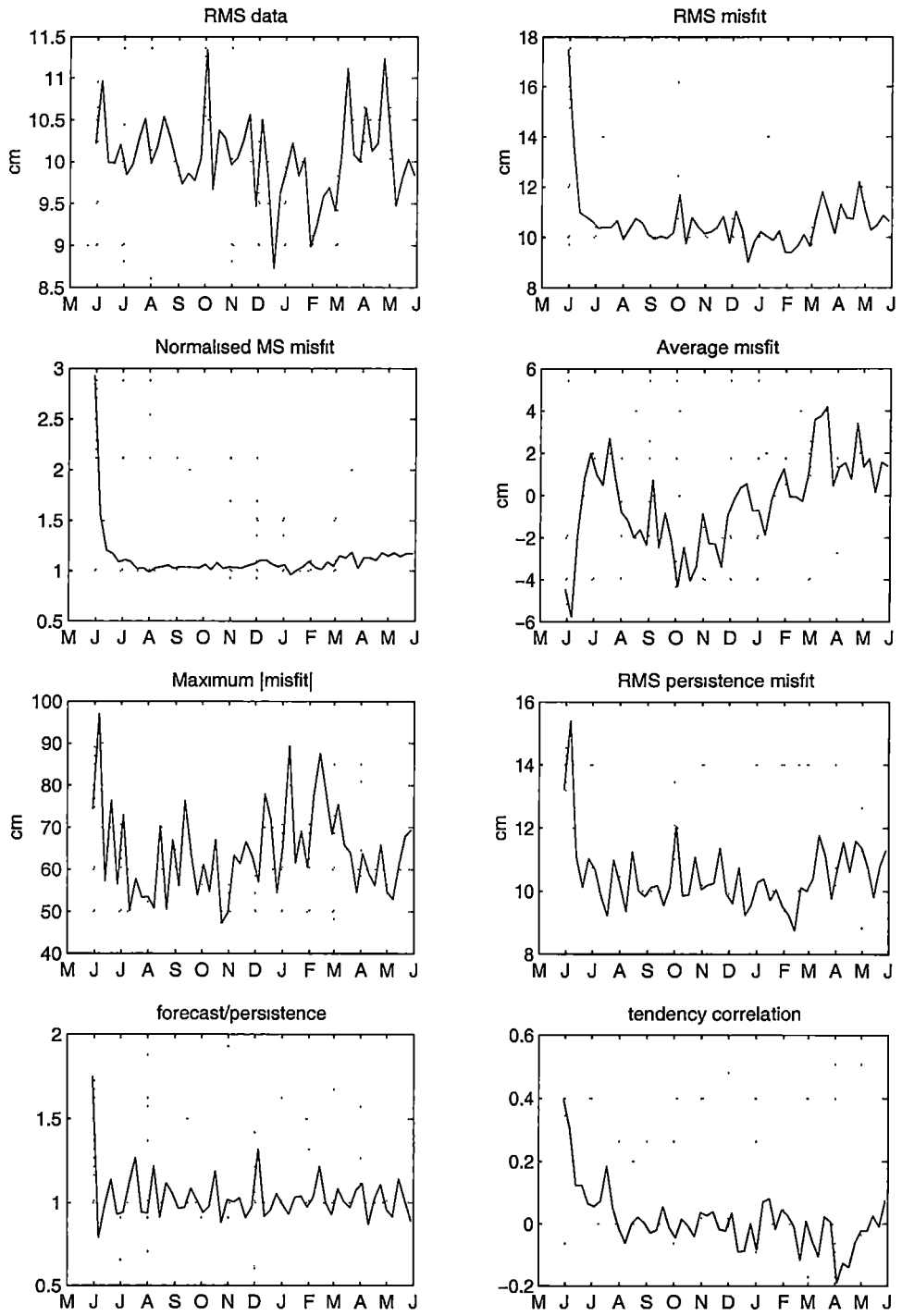


Figure 5.20: SOFA statistics for unassimilated run.

The assimilation runs were all initialised from a state after six months of a spin-up run (*i.e.* end of June, Figure 5.19). They were then integrated, however, using climatological forcing fields from the beginning of June. This one month discontinuity in surface forcing lead to large transients in sea-level during the first week or two for all the experiments presented here. Apart from this initial spike, however, Figure 5.20 shows that for the unassimilated run the data–model misfits are consistently at least as large as the observed altimetry (around 10 cm RMS) throughout the run. A non-zero average misfit may indicate model biases. There are two factors contributing to this in the unassimilated run. First, the large degree of model drift discussed in chapter 4 will lead to steady drift between the modeled and observed sea-level anomalies. This is likely to be masked, however, in the average misfit since the model is volume-conserving: a positive trend in data–model misfits at one location is balanced by a negative trend in the misfits elsewhere. From the plot of the average misfit (Figure 5.20), it is apparent that a more significant factor is a difference between the observed and modeled seasonal cycle. There is a clear annual signal indicated, with an amplitude of around 6 cm. From the remaining statistics, the model, as expected, has no forecast skill (forecast/persistence error around one), nor is its variability correlated with the observed variability (tendency correlation around zero).

Figure 5.21 shows the median values of several of these statistics mapped in $10° \times 5°$ boxes over the model domain. In this unassimilated run there is no spatial variation in the data–model misfits or forecast performance.

Sea-level variability

Sea-level variability was examined in chapter 4 during successive five-year periods in a 40 year run of the model. It is shown here (Figure 5.22) for the present one year unassimilated run. The variability bears a close resemblance to that seen in the first five-year period (of which it is a part) of the 40 year run (Figure 4.13). While the patterns of variability in these first few years are similar to that observed by altimeter (Figure 4.15), the level is considerably lower. Thus, in the unassimilated run here, sea-level variability is around 50%, or lower, of that observed.

Meridional heatflux

The mean, and standing and transient eddy components of meridional heatflux for the unassimilated run are shown in Figure 5.23. These are again essentially the same as calculated earlier in the first five year period of the 40 year spinup run (Figures 4.9 and 4.16). The transient eddy component of meridional heatflux at latitudes of the ACC (45°–60°S) is lower than expected at less than 0.02 PW for most of this latitude band (see discussion in sections 1.2 and 1.2.1). For instance, if de Szoeke and Levine's (1981) circumpolar estimate of 0.3–0.45 PW were uniformly distributed around the globe, we would expect between 0.07 PW and 0.10 PW over the model domain. If the spot observations of Phillips and Rintoul (2000) south of Australia (11 kW/m^2) were typical, an elevated figure of 0.2 PW at 50°S for this domain would be expected.

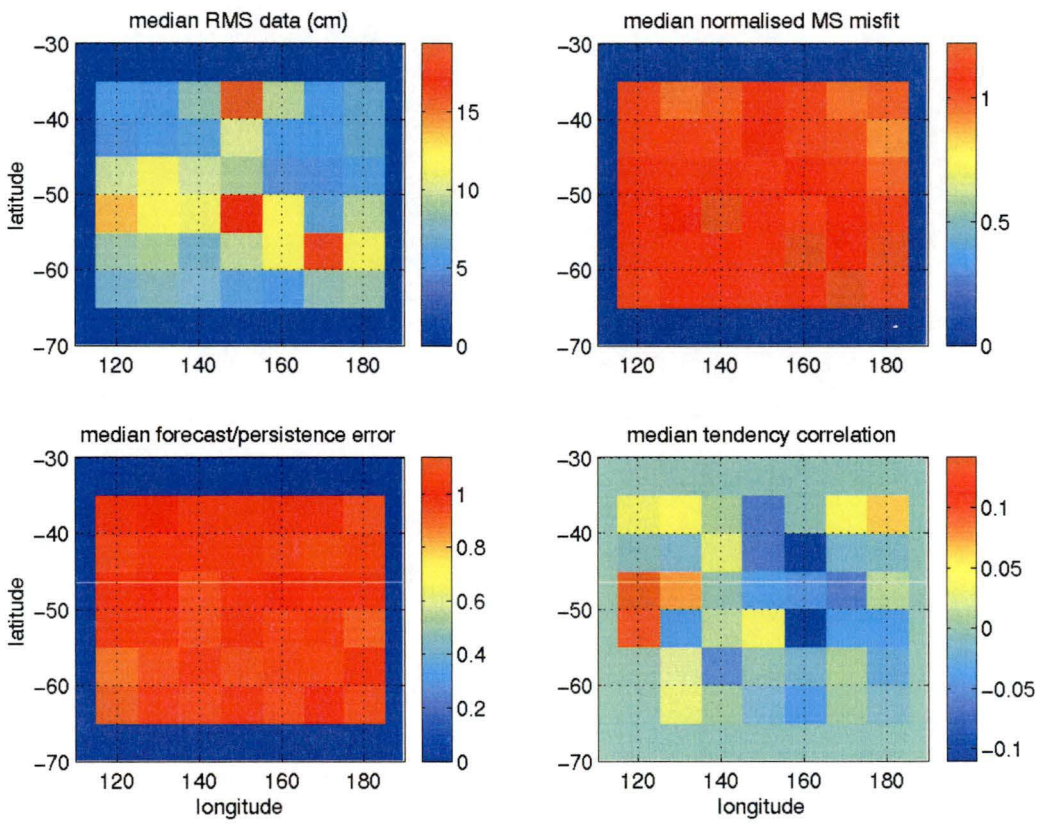


Figure 5.21: Regional statistics (median for $5° \times 10°$ boxes) for unassimilated run.

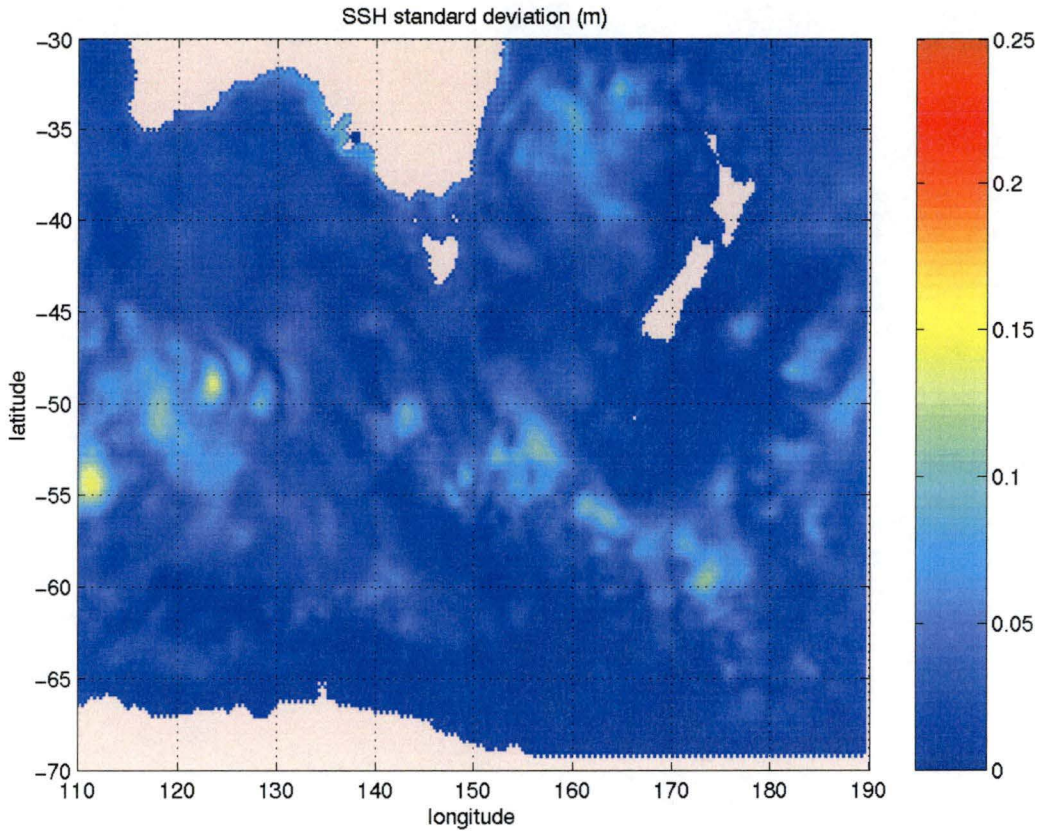


Figure 5.22: Sea-level variability for unassimilated run.

We now move on to results of the various assimilation experiments.

5.4.2 Baseline assimilation run

As with the set of experiments in chapter 3 trialling various open boundary conditions, it was impossible to perform a set of assimilation experiments covering the full parameter space of the assimilation algorithm. The approach taken was first to make an informed choice on a 'baseline' set of parameters, and then to consider a few variations away from this set.

The set of parameters chosen for the baseline assimilation run is listed in Table 5.7. These were defined earlier in Table 5.6. An assimilation period of seven days was used. Since this was different to the TOPEX/POSEIDON repeat cycle of 10 days, it ensured that the most recent observations were at different locations for different analysis times. The altimetry was averaged and decimated by a factor of three, providing observations about every 21 km along track. The forecast error correlation was taken as a positive function with a length-scale of 200 km and time-scale of 10 days. Both Stammer (1997b) and Leeuwenburgh (2000) found autocorrelation length scales of 100–200 km for sea-surface height from altimetry measurements. The observation error correlation was taken as a delta function in time and space. This correlation structure was also used in SOFA by Gavart et al. (1999) for their assimilation of altimetry in a model of the Azores Current region. The forecast and obser-

Table 5.7: SOFA parameters used for baseline assimilation experiment.

parameter	*value*	*brief description*
TINT	seven days	assimilation period
IFREQ	3	altimetry decimation
ICMODEL(1)	1	forecast error correlation function with no negative lobe
TAU	0	no propagation of analysis error
FNOISE	0.5	normalised forecast error variance
RCX1/2(1)	200 km	forecast error covariance length scales
RCT(1)	10 days	forecast error covariance time scale
ICMODEL(2)	0	delta function for observation error correlation
ENOISE	0.1	normalised observational error variance
RCX1/2(2)	–	*NA*
RCT(2)	–	*NA*
RIX1/2	300 km	spatial data selection radii
RIT	15 days	temporal data selection radii
NSEL	33	maximum number of data selected per analysis gridpoint
RATIO	9	data pruning parameter

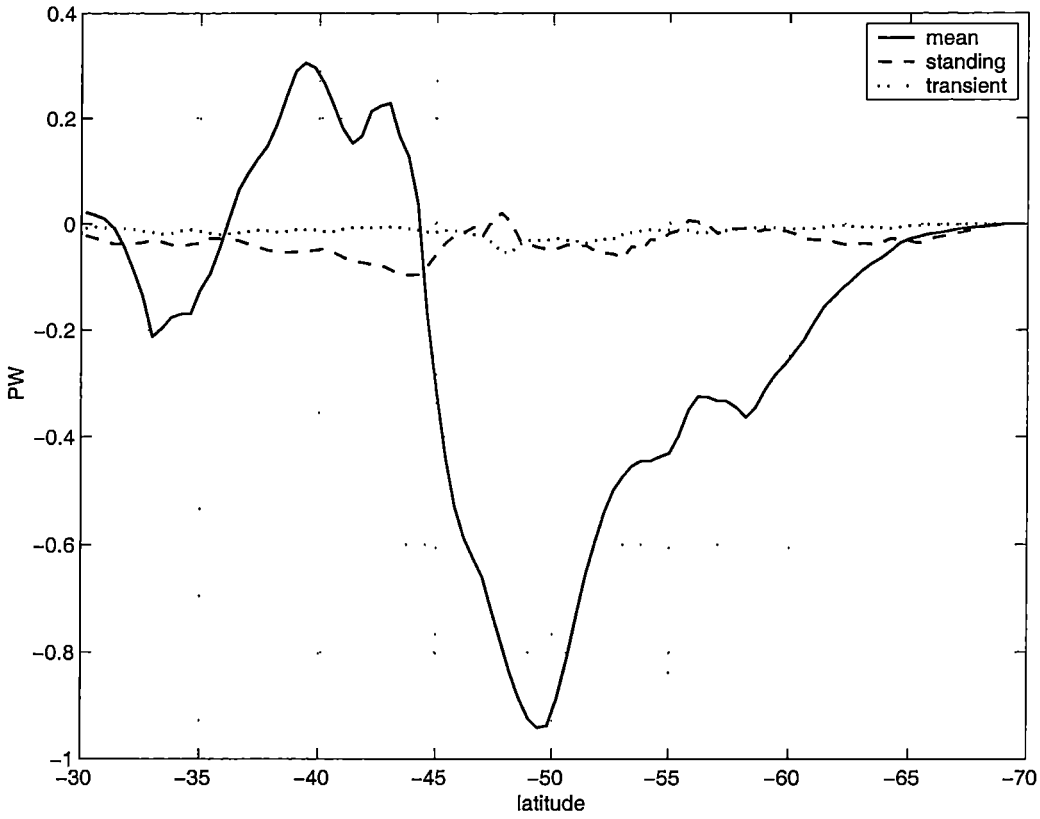


Figure 5.23: Meridional heatflux for unassimilated run.

vation error variances were taken as 50% and 10%, respectively, of the data (altimetry) variance. No propagation of analysis error for the next *a-priori* forecast error was used. (The forecast error was a spatially homogeneous field each analysis time.) Finally, the data selection procedure used a window of 300 km in space and 15 days in time, further pruning selected data to 33 data points for each analysis point. These suboptimal selection parameters were used successfully in SOFA by Gavart et al. (1999).

For the vertical projection of analysed sea-level anomaly fields, the 'purely baroclinic' scheme of section 5.2.1 was used. That is, the entire analysed sea-level correction was used to calculate a corresponding vertical displacement of the water column, with no portion used to increment barotropic velocities. For both the vertical gradients and steric expansion coefficients used in equation (5.17) to calculate the amount of displacement, climatological profiles were used. In practice, this was found to give similar results to using synoptic profiles and avoided occasional instabilities that could result from using synoptic model profiles. In addition, the projection was only performed in water below 300 m depth. As discussed earlier (section 5.2.4) uniform vertical displacement of isopycnals does not occur in the surface mixed layer. No update of model velocities was performed at analysis time, leaving the model to adjust dynamically to the modified density field. The development of the vertical projection scheme assumed a small-amplitude isopycnal displacement (section 5.2). To avoid excessively large displacements in the assimilation, the analysed

sea-level correction was clamped to a maximum amplitude of 0.5 m on the rare occasions that it was greater.

Various diagnostics from the baseline assimilation run are now discussed.

Assimilation statistics

Figure 5.24 plots the various assimilation statistics for the whole domain. The RMS data is simply the root mean square altimeter observations, and therefore remains unchanged in all assimilation runs. Importantly, it is clear that assimilation has now reduced the data–model misfit variance to below the level of the observations ('normalised MS misfit') in most assimilation periods. By examining Figure 5.25 mapping the statistics over the domain, it can be seen that assimilation has successfully reduced the misfits over most of the domain, with the exception of the high latitudes. This is where, from both a model analysis (section 4.2.3) and hydrographic data (section 5.2.4), we expect there to be a greater proportion of barotropic variability.

The seasonal signal seen in the average misfit for the unassimilated run has also been removed (Figure 5.24) — assimilation has improved the model's representation of the seasonal cycle in sea-level.

The maximum misfit over the domain has been reduced, mainly over the second half year of assimilation. Neither the model's forecast performance nor the tendency correlation have been significantly improved, however, according to these global statistics.

Sea-level variability

Figure 5.26 shows the sea-level variability for the baseline assimilation run. Both the spatial location and level of variability are much improved with respect to the unassimilated run (Figure 5.22). The standard deviation of sea-level in the regions of greatest variability have increased from around 11 cm to around 18 cm. The spatial location of these high eddy-energy regions is much closer to the raw variability seen by the altimeter (Figure 4.15). This is hardly surprising, of course, since we expect the sea-level variability observed by the altimeter to be reflected in the model during assimilation. There is, however, an unexpected high level of variability in the southwest of the domain (centred around 128°E and 58°S). Once more, this is in a region of reduced stratification where barotropic variability is expected to be significant. It is possible that forcing subsurface variability in this region to be entirely baroclinic in nature leads to unrealistic dynamics with elevated sea-level variability. It will be seen in later experiments that variability in this region is reduced by applying a barotropic projection of the sea-level analyses. Finally, the effect of limiting altimetry observations to the interior away from the open boundaries is clear; the boundary regions maintain low levels of variability in the assimilation run.

Comparison of SOFA analysis with CLS gridded sea-level anomaly

To examine in more detail the sea-level analysis produced by the SOFA scheme, Figure 5.27 compares the sea-level analysis produced for 26 December 1993

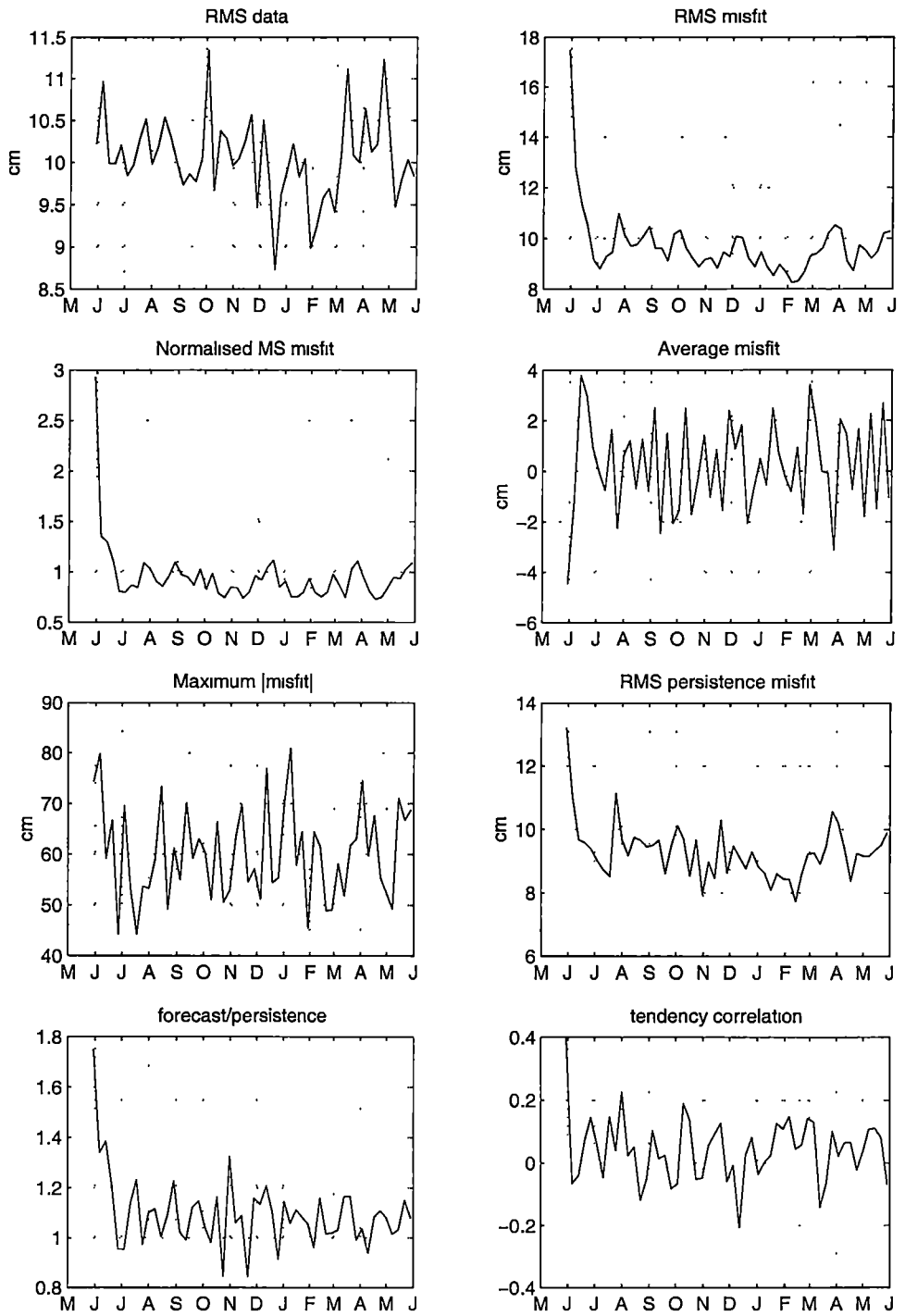


Figure 5.24: Assimilation statistics for baseline assimilation run.

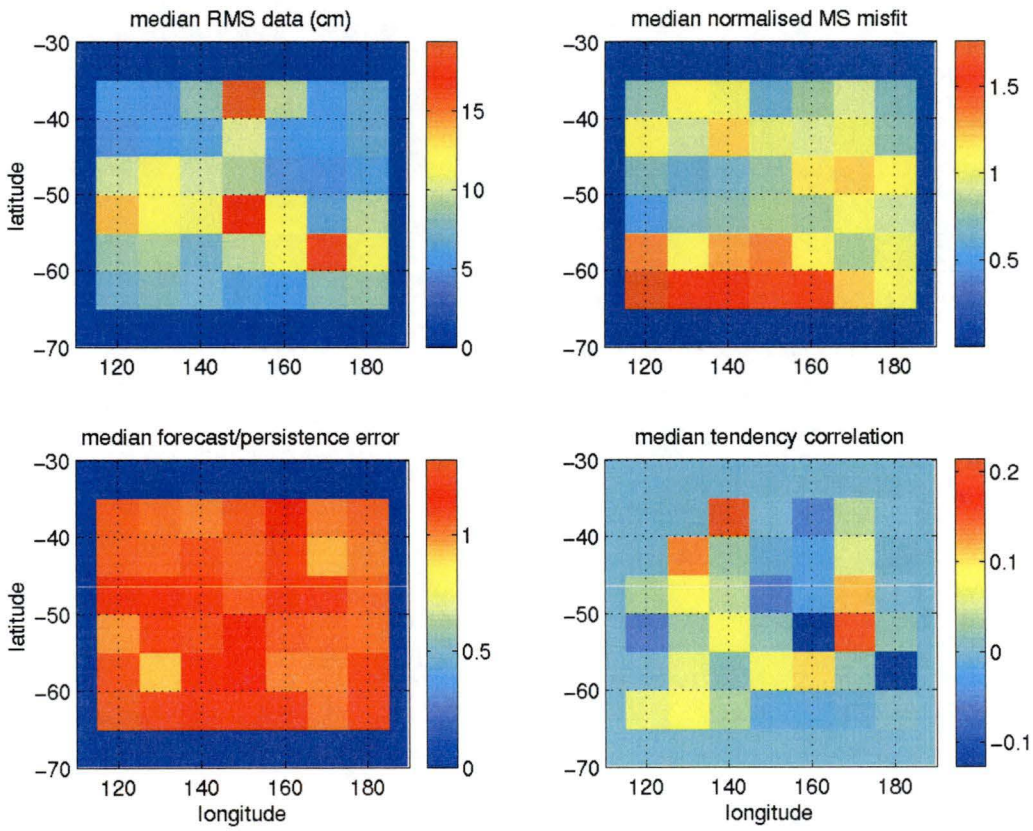


Figure 5.25: Regional statistics (median for $5° \times 10°$ boxes) for baseline assimilation run.

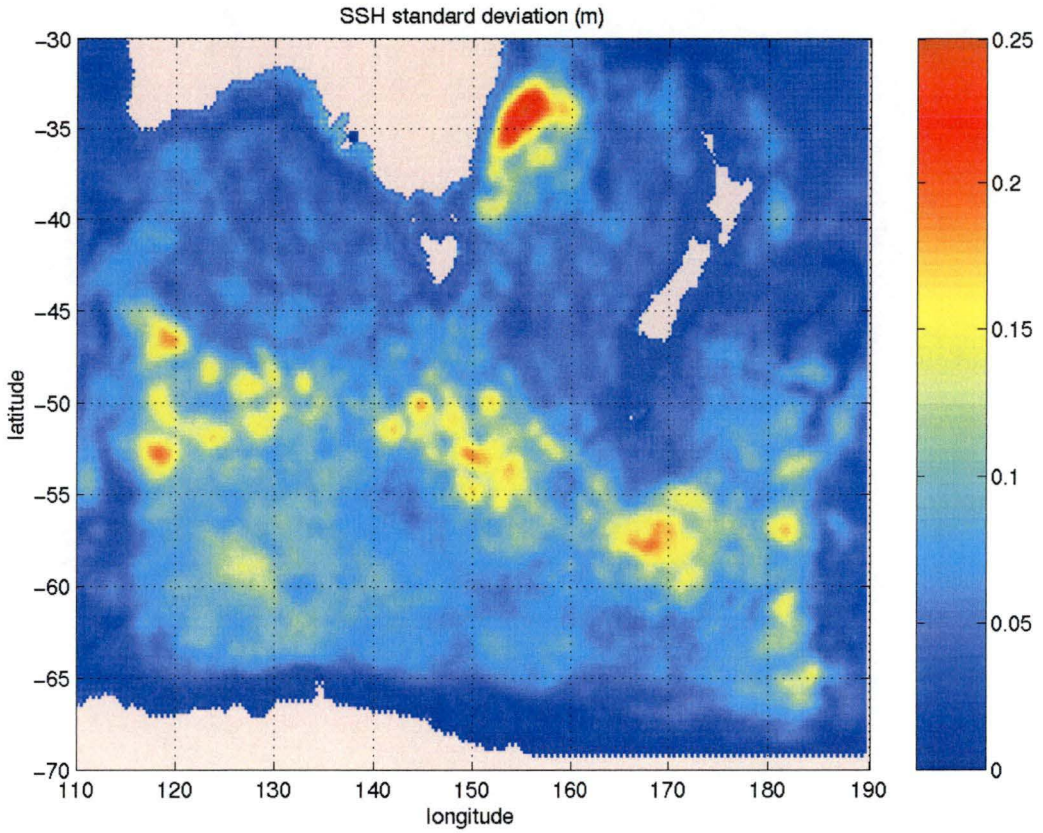


Figure 5.26: Sea-level variability for baseline assimilation run.

with the gridded field produced by the French CLS laboratory (Le Traon et al. 1998). These gridded fields were used earlier (section 5.2.4) to compare the steric sea-level anomaly along repeats of the WOCE SR3 section with that observed with altimetry. This particular date was chosen because there happened to be both a CLS gridded field and an analysis in the current assimilation experiment available at the same time, and because it is also close to the time of the WOCE SR3 cruise AU9407 (Table 5.4). The respective panels of the figure show the model forecast (*i.e.* just prior to analysis) sea-level anomaly (top left), the analysed correction from SOFA (top right), the sum of these (bottom left), and the corresponding CLS gridded sea-level anomaly (bottom right). The model sea-level anomaly is with respect to the mean sea-level used throughout the assimilation experiments, taken from the mean of the one year reference run (Figure 5.19). The CLS field should be compared with the total *a-posteriori* sea-level anomaly from the SOFA analysis (bottom left). It is evident that the analysed correction has markedly improved the model sea-level forecast with respect to the CLS data. This is most evident with the large positive sea-level anomaly located around (152°E, 54°S), and also off the east coast of Australia. In both of these locations SOFA has produced a sea-level correction which drives the model sea-level anomaly much closer to that estimated in the CLS gridded fields. The CLS fields use future, as well as past, altimetry data and combine ERS-1 observations with TOPEX/POSEIDON. The inclusion of ERS-1 observations, with smaller track

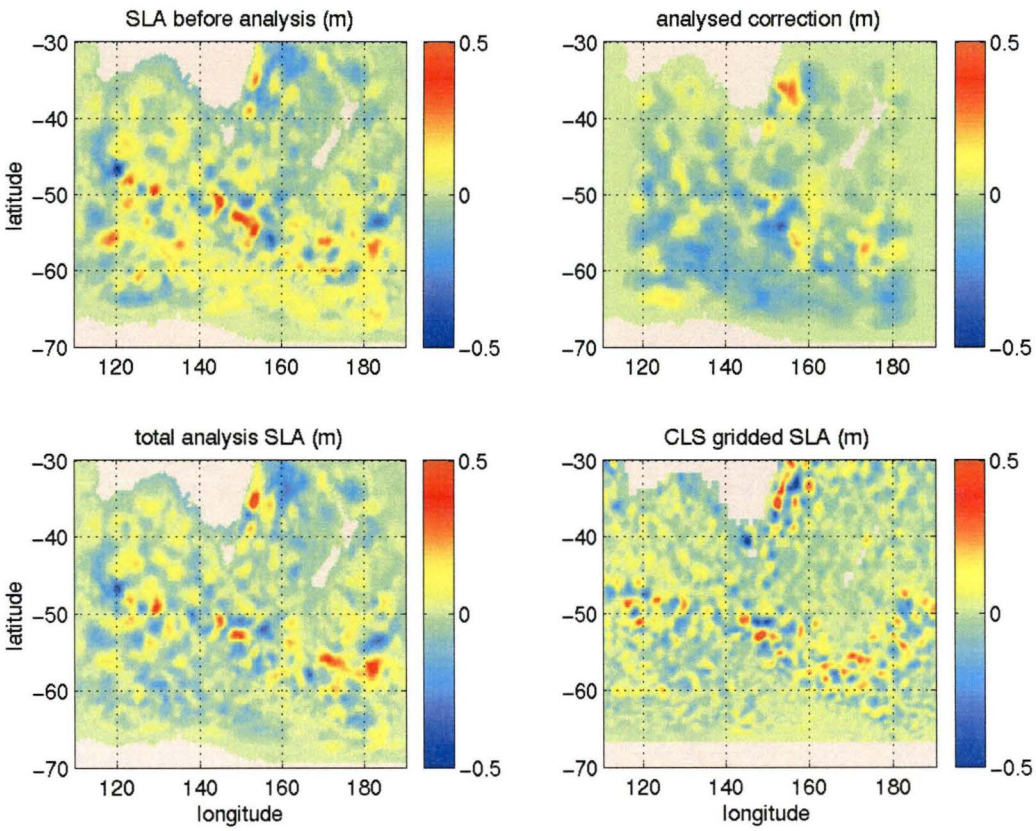


Figure 5.27: SOFA sea-level analysis (baseline assimilation run) compared with CLS gridded sea-level anomaly for 26 December 1993. (Top left) model forecast, (top right) SOFA sea-level correction, (bottom left) total sea-level anomaly, (bottom right) CLS gridded sea-level anomaly.

spacing than TOPEX/POSEIDON, permits mapped data with smaller spatial scales than possible with TOPEX/POSEIDON alone. The suboptimal interpolation method used by CLS employs a correlation function with a length scale that decreases from 250 km at latitude 14° to 90 km at 60° (Le Traon et al. 1998). This is smaller than the forecast error correlation function used in the baseline assimilation run here, which had a length scale of 200 km throughout the domain. The CLS analysis shown in Figure 5.27 does indeed have smaller spatial scales than the corresponding SOFA analysis. The CLS maps represent one of the best suboptimal sea-level analysis products available and have been used, for example, by Fox et al. (2000b) for assimilation in place of a scheme like SOFA. Comparison with the SOFA analyses provides an important independent validation of the present results. While the correspondence between the two in Figure 5.27 is quite good, there are also significant differences. For instance, the large negative anomaly in Bass Strait in the CLS field is completely absent in the SOFA analysis. This is most likely a result of the bogus observations used at locations shallower than 300 m in these assimilation experiments. A sea-level correction of several tens of centimetres in Bass Strait leads to vertical displacements in the 'baroclinic' projection scheme of several hundred metres, producing unphysical values of temperature and salinity, and causing model failure. Bogus 'no-change' observations were used at all locations shallower than 300 m to prevent this happening. From the analysis of subsurface variability (section 4.2.3) it was concluded that variability in Bass Strait was almost exclusively barotropic, and so a more sophisticated approach would apply a barotropic projection of analysed sea-level correction in this region, rather than artificially constraining it to be small. On the other hand, there are likely to be large errors in any altimeter measurements in Bass Strait due to inaccuracies in the tidal and path-length corrections applied to the raw altimeter data in this region (White et al. 1999; Lefevre et al. 2000a; Lefevre et al. 2000b). Artificially constraining the SOFA analysis to be small in Bass Strait using bogus observations is reasonable under these circumstances.

Vertical projection of analysis

As well as examining the sea-level analysis itself, it is important to examine whether the vertical projection scheme successfully introduces the sea-level analysis into the model. From the idealised perturbation experiments of section 5.2.3 we expect that this should be the case.

Figure 5.28 shows, for the same analysis time as above (26 December 1993), the analysed sea-level correction and its estimated error, together with the actual model sea-level difference from the forecast at four and 24 hours after the assimilation. It can be seen that the model has successfully retained most of the features of the analysis after one day. From the results in section 5.2.3 we have seen already that an idealised sea-level anomaly is retained for up to 10 days by the vertical projection scheme used in this experiment. There are some differences between the analysed correction and the post-assimilation model sea-level, however. In general, the retained sea-level anomaly is smaller than the analysed correction. There is also some variation between four hours

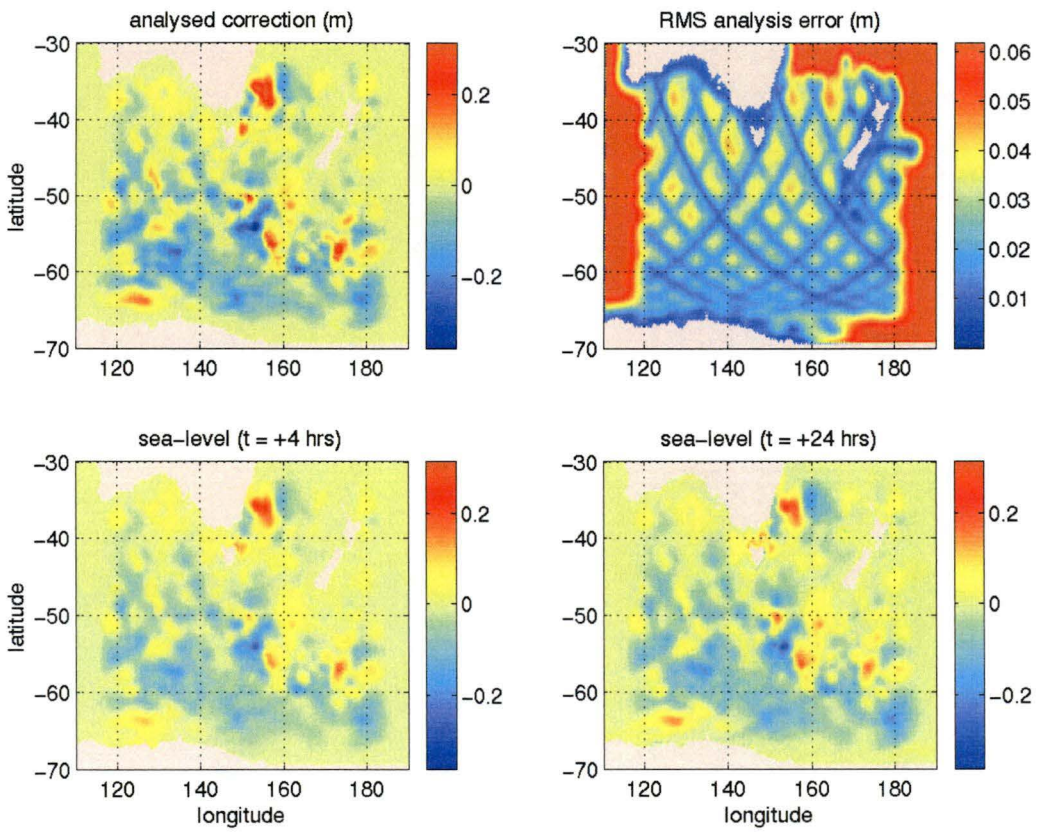


Figure 5.28: Analysed sea-level correction (top left) and RMS analysis error (top right) for baseline assimilation run. Post-analysis difference from forecast at four hours (bottom left) and 24 hours (bottom right) after assimilation.

and 24 hours after the assimilation. Both of these are expected from the experiments in section 5.2.3. Because no update of barotropic velocities is performed at analysis time, we have seen that the sea-level cannot attain the desired correction. We also expect variation associated both with initial geostrophic adjustment and also excitation of barotropic Rossby waves.

The effects of the altimetry data sampling pattern are evident in the analysis error. This is smallest at gridpoints closest to the observations in time and space. The temporal impact of the data is particularly evident. Comparing the analysis error with the TOPEX/POSEIDON ground tracks (Figure 4.15) it is clear that not all data within a 10 day altimeter repeat cycle have contributed significantly to the analysis. This is due to the 10 day e-folding time scale used for the forecast error correlation. It means that in forming an analysis, greater weight is given to the model state than observations if those observations are more than several days distant from the analysis time. It is for this reason that an analysis period different to 10 days was used. If the analysis period coincides with the repeat cycle, the regions of large analysis error would be the same each analysis time. (Assimilation experiments which varied the analysis period are mentioned, but not discussed in detail, in section 5.4.5.) The limiting error in regions of no data (*e.g.* the boundary regions) is simply the *a-priori* forecast error, specified in this run to be half the average data variance (Table 5.7).

Finally, the low analysis error over regions of shallow topography is due to the bogus observations used in those regions to ensure a small sea-level correction.

Comparison with SR3 section

The assimilation run encompassed the period of WOCE SR3 cruise AU9407, which was occupied during January 1994. An important test of the assimilation is its ability to capture the subsurface variability of the real ocean. To enable a comparison between the hydrographic section and the model, a model equivalent section was interpolated in space and time from full model fields stored every five days during the assimilation. The resulting comparison is shown in Figure 5.29. The result is rather disappointing — it is clear that the large eddy evident in the synoptic hydrographic section at around 48°S is barely present in the model. There is only the slightest hint of raised isopycnals in the model equivalent section.

To probe this in more detail, the actual model sea-level and analysed correction must be examined. The eddy in question was traversed by the cruise on around 5 January, 1994 (Rosenberg et al. 1995a). The closest SOFA analysis to this date in the present assimilation run was 2 January, 1994. Figure 5.30 shows the sea-level anomaly in the region of the eddy for this date.

The first panel of the figure plots the raw altimeter data together with the model sea-level anomaly for the altimeter track through the eddy on 28 December, 1993 (marked with red dots in other panels). This is the altimeter pass along the track closest in time to the analysis date. The most significant features along this part of the altimeter track are two peaks — a sea-level depression of around 28 cm at 48.6°S, and a rise of around 18 cm at 50°S.

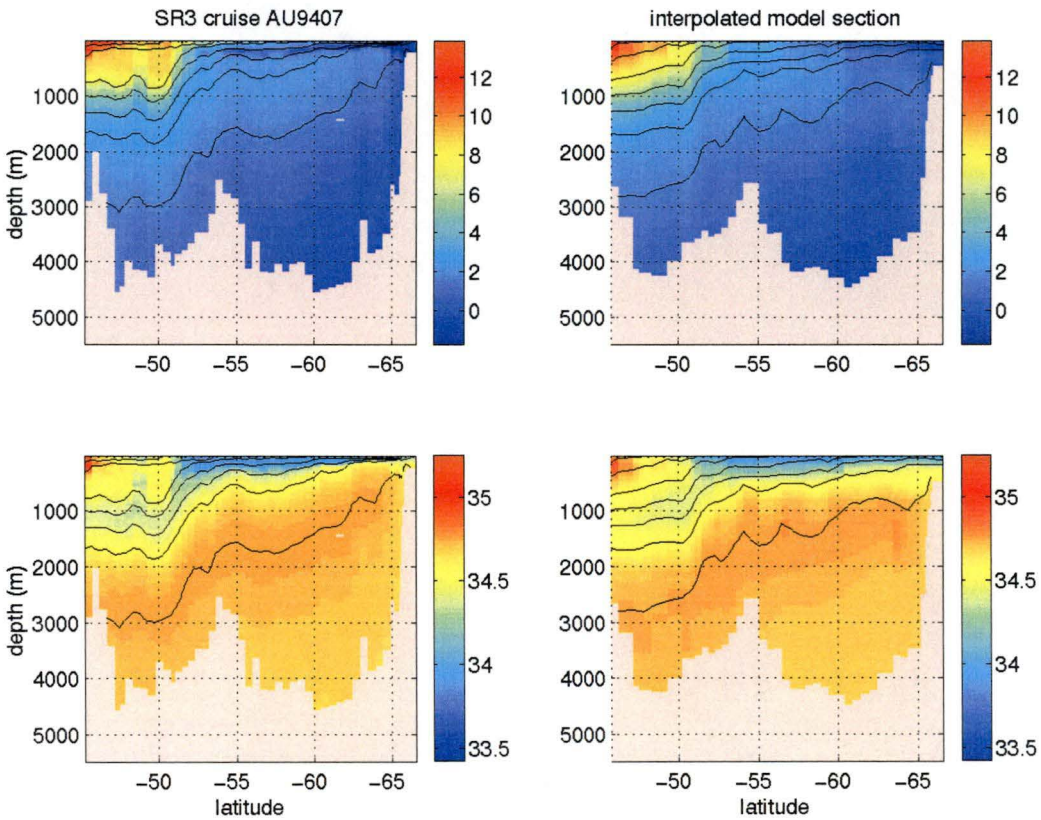


Figure 5.29: Comparison between SR3 cruise AU9407 hydrographic sections and model equivalents for baseline assimilation run: potential temperature (top panel), salinity (lower panel). Sections are overlaid with contours of potential density.

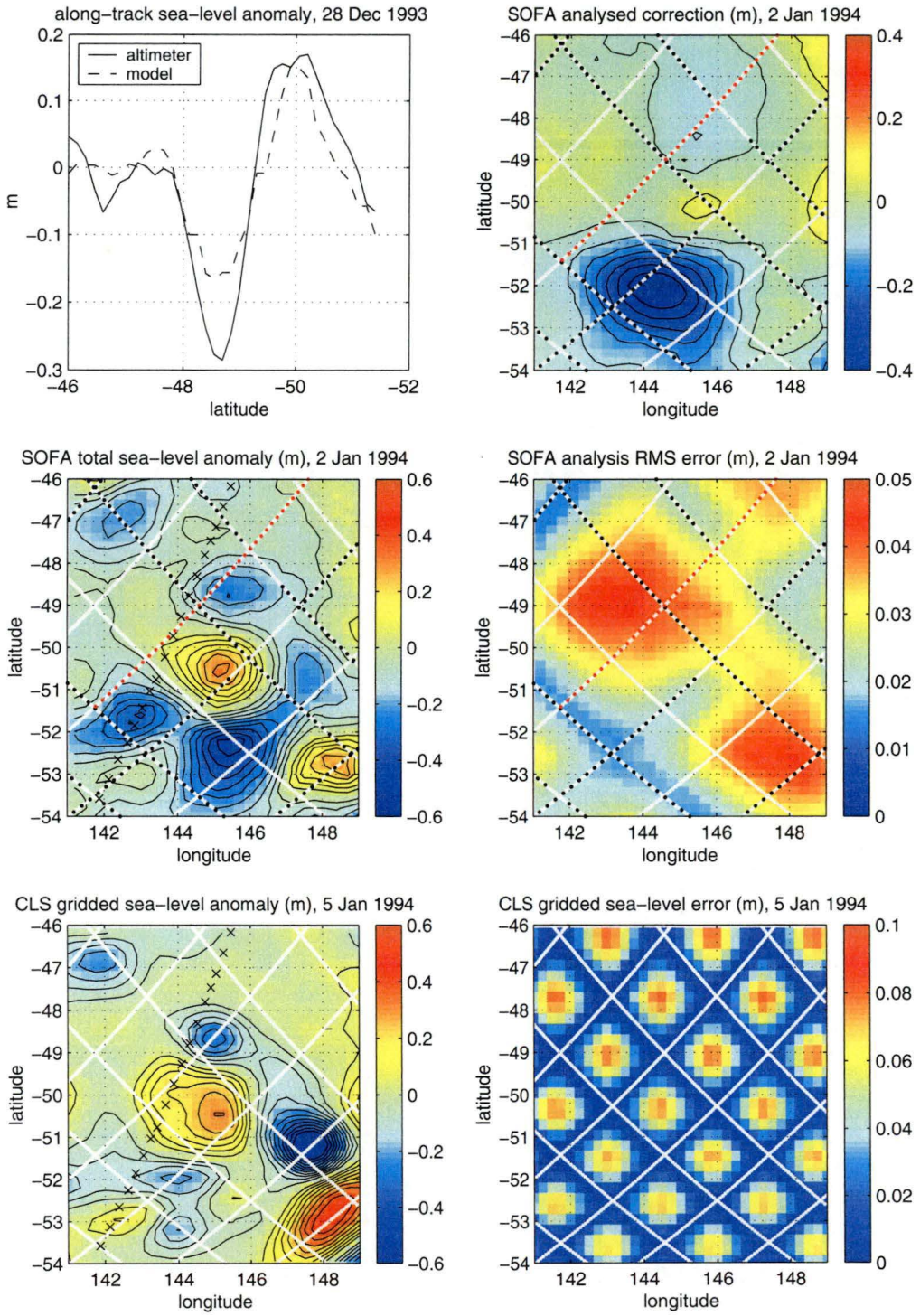


Figure 5.30: Details of SOFA analysis (baseline assimilation run) for 2 January, 1994. Along-track altimetry data and misfits on 28 December, 1993 (top left); analysed sea-level correction on 2 January, 1994 (top right); total post-analysis sea-level anomaly (middle left) and error (middle right); CLS gridded sea-level anomaly for 5 January, 1994 (lower left) and error (lower right). The altimeter data points used by SOFA to form the analysis are marked with dots; those in red correspond to the data in the first panel. Locations of SR3 hydrographic stations are marked with crosses.

Both of these were already present in the model to some degree. However the depression at 48.6°S in the model has little more than half the observed magnitude, while the rise centred at 50°S has only half the width of the observed peak along the altimeter track. While the altimeter track does not coincide exactly with the SR3 cruise track, the sea-level anomaly associated with the hydrographic section is, nevertheless, similar to that observed by the altimeter. This anomaly was shown earlier, in Figure 5.14(f) (red curve), where it was compared with the CLS gridded altimetry product. Like the altimeter data, the hydrographic sea-level anomaly has a strong negative peak at around 48.6°S and a broad positive peak at 50°S. Examining Figure 5.14(c), it is clear that the very obvious eddy feature at 48.6°S appears as pronounced as it is because of the lowered isopycnals over around two degrees of latitude south of this location. These lowered isopycnals are associated with the broad positive sea-level anomaly centred at 50°S.

The second panel of Figure 5.30 shows the analysed correction from SOFA, valid on 2 January, 1994. Non-zero contours are at intervals of 5 cm. This should be compared with the along-track misfit for the track in question (marked with red dots). The along-track misfit is just the difference between the two curves in the first panel. Although the misfit was measured five days earlier than the analysis date, it is expected to remain similar over this time interval. Following the analysis along the track, the eddy at 48.6°S is captured with a peak correction of -10 cm which is about the right magnitude. It is clear that the correction is smoother than the misfit, however. The negative correction from SOFA extends at least from around 46.5°S to 49°S in a broad peak of over 5 cm magnitude. The misfit, however, is not that smooth. From the first panel, the data–model misfit is zero at around 47.2°S and also at 48°S, and smaller than 5 cm everywhere north of around 48.2°S. Similarly, the first panel shows positive data–model misfits of between five and 10 cm along the altimeter track in two narrow bands either side of 50°S. Both of these are missing in the analysed correction. This smoothness is expected because the forecast error covariance had a spatial correlation length of 200 km — considerably larger than the small-scale features seen in the along-track data–model misfits in the first panel.

The altimeter tracks shown on the diagrams include both the full set of ground tracks (solid lines) and those observation points used for the analysis (black and red dots). The data selection window in SOFA is limited in time to the shorter of either the analysis period or the window specified with the `RIT` parameter (Table 5.7). In this case, the analysis interval of seven days is shorter than the specified time window of 15 days, so no observations more than one week distant from the analysis date are used. Thus some of the ground tracks have contributed no data at all to the analysis.

The sum of the model forecast sea-level anomaly and the analysed correction for 2 January, 1994, is shown in the third panel, with the RMS analysis error in the fourth. Comparing the total analysis with the observed altimetry from five days earlier (solid line, first panel), some differences are apparent. The two main peaks in the altimetry of 28 December, 1993, measure around -28 cm and 15 cm. The total post-analysis sea-level anomaly captures these

at -22 cm and 5 cm, respectively, along the relevant altimeter track. Between 50.2°S and 51°S, the total post-analysis anomaly is negative (up to -15 cm), while the altimeter measurement is positive.

Locations of hydrographic stations on the AU9407 cruise are marked with crosses in Figure 5.30. The discrepancies between altimeter observations and the SOFA analysis just discussed suggest a similar interpretation of the differences between the cruise hydrography and the model equivalent (Figure 5.29). The negative sea-level anomaly at 48.6°S is of roughly the same magnitude in the analysis as the altimetry (-22 cm and -28 cm, respectively). Similarly, the isopycnals at this latitude (Figure 5.29) are displaced vertically by about the same amount in the model equivalent hydrography, compared with the cruise data. However, the significant positive sea-level anomaly measured by the altimeter at around 50°S is largely absent along-track in the resulting analysis (Figure 5.30). Similarly, the model isopycnals at this latitude are not displaced sufficiently downwards in the water column to reproduce the strong eddy signal seen in the hydrography (Figure 5.29). Had the observed altimetry been better reproduced in the SOFA analysis, the model equivalent hydrographic sections would have appeared more like the AU9407 sections. The differences between the altimetry and the sea-level analysis are at least partly due to the 200 km forecast error correlation length scales used in this assimilation run.

The analysis error field (fourth panel) shows most obviously the effects of the temporal sampling of the altimeter data relative to the analysis time. The error is largest where data on altimeter tracks have not been used in the analysis. It is also large where the data are several days away from the analysis time. The forecast error correlation function used in this assimilation run had an e-folding time (RCT, Table 5.7) of 10 days. Thus, even though the altimeter data from 28 December (red dots, and first panel) were used in forming the 2 January analysis, their weighting was low so that the analysis error has not reduced much along this particular track. Indeed, comparing the second and fourth panels, it is clear that the analysed sea-level correction is small wherever the analysis error is large. The reduced weighting of the 28 December altimetry data is therefore another likely reason for the discrepancies between the observations and the analysis of 2 January along the satellite groundtrack. A greater weighting would have been given to the 28 December altimetry if a longer timescale had been used in the forecast error correlation function.

For comparison, both the CLS gridded sea-level anomaly and its associated error for 5 January, 1994 are also shown. The CLS field appears to have placed a significant weight on the 28 December TOPEX/POSEIDON track through the eddy. At both the 48.6°S and 50°S peaks along this track, the CLS analysis is much closer to the observed altimetry from 28 December, 1993 (first panel), than is the total post-analysis sea-level anomaly from SOFA (third panel). The CLS mapping procedure used an e-folding timescale of 15 days (Le Traon et al. 1998), longer than that used in the assimilation run here. In addition, the CLS maps incorporate measurements from the ERS-1 satellite, with groundtracks much closer together than the TOPEX/POSEIDON satellite. Thus, the CLS gridded field in the fifth panel of Figure 5.30 has considerably smaller spatial scales than the analysed sea-level correction from

SOFA in the second panel.

The discussion so far has focussed in detail on the single SOFA analysis of 2 January, 1994. This is the analysis date closest in time to the sampling of the eddy by the AU9407 cruise. Full model output fields were stored every five days throughout the assimilation run. Figure 5.31 examines the temporal evolution of the eddy field in these individual model snapshots. The figure shows both the model sea-level anomaly and the SR3 equivalent salinity section on each of the model output dates over the duration of the AU9407 cruise. It is from these sections that the model-equivalent AU9407 section of Figure 5.29 was interpolated. It is clear that the sea-level anomaly peaks responsible for the main eddy feature in the cruise data remain relatively stationary in the model for the duration of the cruise. Nevertheless, the level of the positive sea-level anomaly at 50°S along the cruise track varies considerably between the four model snapshots. It is a maximum of 16–17 cm on the 10 January and 15 January model output dates. The SR3 equivalent salinity sections for both of these dates shows the eddy feature between 48°S and 50°S significantly enhanced compared with the time-interpolated section of Figure 5.30. This strengthens the conclusion that the model equivalent tracer sections would have appeared much more realistic had the SOFA analysis of 2 January, 1994 better reproduced the along-track altimetry through the eddy of 28 December, 1993, and the positive feature at 50°S in particular.

The eddy observed on WOCE SR3 cruise AU9407 is a very significant feature in the hydrographic sections. It is disappointing that the assimilation did not better reproduce the subsurface tracer fields along this section. The detailed analysis above suggests that this was due to a combination of factors. First, the 10 day time scale used in the forecast error correlation function meant that the altimeter track through the eddy (28 December, 1993) was only weakly weighted in forming the analysis of 2 January, 1994. Second, the length scales used in the forecast error correlation function were too large (200 km) to enable small-scale adjustments to be made to the model's representation of the eddy.

Meridional heatflux

The meridional heatflux was shown for an unassimilated model run in Figure 5.23. It is shown in Figure 5.32 for the baseline assimilation run. Assimilation has increased the poleward transport of heat at most latitudes. In particular, the transient eddy component has increased from small values of less than 0.02 PW to around 0.1 PW over the latitude band of the ACC from 45°S to 60°S. Taking an average depth of 3000 m, this corresponds to a uniform poleward flux of around 6 kW/m^2 at 50°S. This compares much more favourably than the unassimilated run with the figure of 11 kW/m^2 measured by Phillips and Rintoul (2000) in the ACC south of Australia (section 1.2.1).

Model forecast ability

The SOFA statistics averaged over the domain (shown earlier, Figure 5.24) demonstrate that assimilation has not substantially improved the model's fore-

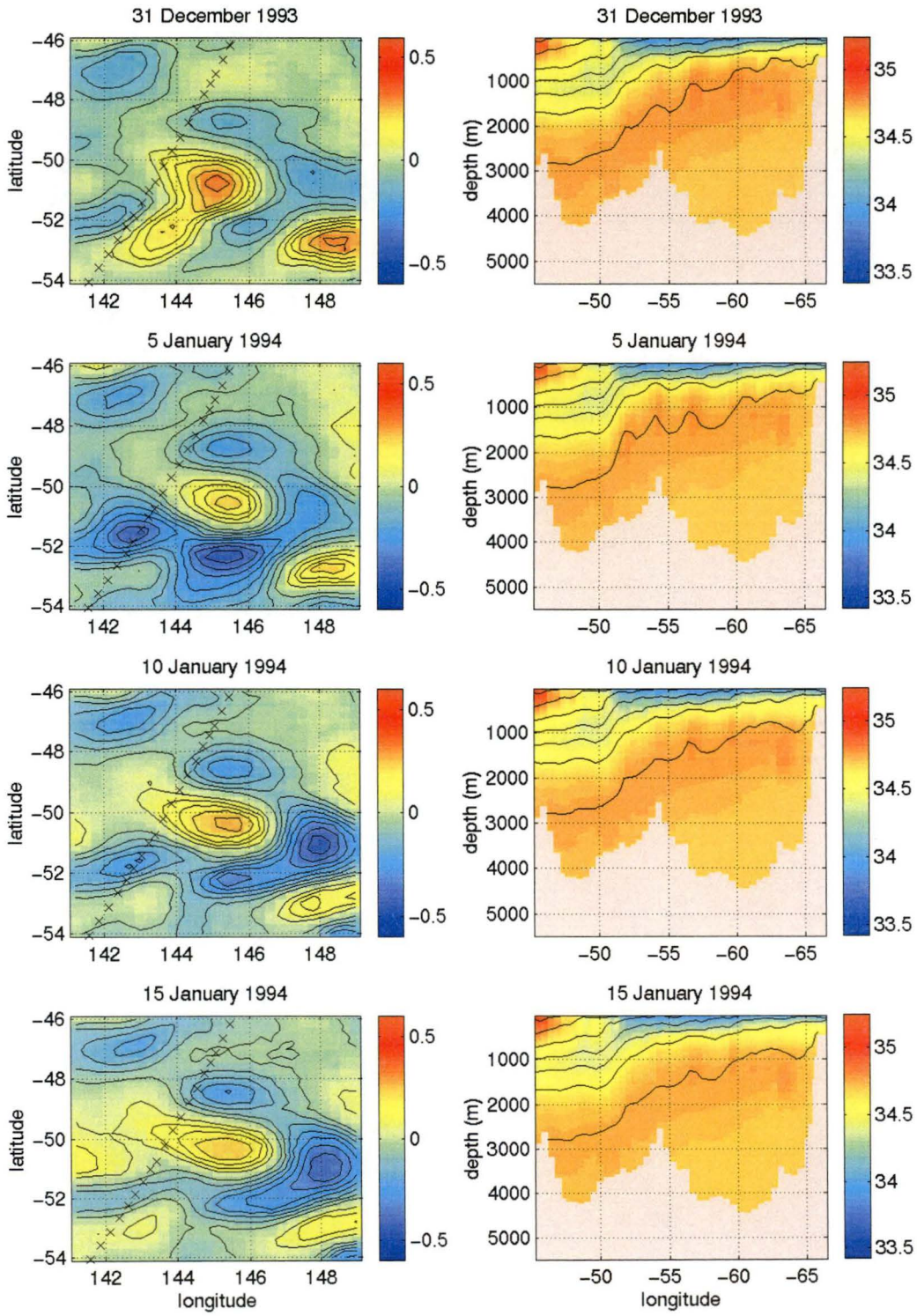


Figure 5.31: Model sea-level anomaly and equivalent SR3 salinity section from the baseline assimilation run every five days between 31 December, 1993, and 15 January, 1994.

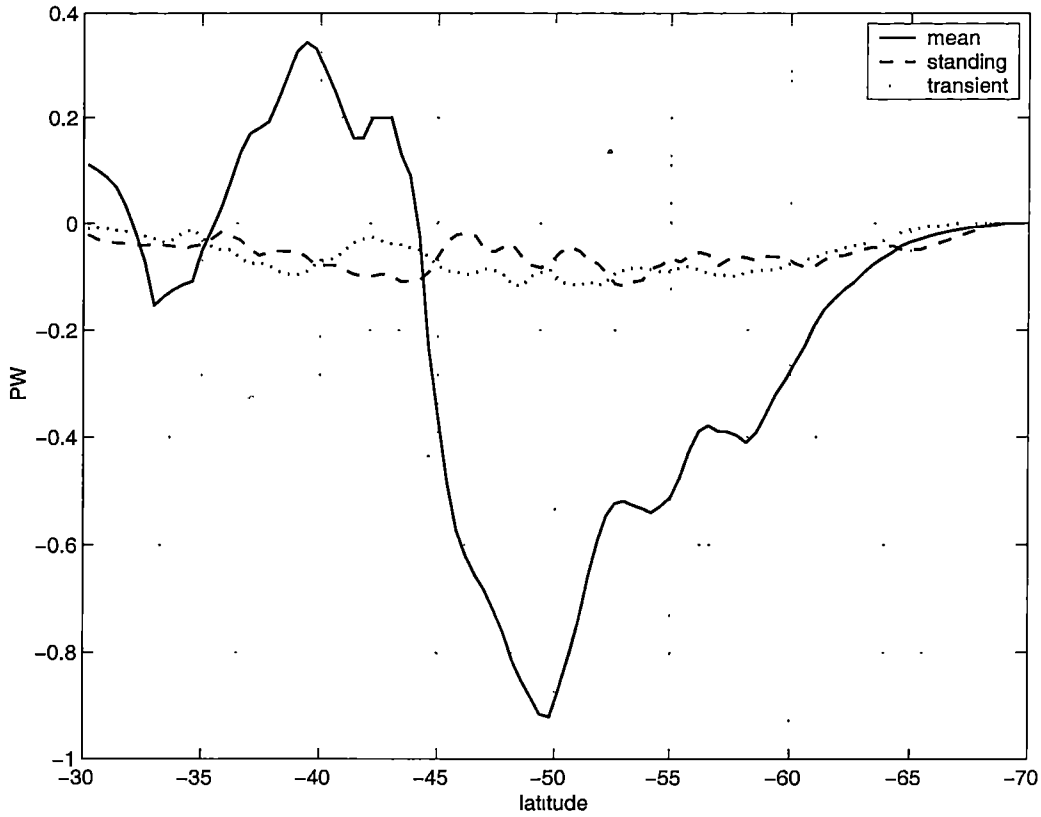


Figure 5.32: Meridional heatflux for baseline assimilation run.

casting ability. For most assimilation periods, the variance of the synoptic misfit between data and model is no lower than the misfit between the data and the previous analysis (persistence). The forecast/persistence statistic is calculated by SOFA over the second half of each assimilation interval. An alternative measure of forecast ability is to compare a given analysis with both the model state just prior to analysis (forecast) and the previous analysis (persistence). If the model, together with assimilation, is successfully tracking the real ocean, then the model forecast should be a better estimate of the new analysis than should the previous analysis. Figure 5.33 compares these differences for the analysis of 9 January, 1994. While not overwhelming, the forecast error is lower than the persistence error in a number of locations. Consistent with the statistics, however, the improvement is not great. Similar results are found for other analysis dates.

Selected results from some other assimilation runs are now described.

5.4.3 Reduced forecast error correlation length scales

From the analysis of the baseline assimilation run above, it is apparent that the length scales in analyses is greater than some oceanic features of interest. As mentioned in section 5.1, it is the forecast error covariance structure that is mainly responsible for spreading observations in time and space in Kalman filters and optimal interpolation.

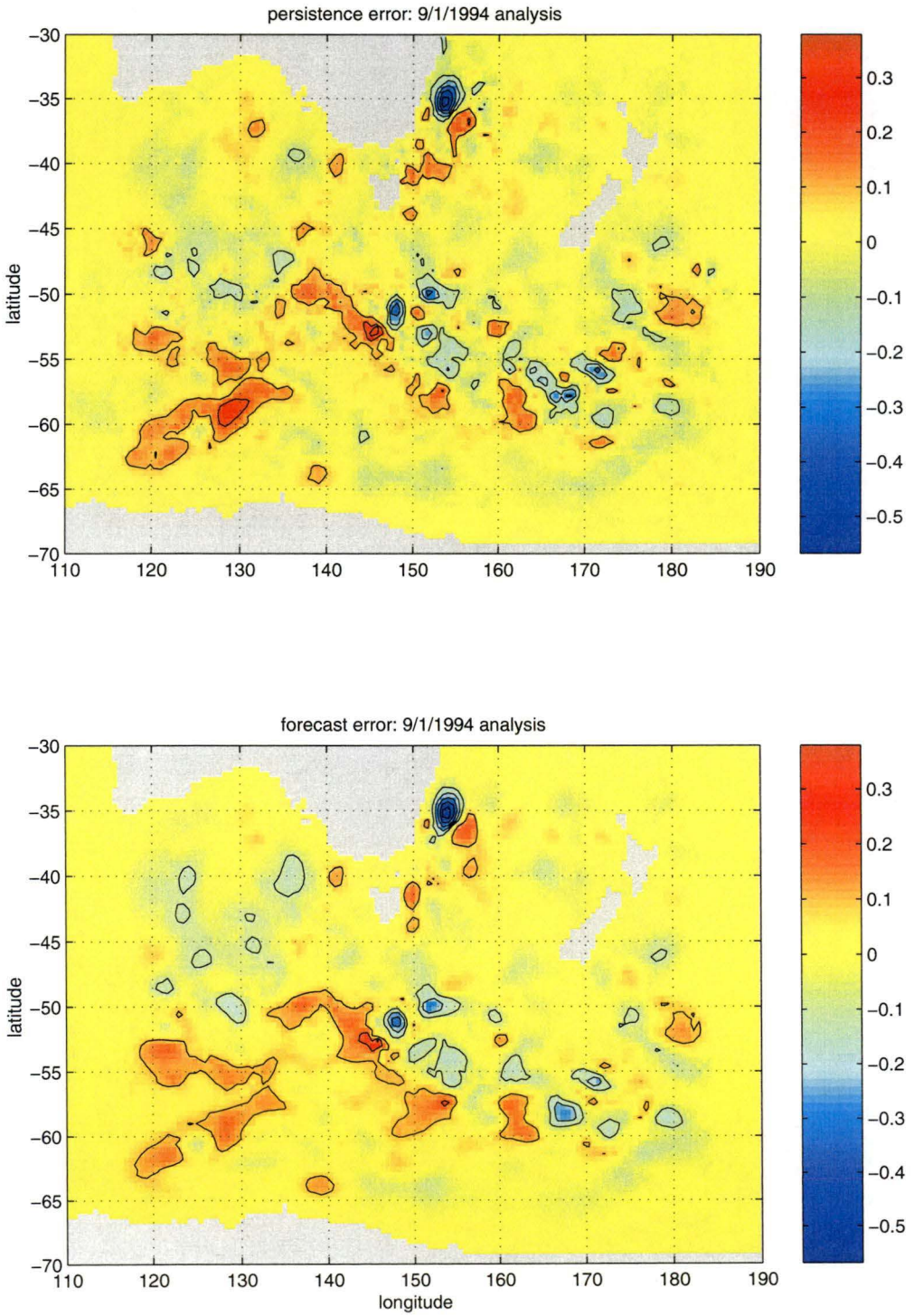


Figure 5.33: Error (in m) of persistence (top panel) and forecast (lower panel) for sea-level analysis of 9 January, 1994, in the baseline assimilation run.

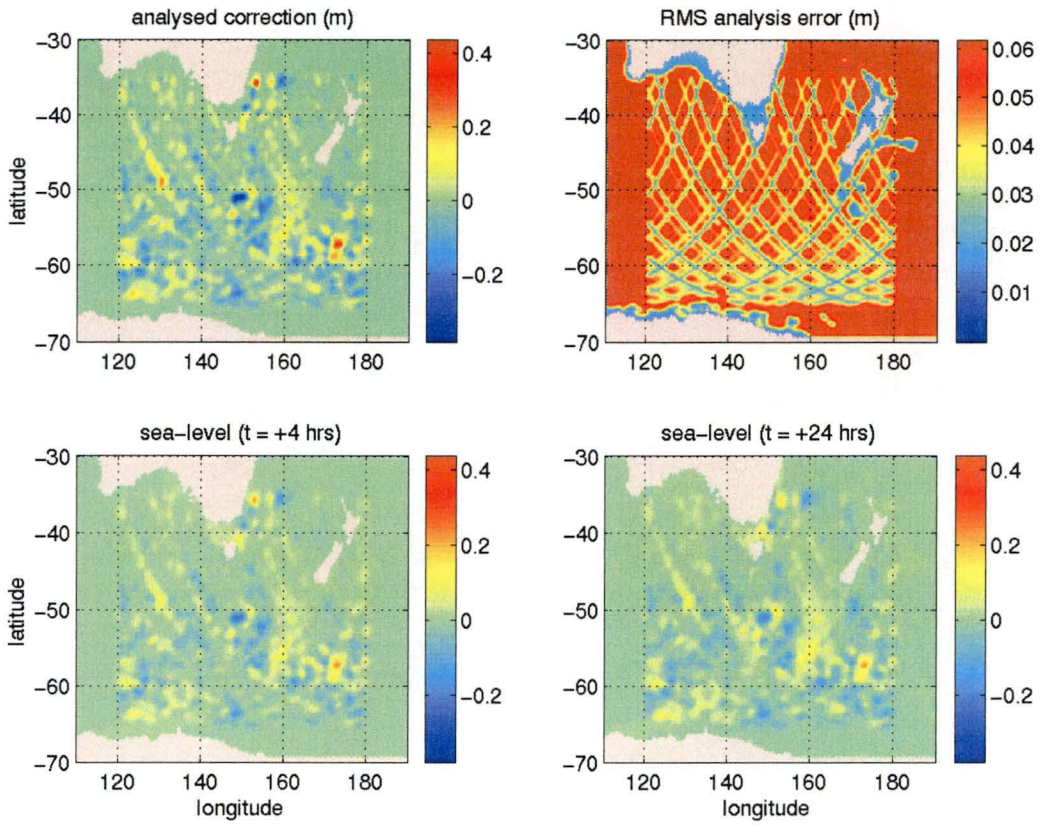


Figure 5.34: 26 December, 1993 analysed sea-level correction (top left) and RMS analysis error (top right) for assimilation with 50 km forecast error correlation length scales. Post-analysis difference from forecast at four hours (bottom left) and 24 hours (bottom right) after assimilation.

To investigate whether better assimilation performance is obtained with smaller length scales, two assimilation runs were performed using forecast error correlation scales (SOFA parameter `RCX1` and `RCX2`) of 100 km and 50 km.

Analysed correction and analysis error

Figure 5.34 shows the analysed sea-level correction and associated error for 26 December, 1993, as well as the post-assimilation sea-level anomaly at four and 24 hours after the analysis time. This should be compared with Figure 5.28 for the baseline assimilation run. The analysed corrections for the two runs are not expected to be the same, and simply reflect the different model forecasts at this analysis time in the two runs. However, in addition, the spatial scales in the analysis certainly are smaller in the present assimilation run (Figure 5.34) compared with the baseline run (Figure 5.28). The length scales in the forecast error covariance are impacting as expected on the assimilation. Another effect of the smaller scales is seen in the analysis error. This is larger away from the altimeter tracks in the present run. The smaller length scales mean that only near an altimeter track are the observations strongly weighted in forming an analysis. Indeed the analysed correction in Figure 5.34 shows artifacts of

the altimeter tracks to a larger degree than the corresponding analysis in the baseline assimilation run (Figure 5.28).

Assimilation statistics

Figure 5.35 shows the global statistics for the run using 50 km length scales in the forecast error correlation function. These are largely similar to the statistics for the baseline assimilation run (Figure 5.24), with the exception of the normalised mean square data–model misfit, which has been improved. It is now below one throughout the assimilation, and almost as low as 0.5 in several assimilation periods. Comparing the two assimilation runs with shorter length scales, the assimilation statistics were better for the run using 50 km than for the run using 100 km (not shown here).

Comparison with SR3 section

Figure 5.36 shows the comparison between the SR3 sections from cruise AU9407 with the model-equivalent sections. There is little improvement compared with the baseline assimilation run (Figure 5.29). The isopycnals are slightly lower in the water column at 50°S in comparison with the baseline run, but the impact is marginal.

In other respects, also, the results are not substantially improved over the baseline assimilation run.

5.4.4 Increased barotropic vertical projection

In earlier parts of this thesis, it has been demonstrated that a significant component of the sea-level variability in this part of the ocean is associated with barotropic flows. The vertical projection scheme used for the previous assimilation runs was completely 'baroclinic', with the entire analysed sea-level correction used to displace the density profile through a column of water. No modification of barotropic velocities was performed.

An alternative assimilation procedure would use some portion of the analysed sea-level correction to update barotropic velocities, with the remainder used for vertical isopycnal displacement. A similar approach was attempted by Fox et al. (2000a), who first filtered the sea-level updates. A geostrophic projection of the large-scale signal onto the barotropic mode was attempted, but led to numerical instability. This was likely due to the explicit numerical scheme used for the barotropic system in their model. Such an approach works here because of the implicit numerics of the barotropic system in the HOPE model.

Three experiments were performed using 50%, 75% and 100% of the sea-level updates to modify barotropic velocities. The first of these experiments gave the best results and is reported here. All other assimilation parameters were the same as used in the baseline assimilation run (section 5.4.2).

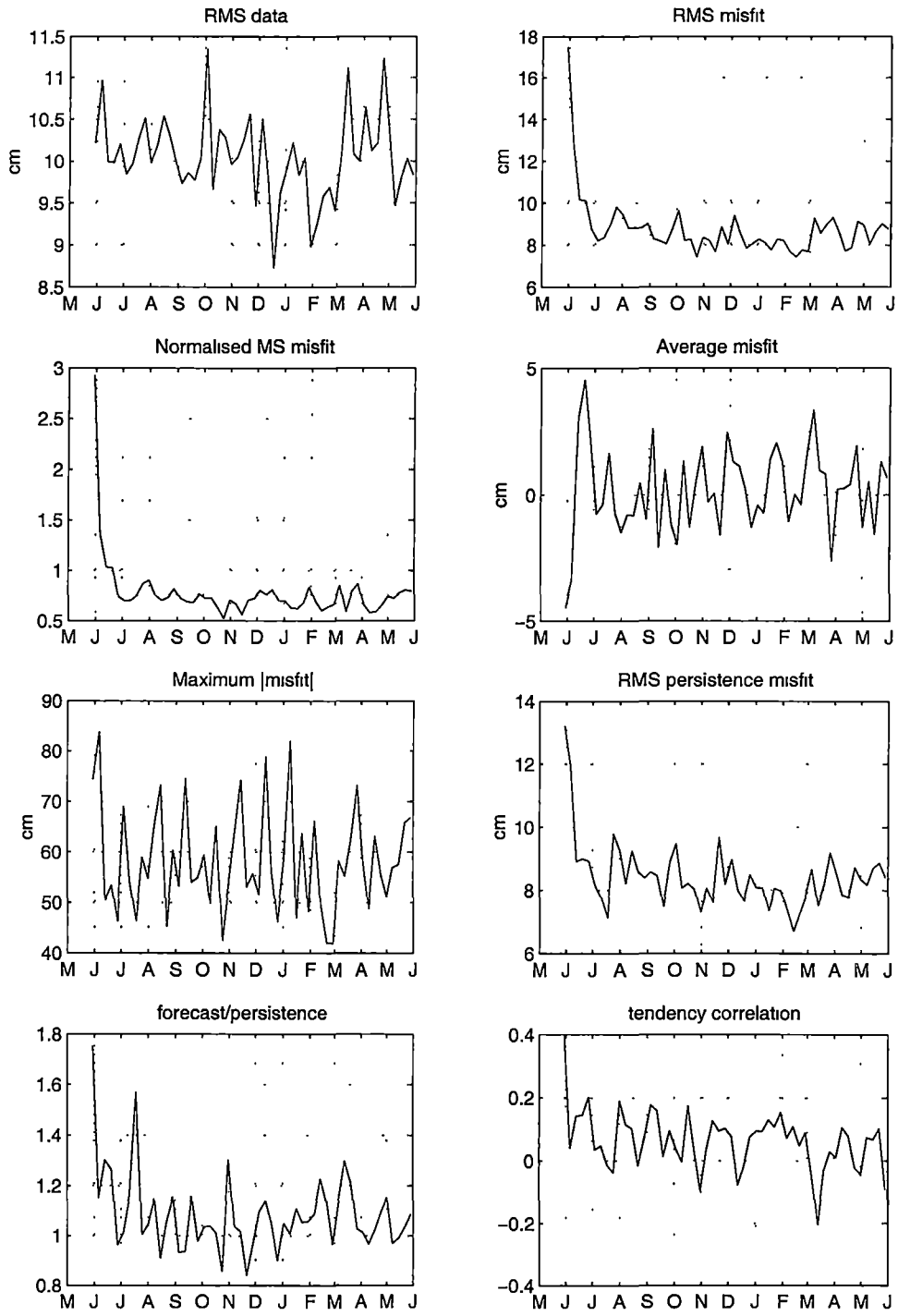


Figure 5.35: Assimilation statistics for assimilation run with 50 km length scales.

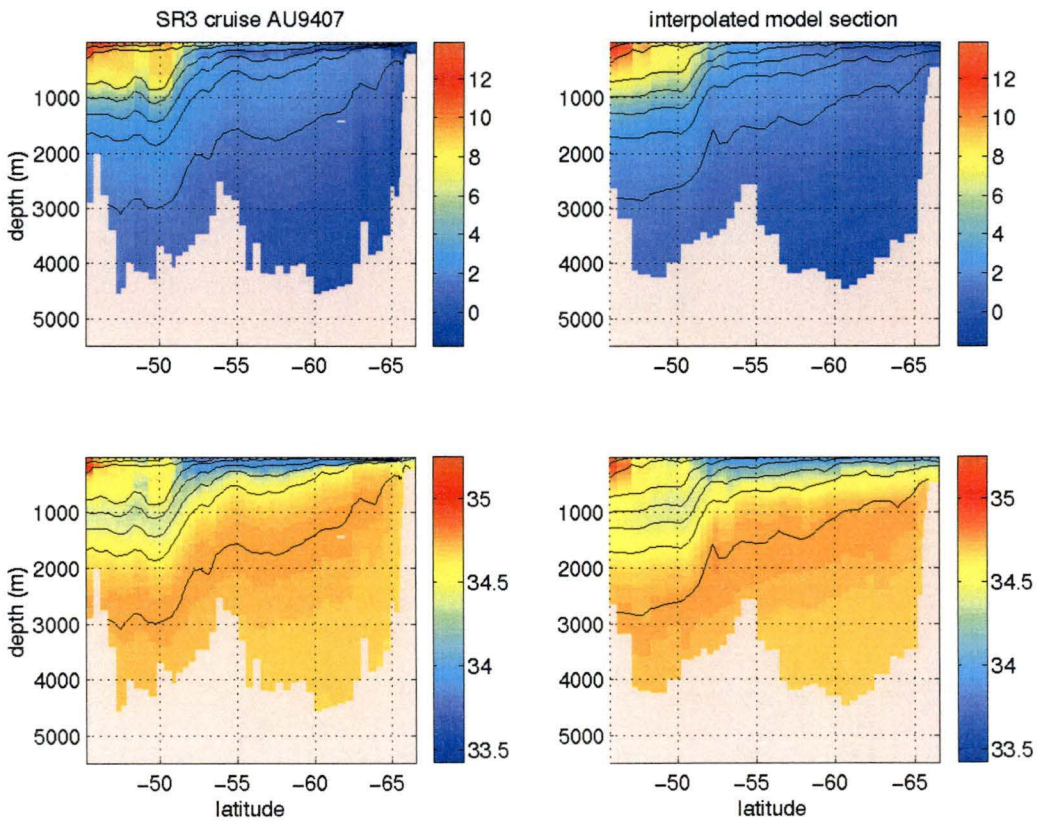


Figure 5.36: Comparison between SR3 cruise AU9407 hydrographic sections and model equivalents for assimilation using 50 km forecast error length scales: potential temperature (top panel), salinity (lower panel). Sections are overlaid with contours of potential density.

Assimilation statistics

Figure 5.37 shows the global statistics for the run applying 50% barotropic projections of the sea-level anomaly, while Figure 5.38 shows some median statistics over various regions of the domain. These should be compared with the equivalents for the baseline assimilation run (Figures 5.24 and 5.25).

The global statistics (Figure 5.37) show improvements over the baseline run in several regards. First, the normalised data–model misfits have been reduced (below one throughout the run). Second, the model's forecast ability has been improved. Both the forecast/persistence error statistic and the tendency correlation show improved forecast skill. In the baseline assimilation run, the ratio of the forecast error to persistence error only occasionally dropped below one, while it is almost always below one for the present assimilation experiment. Also, the tendency correlation of the baseline run was centred around zero, while it has now been raised to above 0.2 during most assimilation periods. A correlation of 0.2 is not large, but the improvement over the baseline run is consistent and clear.

The mapped statistics (Figure 5.38) provide an indication of the regions of improvement. While the pictures are somewhat noisy, there is evidence that an increased barotropic projection of the analysed sea-level anomaly has improved the statistics, primarily in the south of the domain (south of 55°S). Examining first the median normalised mean square misfit, the largest misfits in the baseline run (Figure 5.25) occur at high latitudes, south of around 55°S where they are over 1.5. In the present assimilation run, the misfits are everywhere below 1.5, and similar at lower latitudes (north of 45°S) and higher latitudes. While the misfits are roughly similar in the north of the domain for the two assimilation runs, they have been reduced in the south. The improvement is weaker for the ratio of forecast to persistence error, but in the present run there appears generally to be better performance in the south of the domain, while the reverse is the case in the baseline run.

Comparison with SR3 section

To examine whether the statistical improvements have an impact on the subsurface fields, the SR3 equivalent sections are shown in Figure 5.39. While there remains a considerable discrepancy between the hydrographic sections and the model equivalents, the present assimilation run does appear to have slightly improved the representation of the eddy at around 48°S, particularly in the upper 1000 m. This is somewhat surprising because it is fair to suppose that assimilation by displacement of isopycnals ought to perform better at representing the eddy than a highly barotropic assimilation.

Sea-level variability

The sea-level variability through the assimilation run is shown in Figure 5.40. The equivalent for the baseline assimilation run was shown in Figure 5.26. The main regions of variability along the path of the ACC have a similar level of variability, up to around 18 cm. However, the broad region of elevated

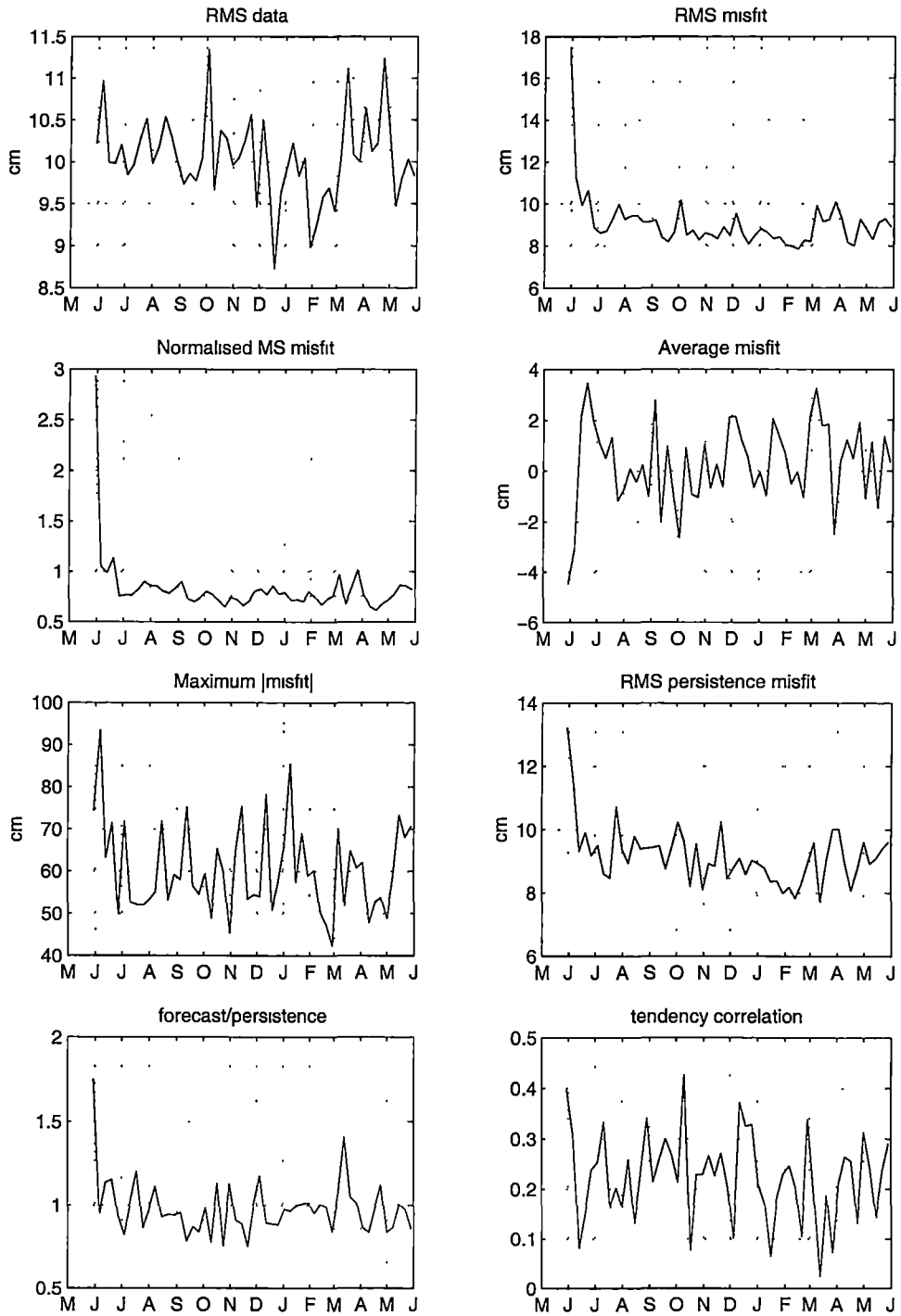


Figure 5.37: Assimilation statistics for assimilation run with 50% barotropic projection.

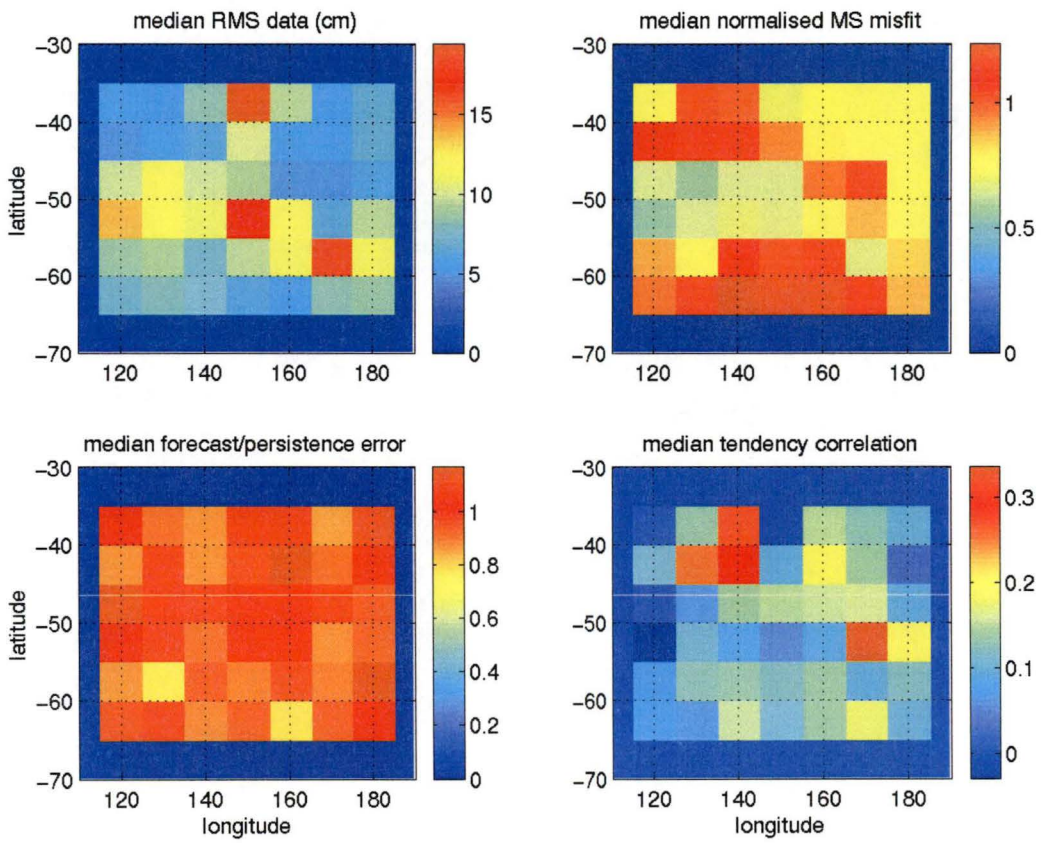


Figure 5.38: Regional statistics (median for 5° × 10° boxes) for assimilation run with 50% barotropic projection.

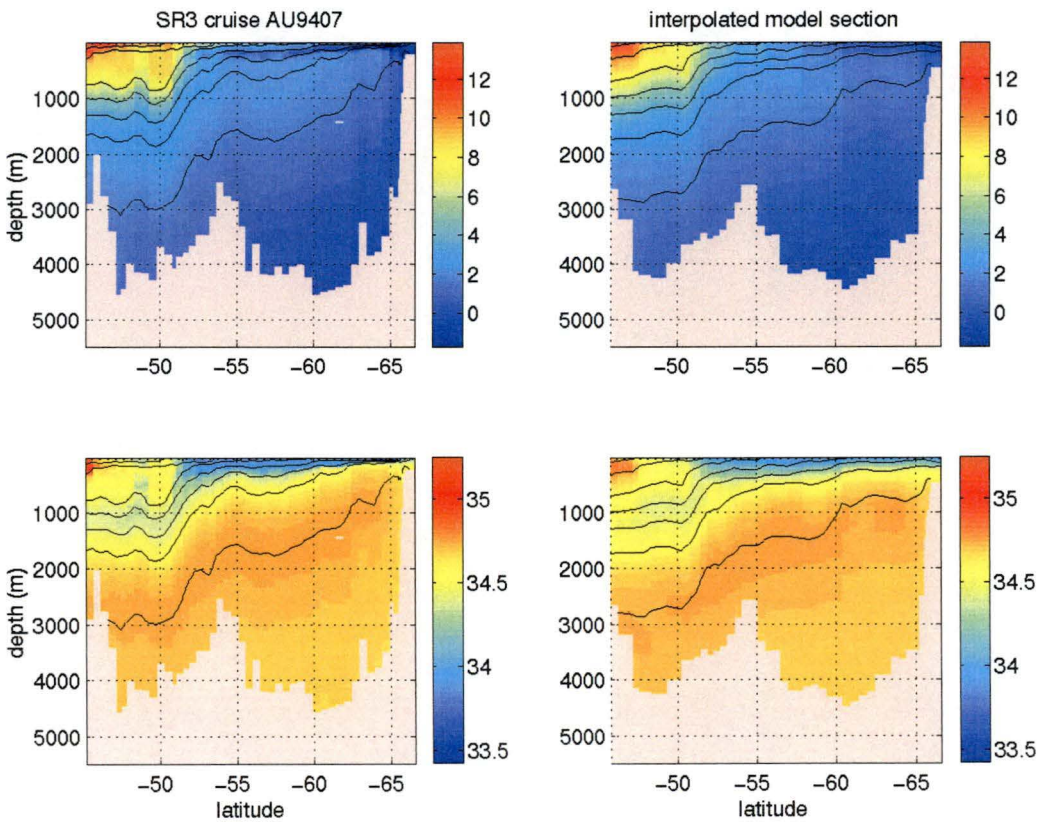


Figure 5.39: Comparison between SR3 cruise AU9407 hydrographic sections and model equivalents for assimilation with 50% barotropic projection: potential temperature (top panel), salinity (lower panel). Sections are overlaid with contours of potential density.

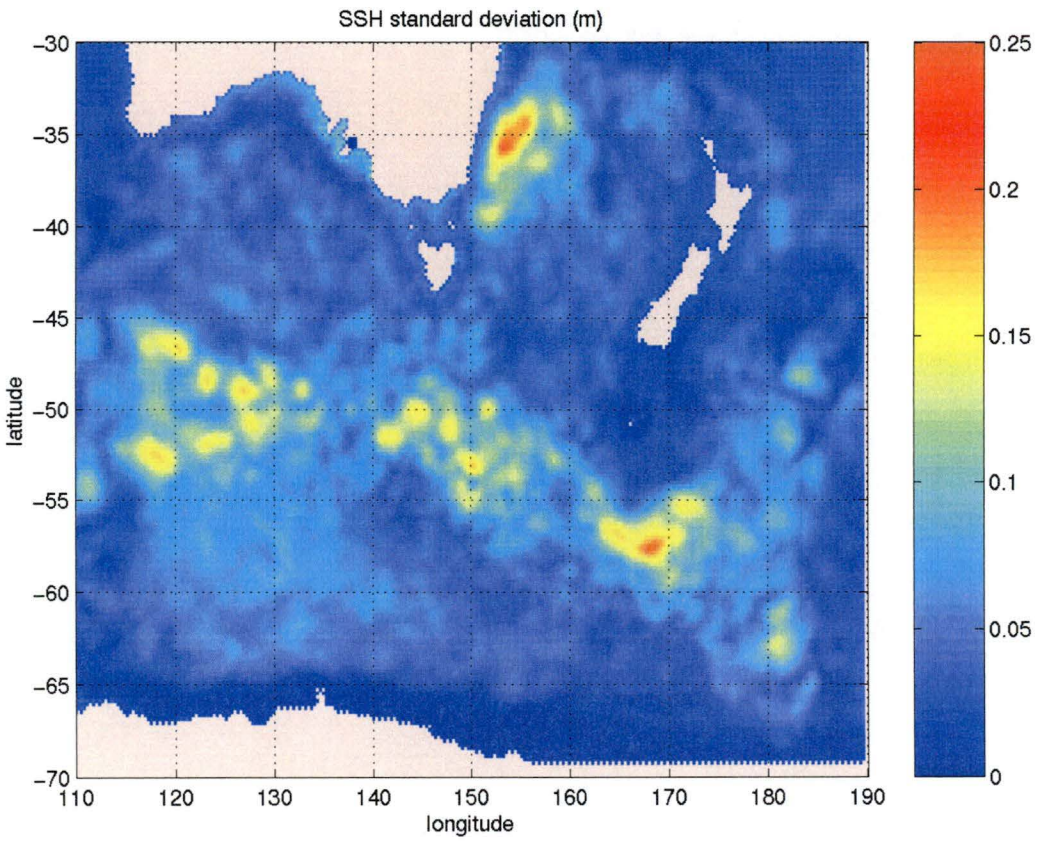


Figure 5.40: Sea-level variability for assimilation run with 50% barotropic projection.

variability (up to 15 cm RMS) in the southwest of the domain in the baseline run has been reduced (below 10 cm RMS) in the present run. If barotropic variability is more representative of the real ocean's behaviour in this region, it may be that applying more barotropic updates in the assimilation procedure produces more realistic variability patterns.

5.4.5 Other assimilation runs

Numerous other assimilation runs were performed, varying other assimilation parameters, *e.g.* the data selection parameters (`RATIO`, `NSEL`), along-track altimetry decimation (`IFREQ`), forecast error covariance time scale (`RCT`), analysis error propagation (`TAU`) and assimilation interval (`TINT`). None of these produced results significantly different to those already shown. In particular, none produced a more realistic subsurface representation of the eddy in the AU9407 section along the WOCE SR3 transect. The statistics calculated by SOFA during the assimilation run were also similar. Eddy variability in terms of both sea-level variance and meridional eddy heatfluxes again were similar to runs already discussed.

Improvements in the representation of subsurface fields will require more substantial investigation of the effects of model physics on the assimilation scheme. It would also be useful to examine the effect of varying the forecast error variance (`FNOISE`) relative to the observational error variance (`ENOISE`) in SOFA.

More sophisticated assimilation would apply the outcomes of the experiments to trial an improved scheme. For instance, based on results reported here, one might expect a scheme to perform better if it combined smaller length scales and longer time scales in the forecast error covariance with a vertical projection that has a more barotropic component at higher latitudes, and where in any case barotropic velocity should be updated to account for the updated depth integrated density gradient.

CHAPTER 6

Summary and discussion

The aim of the research project presented in this thesis was to implement a satellite altimetry data assimilation scheme in a high-resolution regional model of the Southern Ocean south of Australia. It was intended that the resulting model output should be used to investigate in detail the synoptic dynamics of the Antarctic Circumpolar Current in this region, and the role of eddies in heat transport and the ACC momentum balance. From observational studies, both satellite and in-situ (Morrow et al. 1994; Phillips and Rintoul 2000), the region is known to be very active dynamically.

The three steps involved in developing an assimilation system have been detailed in this thesis: the implementation of open boundaries in the numerical model, the verification of model performance in its regional configuration, and the implementation of the assimilation scheme in this regional model. The unforeseen complexities associated with the first of these tasks limited the scope for analysing the results of the third, the altimetry assimilation. Nevertheless, a successful assimilation system was implemented, and statistical performance measures indicated forecast skill better than persistence.

In any research project, even one, like this, that is as much engineering as science, there are as many new questions raised as answered, and numerous perspectives on further work are suggested. It is the purpose of this chapter to summarise the results in this thesis and to review the issues left unexplored. This discussion will mirror the structure of the work (and the thesis).

6.1 Open boundary conditions

It became obvious soon after commencing work on implementing open boundaries in the Hamburg Ocean Primitive Equation (HOPE) model that this is a poorly understood — often misunderstood — area of ocean modelling. What results there are tend either to be abstractly theoretical or blissfully simplistic. The oft-quoted fact that the primitive equations are ill-posed under any pointwise open boundary condition (Oliger and Sundström 1978; Bennett and Kloeden 1978) seems sometimes to be used as something of a guise. Behind it, poor numerical results can be excused by virtue of the problem's insolubility. However, there are theoretical results which are known, and which

may usefully be applied. There are numerous examples in the literature of boundary conditions being overprescribed even in cases, like the shallow-water equations, where there is no fundamental ill-posedness. However, oversight of known results may often be due to the gulf between the mathematical literature containing them and the conventional oceanographic literature.

The approach taken in this thesis was to attempt to tread the middle ground; to apply, as far as possible, accessible theoretical results, but to yield to pragmatism where the theory became too difficult, or non-existent. Thus, for the barotropic system in HOPE, it was recognised that the propagation of characteristic variables in this hyperbolic system required precisely one boundary condition to be specified at each open boundary (section 3.5.1). This result is important, since the author's own early attempts at implementing open boundary conditions succeeded, in ignorance of the theory, at developing solutions where values were (over)prescribed for each of the primitive variables (U,V and η) of the barotropic system. In retrospect, some of the curious boundary effects in those results may have been due to the over-prescription of the boundary conditions (but might, in other circumstances, have been summarily ascribed to the problem's generic ill-posedness). Prescribing values for the incoming characteristic variables at a boundary is trivially well-posed, but proving well-posedness for prescribing either the normal velocity or sea-level alone was not possible using the energy method (section 3.5.1 and Appendix A). Rather than attempt the considerably more difficult normal mode analysis of Kreiss (1970) to prove well-posedness, both cases were implemented in the HOPE model and experiments performed to evaluate their performance (sections 3.5.2 and 3.5.3). It is a well-defined, and probably straightforward (if difficult) task to prove well-posedness of HOPE's barotropic system under various open boundary conditions. Indeed the solution is most likely an application (or extension) of known results for the non-linear shallow-water equations (Oliger and Sundström 1978). Such a proof would be useful for informing the further development of open boundary conditions in HOPE.

Independent of the underlying theoretical considerations, the numerical implementation of a given open boundary condition is a significant task in its own right; and not more so than in a model with implicit numerics, as the HOPE model has. Indeed, the work in this thesis appears to be one of the first investigations of implicit open boundary conditions in an implicit model. There are various numerical methods that may be used for solving a time-implicit finite-difference problem. Relaxation methods (Press et al. 1992) are simple to implement but use an iterative technique to converge on a solution. The method used in HOPE applies a direct matrix solution for the barotropic system (section 2.1.3 and Wolff et al. (1997)). This reduces the computer time required for a solution, at the expense of increased memory requirements and considerably more complex code. With more rapid solutions, it obtains the full benefit of a long timestep in the barotropic system. However, the direct-solution implicit numerics place some important constraints on the boundary conditions that may be implemented. For instance, any boundary condition must be linear and time-invariant (section 3.5). This prevented the consideration here of implicit formulations of Orlanski-type radiation schemes,

for instance.

The analytical form of a boundary condition is easy enough to write down. Its numerical implementation in HOPE's implicit direct-solution barotropic system is relatively straightforward (and systematic), but exceedingly tedious and error-prone (Appendices C and D). In practice, the task of hand-deriving the finite-difference implementation of a given boundary condition in HOPE's barotropic system was spread over several days (and several tens of pages of paper!), with the coding taking several days more. Small bugs in either the algebra or the coding required additional time to eliminate. The process was repeated for every new boundary condition to be examined. In principle the task could vastly be improved using computer algebra and automatic programming technologies. Much more rapid and robust prototyping of open boundary conditions would then be possible. The power of automated programming technology like this is evident in applications like the tangent linear and adjoint model compiler (TAMC) of Giering and Kaminski (1998) which automatically generates the adjoint of a given ocean modelling code.

Several boundary conditions were implemented for HOPE's barotropic system in this thesis. These include conditions on normal velocity (section 3.5.2), on sea-level (section 3.5.3), and on the incoming characteristic variables (section 3.5.4). The performance of different boundary conditions was judged on criteria of numerical stability, realism of solutions, volume conservation, and nonreflectiveness to surface gravity waves. Some of the results appear to contradict published results using similar conditions in explicit models. For instance, Chapman (1985) found both a zero-gradient condition on normal velocity, and a clamped sea-level condition, to be reflective to surface gravity waves; here they were found to be remarkably unreflective. This result is surprising and demonstrates that the behaviour of open boundary conditions depends very much on the numerical differencing of the interior dynamics. It was hypothesized here (sections 3.5.2 and 3.5.3) that the result may be related to the numerical dissipation of implicit differencing schemes. The energetics of the barotropic system were confirmed (Figures 3.17 and 3.18) to be very different, on timescales from hours to days, for time discretisations that varied the implicitness of the barotropic system. This appears to have significant implications for the reflection properties of open boundary conditions. Since open boundary conditions in implicit models have not previously been examined in detail, the different reflection behaviour does not appear to have been reported. Due to time constraints, it was not possible to investigate it further in this thesis, but it clearly has great importance to the general question of open boundary conditions in implicit discretisations of hyperbolic systems. A simple extension of the perturbation experiments used here would take a simple HOPE box model with, say, a clamped sea-level boundary condition and perform a series of Rossby adjustment experiments where the implicitness of the barotropic system was varied (using HOPE parameters α and β).

Of all the conditions examined for HOPE's barotropic system, the best performing was found (section 3.5.4) to be a modification of a condition proposed by Flather (1976). This condition prescribes values for the incoming characteristic variables. These are quantities which take the form $U + c\eta$ for

a western boundary, for instance, where c is the shallow water gravity wave speed, $c = \sqrt{gH}$. Of course, such a scheme guarantees the boundary values of the characteristic variables, but not of the primitive model variables (U, η) of which they are composed. Indeed, in experiments where the prescribed characteristic variables were derived from bottom-referenced geostrophy for the primitive variables, the actual across-boundary transports differed by tens of Sverdrups from the values calculated from geostrophy (Figures 3.34 and 3.35). Despite this, the boundary condition maintained an almost perfect mass balance for the domain, and transmitted almost perfectly surface gravity waves. Flather's solution relied on an iterative scheme to determine the value of the prescribed characteristic variables on the boundaries that would induce the model to reproduce a desired boundary solution for one of the primitive variables. For a continuous model run, such an iterative scheme for determining *a-priori* values for the characteristic variables cannot be used, and so an alternative solution motivated by this technique was trialled. Boundary values of barotropic velocity, U, were prescribed, and combined with earlier values of the model-calculated sea-level η to obtain the prescribed value for the characteristic combination $U+c\eta$. This approach succeeded in producing transport values across the boundaries essentially identical to that prescribed. The ability of the boundary condition to transmit surface gravity waves depended on the timescale on which earlier boundary values of sea-level, η, were incorporated into the prescribed characteristic variable, $U + c\eta$. If the sea-level value from the previous timestep was inserted directly into the prescribed characteristic variable, then the boundary condition became highly reflective (Figure 3.41). If it was inserted into the prescribed characteristic variable with a relaxation timescale of two days, then both surface gravity wave transmission (Figure 3.42) and volume conservation for the domain (Figure 3.44) were maintained perfectly adequately. This represents a generalisation of Flather's boundary condition and ought to be applicable to a wide range of problems from tidally forced coastal models to open ocean models. It would be useful to verify its behaviour under a wider suite of test problems than examined here, and to ensure it works for cases where it is desired to control sea-level on the boundary rather than normal velocity.

The barotropic system in HOPE was the most difficult part to modify for open boundary use. To a large degree this was because of the lack of previous literature implementing open boundary conditions in an implicit barotropic model. For the baroclinic momentum equations, and prognostic tracer equations, results from the literature were more relevant.

For the baroclinic velocities (section 3.6), mathematical theory was of little use due to the ill-posedness of pointwise boundary conditions for the primitive equations. Boundary values were prescribed for both components of internal velocity, u and v, on the northern and eastern boundaries. Values on the western boundary could be calculated as in the interior due to the layout of HOPE's Arakawa E-grid (and values on the southern boundary were not needed since it was land). For the eastern boundary, Miller and Thorpe's (1981) formulation of Orlanski's (1976) radiation condition was used, while a zero-gradient condition was applied on the northern boundary to improve the

model's representation of the East Australia Current. These boundary conditions were found to provide satisfactory solutions in a five year trial run, and no extensive examination of variations was undertaken. It is likely that some difference in results would be obtained, however, if different implementations of the radiation condition were trialled, for example. Tang and Grimshaw (1996) reported on a series of alternate implementations of Orlanski's radiation condition in a barotropic coastal model. It would be of interest to undertake a similar study for HOPE's baroclinic system.

Horizontal advection and diffusion of tracers is a hyperbolic problem in characteristic form. The mathematical theory then suggests an approach, consistent with physical reasoning, that boundary values ought to be prescribed for inflow conditions, and upstream-advected for outflow. This was the approach implemented here (section 3.7). For diffusion on the boundaries, a Neumann condition of zero normal gradient across the boundaries was used to close the problem.

As discussed below, the model performance during a longer term integration degraded considerably. One aspect of this degradation was the failure of the northern boundary condition to maintain the water mass properties of the Tasman Sea. More extensive work on the implementation of the northern boundary would focus on correcting this problem.

6.2 Model performance

Having implemented in the HOPE model boundary conditions that were stable and produced reasonable solutions over a five-year run, a 40 year model integration was performed to examine the model's performance more generally (chapter 4).

While the model remained stable for the duration of the run, there was considerable drift of the solution away from climatology over the first 20–25 years. This was characterised by large-scale volume changes of principal water masses in the domain, and by the migration of major currents away from their mean climatological positions. The most salient illustration of this was the loss of over 4.5 million km^3 (or 10%) of Antarctic Bottom Water (section 4.1.1). Examination of zonally-integrated meridional and vertical volume transports (section 4.1.2) indicated no substantial meridional overturning circulation. The deterioration of the model over 40 years appears most likely to be due to the lack of sea-ice in the model and bottom-water formation around Antarctica. The loss of 4.5 million km^3 of bottom water over 40 years is equivalent to an average rate of around 3.6 Sv over the 80° of longitude covered by the model. Extrapolating this around the globe implies a loss in bottom water production of 16 Sv. This figure is broadly consistent with the range of estimates of bottom water formation from observational and modelling studies (England 1992; Hellmer and Beckmann 2001; Sloyan and Rintoul 2001). The open boundary condition advects in climatological tracer values at points of inflow on the boundary. However, the very small inflows at high latitude on the western boundary were insufficient to replace the dense waters lost through

internal mixing and outflow. The HOPE model incorporates the dynamic-thermodynamic sea-ice model of Hibler (1979), but it was not enabled in this study in order to simplify the altimetry assimilation. The lack of sea-ice and bottom-water formation was not a significant problem for short (one year) assimilation experiments initialised from climatology. An obvious extension of this work, however, would be to examine the model's longer term performance with the sea-ice model enabled.

Horizontal circulation patterns became progressively less realistic over the course of the 40 year run (section 4.1.3). The main mode of evolution was a shift of the ACC southwards, and the development of a strong current extending around the south-eastern corner of Tasmania north along the east coast of Australia. The southwards migration of the ACC is associated with the loss of bottom water and the progression south of the Polar Front. The unrealistic northwards current along the east coast of Australia appears to be due to the loss of warm salty subtropical surface waters in the Tasman Sea. Throughout the development of the open boundary conditions, it was found to be difficult to maintain a realistic East Australia Current. For the barotropic system, an artificial volume transport of 15 Sv had to be imposed in an attempt to maintain a realistic flow. In spite of this imposed volume transport, the inflow at the northern boundary failed to maintain realistic surface water properties in the Tasman Sea. To correct this failure would require more in-depth experiments with the open boundary conditions applied to the baroclinic system and tracers along the northern boundary.

As well as the slow drift in the model, shorter timescale variability was examined. Sea-level variability is initially comparable to that observed in satellite altimetry — both in location and strength. However, after 15 years the maximum variability in the domain is barely one half of that measured (section 4.2.1). It is associated with the major fronts which migrate southwards, and so the locations of maximum variability are south of their observed locations. After 30 years, the maximum variability is concentrated along a narrow meandering jet south of New Zealand, associated with a displaced Polar Front. Analysing the timescales of the sea-level variability shows that over the 40 year run, the largest fraction of the variance is associated with the long term model drift, on timescales longer than a year. Variability at subannual timescales is associated with the strong flow south of New Zealand that persists in a quasi-steady state through the final 15 years of the run. This is the only sea-level variability that can reasonably be associated with mesoscale dynamics. A very strong annual cycle of sea-level variability is observed in the model along the shelf of the Great Australian Bight, most likely associated with the strongly seasonal winds in this region.

The combination of strong drift in the model initially, and very low levels of variability in the wrong places after reaching a quasi-equilibrium subverted the usefulness of a traditional twin experiment (described in section 1.6.2) for evaluating the assimilation scheme. Not only was the mesoscale sea-level variability far too limited spatially and in strength to provide adequate sampling along a satellite-equivalent path, but any results from such an experiment would have provided little reassurance that the scheme would work in the case

of realistic altimetry. If the inclusion of a sea-ice model and improved northern boundary conditions produced a model which maintained realistic flows for a longer period of time, then it would be very informative and useful to perform a twin experiment assimilation. An alternative twin experiment might also be conducted during the initial stages of the model run containing strong drift, but where also mesoscale variability is high and reasonably realistic. What would be required is two different initial states. These could be obtained by adding a small stochastic perturbation to the initial tracer fields of one member of the twins, for instance, or by running both members for a short period with perturbed wind-forcing applied to one. The drift, though strong, should be the same in both cases and be removed during the comparison of the twin results.

Eddy heatflux in the model was poleward at all latitudes (section 4.2.2), consistent with expectations from observation and modelling studies. The values were somewhat low and associated with the regions of strong flow and sea-level variability.

A range of studies (Whitworth and Peterson 1985; Killworth 1992; Pinardi et al. 1995; Phillips and Rintoul 2000) have shown that variability in the Southern Ocean has a significant barotropic component. Examination of the subsurface variability in the Southern Ocean model implemented here provided a similar result (section 4.2.3). The model sea-level was decomposed into steric and bottom-pressure components, and these were linearly regressed against the total sea-level. Significant correlations against both components were found nearly everywhere over the domain. The baroclinic component was correlated positively with the total sea-level, while the barotropic component had a positive correlation at most locations, but a negative correlation in the northeast of the domain, especially in the Tasman Sea. The regression coefficients showed subsurface variability that was primarily baroclinic in nature, but with a significant barotropic component, between 20% and 30% over much of the domain. The detailed spatial distribution of this decomposition is unlikely to accurately characterise the variability of the real ocean, due to the significant drift that occurs, and other model deficiencies. Nevertheless, the general conclusion that sea-level variability in this region has a significant barotropic component is consistent with other studies, and has implications for the implementation of the assimilation scheme. It would be interesting to compare the analysis of subsurface variability in this model with a similar analysis of a hemispheric model such as FRAM.

6.3 Altimetry assimilation

The assimilation scheme implemented was a reduced-order optimal interpolation (ROOI) method (section 5.1). The System for Ocean Forecasting and Analysis (SOFA, section 5.3) of De Mey (2002) was used to perform the optimal interpolation of sea-level corrections based on model forecasts and altimeter measurements.

ROOI applies order reductions in the horizontal and vertical to make

the analysis problem tractable. The vertical order reduction projects analysed sea-level anomalies to depth using linear modes of subsurface variability that have a sea-level signature. The vertical projection scheme used here took account of barotropic variability and enabled a partition of the analysed sea-level correction into barotropic and baroclinic components (section 5.2). The partition of a sea-level correction between these two components was taken as a free parameter in the assimilation experiments.

For the projection of the baroclinic component, the method of Cooper and Haines (1996) was used (section 5.2.1). This applies conservation of potential vorticity and of watermasses on isopycnals to derive a transformation between sea-level anomaly and subsurface tracer fields. The mode of variability implied by these constraints is a uniform vertical displacement of isopycnals (and their embedded tracer profiles), with the sea-level anomaly equal to the resulting steric anomaly.

For the projection of the barotropic component, the sea-level correction was used directly to calculate updates of depth-integrated velocity from geostrophy. In fact such updates are ageostrophic (unless the sea-level correction is constant along contours of f/H), but since the spatial scale of the corrections is much smaller than the barotropic Rossby radius, the imposed velocity anomaly persists in the model, dissipating through the action of bottom torque against the sloping topography.

A dynamical analysis of the vertical projection scheme was performed in some perturbation experiments (section 5.2.3). Initial geostrophic adjustment ensures that sea-level balances both an imposed baroclinic density anomaly and an imposed barotropic velocity anomaly. Subsequent evolution in both cases at timescales of interest is dominated by the excitation of topographic Rossby waves. The baroclinic projection is not 'purely baroclinic' in the sense that it also induces a non-zero depth-integrated velocity anomaly. Because of this (and because sea-level adjusts to balance both density and barotropic velocity) the adjusted sea-level cannot attain the desired value unless the barotropic velocity is compensated, even in the 'baroclinic projection' case. In the perturbation experiments here, the sea-level deficit was up to 20% in the case of a baroclinic projection. The assimilation scheme was not modified to account for this here, nor does it seem to have been in other published applications of the Cooper and Haines scheme. An obvious extension of the method would be to take account of this fact and adjust barotropic velocities to ensure that sea-level adjusts completely to the required value.

Another aspect of the baroclinic projection scheme that does not seem to have been examined in the literature is the role of the Joint Effect of Baroclinicity and Relief (JEBAR) in the dynamics of the scheme. The torque of a depth-integrated density anomaly against sloping topography acts as a forcing term in the barotropic momentum equations. The dynamical effect of this should be examined to fully characterise the behaviour of the Cooper and Haines scheme.

An analysis of five repeat hydrographic sections along the WOCE SR3 line through the model domain showed that a very high fraction of the density variance through the water column can be explained with a uniform displace-

ment model (section 5.2.4). Density anomalies along the sections that were not uniform with depth made only a minor contribution to the total steric sea-level anomaly. In the formalism of ROOI, it is important that the null-space of the transformation be only marginally observable in this manner. A comparison between the steric sea-level anomalies along these sections and the observed anomaly interpolated from altimeter measurements showed a close correspondence, but with evidence of an increasing barotropic (bottom pressure) component at higher latitudes. No attempt was made on this basis to determine a precise relationship for surface sea-level expression of barotropic variability in the Southern Ocean, but this clearly remains an important unresolved issue. The upcoming satellite gravity missions (GRACE and CHAMP) will make major contributions to this question, but until then comparison of in-situ hydrographic data against altimeter measurements may provide the best opportunity to examine the relationship.

A limited set of assimilation experiments was performed in section 5.4. These all assimilated one year of TOPEX/POSEIDON altimeter data taken between June, 1993 and June, 1994. A baseline run was analysed in some detail, and further experiments varied elements of the assimilation procedure relative to parameters used in this baseline run. Assimilation performance was assessed on the basis of various statistics diagnosed in the SOFA package, as well as objective changes in variability (sea-level, eddy heatflux) compared with a reference unassimilated run. In addition, sea-level analyses from SOFA were compared with the suboptimal mapped altimeter fields produced by the French CLS laboratory. Model-equivalent hydrographic sections were interpolated from fields stored every five days and compared with the AU9407 section occupied in January, 1994.

The baseline assimilation run used forecast error correlation length scales of 200 km and an e-folding timescale of 10 days. The forecast and observational error variances were taken as 50% and 10%, respectively, of the mean altimeter variance. A 'purely baroclinic' vertical projection of the analysed sea-level correction was applied at analysis times. Decreases of both maximum and RMS misfit between the model and altimetry were achieved with respect to the unassimilated run, although there was no strong statistical evidence of improved forecast skill. The level and spatial distribution of sea-level variability was improved. Comparison of the SOFA sea-level analyses provided a pleasing correspondence with the CLS mapped altimeter fields. The latter uses both past and future altimeter measurements, and incorporates the spatially denser ERS-1 data in their interpolation scheme. An examination of the post-analysis sea-level anomaly in the baseline assimilation run confirmed that the vertical projection scheme ensured the retention of the sea-level anomaly on timescales longer than the geostrophic adjustment process. The comparison of the AU9407 hydrographic section with the model equivalent was somewhat disappointing as the assimilation scheme failed to reproduce the most significant feature in the hydrography, a large eddy at around 48°S. Detailed examination of the sea-level analysis in the region of the eddy suggested several reasons for the failure. The forecast error correlation length scales seemed to be too large to enable small scale adjustments to eddies in the model, while the forecast

error time scale was too short to ensure that altimetry over a full repeat cycle was included with sufficient weight in each analysis. A more satisfactory sea-level anomaly in the CLS maps was probably due primarily to the inclusion of ERS-1 data. The assimilation scheme significantly increased the eddy heat-flux in the domain, up to levels consistent with those calculated from in-situ measurements south of Australia.

A number of additional assimilation runs were performed, with results reported in this thesis for a reduced forecast error correlation length scale and also a 50% barotropic projection of the analysed sea-level correction. The best results of all the runs were obtained with the 50% barotropic projection. The assimilation statistics demonstrated forecast skill that beat persistence and also an improved subsurface representation of the eddy along the SR3 section.

The set of assimilation experiments reported here must be regarded as a rather cursory investigation into assimilation performance. By no means did it cover the full parameter space of the assimilation scheme, nor did it analyse in detail the performance of the vertical projection scheme under different assimilation parameters. A limited, but important, comparison was made against independent observational data. Clearly a considerable amount of additional work is indicated to fully explore the potential of the assimilation system implemented in this thesis. An obvious first step would be to perform additional assimilation runs making use of the results from the experiments reported here. For instance, a run with smaller length scales and longer timescales in the forecast error correlation function would be expected to better reproduce the eddy seen in the AU9407 hydrographic section. The vertical projection applied in these experiments made a partition between baroclinic and barotropic that was uniform across the domain. The experiments here (and other studies) suggest that this should be spatially varying, with an increased barotropic component at higher latitudes. A better estimate of the level of partitioning between the two components, and its geographical variability, might be obtained through an analysis of a free-surface global or hemispheric model such as FRAM. The inclusion of ERS-1 data in the assimilation scheme is straightforward and ought to significantly improve the representation of small-scale features in the analyses.

Finally, the effect of model resolution on the results obtained in this study is likely to be significant. The grid-spacing is roughly equal to the first baroclinic Rossby radius. Current terminology denotes such models as 'eddy-permitting'. A number of studies have shown that substantial qualitative and quantitative differences occur in numerical simulations as resolution increases, especially upon transition to full eddy-resolution (Schmitz and Thompson 1993; Smith et al. 2000; Megann and New 2001; Oschlies 2002). Typically, kinetic energy increases, sea-level variability better corresponds with altimeter observations, mean circulation improves, and quantities such as heat transport are better simulated. All of these are quantities which the present study has attempted to improve using assimilation. It is obvious that if the model statistics are more realistic prior to assimilation, then the assimilation has a much better chance to correct the synoptic details of variability. Of course, increased resolution comes at a cost. Doubling the resolution typically increases

memory requirements four-fold, and computation time eight-fold. The resolution chosen for this study was the greatest that could be efficiently run on the available resources. A trial run of six months was performed of a model with twice the resolution (0.2° in latitude and 0.3° in longitude). The assimilation code was removed to reduce memory requirements. Visual inspection of surface tracer fields showed significantly greater meandering occured than in the coarser resolution model used in this study, consistent with expectations. While computational resources did not permit it, altimeter assimilation in this higher resolution model would be expected to perform considerably better than found in the present study.

6.4 Conclusion

In conclusion, an altimetry assimilation scheme has been successfully implemented in a limited-area model of the Southern Ocean south of Australia. Significant changes were made to the HOPE model to achieve this. The open boundary condition used was based on characteristic variables and generalised a similar condition used in a coastal model by Flather (1976). The long-term performance of the model was degraded primarily due to a lack of sea-ice thermodynamics. The assimilation scheme significantly improved the variability in the model. A barotropic component in the sea-level projection was needed in order to achieve forecast ability. There remains a considerable amount of further work to fully explore the potential of the assimilation scheme, with the expectation that results could be further improved over the preliminary results presented here.

APPENDIX A

Energy Method for the Shallow Water Equations

Given a hyperbolic initial-boundary value problem in n unknowns and r space dimensions, with boundaries at $x_1 = 0$, $x_1 = 1$,

$$\frac{\partial \boldsymbol{s}}{\partial t} = \sum_{i=1}^{r} \boldsymbol{B}_i \frac{\partial \boldsymbol{s}}{\partial x_i} + \boldsymbol{C}\boldsymbol{s} + \boldsymbol{f} \tag{A.1}$$

$$\boldsymbol{s}(\boldsymbol{x}, 0) = \boldsymbol{h}(\boldsymbol{x}) \tag{A.2}$$

$$\boldsymbol{s}_-(0, \boldsymbol{x}_*, t) = \boldsymbol{Q}_0 \boldsymbol{s}_+(0, \boldsymbol{x}_*, t) + \boldsymbol{g}_0(\boldsymbol{x}_*, t) \tag{A.3}$$

$$\boldsymbol{s}_+(1, \boldsymbol{x}_*, t) = \boldsymbol{Q}_1 \boldsymbol{s}_-(1, \boldsymbol{x}_*, t) + \boldsymbol{g}_1(\boldsymbol{x}_*, t), \tag{A.4}$$

then we make

Definition 1 *The problem (A.1)–(A.4) is* strongly well-posed *if the following* energy inequality *holds for all time intervals $t \in [0, T]$:*

$$\|\boldsymbol{s}\|^2 + \int_0^t \|\boldsymbol{s}\|_\Gamma^2 d\tau \leq K_T \left\{ \|\boldsymbol{h}\|^2 + \int_0^t \left[\|\boldsymbol{g}\|_\Gamma^2 + \|\boldsymbol{f}\|^2 \right] d\tau \right\}, \tag{A.5}$$

where K_T is a constant independent of the data $\boldsymbol{f}$, $\boldsymbol{g}$ and $\boldsymbol{h}$.

Existence and uniqueness of solutions can be shown to follow (Kreiss and Lorenz 1989; Gustafsson et al. 1995). Conceptually, the inequality states that the energy of the solution is bounded by the forcing and boundary data. The L_2 *norm* and *boundary norm* above are defined respectively as follows:

$$\|\boldsymbol{a}\|^2 = \int |\boldsymbol{a}(\boldsymbol{x}, t)|^2 d\boldsymbol{x}$$

$$\|\boldsymbol{a}\|_\Gamma^2 = \int_{x=0} |\boldsymbol{a}(\boldsymbol{x}, t)|^2 d\boldsymbol{x}_* + \int_{x=1} |\boldsymbol{s}(\boldsymbol{x}, t)|^2 d\boldsymbol{x}_*,$$

with $|\boldsymbol{a}|^2 = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ the usual vector norm. In what follows, we will require

Lemma 1 (Gronwall's lemma) *Suppose $a(t)$, $g(t) > 0$, and $h(t) > 0$ are smooth, continuous functions over an interval $t \in [0, T]$, and*

$$a'(t) \leq \lambda a(t) + g(t) - h(t)$$

for some $\lambda \geq 0$. *Then*

$$a(t) + \int_0^t h(\tau)d\tau \leq e^{\lambda t}\left[a(0) + \int_0^t g(\tau)d\tau\right], 0 \leq t \leq T.$$

Proof: Place $z(t) = e^{-\lambda t}a(t)$. Then

$$\begin{aligned} z'(t) &= -\lambda e^{-\lambda t}a(t) + e^{-\lambda t}a'(t) \\ &\leq e^{-\lambda t}g(t) - e^{-\lambda t}h(t). \end{aligned}$$

Now integrate this to yield

$$\begin{aligned} z(t) - z(0) = e^{-\lambda t}a(t) - a(0) &\leq \int_0^t e^{-\lambda\tau}g(\tau)d\tau - \int_0^t e^{-\lambda\tau}h(\tau)d\tau \\ a(t) + \int_0^t e^{-\lambda(\tau-t)}h(\tau)d\tau &\leq e^{\lambda t}\left[a(0) + \int_0^t e^{-\lambda\tau}g(\tau)d\tau\right]. \end{aligned}$$

Now, for $\kappa \geq 0$ and smooth $f(t) > 0$, the following is true:

$$\int_0^t e^{-\kappa\tau}f(\tau)d\tau \leq \int_0^t f(\tau)d\tau \text{ and } \int_0^t e^{\kappa\tau}f(\tau)d\tau \geq \int_0^t f(\tau)d\tau$$

and the required result follows.

Now consider the linearized shallow-water equations (3.11) in characteristic form:

$$\boldsymbol{s}_t = \boldsymbol{\Lambda}\boldsymbol{s}_x + \boldsymbol{B}\boldsymbol{s}_y + \boldsymbol{C}\boldsymbol{s} + \boldsymbol{f}, \tag{A.6}$$

where the characteristic variables

$$\boldsymbol{s} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} v \\ u + \eta^* \\ u - \eta^* \end{pmatrix},$$

and

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\overline{u} & 0 & 0 \\ 0 & -\overline{u} - c & 0 \\ 0 & 0 & -\overline{u} + c \end{pmatrix} = \text{diag}(\lambda_1, \lambda_2, \lambda_3).$$

Initial conditions $\boldsymbol{s}(x, y, 0) = \boldsymbol{h}(x, y)$ are given. Boundaries are at $x = 0$ and $x = 1$ and we assume periodicity in y. We consider the case where $\overline{u} > 0$ so there are two negative eigenvalues, λ_1, λ_2, and one positive eigenvalue λ_3. Thus, two boundary conditions are required at the *inflow boundary*, $x = 0$, and one condition is required at the *outflow boundary*, $x = 1$. The most general form for these (see equations (3.13), (3.14)) is

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} q_{0,1} \\ q_{0,2} \end{pmatrix} s_3 + \begin{pmatrix} g_{0,1} \\ g_{0,2} \end{pmatrix}, \text{ at } x = 0 \tag{A.7}$$

$$s_3 = \begin{pmatrix} q_{1,1} & q_{1,2} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + g_{1,3}, \text{ at } x = 1. \tag{A.8}$$

We wish to determine what values of the coefficients $q_{0,1}$, $q_{0,2}$, $q_{1,1}$, $q_{1,2}$ lead to a well-posed problem according to the definition (A.5). Using the shallow-water equations (A.6), we find

$$\frac{d}{dt}\|\boldsymbol{s}\|^2 = 2(\boldsymbol{s}, \boldsymbol{\Lambda s}_x) + 2(\boldsymbol{s}, \boldsymbol{B s}_y) + 2(\boldsymbol{s}, \boldsymbol{C s}) + 2(\boldsymbol{s}, \boldsymbol{f}), \tag{A.9}$$

where the notation $(\boldsymbol{a}, \boldsymbol{b}) = \int\int \boldsymbol{a}^{\mathrm{T}}(x, y, t)\boldsymbol{b}(x, y, t)dxdy$ has been used for the scalar product of two vector functions. Integrating by parts the first term on the right-hand side, we find

$$2(\boldsymbol{s}, \boldsymbol{\Lambda s}_x) = \int_{x=1} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dy - \int_{x=0} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dy.$$

Similarly, integrating by parts the second term, and assuming periodicity in the $y-$direction,

$$2(\boldsymbol{s}, \boldsymbol{\Lambda s}_y) = \int_{y=y_1} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dx - \int_{y=y_0} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dx = 0.$$

For the third term, we use $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A x} \leq |\boldsymbol{A}|\,|\boldsymbol{x}|^2$ to write

$$2(\boldsymbol{s}, \boldsymbol{C s}) \leq k_1\|\boldsymbol{s}\|^2$$

for some $k_1 > 0$. For the final term, we use the *quadratic inequality* $2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} \leq |\boldsymbol{a}|^2 + |\boldsymbol{b}|^2$ to write

$$2(\boldsymbol{s}, \boldsymbol{f}) \leq \|\boldsymbol{s}\|^2 + \|\boldsymbol{f}\|^2.$$

So, for the *energy evolution equation* (A.9) we now have

$$\frac{d}{dt}\|\boldsymbol{s}\|^2 \leq \int_{x=1} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dy - \int_{x=0} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dy + k_2\|\boldsymbol{s}\|^2 + \|\boldsymbol{f}\|^2,\ k_2 \geq 0. \tag{A.10}$$

If we can show for the *boundary contributions*

$$\int_{x=1} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dy - \int_{x=0} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dy \leq k_3\|\boldsymbol{g}\|_\Gamma^2 - k_4\|\boldsymbol{s}\|_\Gamma^2, \tag{A.11}$$

for some $k_3, k_4 \geq 0$ then the energy inequality may be written

$$\frac{d}{dt}\|\boldsymbol{s}\|^2 \leq k_2\|\boldsymbol{s}\|^2 + \|\boldsymbol{f}\|^2 + k_3\|\boldsymbol{g}\|_\Gamma^2 - k_4\|\boldsymbol{s}\|_\Gamma^2 \tag{A.12}$$

and Gronwall's lemma (Lemma 1) may be applied directly to obtain the required estimate (A.5). Consider first the outflow boundary, $x = 1$. We have

$$\begin{aligned}
\int_{x=1} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Lambda s}\, dy &= \int(\lambda_1 s_1^2 + \lambda_2 s_2^2 + \lambda_3 s_3^2)dy \qquad (A.13)\\
&= \int[\lambda_1 s_1^2 + \lambda_2 s_2^2 + k s_3^2 + (\lambda_3 - k)s_3^2]dy\\
&= \int\left[\lambda_1 s_1^2 + \lambda_2 s_2^2 + k s_3^2 + (\lambda_3 - k)(q_{1,1}s_1 + q_{1,2}s_2 + g_{1,3})^2\right]dy\\
&= \int\left\{[\lambda_1 + (\lambda_3 - k)q_{1,1}^2]s_1^2 + [\lambda_2 + (\lambda_3 - k)q_{1,2}^2]s_2^2 + k s_3^2\right\}dy\\
&\quad + \int(\lambda_3 - k)g_{1,3}^2 dy\\
&\quad + 2\int(\lambda_3 - k)(q_{1,1}q_{1,2}s_1 s_2 + q_{1,1}s_1 g_{1,3} + q_{1,2}s_2 g_{1,3})dy,
\end{aligned}$$

where the boundary condition (A.8) has been used. The approach now will be to use the quadratic inequality to replace "cross terms" in the final integrand with quadratic terms that can be included in the first two integrands of the final expression. Of course, these will add only positive contributions. We already have a few constraints that must be satisfied if the integral is to satisfy (A.11). Since the cross terms will contribute nothing to the coefficient of s_3^2, we must have $k \leq 0$. Note first that if both $q_{1,1} = 0$ and $q_{1,2} = 0$, we trivially satisfy (A.11) so that *prescribing the characteristic variables* certainly provides a well-posed problem. Next, we notice that the coefficient of s_1^2, before adding contributions from cross-terms, is

$$\lambda_1 + (\lambda_3 - k)q_{1,1}^2 = -\overline{u} + (-\overline{u} + c - k)q_{1,1}^2.$$

This is the minimum value that can possibly be attained (it will only increase with cross-term contributions). In order that it be negative as required, we must have

$$q_{1,1}^2 \leq \frac{\overline{u}}{c - \overline{u} - k} \leq \frac{\overline{u}}{c - \overline{u}}.$$

Recalling that the shallow water gravity wave speed $c = \sqrt{gH}$ is typically much greater than a mean current speed $\overline{u}$, this places a severe constraint on the coefficient $q_{1,1}$. With some loss of generality we assume for convenience

$$q_{1,1} = 0 \tag{A.14}$$

and note that, in any case, this is required if we wish the boundary condition (A.8) to prescribe either of the primitive variables u or η^* rather than some combination. With $q_{1,1} = 0$, the final integral reduces to

$$\int 2(\lambda_3 - k)q_{1,2}\, s_2\, g_{1,3}\, dy.$$

Using the quadratic inequality we write

$$2(\lambda_3 - k)q_{1,2}\, s_2\, g_{1,3} \leq p^2 s_2^2 + \frac{(\lambda_3 - k)^2 q_{1,2}^2}{p^2} g_{1,3}^2$$

for any value of the real parameter $p \neq 0$. For the boundary integral (A.13), then, we have finally

$$\begin{aligned} \int_{x=1} \boldsymbol{s}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{s}\, dy \;\leq\; & \int \{\lambda_1 s_1^2 + [\lambda_2 + (\lambda_3 - k)q_{1,2}^2 + p^2]s_2^2 + k s_3^2\}\, dy \\ & + \int [(\lambda_3 - k) + (\lambda_3 - k)^2 q_{1,2}^2/p^2] g_{1,3}^2 dy \end{aligned}$$

Now all coefficients are guaranteed to be of the required sign for (A.11) (negative for $\|\boldsymbol{s}\|^2$, positive for $\|\boldsymbol{g}\|^2$) with the exception of the coefficient for s_2^2. We thus require

$$\lambda_2 + (\lambda_3 - k)q_{1,2}^2 + p^2 \leq 0.$$

Writing this as a constraint for p, we have

$$\begin{aligned} 0 \le p^2 &\le -\lambda_2 - (\lambda_3 - k)q_{1,2}^2 \\ &= \overline{u} + c - (-\overline{u} + c - k)q_{1,2}^2 \\ &= (1 + q_{1,2}^2)\overline{u} + (1 - q_{1,2}^2)c + kq_{1,2}^2 \end{aligned}$$

Without determining precise bounds, we note that $q_{1,2}^2$ may attain the value one, but not much greater. With slight loss of generality, we satisfy ourselves with the bound

$$|q_{1,2}| \le 1 \tag{A.15}$$

and note that $q_{1,2} = 1$ in the boundary condition (A.8) amounts to prescribing the sea-level η^*, while $q_{1,2} = -1$ allows us to prescribe normal velocity u.

Thus, we have proved that for the *outflow* boundary we can obtain well-posed boundary conditions by prescribing either normal velocity u or sea-level η^*.

In similar fashion, consider now the inflow boundary, $x = 0$. The boundary integral

$$\begin{aligned} \int_{x=0} \boldsymbol{s}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{s}\, dy &= \int (\lambda_1 s_1^2 + \lambda_2 s_2^2 + \lambda_3 s_3^2) dy \qquad \text{(A.16)} \\ &= \int [k_1 s_1^2 + k_2 s_2^2 + \lambda_3 s_3^2 + (\lambda_1 - k_1) s_1^2 + (\lambda_2 - k_2) s_2^2] dy \\ &= \int [k_1 s_1^2 + k_2 s_2^2 + \lambda_3 s_3^2 + \\ &\quad (\lambda_1 - k_1)(q_{0,1} s_3 + g_{0,1})^2 + (\lambda_2 - k_2)(q_{0,2} s_3 + g_{0,2})^2] dy \\ &= \int \left\{ k_1 s_1^2 + k_2 s_2^2 + [\lambda_3 + (\lambda_1 - k_1) q_{0,1}^2 + (\lambda_2 - k_2) q_{0,2}^2] s_3^2 \right\} dy \\ &\quad + \int [(\lambda_1 - k_1) g_{0,1}^2 + (\lambda_2 - k_2) g_{0,2}^2] dy \\ &\quad + 2 \int [(\lambda_1 - k_1) q_{0,1} s_3 g_{0,1} + (\lambda_2 - k_2) q_{0,2} s_3 g_{0,2}] dy. \end{aligned}$$

From the condition (A.11), we therefore require in this case that

$$\int_{x=0} \boldsymbol{s}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{s}\, dy \ge -k_3 \|\boldsymbol{g}\|_{x=0}^2 + k_4 \|\boldsymbol{s}\|_{x=0}^2 \tag{A.17}$$

for some positive constants k_3, k_4. Rather than attempt to derive general constraints on the coefficients $q_{0,1}$, $q_{0,2}$, we substitute immediately $q_{0,1} = 0$ and $q_{0,2}^2 = 1$, which are the requirements if the boundary condition is to prescribe the primitive variables v and either u or η^*, rather than a combination. Then the above integral reduces to

$$\begin{aligned} \int_{x=0} \boldsymbol{s}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{s}\, dy &= \int \left\{ k_1 s_1^2 + k_2 s_2^2 + [\lambda_3 + (\lambda_2 - k_2)] s_3^2 \right\} dy \\ &\quad + \int [(\lambda_1 - k_1) g_{0,1}^2 + (\lambda_2 - k_2) g_{0,2}^2] dy \\ &\quad + 2 \int (\lambda_2 - k_2) q_{0,2} s_3 g_{0,2} dy, \end{aligned}$$

with $q_{0,2} = \pm 1$. We again use a quadratic inequality to replace the cross-terms in the final integrand. In this case we must use an inequality of the form $2ab \geq -(a^2 + b^2)$ in order to reproduce the relation (A.17). Thus, only negative terms will be contributed to the coefficients of s_3^2 and $g_{0,2}^2$. However, the coefficient of s_3^2 is

$$\lambda_3 + \lambda_2 - k_2 = -\overline{u} - k_2 < 0.$$

It is impossible, therefore, to obtain a positive coefficient for this term. Thus, we are unable to prove well-posedness at an inflow boundary for conditions which prescribe tangential velocity and either normal velocity or sea-level. This does not prove that such boundary conditions are ill-posed; it simply means that the energy method is incapable of resolving the question. This was noted previously by Oliger and Sundström (1978). A more sophisticated technique is required, such as the *normal mode analysis* of Kreiss (1970). Oliger and Sundström note that Elvius and Kreiss undertook such an analysis in an unpublished work to show that prescribing both v and u is well-posed, while prescribing v and η is ill-posed.

Further examples of the use of the energy method to prove well-posedness of boundary conditions may be found in (Davies 1973a; Davies 1973b; Elvius and Sundström 1973; Gustafsson and Sundström 1978; Hesthaven and Gottlieb 1996). Examples of Kreiss' normal mode analysis to prove well-posedness may be found in (Kreiss 1970; Majda and Osher 1975; Engquist and Majda 1977; Higdon 1994).

Appendix B

The nonreflecting boundary condition of Røed and Cooper

The linear shallow-water equations considered by Røed and Cooper (1987) are written

$$\begin{aligned}\frac{\partial u}{\partial t} &= fv - c^2\frac{\partial \eta}{\partial x}\\ \frac{\partial v}{\partial t} &= -fu - c^2\frac{\partial \eta}{\partial y}\\ \frac{\partial \eta}{\partial t} &= -\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y},\end{aligned}$$

ignoring windstress and bottom friction. In characteristic form these become (for $x-$boundaries)

$$\frac{\partial}{\partial t}(u + c\eta) + c\frac{\partial}{\partial x}(u + c\eta) = fv - c\frac{\partial v}{\partial y} \tag{B.1}$$

$$\frac{\partial}{\partial t}(u - c\eta) - c\frac{\partial}{\partial x}(u - c\eta) = fv + c\frac{\partial v}{\partial y} \tag{B.2}$$

$$\frac{\partial v}{\partial t} = -fu - c^2\frac{\partial \eta}{\partial y}. \tag{B.3}$$

Thus the characteristic variables $(u + c\eta)$ and $(u - c\eta)$ propagate as forced waves along the characteristic curves $dx/dt = c$ and $dx/dt = -c$ respectively while v propagates as a forced wave parallel to the boundaries $dx/dt = 0$.

Hedstrom's (1979) condition for there to be no reflections was simply that the incoming characteristic variables should be constant at each boundary:

$$\begin{aligned}&\frac{\partial}{\partial t}(u + c\eta) = 0 \qquad \text{at a left boundary,}\\ &\frac{\partial}{\partial t}(u - c\eta) = 0 \qquad \text{at a right boundary.}\end{aligned} \tag{B.4}$$

As discussed in section 3.5.4 this reproduces the boundary condition used in this thesis for the barotropic mode. Røed and Cooper proposed, however, (while claiming to base their work on (Hedstrom 1979)) that reflections could

be prevented by formally *setting the slope of the incoming characteristic curves to zero* at the boundaries, in other words setting

$$\begin{aligned} \frac{dx}{dt} &= c \Rightarrow 0 \qquad \text{at a left boundary,} \\ \frac{dx}{dt} &= -c \Rightarrow 0 \qquad \text{at a right boundary.} \end{aligned} \tag{B.5}$$

This amounts to formally setting the term

$$c\frac{\partial}{\partial x}(u + c\eta) \Rightarrow 0 \tag{B.6}$$

in equation (B.1) for a left boundary, for example. Then adding and subtracting equations (B.1) and (B.2) they obtained their nonreflecting conditions

$$\begin{aligned} \frac{\partial u}{\partial t} &= fv + \frac{1}{2}c\frac{\partial}{\partial x}(u - c\eta) \\ \frac{\partial \eta}{\partial t} &= -\frac{\partial v}{\partial y} - \frac{1}{2}\frac{\partial}{\partial x}(u - c\eta) \\ \frac{\partial v}{\partial t} &= -fu - c^2\frac{\partial \eta}{\partial y} \end{aligned} \tag{B.7}$$

at a left boundary, with analogous conditions at a right boundary. The authors offered no justification for their procedure (B.5). Since the characteristic curves arise consequentially from the equations themselves, and are a straightforward property of the system, there seems to be no freedom to arbitrarily alter their slopes.

Jensen (1998) used Røed and Cooper's nonreflecting conditions, but offered an alternative derivation. He first added and subtracted equations (B.1) and (B.2) to obtain

$$\begin{aligned} \frac{\partial u}{\partial t} + c(\mathcal{L}_1 - \mathcal{L}_2) - fv &= 0 \\ \frac{\partial \eta}{\partial t} + (\mathcal{L}_1 + \mathcal{L}_2) + \frac{\partial v}{\partial y} &= 0 \end{aligned}$$

where

$$\begin{aligned} \mathcal{L}_1 &= \frac{1}{2}\frac{\partial}{\partial x}(u + c\eta) \\ \mathcal{L}_2 &= \frac{1}{2}\frac{\partial}{\partial x}(u - c\eta). \end{aligned}$$

He stated imprecisely that this represents (at a right boundary) an outgoing wave propagating with velocity c, and an incoming wave propagating with velocity $-c$; which it certainly does *for the characteristic variables* $(u + c\eta)$ and $(u - c\eta)$ respectively. However, he then claimed that reflections could be eliminated by applying the Røed and Cooper substitution

$$\mathcal{L}_2 = 0,$$

leading to the same boundary conditions

$$\begin{aligned}\frac{\partial u}{\partial t} &= fv - \frac{1}{2}c\frac{\partial}{\partial x}(u + c\eta)\\ \frac{\partial \eta}{\partial t} &= -\frac{\partial v}{\partial y} - \frac{1}{2}c\frac{\partial}{\partial x}(u + c\eta)\end{aligned}$$

for a right boundary.

I show now that for *slowly varying flows*, Røed and Cooper's boundary condition approximately reproduces the correct condition for nonreflection (B.4). Substituting the boundary condition (B.7) in the characteristic equation (B.1) gives

$$\frac{\partial}{\partial t}(u + c\eta) + c\frac{\partial}{\partial x}(u + c\eta) = \left(fv - c\frac{\partial v}{\partial y}\right) + c\frac{\partial}{\partial x}(u + c\eta) = fv - c\frac{\partial v}{\partial y}.$$

From which we have

$$c\frac{\partial}{\partial x}(u + c\eta) = 0 \tag{B.8}$$

$$\frac{\partial}{\partial t}(u + c\eta) = fv - c\frac{\partial v}{\partial y}. \tag{B.9}$$

If the flow is slowly varying in time, then we make the geostrophic approximation

$$\begin{aligned}fv &\approx c^2\frac{\partial \eta}{\partial x}\\ \frac{\partial v}{\partial y} &\approx -\frac{\partial u}{\partial x}.\end{aligned}$$

Substituting these in (B.9) and using (B.8), we find

$$\frac{\partial}{\partial t}(u + c\eta) \approx 0,$$

which is the correct characteristic condition for nonreflection (B.4).

Røed and Cooper's condition appears to be making a recent resurgence in open boundary modelling (Jensen 1998; Palma and Matano 1998; Palma and Matano 2000). It is important to realise it is in fact only an approximately correct construction of Hedstrom's (1979) condition. The dubious assertions required for its derivation would more legitimately be abandoned in favour of the simpler exact condition (B.4), which gives directly the condition used in this thesis. Alternatively, if conditions are required for the primitive variables u, η instead of their characteristic combinations, the nonreflecting condition (B.4) may be combined directly with (B.1)–(B.3) to easily obtain

$$\begin{aligned}\frac{\partial u}{\partial t} &= \frac{1}{2}fv + \frac{1}{2}c\frac{\partial v}{\partial y} + \frac{1}{2}c\frac{\partial}{\partial x}(u - c\eta) \qquad \text{(B.10)}\\ \frac{\partial \eta}{\partial t} &= -\frac{1}{2}\frac{fv}{c} - \frac{1}{2}\frac{\partial v}{\partial y} - \frac{1}{2}\frac{\partial}{\partial x}(u - c\eta)\\ \frac{\partial v}{\partial t} &= -fu - c^2\frac{\partial \eta}{\partial y},\end{aligned}$$

which should be compared with Røed and Cooper's condition (B.7).

Each of the abovementioned authors (Røed and Cooper 1987; Jensen 1998; Palma and Matano 1998; Palma and Matano 2000) applied Røed and Cooper's condition (B.7) to a coastal model, with most of the experiments setting up coastal Kelvin waves. For these waves, the cross-shelf velocity component, v, is zero (Gill 1982, §10.4) and the conditions (B.7) and (B.10) become equivalent. While the condition (B.7) generally performed reasonably in the coastal models, it is expected to be less successful in more general situations.

Appendix C

Implementation of barotropic boundary condition on normal velocity in HOPE

While a messy set of formulas could be provided to encapsulate the method whereby the barotropic system is modified for an open boundary, a more descriptive, but less concise, approach is preferred. The detailed formulas are particularly cumbersome and unenlightening, and unlikely to be of interest to anyone except another HOPE modeller implementing open boundaries who, in any case, would need to understand their derivation in order to suit their particular problem.

Recalling from page 49 the method of solution for the barotropic system, sea-level is calculated prognostically from the continuity equation

$$\frac{\partial \eta}{\partial t} = -\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right) \tag{C.1}$$

using the four surrounding velocity points as shown in Figure C.1. If any one of the four velocity points is on a boundary, then the system matrix (see equations (2.41), (2.42)) must be modified for that central sea-level point. Figure C.2 shows HOPE's E-grid with the notional boundary-lines and affected sea-level points.

Because of HOPE's separation of the E-grid into EVEN and ODD subgrids, there is an extra column of sea-level points outside the western boundary and an extra column of velocity points outside the eastern boundary. These are ignored in the construction of the numerical scheme for the boundary. Their values are set separately (described later) after the interior and boundary has been calculated at each timestep.

Thus, referring to Figure C.2, there are two sets of sea-level points to be dealt with at the open boundaries — those exactly on the boundary, and those just inside the boundary. To simplify the discussion, only the formulation of the western boundary will be described here. An analogous procedure is followed for the eastern and northern boundaries.

Consider first a sea-level point just inside the boundary, Figure C.3. As for the interior, sea-level on this point is calculated prognostically from convergence on the four surrounding velocity points. The difference here is that U at the western velocity point (labelled "W" in Figure C.3) has applied

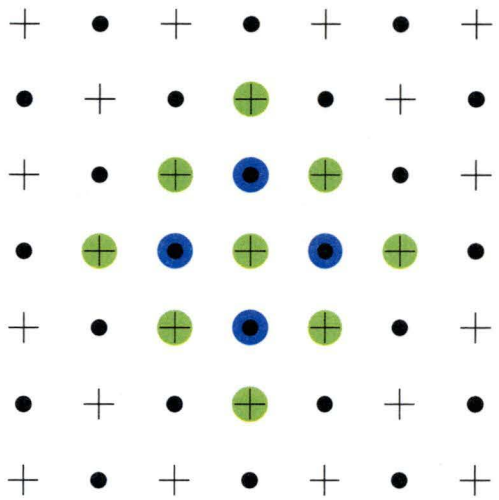


Figure C.1: The central sea-level point is calculated from convergence of four surrounding velocity points, each of which depends implicitly on gradients of surrounding sea-level points.

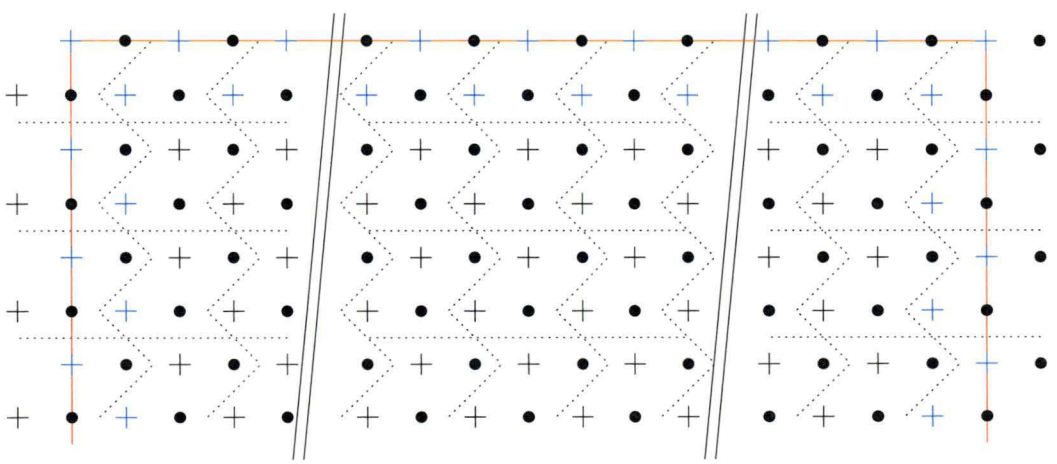


Figure C.2: Model boundary is marked in red, sea-level points affected by the boundary are marked in blue.

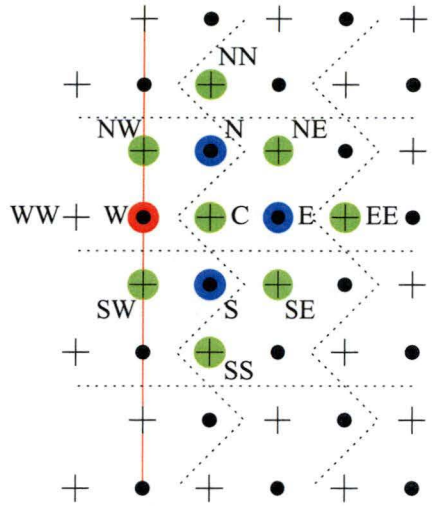


Figure C.3: Sea-level point just inside the western boundary.

to it an *open boundary condition*, rather than being calculated prognostically according to equation (2.38). Thus when the prognostic momentum equations (2.38), (2.39) are substituted into the continuity equation (2.40) to eliminate U and V before solving directly for η, the open boundary condition must be used for the western velocity point, instead of the U-momentum equation (2.38).

This can be described a little more explicitly. Suppose the open boundary condition for the western U point takes the general form (3.21), reproduced here:

$$U_W^{n+1} = \mathcal{L}(U^{n+1}, V^{n+1}, \eta^{n+1}) + \mathcal{G}(U^n, V^n, \eta^n) + \mathcal{F}^{n+1}. \tag{C.2}$$

(We recall from the discussion in section 3.5 that $\mathcal{L}$ must be linear, while $\mathcal{G}$ may be nonlinear and $\mathcal{F}$ allows for prescribing U.) The *implicit term*, $\mathcal{L}$ must first be reduced to an expression containing only η, by substituting the prognostic momentum equations (2.38) and (2.39) for any references to U^{n+1} and V^{n+1} respectively. As an example, suppose the open boundary condition on U was an *implicit gravity-wave radiation* condition:

$$\begin{aligned}
\left(\frac{\partial U}{\partial t}\right)_W^{n+1} &= c\left(\frac{\partial u}{\partial x}\right)_W^{n+1} \\
\frac{U_W^{n+1} - U_W^n}{\Delta t} &= c\frac{U_E^{n+1} - U_W^{n+1}}{\Delta x} \\
U_W^{n+1} &= \frac{1}{1+r}U_W^n + \frac{r}{1+r}U_E^{n+1},
\end{aligned}$$

where $c = \sqrt{gH}$, $r = c\Delta t/\Delta x$ and the subscripts refer to the labels in Figure C.3. One-sided differences have been used for the spatial derivative at the western boundary. Then, U_E^{n+1} in this expression must be replaced using its prognostic momentum equation (2.38), reproduced here:

$$\begin{aligned}
U_E^{n+1} = U_E^n &+ \alpha\Delta t\left[fV^{n+1} - gH\frac{\partial \eta}{\partial x}^{n+1} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial x}^{n+1} dz\right]_E \\
&+ (1-\alpha)\Delta t\left[fV^n - gH\frac{\partial \eta}{\partial x}^{n} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial x}^{n} dz\right]_E,
\end{aligned} \tag{C.3}$$

Table C.1: Usual contributions of surrounding velocity points to a given sea-level coefficient in HOPE's barotropic system matrix.

velocity point	*contribution to system matrix*
N	$\frac{1}{1+(f_N\alpha\Delta t)^2}[f_N H_N \frac{\alpha^2\beta g(\Delta t)^3}{\Delta x_\zeta \Delta y}(\zeta_{NE}^{n+1}-\zeta_{NW}^{n+1}) - H_N \frac{\alpha\beta g(\Delta t)^2(\Delta x_u)_N}{\Delta x_\zeta \Delta y(\Delta y)_N}(\zeta_{NN}^{n+1}-\zeta_C^{n+1})]$
S	$\frac{-1}{1+(f_S\alpha\Delta t)^2}[f_S H_S \frac{\alpha^2\beta g(\Delta t)^3}{\Delta x_\zeta \Delta y}(\zeta_{SE}^{n+1}-\zeta_{SW}^{n+1}) - H_S \frac{\alpha\beta g(\Delta t)^2(\Delta x_u)_S}{\Delta x_\zeta \Delta y(\Delta y)_S}(\zeta_C^{n+1}-\zeta_{SS}^{n+1})]$
E	$\frac{-1}{1+(f_E\alpha\Delta t)^2}[f_E H_E \frac{\alpha^2\beta g(\Delta t)^3}{\Delta x_\zeta \Delta y}(\zeta_{NE}^{n+1}-\zeta_{SE}^{n+1}) + H_E \frac{\alpha\beta g(\Delta t)^2}{\Delta x_\zeta (\Delta x_u)_E}(\zeta_{EE}^{n+1}-\zeta_C^{n+1})]$
W	$\frac{1}{1+(f_W\alpha\Delta t)^2}[f_W H_W \frac{\alpha^2\beta g(\Delta t)^3}{\Delta x_\zeta \Delta y}(\zeta_{NW}^{n+1}-\zeta_{SW}^{n+1}) + H_W \frac{\alpha\beta g(\Delta t)^2}{\Delta x_\zeta (\Delta x_u)_W}(\zeta_C^{n+1}-\zeta_{WW}^{n+1})]$

where the sea-level gradients

$$\left(\frac{\partial\eta}{\partial x}\right)_E = \frac{\eta_{EE}-\eta_C}{\Delta x}.$$

It should be clear that we can write a final expression

$$U_W^{n+1} = \mathcal{L}^*(\eta^{n+1}) + \mathcal{G}^*(U^n, V^n, \eta^n) + \mathcal{F}^{n+1} \tag{C.4}$$

for the open boundary condition on U_W. This expression replaces the prognostic momentum equation (2.38) when U^{n+1} and V^{n+1} are eliminated from the continuity equation (2.40).

So finally, we have the implicit equation for sea-level (2.41),

$$\eta_C^{n+1} + \Gamma^{n+1} = \Gamma^n.$$

This forms a row in the matrix equation

$$\boldsymbol{A}\underline{\eta}^{n+1} = \underline{\Gamma}^n.$$

The net result of applying the boundary condition (C.2) is that coefficients of $\boldsymbol{A}$ in this row will be different from what they would have been had point C been a normal interior sea-level point; also the right-hand side for this row will be different.

To be even more explicit about the differences, Table C.1 lists the contributions of each surrounding velocity point to coefficients in the system matrix $\boldsymbol{A}$ for a normal interior gridpoint C. Table C.2 similarly lists the usual contributions to the right-hand side. For the definitions of symbols, the reader is referred to the HOPE manual (Wolff et al. 1997, §6). For the velocity point "W" on the boundary, the corresponding contributions to both the system matrix and the right-hand side must be replaced with the contributions arising from the boundary condition (C.4): $\mathcal{L}^*$ contributes to the system matrix, while $\mathcal{G}^* + \mathcal{F}$ provides the right-hand side.

So far, the changes to the system matrix and right-hand side have been described for a sea-level point just inside the western boundary. We now

Table C.2: Usual contributions of surrounding velocity points to right-hand side.

velocity point	*contribution to right-hand side*
N	$\frac{\beta\Delta t(\Delta x_u)_N}{\Delta y \Delta x_\zeta} \frac{1}{1+(f_N \alpha \Delta t)^2} (f\alpha\Delta t\Gamma_U - \Gamma_V)_N$
S	$-\frac{\beta\Delta t(\Delta x_u)_S}{\Delta y \Delta x_\zeta} \frac{1}{1+(f_S \alpha \Delta t)^2} (f\alpha\Delta t\Gamma_U - \Gamma_V)_S$
E	$-\frac{\beta\Delta t}{\Delta x_\zeta} \frac{1}{1+(f_E \alpha \Delta t)^2} (\Gamma_U + f\alpha\Delta t\Gamma_V)_E$
W	$\frac{\beta\Delta t}{\Delta x_\zeta} \frac{1}{1+(f_W \alpha \Delta t)^2} (\Gamma_U + f\alpha\Delta t\Gamma_V)_W$

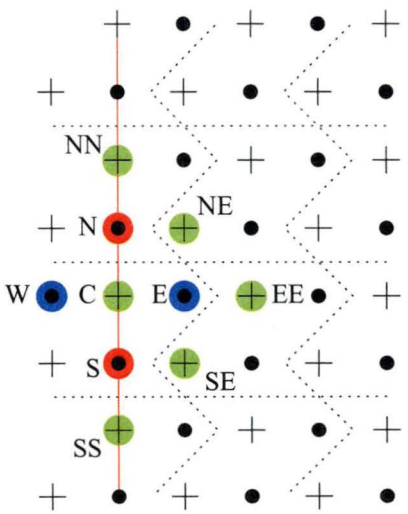


Figure C.4: Sea-level point exactly on the western boundary. Note that velocity point "W" is a *virtual* velocity point.

consider a sea-level point exactly on the boundary, shown in Figure C.4. Recall that the sea-level is calculated prognostically from velocity convergence (C.1). We see that in this case it is the *meridional* components of velocity on the boundaries, V_N and V_S, that will affect the sea-level calculation. Since our boundary condition is being applied to U on the boundary, both V_N and V_S must be calculated using the normal prognostic momentum equation for V (2.39), reproduced here:

$$\begin{aligned} V^{n+1}_{\{N,S\}} &= V^n + \alpha\Delta t\left[-fU^{n+1} - gH\frac{\partial \eta}{\partial y}^{n+1} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial y}^{n+1} dz\right]_{\{N,S\}} \\ &\quad + (1-\alpha)\Delta t\left[-fU^n - gH\frac{\partial \eta}{\partial y}^{n} - \frac{1}{\rho_0}\int_{-H}^{0}\frac{\partial p'}{\partial y}^{n} dz\right]_{\{N,S\}}. \end{aligned} \tag{C.5}$$

We see, however, that the U boundary condition enters via the Coriolis term. Thus, this boundary condition,

$$U^{n+1}_{\{N,S\}} = \mathcal{L}(U^{n+1}, V^{n+1}, \eta^{n+1}) + \mathcal{G}(U^n, V^n, \eta^n) + \mathcal{F}^{n+1}, \tag{C.6}$$

must be substituted into (C.5) before it, in turn, is substituted into the continuity equation (2.40) to obtain the final implicit equation for the sea-level point C (2.41). In this case, we also require an expression for the velocity U^{n+1}_W on the *virtual* velocity point "W" outside the boundary (Figure C.4). It is at this point we implement the averaging between subgrids on the boundary referred to in section 3.5. We therefore write

$$U^{n+1}_W = \frac{1}{2}(U^{n+1}_N + U^{n+1}_S), \tag{C.7}$$

and also

$$U^{n}_W = \frac{1}{2}(U^{n}_N + U^{n}_S),$$

since the stability parameter β allows for a scheme which is not fully implicit (see equation (2.40) and the discussion on page 49). For convenience, the continuity equation for η^{n+1} (2.40) is reproduced here[1]:

$$\begin{aligned} \eta^{n+1} &= \eta^n - \beta\Delta t\left(\frac{\partial U}{\partial x}^{n+1} + \frac{\partial V}{\partial y}^{n+1}\right) - (1-\beta)\Delta t\left(\frac{\partial U}{\partial x}^{n} + \frac{\partial V}{\partial y}^{n}\right) \\ &= \eta^n - \beta\Delta t\left[(U^{n+1}_E - U^{n+1}_W)/\Delta x + (V^{n+1}_N - V^{n+1}_S)/\Delta y)\right] \\ &\quad - (1-\beta)\Delta t\left[(U^{n}_E - U^{n}_W)/\Delta x + (V^{n}_N - V^{n}_S)/\Delta y)\right] \end{aligned}$$

Summarising, the boundary condition (C.6) must be substituted at several places into this: first via the Coriolis terms in the expressions for $V^{n+1}_{\{N,S\}}$ (C.5); and, second, via the expression for U^{n+1}_W (C.7) obtained by averaging adjacent boundary velocity points. Of course, as usual, any references in the boundary condition (C.6) to U^{n+1} or V^{n+1} must also be eliminated by substituting the

[1]For the sake of clarity, factors accounting for convergence of meridians with latitude have been omitted on the V terms in this expression.

relevant prognostic momentum equation, (2.38) or (2.39). Finally, then, we have again an implicit equation for sea-level

$$\eta_C^{n+1} + \Gamma^{n+1} = \Gamma^n,$$

and again the usual contributions to the system matrix (Table C.1) and the right-hand side (Table C.2) for the velocity points "N", "S" and "W" must be replaced by the contributions derived from the above analysis.

So we have now described how the system matrix and right-hand side may be constructed for a boundary condition on normal velocity. In HOPE, the system matrix is built and triangularised in the subroutine `OCTRIAN` once at the beginning of a model run; the right-hand side is constructed each timestep in the routine `OCRHSZ`. Thus, these must be modified as described above in order to implement open boundaries. In addition, subroutine `OCTOPO` numbers all sea-level points at the very start of a model run. This numbering is used to define the ordering of sea-level points within the system matrix $\boldsymbol{A}$. Noting that the EVEN grid has a column of sea-level points outside the western boundary (and that these are not involved in the open boundary barotropic system), it is clear that these should not be numbered and `OCTOPO` must be modified accordingly. Finally, after having solved for the new sea-level points (routine `ZGAUSS`), the subroutine `OCVTRO` is used to calculate the new barotropic velocities from the prognostic momentum equations (2.38) and (2.39). Then this routine must also be modified, to calculate U on the western boundary (U on the eastern, V on the northern) according to the boundary condition rather than the momentum equations.

It remains to describe what can be done for the EVEN sea-level points outside the western boundary and the ODD velocity points outside the eastern boundary. For the sea-level points, since they are not used for the barotropic solution, and since sea-level is not used in the model except in the barotropic system, they may have any values whatsoever and not affect any part of an integration. They may be left at zero for all time, or may be set with a zero-gradient condition (or something else) if they need reasonable values for postprocessing purposes. The barotropic velocities outside the eastern boundary are a different matter, since velocities on these points are used to advect tracers on the boundary, and for viscous mixing and the calculation of vertical velocity. The method for dealing with them is straightforward and obvious from the discussion above — U is simply set to the average of the adjacent velocity points on the boundary. In fact in this case, there is no need for a *virtual* point for U outside the boundary. V on these points may also be set as the average of the boundary points, or using a zero-gradient condition. The latter was done for the runs in this thesis.

APPENDIX D

Implementation of barotropic Flather boundary condition in HOPE

Without going into the level of detail of Appendix C for the implementation of a boundary condition on normal velocity, a few technical details of the implementation of the Flather condition are briefly discussed.

There are three important stages in its development. First, in order to have co-located sea-level and velocity points, 'virtual' velocity points are carried at all boundary sea-level points. Referring to Figure D.1(a) for the western boundary, the (ODD) point C therefore represents both a (real) sea-level point and a (virtual) velocity point. The Flather condition at this point is written

$$U_C^{n+1} + c\eta_C^{n+1} = U_{C_0} + c\eta_{C_0}. \tag{D.1}$$

Recall (page 49) that HOPE solves the barotropic system first by substituting the prognostic momentum equations into the continuity equation to eliminate barotropic velocities at the new time level $n+1$. An implicit equation for sea-level results. The continuity equation is repeated here for convenience:

$$\frac{\partial \eta}{\partial t} = -\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right).$$

In the present situation, this becomes (ignoring the $y-$derivative for the moment, and also ignoring the stability parameter β of equation (2.40) for simplicity)

$$\frac{\eta_C^{n+1} - \eta_C^n}{\Delta t} = -\left(\frac{U_E^{n+1} - U_W^{n+1}}{\Delta x} + \cdots\right). \tag{D.2}$$

The velocity point W, however, is non-existent at the boundary of HOPE's staggered E-grid.

A second technical condition for implementing the Flather condition in HOPE is then to assume that the velocity on this missing point is the same as that on the 'virtual' velocity point C,

$$U_W^{n+1} = U_C^{n+1}.$$

Thus, the discrete Flather condition (D.1) can be substituted directly into (D.2) for U_W^{n+1}. An expression for U_E^{n+1} can be written in the usual way (see

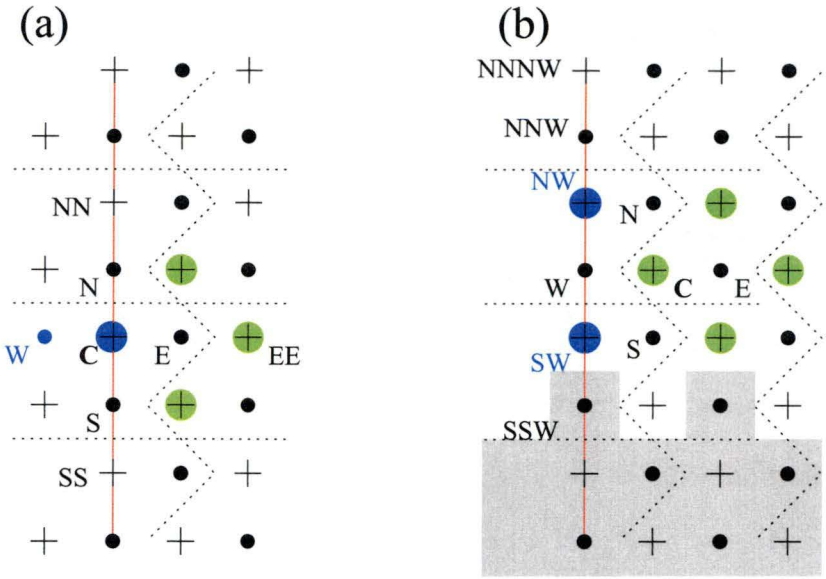


Figure D.1: Grid implementation of Flather condition for sea-level points (a) on, and (b) adjacent to, western boundary.

the HOPE manual (Wolff et al. 1997)) in terms of sea-level at the new time-level on the four coloured sea-level points in Figure D.1(a). Thus we have eliminated the barotropic velocities U_W^{n+1} and U_E^{n+1} at the new time-level from the discretised continuity equation, leaving an implicit equation in sea-level (the $\partial V/\partial y$ term will be dealt with shortly).

Consider now an EVEN sea-level point adjacent to the western boundary, point C in Figure D.1(b). The continuity equation (D.2) is applied at this point. As before, the usual expression for U_E^{n+1} in terms of surrounding sea-level points is used; the Flather boundary condition enters via the term U_W^{n+1}. Thus we come to the third key element in the implementation of the condition. As with the implementation for conditions on normal velocity (Appendix C), we average the transport between the two subgrids, so that we write

$$U_W^{n+1} = \frac{1}{2}(U_{NW}^{n+1} + U_{SW}^{n+1}). \tag{D.3}$$

Again, the points NW and SW are 'virtual' velocity points. Applying the Flather condition to these, we have

$$\begin{aligned} U_{NW}^{n+1} &= -c\eta_{NW}^{n+1} + (U_{NW_0} + c\eta_{NW_0}) \\ U_{SW}^{n+1} &= -c\eta_{SW}^{n+1} + (U_{SW_0} + c\eta_{SW_0}). \end{aligned} \tag{D.4}$$

Using (D.4) together with (D.3), and substituting in the continuity equation (D.2) we have again eliminated U_W^{n+1} and have remaining an implicit equation for sea-level at the new time-level involving now all coloured sea-level points in Figure D.1(b).

Before discussing the term $\partial V/\partial y$ in the continuity equation that was neglected above, an important subtlety of the averaging procedure must be mentioned. Under the procedure just described, the velocity U on EVEN

(real) vector points on the boundary is taken as the average of the two adjacent virtual velocity points. Thus, in Figure D.1(b), U_W is averaged from U_{NW} and U_{SW}. In this way, the total transport into each subgrid would be the same *except for the influence of land points.* Consider, for example, the 'virtual' velocity point NW. It will contribute equally half its transport to both the velocity points NNW and W. On the other hand, the 'virtual' point SW, which would normally contribute equally to both W and SSW, is prevented from doing so because the point SSW is land in this case. Averaged along the western boundary, the transport into the EVEN grid will be lower than the transport into the ODD grid by an amount $U_{SW}/2$ (with a similar deficit at the northwest corner). To correct this, the expression (D.3) must be replaced with

$$U_W^{n+1} = \frac{1}{2}(U_{NW}^{n+1} + 2U_{SW}^{n+1}).$$

Such a replacement must be made wherever the *next* real velocity point on a boundary is land, and at the corners. If this is not done, significant separation of sea-level occurs between the two subgrids over time.

Finally, the meridional divergence term in the continuity equation, $\partial V/\partial y$, must be discussed. For the EVEN sea-level point adjacent to the western boundary in Figure D.1(b) this presents no problems, and the usual formula replacing V^{n+1} by an expression in surrounding sea-level points (Wolff et al. 1997) can be used. For a *boundary* sea-level point, however (Figure D.1(a)), we first note that the prognostic momentum equation for V involves U through the Coriolis term, see equation (2.39). Since we have replaced the prognostic equation for U with the Flather condition (D.1), this must be used together with (2.39) when substituting

$$\frac{\partial V}{\partial y} = \frac{V_N^{n+1} - V_S^{n+1}}{\Delta y}$$

in the continuity equation.

References

Abarbanel, S. S., and E. M. Murman, 1981: Stability of two-dimensional hyperbolic initial boundary value problems for explicit and implicit schemes, In (NASA 1981), pp. 199–207.

Adcroft, A., C. Hill and J. Marshall, 1997: Representation of topography by shaved cells in a height coordinate ocean model, *Monthly Weather Review*, **125**(9), 2293–2315.

Anderson, D. L., and J. Willebrand (eds.), 1989: *Oceanic Circulation Models: Combining Data and Dynamics*, number Series C: Vol. 284 in NATO ASI Series, Kluwer Academic Publishers.

Anderson, D. L., and J. Willebrand, 1992: Recent advances in modelling the ocean circulation and its effects on climate, *Reports on Progress in Physics*, **55**(1), 1–37.

Anderson, D., J. Sheinbaum and K. Haines, 1996: Data assimilation in ocean models, *Reports on Progress in Physics*, **59**(10), 1209–1266.

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I, *Journal of Computational Physics*, **1**, 119–143.

Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model, *Methods in Computational Physics*, **17**, 173–265.

Baker, Jr., D. J., 1982: A note on Sverdrup balance in the Southern Ocean, *Journal of Marine Research*, **40**(supplement), 21–26.

Barnier, B., P. Marchesiello, A. P. de Miranda, J.-M. Molines and M. Coulibaly, 1998: A sigma-coordinate primitive equation model for studying the circulation in the South Atlantic. Part I: Model configuration with error estimates, *Deep-Sea Research*, **45**, 543–572.

Beardsley, R. C., and D. B. Haidvogel, 1981: Model studies of the wind-driven transient circulation in the middle Atlantic Bight. part I: Adiabatic boundary conditions, *Journal of Physical Oceanography*, **11**, 355–375.

Beckers, J.-M., 1999: On some stability properties of the discretisation of damped propagational shallow-water inertia-gravity waves on the Arakawa B-grid, *Ocean Modelling*, **2**(2–4).

Benada, J. R., 1997: *Merged GDR (TOPEX/POSEIDON) Generation B (MGDR-B) Users Handbook, Version 2.0*, Physical Oceanography Distributed Active Archive Center, JPL.

Benada, R., 1993: *Merged GDR (TOPEX/POSEIDON) Users Handbook, Version 1.0*, Physical Oceanography Distributed Active Archive Center, JPL.

Bengtsson, L., M. Ghil, and E. Källén (eds.), 1981: *Dynamic Meteorology: Data Assimilation Methods*, vol. 36 of *Applied Mathematical Sciences*, Springer-Verlag.

Bennett, A., 1976: Open boundary conditions for dispersive waves, *Journal of the Atmospheric Sciences*, **33**, 176–182.

Bennett, A., and B. Chua, 1999: Open boundary conditions for Lagrangian geophysical fluid dynamics, *Journal of Computational Physics*, **153**, 418–436.

Bennett, A., and P. Kloeden, 1978: Boundary conditions for limited-area forecasts, *Journal of the Atmospheric Sciences*, **35**, 990–996.

Bergman, K. H., 1979: Multivariate analysis of temperatures and winds using optimum interpolation, *Monthly Weather Review*, **107**(11), 1423–1444.

Berry, P., and J. Marshall, 1989: Ocean modelling studies in support of altimetry, *Dynamics of Atmospheres and Oceans*, **13**, 269–300.

Best, S., V. Ivchenko, K. Richards, R. Smith and R. Malone, 1999: Eddies in numerical models of the Antarctic Circumpolar Current and their influence on the mean flow, *Journal of Physical Oceanography*, **29**(3), 328–350.

Bindoff, N. L., and J. A. Church, 1992: Warming of the water column in the southwest Pacific Ocean, *Nature*, **357**, 59–62.

Bindoff, N. L., and T. J. McDougall, 1994: Diagnosing climate change and ocean ventilation using hydrographic data, *Journal of Physical Oceanography*, **24**(6), 1137–1152.

Bindoff, N. L., M. A. Rosenberg and M. J. Warner, 2000: On the circulation and water masses over the Antarctic continental slope and rise between 80 and 150°E, *Deep-Sea Research*, **47**(12–13), 2299–2326.

Blumberg, A., and G. Mellor, 1987: A description of a three-dimensional coastal ocean circulation model, in *Three-dimensional coastal ocean models*, edited by N. Heaps, pp. 1–16, American Geophysical Union, Washington, DC.

Blumberg, A. F., and L. H. Kantha, 1985: Open boundary condition for circulation models, *Journal of Hydraulic Engineering*, **111**(2), 237–255.

Boussinesq, J., 1903: *Théorie analytique de la chaleur*, vol. 2, Gauthier-Villars, Paris.

Bouttier, F., and P. Courtier, 1992: Data assimilation concepts and methods, meteorological training course lecture series, [Online] Available: http://www.ecmwf.int/services/training/rcourse_notes/pdf_files/Assim_concepts.pdf.gz [2001, January 3].

Brasseur, P. P., and J. C. Nihoul (eds.), 1994: *Data Assimilation: Tools for Modelling the Ocean in a Global Change Perspective*, vol. 19 of *NATO ASI Series I*, Springer-Verlag.

Broecker, W. S., 1991: The great ocean conveyor, *Oceanography*, **4**(2), 79–89.

Broecker, W. S., 1997: Thermohaline circulation, the Achilles Heel of our climate system: Will man-made CO_2 upset the current balance, *Science*, **278**, 1582–1588.

Bryan, F., 1986: High-latitude salinity effects and interhemispheric thermohaline circulations, *Nature*, **323**, 301–304.

Bryan, F. O., and W. R. Holland, 1989: A high resolution simulation of the wind- and thermohaline- driven circulation in the North Atlantic Ocean, in *Parameterization of small-scale processes, proceedings Aha Hulikoa Hawaiian Winter Workshop*, Honolulu, Hawaii, University of Hawaii, Hawaii Institute of Geophysics.

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

Bryden, H. L., 1979: Poleward heat flux and conversion of available potential energy in Drake Passage, *Journal of Marine Research*, **37**(1), 1–22.

Bryden, H. L., and R. A. Heath, 1985: Energetic eddies at the northern edge of the Antarctic Circumpolar Current in the Southwest Pacific, *Progress in Oceanography*, **14**, 65–87.

Bunyard, P., 1999: How global warming could cause Northern Europe to freeze, *The Ecologist*, **29**(2), 79–80.

Camerlengo, A. L., and J. J. O'Brien, 1980: Open boundary conditions in rotating fluids, *Journal of Computational Physics*, **35**, 12–35.

Carissimo, B., A. Oort and T. V. Haar, 1985: Estimating the meridional energy transports in the atmosphere and oceans, *Journal of Physical Oceanography*, **15**, 82–91.

Carnes, M., D. Fox, R. Rhodes and O. Smedstad, 1996: Data assimilation in a North Pacific Ocean monitoring and prediction system, In Malanotte-Rizzoli

(Malanotte-Rizzoli 1996), pp. 319–345.

Chapman, D. C., 1985: Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model, *Journal of Physical Oceanography*, **15**, 1060–1075.

Charney, J., R. Fjørtoft and J. von Neumann, 1950: Numerical integration of the barotropic vorticity equation, *Tellus*, **2**(4), 237–254.

Chassignet, E. P., and J. Verron (eds.), 1998: *Ocean Modeling and Parameterization*, vol. 516 of *NATO Science Series C*, Kluwer Academic Publishers.

Chelton, D., A. Mestas-Nuñez and M. Freilich, 1990a: Global wind stress and Sverdrup circulation from the Seasat scatterometer, *Journal of Physical Oceanography*, **20**, 1175–1205.

Chelton, D. B., R. A. deSzoeke, M. G. Schlax, K. E. Naggar and N. Siwertz, 1998: Geographical variability of the first baroclinic Rossby radius of deformation, *Journal of Physical Oceanography*, **28**, 433–460.

Chelton, D. B., M. G. Schlax, D. L. Witter and J. G. Richman, 1990b: Geosat altimeter observations of the surface circulation of the Southern Ocean, *Journal of Geophysical Research*, **95**(C10), 17877–17903.

Chiswell, S. M., J. Toole and J. Church, 1997: Transports across the Tasman Sea from WOCE repeat sections: the East Australian Current 1990–94, *New Zealand Journal of Marine and Freshwater Research*, **31**, 469–475.

Chua, B., and A. Bennett, 2001: An inverse ocean modeling system, *Ocean Modelling*, **3**, 137–165.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz and D. Lamich, 1998: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System, *Monthly Weather Review*, **126**, 2913–2926.

Cooper, M., and K. Haines, 1996: Altimetric assimilation with water property conservation, *Journal of Geophysical Research*, **101**(C1), 1059–1077.

Courant, R., and K. Friedrichs, 1976: *Supersonic Flow and Shock Waves*, vol. 21 of *Applied Mathematical Sciences*, Springer-Verlag.

Courant, R., and D. Hilbert, 1962: *Methods of mathematical physics. Volume II Partial differential equations*, Interscience Publishers.

Cox, M., 1984: A primitive equation, three-dimensional model of the ocean, *GFDL Ocean Group Technical Report No. 1*, Geophysical Fluid Dynamics Laboratory, Princeton University.

Daley, R., 1981: Normal mode initialization, *Reviews of Geophysics and Space Physics*, **19**(3), 450–468.

Daley, R., 1991: *Atmospheric Data Analysis*, Cambridge University Press.

Davies, A., and J. Lawrence, 1994: A three-dimensional model of the M_4 tide in the Irish Sea: The importance of open boundary conditions and influence of wind, *Journal of Geophysical Research*, **99**(C8), 16197–16227.

Davies, H. C., 1973a: On the initial-boundary value problems of some geophysical fluid flows, *Journal of Computational Physics*, **13**, 398–422.

Davies, H. C., 1973b: On the lateral boundary conditions for the primitive equations, *Journal of the Atmospheric Sciences*, **30**, 147–150.

De Mey, P., 1998: Scientific feasibility of data assimilation in the MERCATOR project, [Online] Available: http://www.omp.obs-mip.fr/omp/legos/english/rech/circu/pdf/mercassim.pdf [2001, June 1].

De Mey, P., and M. Benkiran, 1998: Mediterranean basin-scale data assimilation at LEGOS: Summary of methodology and experiments, [Online] Available: http://www.omp.obs-mip.fr/omp/legos/english/rech/circu/pdf/medwork.pdf

[2001, June 1].

De Mey, P., and M. Benkiran, 2002: A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation, in *Ocean Forecasting, Conceptual Basis and Applications*, edited by N. Pinardi, and J. Woods, Springer-Verlag.

De Mey, P., M. Benkiran, E. Greiner and C. Maes, 1999: AGORA final report for LEGOS, [Online] Available: http://www.omp.obs-mip.fr/omp/legos/english/rech/circu/pdf/agora.pdf [2001, June 1].

De Mey, P., 1997: Data assimilation at the oceanic meso-scale: A review, *Journal of the Meteorological Society of Japan*, **75**(1B), 415–427.

De Mey, P., 1999: SOFA: A reduced-order optimal interpolation algorithm for ocean data assimilation, *Journal of Atmospheric and Oceanic Technology*, submitted.

De Mey, P., and A. R. Robinson, 1987: Assimilation of altimeter eddy fields in a limited-area quasi-geostrophic model, *Journal of Physical Oceanography*, **17**, 2280–2293.

de Szoeke, R. A., and M. D. Levine, 1981: The advective flux of heat by mean geostrophic motions in the Southern Ocean, *Deep-Sea Research*, **28A**(10), 1057–1085.

Deleersnijder, E., and J.-M. Campin, 1993: On a peculiarity of the B-grid, *Ocean Modelling*.

Derber, J., and A. Rosati, 1989: A global ocean data assimilation system, *Journal of Physical Oceanography*, **19**, 1333–1347.

Dombrowsky, E., and P. De Mey, 1992: Continuous assimilation in an open domain of the Northeast Atlantic 1. Methodology and application to AthenA-88, *Journal of Geophysical Research*, **97**(C6), 9719–9731.

Dukowicz, J. K., and R. D. Smith, 1994: Implicit free-surface method for the Bryan-Cox-Semtner ocean model, *Journal of Geophysical Research*, **99**(C4), 7991–8014.

Durran, D. R., M.-J. Yang, D. N. Slinn and R. G. Brown, 1993: Toward more accurate wave-permeable boundary conditions, *Monthly Weather Review*, **121**, 604–620.

Ellis, J., T. V. Haar, S. Levitus and A. Oort, 1978: The annual variation in the global heat balance of the earth, *Journal of Geophysical Research*, **83**(C4), 1958–1962.

Elvius, T., and A. Sundström, 1973: Computationally efficient schemes and boundary conditions for a fine-mesh barotropic model based on the shallow-water equations, *Tellus*, **25**(2), 132–156.

England, M. H., 1992: On the formation of Antarctic Intermediate and Bottom Water in ocean circulation models, *Journal of Physical Oceanography*, **22**, 918–926.

England, M. H., 1993: Representing the global-scale water masses in ocean general circulation models, *Journal of Physical Oceanography*, **23**, 1523–1552.

Engquist, B., and A. Majda, 1977: Absorbing boundary conditions for the numerical simulation of waves, *Mathematics of Computation*, **31**(139), 629–651.

Evensen, G., and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the Ensemble Kalman Filter with a quasigeostrophic model, *Monthly Weather Review*, **124**, 85–96.

Ezer, T., and G. L. Mellor, 1994: Continous assimilation of Geosat altimeter data into a three-dimensional primitive equation Gulf Stream model, *Journal of Physical Oceanography*, **24**, 832–847.

Ezer, T., and G. L. Mellor, 1997: Simulations of the Atlantic Ocean with a free surface sigma coordinate model, *Journal of Geophysical Research*, **102**(C7), 15647–

15657.

Faucher, P., M. Gavart and P. D. Mey, 2002: Isopycnal EOFs and their observability from altimetry in the North and Tropical Atlantic, *Journal of Physical Oceanography*, in press.

Flather, R., 1976: A tidal model of the north-west European continental shelf, in *Continental shelf dynamics : Seventh Liege Colloquium on Ocean Hydrodynamics*, edited by J. C. Nihoul, number 6 in Memoires Société Royale des Sciences de Liège, 6^e série, pp. 141–164, Elsevier Scientific Publishing Company.

Florenchie, P., and J. Verron, 1998: South Atlantic ocean circulation — simulation experiments with a quasi-geostrophic model and assimilation of TOPEX/POSEIDON and ERS-1 altimeter data, *Journal of Geophysical Research*, **103**(C11), 24737–24758.

Fofonoff, N., and R. Millard, 1983: Algorithms for computation of fundamental properties of seawater, *UNESCO Technical Papers in Marine Science 44*, UNESCO, Paris.

Forsythe, G. E., and W. R. Wasow, 1960: *Finite-difference methods for partial differential equations*, John Wiley and Sons, Inc.

Fox, A., K. Haines, B. de Cuevas and D. Webb, 2000a: Altimeter assimilation in the OCCAM global model, Part I: A twin experiment, *Journal of Marine Systems*, **26**(3-4), 303–322.

Fox, A., K. Haines, B. de Cuevas and D. Webb, 2000b: Altimeter assimilation in the OCCAM global model, Part II: TOPEX/POSEIDON and ERS-1 assimilation, *Journal of Marine Systems*, **26**(3-4), 323–347.

Freeman, J. C., 1951: The solution of nonlinear meteorological problems by the method of characteristics, in *Compendium of Meteorology*, edited by T. F. Malone, pp. 421–433, American Meteorological Society, Boston, Massachusetts.

Fu, L.-L., and A. Cazenave (eds.), 2000: *Satellite Altimetry and the Earth Sciences: A Handbook of Techniques and Applications*, Academic Press.

Fu, L.-L., and R. E. Cheney, 1995: Application of satellite altimetry to ocean circulation studies: 1987-1994, *Review of Geophysics, Supplement*, pp. 213–223.

Fu, L.-L., and I. Fukumori, 1993: Fitting dynamic models to the Geosat sea level observations in the tropical Pacific Ocean. Part II: A linear, wind-driven model, *Journal of Physical Oceanography*, **23**, 2162–2181.

Fukumori, I., 2000: Data assimilation by models, In Fu and Cazenave (Fu and Cazenave 2000).

Fukumori, I., 1995: Assimilation of TOPEX sea level measurements with a reduced-gravity shallow water model of the tropical Pacific Ocean, *Journal of Geophysical Research*, **100**(C12), 25027–25040.

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, *Journal of Physical Oceanography*, **23**(8), 1831–1855.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model, *Journal of Geophysical Research*, **100**(C4), 6777–6793.

Gan, J., L. A. Mysak and D. N. Straub, 1998: Simulation of the South Atlantic Ocean circulation and its seasonal variability, *Journal of Geophysical Research*, **103**(C5), 10241–10251.

Garabedian, P., 1964: *Partial Differential Equations*, John Wiley and Sons, Inc.

Gargett, A., 1986: Small-scale parametrization in large-scale ocean models, In O'Brien (O'Brien 1986a), pp. 145–154.

Gavart, M., and P. D. Mey, 1997: Isopycnal EOFs in the Azores Current region: A statistical tool for dynamical analysis and data assimilation, *Journal of Physical Oceanography*, **27**, 2146–2157.

Gavart, M., P. D. Mey and G. Caniaux, 1999: Assimilation of satellite altimeter data in a primitive-equation model of the Azores-Madeira region, *Dynamics of Atmospheres and Oceans*, **29**, 217–254.

Gelb, A. (ed.), 1974: *Applied optimal estimation*, The M.I.T. Press.

Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models, *Journal of Physical Oceanography*, **20**, 150–155.

Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework, *Dynamics of Atmospheres and Oceans*, **13**, 171–218.

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

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

Gill, A. E., 1982: *Atmosphere-Ocean Dynamics*, Academic Press.

Gill, A., and P. Niiler, 1973: The theory of seasonal variability in the ocean, *Deep-Sea Research*, **20**, 141–177.

Gille, S. T., 1995: Dynamics of the Antarctic Circumpolar Current: Evidence for topographic effects from altimeter data and numerical model output, Ph.D. thesis, Massachusetts Institute of Technology, Woods Hole Oceanographic Institute.

Gille, S. T., 1997: The Southern Ocean momentum balance: Evidence for topographic effects from numerical model output and altimeter data, *Journal of Physical Oceanography*, **27**, 2219–2232.

Gille, S. T., and K. A. Kelly, 1996: Scales of spatial and temporal variability in the Southern Ocean, *Journal of Geophysical Research*, **101**(C4), 8759–8773.

Godfrey, J., 1989: A Sverdrup model of the depth-integrated flow for the world ocean allowing for island circulations, *Geophysical and Astrophysical Fluid Dynamics*, **45**, 89–112.

Gordon, A., and H. Taylor, 1975: Heat and salt balance within the cold waters of the world ocean, in *Numerical models of ocean circulation*, National Academy of Science.

Gordon, A. L., and W. B. Owens, 1987: Polar oceans, *Reviews of Geophysics*, **25**(2), 227–233.

Gottlieb, D., M. Gunzburger and E. Turkel, 1982: On numerical boundary treatment of hyperbolic systems for finite difference and finite element methods, *SIAM Journal on Numerical Analysis*, **19**(4), 671–682.

Griffies, S. M., C. Böning, F. O. Bryan, E. P. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A.-M. Treguier and D. Webb, 2000: Developments in ocean climate modelling, *Ocean Modelling*, **2**(3–4), 123–192.

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

Griffiths, C., 1995: A fine resolution numerical model of the Iceland-Farœ front with open boundary conditions, *Journal of Geophysical Research*, **100**(C8), 15915–15931.

Grotov, A., D. Nechaev, G. Panteleev and M. Yaremchuk, 1998: Large scale circulation in the Bellingshausen and Amundsen seas as a variational inverse of climatological data, *Journal of Geophysical Research*, **103**(C6), 13011–13022.

Gustafsson, B., and H.-O. Kreiss, 1979: Boundary conditions for time dependent problems with an artificial boundary, *Journal of Computational Physics*, **30**, 333–351.

Gustafsson, B., H.-O. Kreiss and J. Oliger, 1995: *Time dependent problems and difference methods*, John Wiley and Sons, Inc.

Gustafsson, B., H.-O. Kreiss and A. Sundström, 1972: Stability theory of difference approximations for mixed initial boundary value problems. II, *Mathematics of Computation*, **26**(119), 649–686.

Gustafsson, B., and A. Sundström, 1978: Incompletely parabolic problems in fluid dynamics, *SIAM Journal on Applied Mathematics*, **35**(2), 343–357.

Gustafsson, N., 1981: A review of methods for objective analysis, In Bengtsson et al. (Bengtsson et al. 1981), pp. 17–76.

Haidvogel, D., A. Robinson and E. Schulman, 1980: The accuracy, efficiency, and stability of three numerical models with application to open ocean problems, *Journal of Computational Physics*, **34**, 1–53.

Haidvogel, D., J. Wilkin and R. Young, 1991: A semi-spectral primitive equation ocean circulation model using vertical sigma and orthogonal curvilinear horizontal coordinates, *Journal of Computational Physics*, **94**, 151–185.

Haines, K., 1991: A direct method for assimilating sea surface height data into ocean models with adjustments to the deep circulation, *Journal of Physical Oceanography*, **21**(6), 843–868.

Haines, K., 1994: Dynamics and data assimilation in oceanography, In Brasseur and Nihoul (Brasseur and Nihoul 1994).

Hall, M., and D. Cacuci, 1983: Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models, *Journal of the Atmospheric Sciences*, **40**, 2537–2546.

Haney, R., 1971: Surface thermal boundary conditions for ocean circulation models, *Journal of Physical Oceanography*, **1**, 241–248.

Hastenrath, S., 1980: Heat budget of tropical ocean and atmosphere, *Journal of Physical Oceanography*, **10**, 159–170.

Hedstrom, G., 1979: Nonreflecting boundary conditions for nonlinear hyperbolic systems, *Journal of Computational Physics*, **30**, 222–237.

Hellerman, S., and M. Rosenstein, 1983: Normal monthly wind stress over the world ocean with error estimates, *Journal of Physical Oceanography*, **13**(7), 1093–1104.

Hellmer, H. H., and A. Beckmann, 2001: The Southern Ocean: A ventilation contributor with multiple sources, *Geophysical Research Letters*, **28**(15), 2927–2930.

Hesthaven, J., and D. Gottlieb, 1996: A stable penalty method for the compressible Navier-Stokes equations: I. Open boundary conditions, *SIAM Journal on Scientific Computing*, **17**(3), 579–612.

Hibler, III, W., 1979: A dynamic thermodynamic sea ice model, *Journal of Physical Oceanography*, **9**, 815–846.

Hidaka, K., and M. Tsuchiya, 1953: On the Antarctic Circumpolar Current, *Journal of Marine Research*, **12**(2), 214–222.

Higdon, R. L., 1994: Radiation boundary conditions for dispersive waves, *SIAM Journal on Numerical Analysis*, **31**(1), 64–100.

Hoffert, M. I., 1990: Climatic change and ocean bottom water formation: Are we missing something?, in *Climate-Ocean Interaction*, edited by M. Schlesinger, pp. 295–317, Kluwer Academic Publishers.

Holland, W. R., 1978: The role of mesoscale eddies in the general circulation of the ocean — numerical experiments using a wind-driven quasi-geostrophic model,

Journal of Physical Oceanography, **8**, 363–392.

Holland, W. R., 1986: Quasigeostrophic modelling of eddy-resolved ocean circulation, In O'Brien (O'Brien 1986a), pp. 203–231.

Holland, W., and P. Malanotte-Rizzoli, 1989: Assimilation of altimeter data into an ocean model: Space versus time resolution studies, *Journal of Physical Oceanography*, **19**, 1507–1534.

Hughes, C. W., M. P. Meredith and K. J. Heywood, 1999: Wind-driven transport fluctuations through Drake Passage: A southern mode, *Journal of Physical Oceanography*, **29**, 1971–1992.

Hughes, C., 1995: Rossby waves in the Southern Ocean: A comparison of TOPEX/POSEIDON altimetry with model predictions, *Journal of Geophysical Research*, **100**(C8), 15933–15950.

Hughes, C., 1997: Comments on "On the obscurantist physics of 'form drag' in theorizing about the Circumpolar Current", *Journal of Physical Oceanography*, **27**, 209–210.

Hurlburt, H. E., 1986: Dynamic transfer of simulated altimeter data into subsurface information by a numerical ocean model, *Journal of Geophysical Research*, **91**(C2), 2372–2400.

Huthnance, J., 1984: Slope currents and "JEBAR", *Journal of Physical Oceanography*, **14**, 795–810.

IPCC, 1995: *IPCC Second Assessment: Climate Change 1995, A report of the Intergovernmental Panel on Climate Change*, Intergovernmental Panel on Climate Change, Geneva, SWITZERLAND.

Ishikawa, Y., T. Awaji, K. Akitomo and B. Qiu, 1996: Successive correction of the mean sea surface height by the simultaneous assimilation of drifting buoy and altimetric data, *Journal of Physical Oceanography*, **26**, 2381–2397.

Israeli, M., and S. A. Orszag, 1981: Approximation of radiation boundary conditions, *Journal of Computational Physics*, **41**, 115–135.

Ivchenko, V. O., K. J. Richards and D. P. Stevens, 1996: The dynamics of the Antarctic Circumpolar Current, *Journal of Physical Oceanography*, **26**, 753–774.

Ivchenko, V., A. Treguier and S. Best, 1997: A kinetic energy budget and internal instabilities in the Fine Resolution Antarctic Model, *Journal of Physical Oceanography*, **27**, 5–22.

Jacobs, G., W. Emery and G. Born, 1993: Rossby waves in the Pacific Ocean extracted from Geosat altimeter data, *Journal of Physical Oceanography*, **23**, 1155–1175.

Jensen, T. G., 1998: Open boundary conditions in stratified ocean models, *Journal of Marine Systems*, **16**, 297–322.

Jiang, S., P. Stone and P. Malanotte-Rizzoli, 1999: An assessment of the Geophysical Fluid Dynamics Laboratory ocean model with coarse resolution: Annual-mean climatology, *Journal of Geophysical Research*, **104**(C11), 25623–25645.

Johnson, G. C., and H. L. Bryden, 1989: On the size of the Antarctic Circumpolar Current, *Deep-Sea Research*, **36**(1), 39–53.

Johnson, T. J., R. H. Stewart, C. Shum and B. D. Tapley, 1992: Distribution of Reynolds stress carried by mesoscale variability in the Antarctic Circumpolar Current, *Geophysical Research Letters*, **19**(12), 1201–1204.

Jones, I. S., Y. Sugumori, and R. Stewart (eds.), 1993: *Satellite Remote Sensing of the Oceanic Environment*, Seibutsu Kenkyusha, Tokyo, JAPAN.

Kalman, R., 1960: A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, **82**, 35–45.

Kantha, L. H., and C. A. Clayson, 2000: *Numerical Models of Oceans and Oceanic Processes*, Academic Press.

Kar, S. K., and R. P. Turco, 1995: Formulation of a lateral sponge layer for limited-area shallow-water models and an extension for the vertically stratified case, *Monthly Weather Review*, **123**, 1542–1559.

Killworth, P. D., 1992: An equivalent-barotropic mode in the Fine Resolution Antarctic Model, *Journal of Physical Oceanography*, **22**, 1379–1387.

Killworth, P. D., and M. M. Nanneh, 1994: Isopycnal momentum budget of the Antarctic Circumpolar Current in the Fine Resolution Antarctic Model, *Journal of Physical Oceanography*, **24**, 1201–1223.

Killworth, P. D., D. A. Smeed and A. G. Nurser, 2000: The effects on ocean models of relaxation toward observations at the surface, *Journal of Physical Oceanography*, **30**(1), 160–174.

Killworth, P. D., D. Stainforth, D. J. Webb and S. M. Paterson, 1991: The development of a free-surface Bryan-Cox-Semtner model, *Journal of Physical Oceanography*, **21**, 1333–1348.

Kim, S., and T. Crowley, 2000: Increased Pliocene North Atlantic Deep Water: Cause or consequence of Pliocene warming?, *Paleoceanography*, **15**(4), 451–455.

Kim, S., T. Crowley and A. Stössel, 1998: Local orbital forcing of Antarctic climate change during the last interglacial, *Science*, **280**, 728–730.

Kim, S., and A. Stössel, 1998: On the representation of the Southern Ocean water masses in an ocean climate model, *Journal of Geophysical Research*, **103**(C11), 24891–24906.

Kowalik, Z., and T. Murty, 1993: *Numerical Modeling of Ocean Dynamics*, World Scientific Publishing Co.

Kreiss, H.-O., 1971: Difference approximations for initial boundary-value problems, *Proceedings of the Royal Society of London*, **A323**, 255–261.

Kreiss, H.-O., 1970: Initial boundary value problems for hyperbolic systems, *Communications on Pure and Applied Mathematics*, **23**, 277–298.

Kreiss, H.-O., and J. Lorenz, 1989: *Initial-Boundary Value Problems and the Navier-Stokes Equations*, vol. 136 of *Pure and Applied Mathematics*, Academic Press, Inc., San Diego, California.

Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations, *Tellus*, **38A**(2), 97–110.

Le Traon, P., P. Gaspar, F. Bouyssel and H. Makhmara, 1995: Using Topex/Poseidon data to enhance ERS-1 data, *Journal of Atmospheric and Oceanic Technology*, **12**(1), 161–170.

Le Traon, P., F. Nadal and N. Ducet, 1998: An improved mapping method of multisatellite altimeter data, *Journal of Atmospheric and Oceanic Technology*, **15**(2), 522–534.

Le Traon, P., and F. Ogor, 1998: ERS-1/2 orbit improvement using TOPEX/POSEIDON: the 2 cm challenge, *Journal of Geophysical Research*, **103**(C4), 8045–8057.

LeBlond, P. H., and L. A. Mysak, 1978: *Waves in the Ocean*, Elsevier Scientific Publishing Company.

Lee, H. J., K. T. Jung, M. Foreman and J. Y. Chung, 2000: A three-dimensional mixed finite-difference Galerkin function model for the oceanic circulation in the Yellow Sea and the East China Sea, *Continental Shelf Research*, **20**(8).

Lee, T., and J. Marotzke, 1997: Inferring meridional mass and heat transports of the Indian Ocean by fitting a general circulation model to climatological data,

Journal of Geophysical Research, **102**(C5), 10585–10602.

Leeuwenburgh, O., 2000: Covariance modeling for merging of multi-sensor ocean surface data, in *Methods and applications of inversion*, edited by P. Hansen, B. Jacobsen, and K. Mosegaard, Springer-Verlag.

Lefevre, et al., 2000a: FES99: a tide finite element solution assimilating tide gauge and altimetry, *Journal of Geophysical Research*, submitted.

Lefevre, et al., 2000b: A new global tide finite element solution independent of altimetry, *Geophysical Research Letters*, submitted.

Lemoine, F., D. Smith, L. Kunz, R. Smith, E. Pavlis, N. Pavlis, S. Klosko, D. Chinn, M. Torrence, R. Williamson, C. Cox, K. Rachlin, Y. Wang, S. Kenyon, R. Salman, R. Trimmer, R. Rapp and R. Nerem, 1997: The development of the NASA GSFC and NIMA joint geopotential model, in *Gravity, Geoid and Marine Geodesy*, edited by J. Segawa, H. Fujimoto, and S. Okubo, number Vol. 117 in International Association of Geodesy Symposia, pp. 461–469, Springer, International symposium, Tokyo, Japan, September 30-October 5, 1996.

Levitus, S., and T. Boyer, 1994: World ocean atlas 1994, vol. 4: Temperature, *Tech. rep.*, National Oceanographic Data Centre, NOAA, US Government Printing Office, Wahington, D.C.

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme, *Monthly Weather Review*, **109**, 701–721.

Macdonald, A. M., and C. Wunsch, 1996: An estimate of global ocean circulation and heat fluxes, *Nature*, **382**, 436–439.

Maes, C., M. Benkiran and P. De Mey, 1999: Sea level comparison between TOPEX/POSEIDON altimetric data and a global ocean general circulation model from an assimilation perspective, *Journal of Geophysical Research*, **104**(C7), 15575–15585.

Mahadevan, A., and D. Archer, 1998: Modeling a limited region of the ocean, *Journal of Computational Physics*, **145**(2), 555–574.

Maier-Reimer, E., U. Mikolajewicz and K. Hasselmann, 1993: Mean circulation of the Hamburg LSG OGCM and its sensitivity to the thermohaline surface forcing, *Journal of Physical Oceanography*, **23**, 731–757.

Majda, A., and S. Osher, 1975: Initial-boundary value problems for hyperbolic equations with uniformly characteristic boundary, *Communications on Pure and Applied Mathematics*, **28**, 607–675.

Malanotte-Rizzoli, P. (ed.), 1996: *Modern approaches to data assimilation in ocean modeling*, Elsevier Science B.V.

Malanotte-Rizzoli, P., and E. Tziperman, 1996: The oceanographic data assimilation problem: Overview, motivation and purposes, In Malanotte-Rizzoli (Malanotte-Rizzoli 1996), pp. 3–17.

Malanotte-Rizzoli, P., R. E. Young and D. B. Haidvogel, 1989: Initialization and data assimilation experiments with a primitive equation model, *Dynamics of Atmospheres and Oceans*, **13**(3-4), 349–378.

Maltrud, M., R. Smith, A. Semtner and R. Malone, 1998: Global eddy-resolving ocean simulations driven by 1985-1995 atmospheric winds, *Journal of Geophysical Research*, **103**(C13), 30825–30853.

Marchuk, G., and A. Sarkisyan, 1988: *Mathematical Modelling of Ocean Circulation*, Springer-Verlag.

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, *Journal of Geophysical Research*, **104**(C12),

29529–29547.

Marotzke, J., and J. Willebrand, 1991: Multiple equilibria of the global thermohaline circulation, *Journal of Physical Oceanography*, **21**, 1372–1385.

Marotzke, J., and C. Wunsch, 1993: Finding the steady state of a general circulation model through data assimilation: Application to the North Atlantic Ocean, *Journal of Geophysical Research*, **98**(C11), 20149–20167.

Marsh, R., A. G. Nurser, A. P. Megann and A. L. New, 2000: Water mass transformation in the Southern Ocean of a global isopycnal coordinate GCM, *Journal of Physical Oceanography*, **30**, 1013–1045.

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, *Journal of Geophysical Research*, **102**, 5753–5766.

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

Marshall, J., and F. Schott, 1999: Open-ocean convection: Observations, theory and models, *Reviews of Geophysics*, **37**(1), 1–64.

Marsland, S., and J.-O. Wolff, 1998: East Antarctic seasonal sea-ice and ocean stability: a model study, *Annals of Glaciology*, **27**, 477–482.

McDougall, T. J., 1987: Neutral surfaces, *Journal of Physical Oceanography*, **17**(11), 1950–1964.

McWilliams, J. C., W. R. Holland and J. H. Chow, 1978: A description of numerical Antarctic Circumpolar Currents, *Dynamics of Atmospheres and Oceans*, **2**, 213–291.

McWilliams, J., and J. Chow, 1981: Equilibrium geostrophic turbulence. I: A reference solution in a β-plane channel, *Journal of Physical Oceanography*, **11**, 921–949.

Megann, A., and A. New, 2001: The effects of resolution and viscosity in an isopycnic-coordinate model of the equatorial Pacific, *Journal of Physical Oceanography*, **31**, 1993–2018.

Mellor, G. L., and T. Ezer, 1991: A Gulf Stream model and an altimetry assimilation scheme, *Journal of Geophysical Research*, **96**(C5), 8779–8795.

Mertz, G., and D. G. Wright, 1992: Interpretations of the JEBAR term, *Journal of Physical Oceanography*, **22**, 301–305.

Mesinger, F., and A. Arakawa, 1976: *Numerical Methods Used in Atmospheric Models*, number No. 17 in GARP Publications Series, World Meteorological Organization.

Mestas-Nuñez, A. M., D. B. Chelton and R. A. deSzoeke, 1992: Evidence of time-dependent Sverdrup circulation in the South Pacific from the Seasat scatterometer and altimeter, *Journal of Physical Oceanography*, **22**, 934–943.

Meuer, H., E. Strohmaier, J. Dongarra and H. D. Simon, 2001: TOP500 supercomputer sites, [Online] Available: http://www.top500.org [2001, January 3].

Miller, M., and A. Thorpe, 1981: Radiation conditions for the lateral boundaries of limited-area numerical models, *Quarterly Journal of the Royal Meteorological Society*, **107**, 615–628.

Moore, A., N. Cooper and D. Anderson, 1987: Initialization and data assimilation in models of the Indian Ocean, *Journal of Physical Oceanography*, **17**, 1965–1977.

Moore, A. M., 1991: Data assimilation in a quasi-geostrophic open-ocean model of the Gulf Stream using the adjoint method, *Journal of Physical Oceanography*, **21**, 398–427.

Moran, J. M., and M. D. Morgan, 2002: *Meteorology : the atmosphere and the science of weather*, Prentice Hall.

Moretti, G., 1969: Importance of boundary conditions in the numerical treatment of hyperbolic equations, *The Physics of Fluids*, **Supplement II**, 13–20.

Moretti, G., 1981: A physical approach to the numerical treatment of boundaries in gas dynamics, In (NASA 1981), pp. 73–95.

Morrow, R., and P. D. Mey, 1995: An adjoint assimilation of altimetric, surface drifter and hydrographic data in a QG model of the Azores current, *Journal of Geophysical Research*, **100**(C12), 25007–25025.

Morrow, R., J. Church, R. Coleman, D. Chelton and N. White, 1992: Eddy momentum flux and its contribution to the Southern Ocean momentum balance, *Nature*, **357**, 482–484.

Morrow, R., R. Coleman, J. Church and D. Chelton, 1994: Surface eddy momentum flux and velocity variances in the Southern Ocean from Geosat altimetry, *Journal of Physical Oceanography*, **24**, 2050–2071.

MPI, 2001: C-hope, [Online] Available: http://www.mpimet.mpg.de/Depts/Klima/natcli/C-HOPE/C-HOPE.html [2001, February 20].

Mulhearn, P., 1987: The Tasman Front – a study using satellite infrared imagery, *Journal of Physical Oceanography*, **17**, 1148–1155.

Müller, P., 1995: Ertel's potential vorticity theorem in physical oceanography, *Reviews of Geophysics*, **33**, 67–97.

Munk, W., 1950: On the wind-driven ocean circulation, *Journal of Meteorology*, **7**, 79–93.

Munk, W., and E. Palmén, 1951: Note on the dynamics of the Antarctic Circumpolar Current, *Tellus*, **3**, 53–55.

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

NASA, 1981: *Numerical Boundary Condition Procedures*, NASA Conference Publication 2201, NASA.

Nerem, R., E. Schrama, C. Koblinsky and B. Beckley, 1994: A preliminary evaluation of ocean topography from the TOPEX/POSEIDON mission, *Journal of Geophysical Research*, **99**(C12), 24565–24583.

NGDC, 1988: *Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth*, NOAA, National Geophysical Data Center, Boulder, Colorado.

Nordström, J., 1995: The use of characteristic boundary conditions for the Navier-Stokes equations, *Computers and Fluids*, **24**(5), 609–623.

Nowlin, Jr., W., S. Worley and T. Whitworth, III, 1985: Method for making point estimates of eddy heat flux as applied to the Antarctic Circumpolar Current, *Journal of Geophysical Research*, **90**(C2), 3305–3324.

Nowlin, Jr., W. D., and J. M. Klinck, 1986: The physics of the Antarctic Circumpolar Current, *Reviews of Geophysics*, **24**(3), 469–491.

Oberhuber, J. M., 1993: Simulation of the Atlantic circulation with a coupled sea ice-mixed layer-isopycnal general circulation model. Part I: Model description, *Journal of Physical Oceanography*, **23**, 808–829.

O'Brien, J. J. (ed.), 1986a: *Advanced Physical Oceanographic Numerical Modelling*, vol. 186 of *NATO ASI Series C*, D. Reidel Publishing Company.

O'Brien, J. J., and R. O. Reid, 1967: The non-linear response of a two-layer, baroclinic ocean to a stationary, axially-symmetric hurricane: Part I. Upwelling induced by momentum transfer, *Journal of the Atmospheric Sciences*, **24**, 197–207.

O'Brien, J., 1986b: Time integration schemes, In O'Brien (O'Brien 1986a), pp.

155–163.

Oey, L.-Y., and P. Chen, 1992: A model simulation of circulation in the northeast Atlantic shelves and seas, *Journal of Geophysical Research*, **97**(C12), 20087–20115.

Oke, P. R., R. N. Miller, J. Allen, G. D. Egbert and P. M. Kosro, 2000: Assimilation of surface velocity data into primitive equation coastal ocean models, in preparation.

Olbers, D., V. Gouretski, G. Seiß and J. Schröter, 1992: Hydrographic atlas of the Southern Ocean, *Tech. rep.*, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany.

Olbers, D., 1998: Comments on "On the obscurantist physics of 'form drag' in theorizing about the Circumpolar Current", *Journal of Physical Oceanography*, **28**, 1647–1654.

Oliger, J., and A. Sundström, 1978: Theoretical and practical aspects of some initial-boundary value problems in fluid dynamics, *SIAM Journal on Applied Mathematics*, **35**(3), 419–446.

OPA 8.1 Ocean General Circulation Model reference manual, *Note du Pôle de modélisation.*, 1998: *Tech. rep.*, Institut Pierre-Simon Laplace, France.

Orlanski, I., 1976: A simple boundary condition for unbounded hyperbolic flows, *Journal of Computational Physics*, **21**, 251–269.

Oschlies, A., 2002: Improved representation of upper-ocean dynamics and mixed layer depths in a model of the North Atlantic on switching from eddy-permitting to eddy-resolving grid resolution, *Journal of Physical Oceanography*, **32**, 2277–2298.

Oschlies, A., and J. Willebrand, 1996: Assimilation of Geosat altimeter data into an eddy-resolving primitive equation model of the North Atlantic Ocean, *Journal of Geophysical Research*, **101**(C6), 14175–14190.

Pacanowski, R., K. Dixon and A. Rosati, 1991: The GFDL modular ocean model user guide, *GFDL Ocean Group Technical Report No. 2*, Geophysical Fluid Dynamics Laboratory, Princeton University.

Pacanowski, R., and S. Philander, 1981: Parametrization of vertical mixing in numerical models of tropical oceans, *Journal of Physical Oceanography*, **11**, 1443–1451.

Palma, E., and R. Matano, 2000: On the implementation of open boundary conditions for a general circulation model: The three-dimensional case, *Journal of Geophysical Research*, **105**(C4), 8605–8627.

Palma, E. D., and R. P. Matano, 1998: On the implementation of passive open boundary conditions for a general circulation model: The barotropic mode, *Journal of Geophysical Research*, **103**(C1), 1319–1341.

Pedlosky, J., 1964: The stability of currents in the atmosphere and the ocean: Part I, *Journal of the Atmospheric Sciences*, **21**, 201–219.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics*, Springer-Verlag, second edition.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*, American Institute of Physics.

Phillips, H. E., 2000: Mean flow, eddy variability and energetics of the Subantarctic Front south of Australia, Ph.D. thesis, University of Tasmania.

Phillips, H. E., and S. R. Rintoul, 2000: Eddy variability and energetics from direct current measurements in the Antarctic Circumpolar Current south of Australia, *Journal of Physical Oceanography*, **30**, 3050–3076.

Pinardi, N., A. Rosati and R. Pacanowski, 1995: The sea surface pressure formulation of rigid lid models. Implications for altimetric data assimilation studies,

Journal of Marine Systems, **6**, 109–119.

Platzman, G. W., 1972: Two-dimensional free oscillations in natural basins, *Journal of Physical Oceanography*, **2**(2), 117–138.

Ponte, R. M., and R. D. Rosen, 1994: Oceanic angular momentum and torques in a general circulation model, *Journal of Physical Oceanography*, **24**, 1966–1977.

Power, S., and R. Kleeman, 1993: Multiple equilibria in a global general circulation model, *Journal of Physical Oceanography*, **23**, 1670–1681.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992: *Numerical recipes in Fortran 77: the art of scientific computing*, Cambridge University Press, second edition.

Rahmstorf, S., 1997: Ice-cold in Paris, *New Scientist*, pp. 26–30.

Randall, D. A., 1994: Geostrophic adjustment and the finite-difference shallow-water equations, *Monthly Weather Review*, **122**, 1371–1377.

Reason, C., and A. Pearce, 1996: Comparison of the Semtner and Chervin eddy-resolving global ocean model with LUCIE and satellite observations in the Leeuwin Current region, *Marine and Freshwater Research*, **47**, 509–516.

Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate rotation, *Journal of Physical Oceanography*, **12**, 1154–1158.

Rhines, P. B., 1986: Vorticity dynamics of the oceanic general circulation, *Annual Review of Fluid Mechanics*, **18**, 433–497.

Richtmyer, R., and K. Morton, 1967: *Difference Methods for Initial Value Problems*, Wiley-Interscience.

Ridgway, K., and J. Godfrey, 1994: Mass and heat budgets in the East Australian Current: a direct approach, *Journal of Geophysical Research*, **99**(C2), 3231–3248.

Ridgway, K., and J. Godfrey, 1997: Seasonal cycle of the East Australian Current, *Journal of Geophysical Research*, **102**(C10), 22921–22936.

Riley, K., 1974: *Mathematical Methods for the Physical Sciences: An informal treatment for students of physics and engineering*, Cambridge University Press.

Rintoul, S., and J. Bullister, 1999: A late winter hydrographic section from Tasmania to Antarctica, *Deep-Sea Research*, **46**(8), 1417–1454.

Rintoul, S., and S. Sokolov, 2001: Baroclinic transport variability of the Antarctic Circumpolar Current south of Australia (WOCE repeat section SR3), *Journal of Geophysical Research*, **106**(C2).

Robinson, I. S., 1994: *Satellite oceanography: an introduction for oceanographers and remote-sensing scientists*, Wiley Press.

Rodgers, K., M. Latif and S. Legutke, 2000: Sensitivity of equatorial Pacific and Indian Ocean watermasses to the position of the Indonesian throughflow, *Geophysical Research Letters*, **27**(18), 2941–2944.

Røed, L. P., and J. J. O'Brien, 1983: A coupled ice-ocean model of upwelling in the marginal ice zone, *Journal of Geophysical Research*, **88**(C5), 2863–2872.

Røed, L., and C. Cooper, 1986: Open boundary conditions in numerical ocean models, In O'Brien (O'Brien 1986a), pp. 411–436.

Røed, L., and C. Cooper, 1987: A study of various open boundary conditions for wind-forced barotropic numerical ocean models, in *Three-dimensional models of marine and estuarine dynamics*, edited by J. Nihoul, and B. Jamart, pp. 305–335, Elsevier, New York.

Rosenberg, M., S. Bray, N. Bindoff, N. Johnston, S. Bell and P. Towler, 1997: Aurora Australis marine science cruises AU9501, AU9604, and AU9601 — oceanographic field measurements and analysis, inter-cruise comparisons and data quality notes, *Research Report 12*, Antarctic Cooperative Research Centre, Hobart, Australia.

Rosenberg, M., R. Eriksen, S. Bell, N. Bindoff and S. Rintoul, 1995a: Aurora Australis marine science cruise AU9407 — oceanographic field measurements and analysis, *Research Report 6*, Antarctic Cooperative Research Centre, Hobart, Australia.

Rosenberg, M., R. Eriksen, S. Bell and S. Rintoul, 1995b: Aurora Australis marine science cruise AU9404 — oceanographic field measurements and analysis, *Research Report 8*, Antarctic Cooperative Research Centre, Hobart, Australia.

Rosenberg, M., R. Eriksen and S. Rintoul, 1994: Aurora Australis marine science cruise AU9309/AU9391 — oceanographic field measurements and analysis, *Research Report 2*, Antarctic Cooperative Research Centre, Hobart, Australia.

Rossby, T., 1996: The North Atlantic Current and surrounding waters: At the crossroads, *Reviews of Geophysics*, **34**(4), 463–481.

Rummel, R., and F. Sansò (eds.), 1993: *Satellite altimetry in geodesy and oceanography*, Springer-Verlag.

Sarmiento, J., and K. Bryan, 1982: An ocean transport model for the North Atlantic, *Journal of Geophysical Research*, **87**, 394–408.

Saunders, P. M., A. C. Coward and B. A. de Cuevas, 1999: Circulation of the Pacific Ocean in a global ocean model OCCAM, *Journal of Geophysical Research*, **104**(C8), 18281–18299.

Scharoo, R., and P. Visser, 1998: Precise orbit determination and gravity field improvement for the ERS satellites, *Journal of Geophysical Research*, **103**(C4), 8113–8127.

Schmitz, Jr., W. J., 1995: On the interbasin-scale thermohaline circulation, *Reviews of Geophysics*, **33**(2), 151–173.

Schmitz, Jr., W. J., and J. D. Thompson, 1993: On the effects of horizontal resolution in a limited-area model of the Gulf Stream system, *Journal of Physical Oceanography*, **23**, 1001–1007.

Schröter, J., 1989: Driving of non-linear time-dependent ocean models by observation of transient tracers — a problem of constrained optimization, In Anderson and Willebrand (Anderson and Willebrand 1989), pp. 257–285.

Schröter, J., 1994: A sensitivity analysis of variational data assimilation, In Brasseur and Nihoul (Brasseur and Nihoul 1994), pp. 59–83.

Schröter, J., U. Seiler and M. Wenzel, 1993: Variational assimilation of Geosat data into an eddy-resolving model of the Gulf Stream extension area, *Journal of Physical Oceanography*, **23**, 925–953.

Sciremammano, Jr., F., 1980: The nature of the poleward heat flux due to low-frequency current fluctuations in Drake Passage, *Journal of Physical Oceanography*, **10**, 843–852.

Seiß, G., J. Schröter and V. Gouretski, 1997: Assimilation of Geosat altimeter data into a quasigeostrophic model of the Antarctic Circumpolar Current, *Monthly Weather Review*, **125**(7), 1598–1614.

Semtner, A. J., 1995: Modeling ocean circulation, *Science*, **269**, 1379–1385.

Semtner, A. J., and R. M. Chervin, 1992: Ocean general circulation from a global eddy-resolving model, *Journal of Geophysical Research*, **97**(C4), 5493–5550.

Shapiro, M., and J. J. O'Brien, 1970: Boundary conditions for fine-mesh limited-area forecasts, *Journal of Applied Meteorology*, **9**, 345–349.

Shulman, I., 1997: Local data assimilation in specification of open boundary conditions, *Journal of Atmospheric and Oceanic Technology*, **14**, 1409–1419.

Sirkes, Z., E. Tziperman and W. Thacker, 1996: Combining data and a global primitive equation ocean general circulation model using the adjoint method, In

Malanotte-Rizzoli (Malanotte-Rizzoli 1996), pp. 119–145.

Sirkes, Z., and E. Tziperman, 1997: Finite difference of adjoint or adjoint of finite difference?, *Monthly Weather Review*, **125**, 3373–3378.

Slørdal, L. H., and J. E. Weber, 1996: Adjustment to JEBAR forcing in a rotating ocean, *Journal of Physical Oceanography*, **26**, 657–670.

Sloyan, B. M., and S. R. Rintoul, 2001: The southern ocean limb of the global deep overturning circulation, *Journal of Physical Oceanography*, **31**, 143–173.

Sloyan, B., and S. Rintoul, 2000: Estimates of area-averaged diapycnal fluxes from basin-scale budgets, *Journal of Physical Oceanography*, **30**(9).

Smith, N., and M. Lefebvre, 1998: The Global Ocean Data Assimilation Experiment, in *Monitoring the Oceans in the 2000s: An Integrated Approach. International Symposium held at Biarritz, France, 15–17 Oct, 1998.*

Smith, R. D., M. E. Maltrud, F. O. Bryan and M. W. Hecht, 2000: Numerical simulation of the North Atlantic Ocean at $\frac{1}{10}^\circ$, *Journal of Physical Oceanography*, **30**, 1532–1561.

Sod, G. A., 1985: *Numerical methods in fluid dynamics: initial and initial boundary-value problems*, Cambridge University Press.

Sommerfeld, A., 1949: *Partial differential equations in physics*, Academic Press, New York.

Sørenson, B., 1979: *Renewable Energy*, Academic Press.

Speedy, C., R. Brown and G. Goodwin, 1970: *Control Theory: Identification and Optimal Control*, Oliver and Boyd, Tweeddale Court, Edinburgh.

Spiegel, E., and G. Veronis, 1960: On the Boussinesq approximation for a compressible fluid, *The Astrophysical Journal*, **131**, 442–447.

Stammer, D., 1997a: Geosat data assimilation with application to the eastern North Atlantic, *Journal of Physical Oceanography*, **27**, 41–61.

Stammer, D., 1997b: Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements, *Journal of Physical Oceanography*, **27**, 1743–1769.

Stammer, D., 1998: On eddy characteristics, eddy transports, and mean flow properties, *Journal of Physical Oceanography*, **28**, 727–739.

Stammer, D., R. Tokmakian, A. Semtner and C. Wunsch, 1996: How well does a 1/4° global circulation model simulate large-scale oceanic observations?, *Journal of Geophysical Research*, **101**(C10), 25779–25811.

Stevens, D. P., 1991: The open boundary condition in the United Kingdom Fine-Resolution Antarctic Model, *Journal of Physical Oceanography*, **21**, 1494–1499.

Stevens, D. P., and V. O. Ivchenko, 1997: The zonal momentum balance in an eddy-resolving general-circulation model of the Southern Ocean, *Quarterly Journal of the Royal Meteorological Society*, **123**, 929–951.

Stewart, R. H., 1985: *Methods of Satellite Oceanography*, University of California Press.

Stommel, H., 1957: A survey of ocean current theory, *Deep-Sea Research*, **4**, 149–184.

Stommel, H., 1948: The westward intensification of wind-driven ocean currents, *Transactions, American Geophysical Union*, **29**, 202–206.

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow, *Tellus*, **13**(2), 224–230.

Stössel, A., S. Kim and S. Drijfhout, 1999: The impact of Southern Ocean sea ice in a global ocean model, *Journal of Physical Oceanography*, **28**(10), 1999–2018.

Stutzer, S., and W. Krauss, 1998: Mean circulation and transports in the South

Atlantic Ocean: Combining model and drifter data, *Journal of Geophysical Research*, **103**(C13), 30985–31002.

Sverdrup, H., 1947: Wind-driven currents in a baroclinic ocean, with application to the equatorial currents of the eastern Pacific, *Proceedings of the National Academy of Sciences*, **33**, 318–326.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. part I: Theory, *Quarterly Journal of the Royal Meteorological Society*, **113**, 1311–1328.

Tang, Y., and R. Grimshaw, 1996: Radiation boundary conditions in barotropic coastal ocean numerical models, *Journal of Computational Physics*, **123**, 96–110.

Thacker, W. C., 1988: Fitting models to inadequate data by enforcing spatial and temporal smoothness, *Journal of Geophysical Research*, **93**(C9), 10655–10665.

Thacker, W. C., 1989: The role of the Hessian matrix in fitting models to measurements, *Journal of Geophysical Research*, **94**(C5), 6177–6196.

Thacker, W. C., and R. B. Long, 1988: Fitting dynamics to data, *Journal of Geophysical Research*, **93**(C2), 1227–1240.

The FRAM Group, 1991: An eddy-resolving model of the Southern Ocean, *Eos, Transactions, American Geophysical Union*, **72**(15), 169–175.

Thompson, S. R., 1983: Estimation of the transport of heat in the Southern Ocean using a fine-resolution numerical model, *Journal of Physical Oceanography*, **23**, 2493–2497.

Tomczak, M., and J. S. Godfrey, 1994: *Regional oceanography: an introduction*, Pergamon.

Trefethen, L. N., 1983: Group velocity interpretation of the stability theory of Gustaffson, Kreiss, and Sundström, *Journal of Computational Physics*, **49**, 199–217.

Treguier, A.-M., et al., 1999: The CLIPPER project: High resolution modelling of the Atlantic, *International WOCE Newsletter*, (36), 3–5.

Treguier, A., B. Barnier, A. de Miranda, J. Molines, N. Grima, M. Imbard, G. Madec, C. Messager, T. Reynaud and S. Michel, 2000: An eddy permitting model of the Atlantic circulation: evaluating open boundary conditions, *Journal of Geophysical Research*, submitted.

Treguier, A., and J. McWilliams, 1990: Topographic influences on wind-driven, stratified flow in a β-plane channel: An idealized model for the Antarctic Circumpolar Current, *Journal of Physical Oceanography*, **20**, 321–343.

Trenberth, K., 1979: Mean annual poleward transports by the oceans in the southern hemisphere, *Dynamics of Atmospheres and Oceans*, **4**, 57–64.

Tsimplis, M., S. Bacon and H. Bryden, 1998: The circulation of the subtropical South Pacific derived from hydrographic data, *Journal of Geophysical Research*, **103**(C10), 21443–21468.

Tziperman, E., and W. C. Thacker, 1989: An optimal-control/adjoint-equations approach to studying the oceanic general circulation, *Journal of Physical Oceanography*, **19**, 1471–1485.

Tziperman, E., W. C. Thacker, R. B. Long and S.-M. Hwang, 1992: Oceanic data analysis using a general circulation model. Part I: Simulations, *Journal of Physical Oceanography*, **22**, 1434–1457.

Tziperman, E., J. Toggweiler, Y. Feliks and K. Bryan, 1994: Instability of the thermohaline circulation with respect to mixed boundary conditions: Is it really a problem for realistic models?, *Journal of Physical Oceanography*, **24**, 217–232.

van Oldenborgh, G., 2000: What caused the onset of the 1997-98 El Niño?, *Monthly*

Weather Review, **128**(7), 2601–2607.

van Oldenborgh, G., G. Burgers, S. Venzke, C. Eckert and R. Giering, 1999: Tracking down the ENSO delayed oscillator with an adjoint OGCM, *Monthly Weather Review*, **127**(7), 1477–1495.

Varah, J., 1971: Stability of difference approximations to the mixed initial boundary value problems for parabolic systems, *SIAM Journal on Numerical Analysis*, **8**(3), 598–615.

Venzke, S., M. Latif and A. Villwock, 2000a: The coupled GCM ECHO-2. Part II: Indian ocean response to ENSO, *Journal of Climate*, **13**(8), 1371–1383.

Venzke, S., M. Munnich and M. Latif, 2000b: On the predictability of decadal changes in the North Pacific, *Climate Dynamics*, **16**(5), 379–392.

Verron, J., 1992: Nudging satellite altimeter data into quasi-geostrophic ocean models, *Journal of Geophysical Research*, **97**(C5), 7479–7491.

Vogeler, A., and J. Schröter, 1999: Fitting a regional ocean model with adjustable open boundaries to TOPEX/POSEIDON data, *Journal of Geophysical Research*, **104**(C9), 20789–20799.

Vogeler, A., and J. Schröter, 1995: Assimilation of satellite altimeter data into an open ocean model, *Journal of Geophysical Research*, **100**(C8), 15951–15963.

Wallace, J. M., and P. V. Hobbs, 1977: *Atmospheric science : an introductory survey*, Academic Press.

Wang, H.-H., and P. Halpern, 1970: Experiments with a regional fine-mesh prediction model, *Journal of Applied Meteorology*, **9**, 545–553.

Wang, X. H., and J. H. Jungclaus, 1996: Open boundary conditions in a three dimensional coastal ocean model, *Working Paper 1996/1*, School of Geography and Oceanography, Australian Defence Force Academy.

Warren, B. A., J. H. LaCasce and P. E. Robbins, 1996: On the obscurantist physics of "form" drag in theorizing about the Circumpolar Current, *Journal of Physical Oceanography*, **26**, 2297–2301.

Warren, B. A., J. H. LaCasce and P. E. Robbins, 1997: Reply, *Journal of Physical Oceanography*, **27**, 211–212.

Warren, B. A., J. H. LaCasce and P. E. Robbins, 1998: Reply, *Journal of Physical Oceanography*, **28**, 1655–1658.

Weaver, A. J., and E. Sarachik, 1991: The role of mixed boundary conditions in numerical models of the ocean's climate, *Journal of Physical Oceanography*, **21**, 1470–1493.

Webb, D. J., P. D. Killworth, A. C. Coward and S. R. Thompson, 1991: *The FRAM Atlas of the Southern Ocean*, Natural Environment Research Council, Swindon.

Webb, D., A. Coward, B. de Cuevas and C. Gwilliam, 1997: A multiprocessor ocean general circulation model using message passing, *Journal of Atmospheric and Oceanic Technology*, **14**(1), 175–183.

Wenzel, M., J. Schröter and D. Olbers, 2001: The annual cycle of the global ocean circulation as determined by 4D VAR data assimilation, *Progress in Oceanography*, **48**, 73–119.

White, N., R. Coleman, J. Church, P. Morgan and S. Walker, 1999: A southern hemisphere verification for the TOPEX/Poseidon satellite altimeter mission, *Journal of Geophysical Research*, **99**(C12), 24505–24516.

Whitworth, III, T., W. Nowlin, Jr. and S. Worley, 1982: The net transport of the Antarctic Circumpolar Current through Drake Passage, *Journal of Physical Oceanography*, **12**, 960–971.

Whitworth, III, T., and R. Peterson, 1985: Volume transport of the Antarctic

Circumpolar Current from bottom pressure measurements, *Journal of Physical Oceanography*, pp. 810–816.

Wilkin, J. L., and R. A. Morrow, 1994: Eddy kinetic energy and momentum flux in the Southern Ocean: Comparison of a global eddy-resolving model with altimeter, drifter, and current-meter data, *Journal of Geophysical Research*, **99**(C4), 7903–7916.

Witter, D. L., and D. B. Chelton, 1998: Eddy-mean flow interaction in zonal oceanic jet flow along zonal ridge topography, *Journal of Physical Oceanography*, **28**, 2019–2039.

Wolff, J.-O., 1994: Ocean modelling efforts in the global climate system, *Australian Meteorological Magazine*, **43**, 263–281.

Wolff, J.-O., 1999: Modelling the Antarctic Circumpolar Current: Eddy-dynamics and their parametrization, *Environmental Modelling and Software*, **14**(4), 317–326.

Wolff, J.-O., E. Maier-Reimer and S. Legutke, 1997: The Hamburg Ocean Primitive Equation model, *Technical Report 13*, Deutsches Klimarechenzentrum (DKRZ), Hamburg.

Wolff, J.-O., E. Maier-Reimer and D. J. Olbers, 1991: Wind-driven flow over topography in a zonal β-plane channel: A quasi-geostrophic model of the Antarctic Circumpolar Current, *Journal of Physical Oceanography*, **21**(2), 236–264.

Wong, A. P., N. L. Bindoff and J. A. Church, 1999: Large-scale freshening of intermediate waters in the Pacific and Indian oceans, *Nature*, **400**, 440–443.

Worthington, L., 1981: The water masses of the world ocean: Some results of a fine-scale census, in *Evolution of Physical Oceanography: Scientific Surveys in Honor of Henry Stommel*, edited by B. A. Warren, and C. Wunsch, pp. 42–69, The MIT Press.

Wright, D. G., W. P. O'Connor and J. Sheng, 2000: CANDIE: A 3D primitive equation model: Manual and model code, *Tech. rep.*, Department of Oceanography, Dalhousie University.

Wunsch, C., 1999: Where do ocean eddy heat fluxes matter?, *Journal of Geophysical Research*, **104**(C6), 13235–13249.

Wunsch, C., and D. Stammer, 1998: Satellite altimetry, the marine geoid, and the oceanic general circulation, *Annual Reviews of Earth and Planetary Sciences*, **26**, 219–253.

Wurtele, M., J. Paegle and A. Sielecki, 1971: The use of open boundary conditions with the storm-surge equations, *Monthly Weather Review*, **99**(6), 537–544.

Xu, W., T. Barnett and M. Latif, 1998: Decadal variability in the North Pacific as simulated by a hybrid coupled model, *Journal of Climate*, **11**(3), 297–312.

Yaremchuk, M., N. Bindoff, J. Schröter, D. Nechaev and S. Rintoul, 2001: On the zonal and meridional circulation and ocean transports between Tasmania and Antarctica, *Journal of Geophysical Research*, **106**(C2), 2795–2814.

Zieger, A. R., D. W. Hancock, G. S. Hayne and C. L. Purdy, 1991: NASA radar altimeter for the TOPEX/POSEIDON project, *Proceedings of the IEEE*, **79**(6), 810–826.