

Coordinated Ocean-ice Reference Experiments (COREs)

Stephen M. Griffies^a, Arne Biastoch^b, Claus Böning^b, Frank Bryan^c,
Gokhan Danabasoglu^c, Eric P. Chassignet^d, Matthew H. England^e, Rüdiger Gerdes^f,
Helmuth Haak^g, Robert W. Hallberg^a, Wilco Hazeleger^h, Johann Jungclaus^g,
William G. Large^c, Gurvan Madecⁱ, Anna Pirani^j, Bonita L. Samuels^a, Markus Scheinert^b,
Alex Sen Gupta^e, Camiel A. Severijns^h, Harper L. Simmons^k, Anne Marie Treguier^l,
Mike Winton^a, Stephen Yeager^c, Jianjun Yin^d

^a*NOAA Geophysical Fluid Dynamics Laboratory, Princeton, USA*

^b*Leibniz IfM-GEOMAR, Kiel, Germany*

^c*National Center for Atmospheric Research, Boulder, USA*

^d*Center For Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, USA*

^e*Climate Change Research Centre, University of New South Wales, Sydney, Australia*

^f*Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven, Germany*

^g*Max-Planck-Institut für Meteorologie, Hamburg, Germany*

^h*Royal Netherlands Meteorological Institute (KNMI), de Bilt, Netherlands*

ⁱ*Laboratoire d'Océanographie et du Climat: Expérimentation et Approches Numériques, CNRS-UPMC-IRD, Paris, France*

^j*International CLIVAR, and Princeton University AOS Program, Princeton USA*

^k*International Arctic Research Center, University of Alaska, USA*

^l*Laboratoire de Physique de Oceans, CNRS-IFREMER-UBO, Plouzané, France*

Abstract

Coordinated Ocean-ice Reference Experiments (COREs) are proposed as a tool to explore the behaviour of global ocean-ice models under forcing from a common atmospheric dataset. We highlight issues arising when designing coupled global ocean and sea ice experiments, such as difficulties formulating a consistent forcing methodology and experimental protocol. Particular focus is given to the hydrological forcing, the details of which are key to realizing simulations with stable overturning circulations. As an outcome of this analysis, we introduce a grid resolution hypothesis: models with sufficient resolution to capture key subpolar Atlantic processes are hypothesized to possess an overturning circulation that is less sensitive to hydrological cycle uncertainties.

The atmospheric forcing from Large and Yeager (2004) was developed for coupled ocean and sea ice models. We found it to be suitable for our purposes, even though its evaluation originally focussed more on the ocean than the sea-ice. Simulations with this atmospheric forcing are presented from seven global ocean-ice models using the CORE-I design. These simulations test the hypothesis that global ocean-ice models run under the same atmospheric state produce qualitatively similar simulations. The validity of this hypothesis is shown to be dependent on the chosen diagnostic.

Given the broad selection of models participating in this study, the simulations presented here can provide some feedback to the fidelity of the atmospheric forcing. That is, places where each model produces a similar behaviour that is biased relative to observations may signal a problem with the atmospheric dataset, thus suggesting areas requiring reexamination. The common bias could, in contrast, indicate a common problem amongst the full suite of models that may highlight problems in the model fundamentals and/or configuration. Analogously, in the case where a single model produces a widely varying behaviour, this outlier model may result from problems in the model's fundamentals and/or configuration. In general, the identification of problem areas promotes avenues for model development aimed at reducing the bias.

1. Introduction

Simulations with coupled ocean-ice models can be used to assist in understanding climate dynamics, and as a step towards the development of more complete coupled earth system models. Unfortunately, there is little consensus in the modelling community regarding the design of ocean-ice experiments, especially those run for centennial and longer time scales. In particular, there is no widely agreed method to force the models. Furthermore, some relatively small differences in forcing methods can lead to large deviations in circulation behaviour and sensitivities. Such difficulties create practical barriers to comparing simulations from different modelling groups.

1.1. Purpose of this paper

A central purpose of this paper is to propose Coordinated Ocean-ice Reference Experiments (COREs). COREs provide a common reference point for research groups developing and analyzing ocean-ice models. They do so by establishing a *standard practice* for the design of a baseline set of experiments that is useful for model development and ocean-ice research. By standard practice, we envision an experimental protocol that satisfies the following goals:

- Provides model simulations that can be tested directly against a broad suite of ocean and sea ice observations;
- Is not specific to a particular model or model framework, facilitating cooperation between groups and model communities;
- Is not so complex or computationally expensive so as to make it too onerous for smaller groups to implement;
- Can be incorporated into a more comprehensive model development or research program; e.g, by providing spun-up initial conditions for fully coupled climate simulations or control experiments in sensitivity studies;
- Facilitates sharing of expertise and reduces redundant efforts in forcing data set design.

It might be argued that prior to the availability of atmospheric reanalysis products, a *de facto* standard practice existed in the ocean modelling community: wind stress was prescribed by the only widely available global dataset (Hellerman and Rosenstein, 1983), and surface temperature and salinity were damped toward observed conditions (see Section 3.1). With the emergence of more comprehensive and realistic atmospheric reanalysis and remote sensing products, the choices have expanded but also become more complex. Our proposal for COREs does not provide the definitive resolution of these forcing issues, but can provoke discussion and debate leading to improved scientific convergence onto a common experimental protocol.

We distinguish the research focus of COREs from that of model intercomparison projects. In an intercomparison project, simulations follow a strict protocol and output is generated for analyses by a broad community. Projects, such as the Atmospheric Model Intercomparison Project (AMIP) (Gates, 1993), help document model similarities and differences, and can be of great use for various research and development purposes. Prior to deciding whether an analogous global ocean-ice model intercomparison project (i.e., an OMIP) would be a useful exercise, it is important for the research community to converge to a baseline experimental design. We believe that COREs provide a useful step toward this convergence.

A related purpose of this paper is to present details appropriate for designing an ocean-ice simulation. These details are rarely described in the peer-review literature, and thus are often rediscovered after much trial and error as research groups undertake the task of building ocean-ice models. It is essential for model groups to describe these details in order to design a controlled suite of comparison experiments and to identify areas where differences may be relevant for characterizing varying simulations. Although it is admittedly very difficult to get all groups to agree on the multitude of choices, it remains important to discuss and debate the various methods.

This paper contributes to both the design aspects of ocean-ice climate modelling, and to methods for analyzing these simulations. This analysis provides guidance for others who employ the CORE experimental design. It furthermore allows for an exami-

* Corresponding author.

Email address: Stephen.Griffies@noaa.gov (Stephen M. Griffies).

nation of the hypothesis that similarly forced models produce similar simulations. This hypothesis is shown to be quite dependent on the chosen diagnostic.

The identification of simulation differences prompts many questions regarding mechanisms accounting for the differences. The questions raised represent a central outcome of a model comparison, since in the absence of the comparison, many important questions tend to never get asked. Focused research under more controlled settings than available in the present study is required to uncover the mechanisms for model differences.

Given the broad selection of models participating in this study, the simulations presented here can provide some feedback to the fidelity of the atmospheric forcing. That is, places where each model produces a similar behaviour that is biased relative to observations may signal a problem with the atmospheric dataset, thus suggesting areas requiring reexamination. The common bias could, in contrast, indicate a common problem amongst the full suite of models that may highlight problems in the model fundamentals and/or configuration. Analogously, in the case where a single model produces a widely varying behaviour, this outlier model may result from problems in the model's fundamentals and/or configuration. In general, the identification of problem areas promotes avenues for model development aimed at reducing the bias.

1.2. Contents of this paper

This paper can be roughly split into three parts, each consisting of a few sections. The first part documents the state of the art in global ocean-ice coupled modelling. This part starts in Section 2, which highlights some uses of ocean-ice models, thus further arguing for the relevance of a reference experimental design. Section 3 reviews methods used to force the ocean-ice models, with emphasis on limitations of these methods. Section 4 then presents our proposal for COREs.

The second part of the paper presents a selection of diagnostics from simulations run with the CORE-I (repeating annual cycle) forcing. We start this part in Section 5, which presents a discussion of the globally averaged temperature and salinity as a function of time over the duration of the 500 year experiment. Section 6 follows up with horizontally averaged temperature and salinity biases for the 500 years, and

Section 7 shows the biases in surface temperature for the final decade of the simulations. Section 8 then discusses the annual cycle in upper ocean heat content at Ocean Weathership Echo ($48^{\circ}W, 35^{\circ}N$) in both observations and models. Section 9 considers the sea ice spin-up as seen in ice area, and then presents maps for the ice concentration in March and September averaged over simulation years 491-500. Section 10 focuses attention onto the Tropical Pacific thermocline and undercurrent, both of which are critical features for model studies of the El Niño Southern Oscillations (ENSO). Section 11 discusses the maximum monthly mixed layer depths found in the simulations over years 491-500. Section 12 presents the zonally averaged anomalies of temperature and salinity during the final decade of the 500 year simulations. Section 13 exhibits the time series for volume transport through the Drake Passage. Section 14 presents the global poleward heat transport. Sections 15 and 16 then consider the meridional overturning circulation (MOC), with emphasis on its sensitivity to hydrological forcing. Section 17 closes the main portion of the paper with discussion and conclusions.

The third part of this paper is presented in various appendices that detail aspects of the models used in this study, the experimental protocol, the methods used to force the models, diagnostics used to evaluate the simulations, and a list of acronyms used in the manuscript. Much of this material is often omitted in modelling papers. The absence of such details prevents other groups from unambiguously testing the robustness of published results, and this represents an example of irreproducible research. To further the science of ocean-ice modelling, such details should be documented and scrutinized within the peer-review literature.

2. Uses of ocean-ice models

To study the earth's climate, and possible climatic changes due to anthropogenic forcing, various research teams have successfully built realistic global climate or earth system models with interactive ocean, sea ice, land, atmosphere, biogeochemical, and ecosystem components (referred to as *climate models* in the following). These models are generally built incrementally, with components considered initially in isolation, then sub-groups of components are coupled, and finally the full set of components are brought together in the climate

model. This process requires a wide suite of scientific and engineering methods, from reductionist process physics and biogeochemical modelling, to wholistic climate systems science methods.

Ice covered regions of the polar and sub-polar oceans are of particular importance for the large scale circulation of the global oceans. In particular, sea ice melt and formation alter the thermohaline fluxes across the surface ocean, and greatly alter the buoyancy forcing affecting deep water formation and thus the large scale overturning circulation. Additionally, the presence of sea ice greatly alters the fluxes entering the ocean, due to the large insulating effects of ice cover relative to open ocean. Hence, realistic modelling studies of global ocean climate include a realistic interactive sea ice model coupled to the ocean.

Coupled ocean-ice models form an important subgroup in the climate system. They are often developed together prior to coupling to other components such as the land and atmosphere. From the perspective of a global climate modeller, the absence of an atmosphere and land component allows for a more focused assessment of the successes and limitations of the ocean-ice components. From the perspective of a global ocean modeller, introducing a sea ice model provides a physically based interactive method to determine high latitude ocean-ice fluxes, rather than the *ad hoc* approaches needed in global ocean-only models. Ocean-ice models also admit more dynamical degrees of freedom than possible in ocean-only simulations. In turn, running ocean-ice models is much more complex than ocean-only simulations, as they place a greater need on the accuracy required from surface boundary forcing, especially due to the ice-albedo feedback, whereby higher albedos arising from too much snow and ice reduces solar heating, thus further increasing the albedo.

The incremental methodology of climate model development is largely pragmatic. Namely, the fully coupled system is far more complicated, computationally expensive, and the ocean and sea ice components reflect errors in the modelled atmosphere. Thus, a component-wise method of development is essential, especially at the early stages. Additionally, for many research groups, ocean-ice models represent the final stage in the development of a tool of use for addressing certain scientific questions. For example, ocean-ice models form the basis for many refined resolution simulations in the high latitudes due to their reduced cost and complexity, with the re-

gional Arctic Ocean Model Intercomparison Project (AOMIP) providing one example with significant scientific impact (Proshutinsky et al., 2001; Holloway et al., 2007). In general, it is hoped that research and development efforts focused on ocean-ice simulations successfully assist in understanding the behaviour of the more complete climate system.

Although many useful insights can be garnered from studies with ocean-ice models, it is critical to understand their limitations. Namely, it often remains difficult to ensure that results from the ocean-ice subsystem carry over to the full climate system, where climate model behaviour, such as sensitivities to perturbations, can prove distinct from ocean-ice models. An example is provided in Section 16.1. Quite often, problems with ocean-ice models stem from unrealistic aspects of surface forcing from a non-interactive atmosphere (Section 3). Nonetheless, even with their limitations, ocean-ice models remain a valuable climate science tool, and so can be used for fruitful scientific research and model development purposes. We summarise here a few uses that motivate us to propose a standard practice for running these models.

- Being less expensive than climate models, ocean-ice models can be formulated with refined grid resolutions thus promoting superior representations of key physical, chemical, and biological processes as well as geographic features. Alternatively, they can be run with a broader suite of algorithms and parameterisations, which helps to develop an understanding of simulation sensitivity to model fundamentals.
- They provide a tool to study interactions between the ocean and sea ice as isolated from the complexities of atmospheric feedbacks and from biases that arise when coupling to a potentially inaccurate atmospheric model.
- Ocean-ice models forced with different atmospheric forcing provide a means to assess implications on the ocean and sea ice climate of various atmospheric reanalysis or observational products. As a complement, many models run using the same atmospheric forcing, and which show similar ocean biases, may suggest that there are problems with the atmospheric forcing. In these ways, models can provide feedback onto the development of atmospheric datasets used to force ocean-ice models (e.g., Large and Yeager, 2008). Alternatively, common model biases may indicate common model problems requiring new efforts at model formulation and/or parameterisations.

- Bulk formulae are needed to produce ocean-ice fluxes given an atmospheric state and ocean-ice state. Ocean-ice models run with the same atmospheric state yet with different bulk formulae allow one to assess the sensitivity of the simulation to the chosen bulk formulae.
- Run under realistic atmospheric forcing, models can be used to reproduce the history of ocean and sea ice variables and help to interpret observations that are scarce in space and time (e.g., Gerdes et al., 2005b). This approach provides a method for ocean *reanalysis* unavailable with fully coupled climate models. Notably, there are nontrivial issues of initial conditions and ocean drifts that need to be resolved before obtaining unambiguous results from such reanalysis studies.
- One can select particular temporal or spatial scales from within the forcing data for use in running ocean-ice models for purposes of understanding variability mechanisms.
- There is great utility for model development by comparing simulations from different ocean-ice models using the same atmospheric forcing. For example, comparisons often highlight deficiencies in the representation of physical processes, which then guide efforts to improve simulation integrity.
- Coupled ocean-ice models provide a valuable engineering step towards the development of more complete climate models. For example, many tools and methods needed to build climate models are more easily prototyped in the simpler ocean-ice models.

3. Boundary fluxes for ocean-only and ocean-ice models

A coupled ocean-ice model requires momentum, thermal and hydrological exchanges with the atmosphere to drive the simulated ocean and ice fields. These exchanges take the form of stress from atmospheric winds, of radiative and turbulent fluxes of heat, and of precipitation, continental runoff and evaporation. Notably, evaporation has an associated turbulent latent heat flux which links the thermal and hydrological fluxes. When decoupling the ocean and sea ice models from the atmosphere, one must introduce a method to generate these fluxes. We briefly review certain points related to this issue, highlighting problems that arise with various approaches.

3.1. Thermohaline fluxes from restoring SST and SSS

Perhaps the simplest and oldest approach to developing fluxes for ocean-only models is to specify a wind stress and to damp the model’s upper layer temperature (SST) and upper layer salinity (SSS)¹ to prescribed values (Cox and Bryan, 1984), such as from the climatologies of Levitus (1982), Conkright et al. (2002), or Steele et al. (2001). The thermohaline fluxes are thus generated without atmospheric information. However, fluxes are non-zero only when model predicted SST and/or SSS differ from observations. Furthermore, there is no direct link between the thermal and hydrological forcing present with latent heating and evaporation. Nonetheless, these fluxes provide a strong negative feedback that limits the errors that can be realized in the simulated surface ocean properties. Hence, this approach has been found to render a useful leading order understanding of the simulated ocean circulation, and to help identify egregious problems with ocean model fundamentals. It has thus been commonly employed by ocean modellers for many decades.

Damping the model predicted SST and SSS fields to prescribed values generates a restoring thermohaline flux for the ocean model. Unfortunately, the resulting fluxes can be quite unrealistic (Killworth et al., 2000), especially the freshwater fluxes (Large et al., 1997). It can also produce distortions in the simulated annual cycle (Killworth et al., 2000). Thermohaline damping is typically associated with rather short damping time scales (i.e., *strong restoring*), which can suppress potentially interesting internal modes of variability such as mesoscale eddies represented in refined resolution models. Damping becomes more problematic for a coupled ocean-ice model, because there is no proven analogue for driving a sea-ice model, and it is ambiguous how to restore to SST and SSS in regions with ice. Hence, thermohaline restoring with relatively strong damping is not an ideal means for generating thermohaline fluxes for ocean-ice climate modelling. An alternative should be considered.

¹ Modellers tend to equate the temperature and salinity in the upper model grid cell with the sea surface temperature and sea surface salinity. This equality is not precise, as the model grid cell values represent a grid cell averaged value, and so do not precisely reflect the surface skin values measured, say, from a satellite. See Robinson (2005) for more discussion.

3.2. Undamped thermohaline fluxes

Applying undamped thermohaline fluxes is a complementary method to the previous approach of damping SST and SSS. Consequently, it possesses complementary attributes, such as allowing surface tracers to evolve freely with no damping. Also, the prescribed surface fluxes can be adjusted to yield zero net gain of heat and freshwater by the ocean-ice system, and to give a desired equilibrium oceanic transport of heat and freshwater.

When using undamped fluxes, one must be more mindful of details than in the restoring case. Here, there are three types of thermohaline fluxes to consider:

- turbulent fluxes for heat (sensible and latent), water (evaporation), and momentum (wind stress);
- radiative heat fluxes (shortwave and longwave);
- water fluxes such as precipitation, river runoff, and sea ice formation/melt.

Unfortunately, fluxes from observations and/or reanalysis products have nontrivial uncertainties (Taylor, 2000; Large and Yeager, 2004). Running ocean-ice models for decades or longer with such large uncertainties can lead to unacceptable model drift in surface temperature and salinity (Rosati and Miyakoda, 1988). Additionally, SST anomalies *do* experience a negative feedback in the climate system, whereby they are damped by interactions with the atmosphere. Hence, SST restoring is based on physical interactions (Haney, 1971), and the lack of a negative feedback exacerbates problems with the undamped fluxes. Consequently, the undamped flux forced simulations can experience unacceptable drift associated with errors in the undamped fluxes and/or model errors, as well as the absence of a feedback mechanism to suppress drift. It is therefore generally not feasible nor physically relevant to run global ocean-ice models with undamped thermohaline fluxes for more than a few years.

3.3. Turbulent fluxes from bulk formulae

The turbulent sensible heat flux lost from the ocean is proportional to the sea-air temperature difference. As this difference increases (decreases), there is more ocean heat loss (gain) through the latent heat flux. Thus, the air-sea interaction represented by the turbulent heat fluxes tends to damp SST differences from the air temperature. The damping strength can be determined by nu-

merically linearizing the thermal boundary condition (Haney, 1971; Barnier et al., 1995; Rivin and Tziperman, 1997; Barnier, 1998). It can be quite strong in regions of strong winds such as the Southern Ocean and North Atlantic, where piston velocities can reach $1 - 2 \text{ m day}^{-1}$, which corresponds to a coupling strength of $50 - 100 (\text{W m}^{-2})/^\circ\text{K}$. More generally, Rahmstorf and Willebrand (1995) point out the scale dependence of the ocean-atmosphere heat flux coupling. Basin scale SST anomalies are damped at a much slower rate ($\sim 5 (\text{W m}^{-2})/^\circ\text{K}$), that is set by outgoing long wave radiation. They propose an approach with scale dependent bulk formulae for the ocean-atmosphere heat flux.

This feedback between the SSTs and the atmospheric state provides a nontrivial space-time dependent damping of SSTs that acts to reduce model drift.² As a means to model this and other air-sea interactions, in the absence of an interactive atmospheric model, a compromise can be made between the damped and undamped approaches by prognostically computing turbulent fluxes for heat, moisture, and momentum using the evolving ocean surface state (SST and surface currents). In this case, turbulent fluxes are computed from bulk formulae, given a prescribed, time evolving atmospheric state (air temperature, humidity, sea level pressure, and wind velocity). This approach directly corresponds to that used in climate models, where the atmospheric state is provided by a prognostic atmospheric model. In this way, the bulk formulae forced ocean-ice models are much more directly relevant to the coupled models than the other methods. They also properly link the latent heat flux and evaporation.

3.4. Problems with ocean-ice models forced by a prescribed atmosphere

The basic assumption made when using an atmospheric dataset to force an ocean-ice model is that changes in the prescribed near surface atmospheric state accurately reflect the surface turbulent heat and moisture fluxes across the ocean-ice surface, plus the divergence of all near surface internal atmospheric transport processes. The fundamental problem with the proposed bulk formulae approach is that in general this assumption is not

² Note that drifts in SST due to errors in the atmospheric forcing may actually lead to model drift.

valid, because of errors in the ocean-ice models, errors in the bulk formulae, and errors in atmospheric datasets. The latter represent only an approximation to Nature, and the uncertainties can be large. Furthermore, there is no unambiguous way of separating model error from forcing error in the simulated ocean-ice system, and errors can be both compensating and additive.

Even a perfect ocean-ice model is exposed to limitations inherent in the forcing and in the problems with decoupling from an interactive atmosphere. For example, a prescribed wind precludes atmospheric feedbacks that, in particular, contribute to the development and evolution of ENSO. Additionally, a prescribed air temperature results in an atmosphere acting as a fluid with infinite heat capacity, which is the opposite of the physically relevant limit where the ocean is more appropriately approximated as the slow climate component with a huge heat capacity. We now detail further problems associated with thermohaline forcing. These problems are intimately related, but we expose them here as separate mechanisms for clarity.

3.4.1. *Mixed boundary conditions and corruption of the temperature negative feedback*

The first problem relates to anticipated errors in the surface fluxes for salinity or fresh water, especially precipitation. These errors will force erroneous drift in ocean salinity. A relatively strong salinity restoring, analogous to the effective restoring of SSTs arising from bulk formulae, can control this drift in the ocean-ice simulations. However, salinity restoring has no physical basis. That is, precipitation fluxes do not depend on local salinity, and so there is no local negative feedback to mitigate the accumulation of flux errors. It is thus desirable physically to use at most a weak salinity restoring. Weak restoring rather than strong restoring allows increased, and typically more realistic, variability in the surface salinity and deep circulation. Furthermore, the weak restoring can be regarded as a correction to the precipitation.

The widely differing time scales determining surface moisture and heat fluxes leads to the term *mixed boundary condition* thermohaline fluxes. As emphasized by Stommel (1961), the differing effects of temperature and salinity on ocean density, as well as their distinct air-sea interactions, present the ocean's thermohaline circulation with the possibility for multiple regimes of circulation

(Bryan, 1986) and/or strong nonlinear oscillations (Zhang et al., 1993; Greatbatch and Peterson, 1996; De Verdière and Huck, 1999). Consider a positive salinity anomaly moving into the subpolar gyre region of the North Atlantic. This anomaly creates a positive density anomaly, which generally acts to support the large-scale overturning circulation with deep water formation in the northern part of the North Atlantic. This positive feedback from salinity is counteracted by a negative feedback from enhanced warm water advected northward, creating a negative density anomaly. We illustrate these feedbacks in Figure 1. Unfortunately, the negative temperature feedback is altered or removed when presenting the ocean with a prescribed atmospheric state which does not respond to the temperature anomaly. In turn, the positive salinity feedback plays a spuriously large role in the ocean-ice simulations using a prescribed atmosphere. The sensitivity of simulations to these altered feedbacks is detailed in Rahmstorf and Willebrand (1995) and Lohmann et al. (1996).

The ability of any ocean or ocean-ice model, in the absence of coupled feedbacks with the atmosphere, to represent the oceanic adjustment in the more complete climate system can be called into question. Consider the deep water formation regions of the North Atlantic, for example. Mixed boundary condition simulations with strong temperature restoring can be susceptible to unrealistically large amplitude thermohaline oscillations, as well as a polar halocline catastrophe, in which a fresh cap develops in high latitudes of the North Atlantic and shuts down the overturning circulation (Zhang et al., 1993; Rahmstorf and Willebrand, 1995; Rahmstorf et al., 1996; Lohmann et al., 1996).

3.4.2. *Absence of an atmospheric response as the ice edge or halocline moves*

The second problem relates to the lack of feedback onto a prescribed atmosphere when the model and forcing errors conspire to cause the simulated sea ice coverage to deviate from that used to produce the atmospheric state. Windy, cold, and dry air is often found near the sea ice edge in Nature. Interaction of this air with the ocean leads to large fluxes of latent and sensible heat which cool the surface ocean, as well as evaporation which increases salinity. This huge buoyancy loss increases surface density, which provides a critical element in the downward branch of the thermohaline circulation (e.g., Marshall and

Atm/Ocn feedbacks affecting the MOC

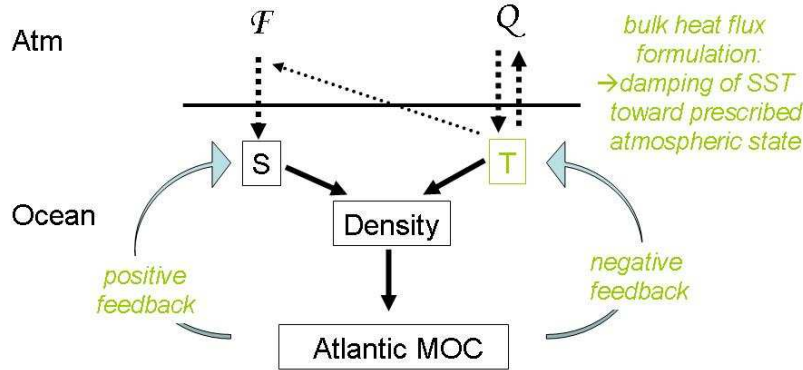


Fig. 1. This is a schematic of the large-scale atmosphere-ocean feedbacks related to temperature and salinity that affect the Atlantic meridional overturning circulation. The air-sea fluxes (F) affecting ocean salinity are nearly one-way, in that precipitation and evaporation are not affected by ocean salinity. In contrast, key pieces of the air-sea heat flux (Q), such as latent, sensible, and longwave fluxes, are directly determined by SSTs, and these fluxes damp oceanic surface temperature anomalies. Within the ocean, positive salinity anomalies advected into the northern North Atlantic strengthen the overturning (a positive feedback onto the overturning circulation), whereas positive temperature anomalies weaken the circulation (a negative feedback onto the overturning circulation). Altering the air-sea feedbacks on temperature, as when the atmospheric state is prescribed, can diminish the negative temperature feedback on the overturning. The result is a system that is overly sensitive to the salinity, as well as fluxes affecting salinity.

Schott, 1999). In contrast, the ocean under sea-ice is very effectively insulated from atmospheric cooling and buoyancy loss.

Suppose the modelled ice edge is too extensive. Then the air-sea cooling and evaporation can be spuriously shut down in the wrongly ice-covered region. Over the extended ice edge, the near freezing water will be under relatively warm air and weaker winds, so there will be less overall buoyancy loss to drive vertical mixing and convection in the ocean, as well as a negative feedback effect tending to melt back the ice. As a result the water column can become prone to freshwater pooling at the surface, which could provide a positive feedback on the reduced buoyancy loss. This process may be similar to the polar halocline catastrophe of mixed boundary condition models described above. In the opposite case of the modelled ice edge not being extensive enough, there would be excess buoyancy loss, a tendency for ice formation or overly strong vertical mixing and convection. The net effect on the simulated thermo-

haline circulation would be a weakening if the ice edge were overall too extensive, and a strengthening if too contracted. We illustrate these situations in Figure 2. The lack of feedback that exists in ocean-ice simulations can be largely eliminated in ocean-only simulations which use observed sea ice distributions to determine surface fluxes (Large et al., 1997; Large and Yeager, 2004). Unfortunately, there are no existing hindcast simulations to demonstrate the importance of this feedback on model solutions.

Lohmann and Gerdes (1998) and Jayne and Marotzke (1999) discuss other feedback processes involving sea ice that affect that sensitivity of the overturning circulation. In particular, errors in sea ice area affect the radiation balance of the earth, and thus the total meridional heat transport in the climate system and its partitioning between the atmosphere and ocean. In the ocean-ice system the feedback is positive with too much ice reducing the solar energy input, and too little ice increasing the solar input.

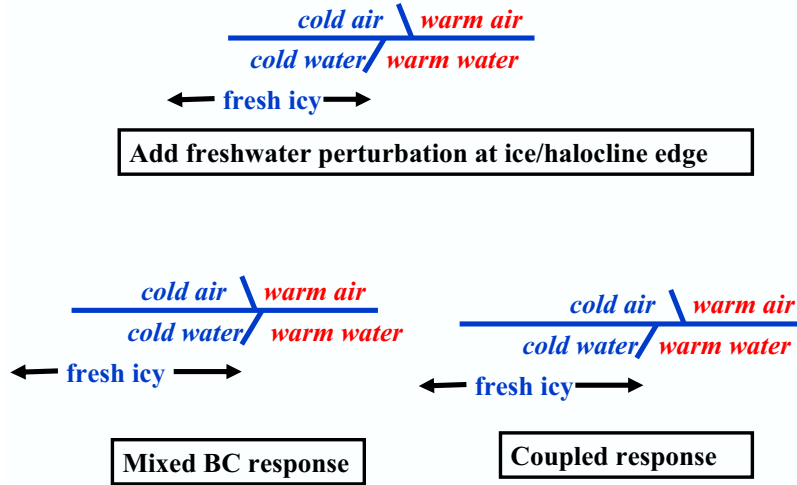


Fig. 2. This schematic illustrates the different responses of the ocean-ice system to changes in surface boundary fluxes. Top panel: Consider a cold-air outbreak (coming from the left and moving to the right in this schematic), such as occurs with synoptic activity over sub-polar regions. This outbreak results in cold and dry air occupying some area over relatively warm water, and this region often occurs near the sea ice or halocline edge. This situation results in huge air-sea fluxes of heat and evaporation in the region where cold and dry air is above relatively warm water. These fluxes drive deep water formation as a result of the huge loss of ocean buoyancy. The sea ice or halocline edge moves (towards the right in this schematic) in response to the large ocean buoyancy loss. With a prescribed atmospheric state (lower left panel), when the ice edge or halocline moves to the right, the prescribed atmospheric state does not feel this motion. The result is a removal of the region where large air-sea fluxes occur, thus rendering an unrealistic shut down of the air-sea fluxes that drive deep water formation. In a coupled climate model (bottom right panel), the atmosphere predominantly follows the ocean, so that as the ice edge or halocline moves to the right, there can remain a nontrivial deep-water formation region in front of the ice edge or halocline, just as in the top panel.

3.4.3. Salinity and/or fresh water normalisation

An issue related to fresh water or salinity forcing is the need to maintain a global balance. In the absence of an interactive atmosphere and land model, the global balance of water in the ocean-ice system will need to be maintained through a normalisation procedure. Namely, with a nonzero restoring applied to surface salinity, and with the diagnosed evaporation based on the bulk formulae and evolving SST, there is no guarantee that the hydrological cycle balances without a normalisation. Such is a limitation of the ocean-ice system absent an interactive land or atmospheric component. In this study, various groups perform this normalisation differently, with details provided in Appendix B3.

4. A proposal for COREs

The previous section highlights some issues that arise when decoupling the ocean and sea ice components from the rest of the climate system, in particular from an interactive atmosphere. Quite simply, it is ambiguous how one specifies interactions with unrepresented components, and these ambiguities can introduce nontrivial and often unphysical sensitivities. It is thus important to recognize the limitations of ocean-ice models, as no methodology for specifying interactions with missing components is immune to difficulties and ambiguities. Nonetheless, working under the assumption that we wish to conduct productive research and development with ocean-ice models, we seek a standard modelling practice for use in establishing benchmark simulations, thus facilitating comparisons and further refinements to the prescribed atmospheric states and experimental

design.

The standard practice we propose is termed a *Coordinated Ocean-ice Reference Experiment* (CORE). There are three COREs thus far comprising our suite of experiments. We emphasize the research nature of each CORE, and our goal here is both illustrative and provocative. Choices for experimental design largely depend on research goals. Underlying the present proposal is the goal to develop a protocol for ocean-ice simulations that leads to behaviour in reasonable alignment with corresponding climate model simulations using the same ocean-ice model components. Due to difficulties described above, we only partially succeed in this goal.

4.1. *The Large and Yeager dataset*

The critical element facilitating the CORE proposal is the existence of a comprehensive atmospheric data set that is useful across a range of ocean-ice model studies. In order to be widely applicable in global ocean-ice modelling, a dataset should produce near zero global mean heat and freshwater fluxes when used in combination with observed SSTs. This criteria precludes the direct use of atmospheric reanalysis products. Instead, the CORE experiments proposed here employ the dataset prepared by Large and Yeager (2004). Although primarily utilized and evaluated as forcing for the ocean component of the Community Climate System Model, NCAR-POP (Danabasoglu et al., 2006), we aim to use this dataset across a suite of ocean-ice models in this study. The Large and Yeager (2004) dataset is well documented, fully supported, periodically updated (Large and Yeager, 2008), and freely available, thus facilitating its use by the international climate modelling community.

As discussed in Taylor (2000), a combination of reanalysis and remote sensing products probably provides the best available choice to force ocean-ice models. That is the approach taken by Large and Yeager (2004). Their report (as well as the paper Large and Yeager (2008)) details methods to merge and "correct" various reanalysis and remote sensing data products to produce a comprehensive dataset for use in running ocean-ice models. This dataset, or earlier versions, have been used for many refined resolution ocean models (Maltrud and McClean, 2005; Marsh et al., 2005), and a similar approach has been used by other groups (e.g., Timmermann et al., 2005).

Furthermore, it is desirable for many research purposes to provide both a repeating "normal" year forcing (NYF) as well as an interannually varying forcing. Large and Yeager (2004) provide both, with the NYF consistently derived from the 43 years of interannual varying atmospheric state. The NYF allows one to focus on longer term signals, trends, the approach to equilibrium, and internally generated ocean variability on long time scales. Such is especially of interest for centennial scale simulations.

The normal year has been constructed to retain synoptic variability (i.e., atmospheric storms), with a seamless transition from 31 December to 1 January. However, air temperature, and humidities in the version of NYF used in this study were constructed to give appropriate atmospheric forcing of the ocean, but not necessarily of the sea-ice.

Both the normal year forcing and interannual varying data contain the following fields on a spherical grid of 192 longitude cells and 94 latitude cells.

- climatological annual mean continental runoff;
- monthly varying precipitation;
- daily varying shortwave and longwave radiative fluxes which are self-consistent, thus reducing cloud errors;
- six-hourly varying meteorological fields: 10m air temperature, humidity, air density, zonal wind, meridional wind, and sea level pressure.

There is no diurnal cycle of either radiation or precipitation, and no leap years.

Access to the dataset, Fortran code for the bulk formulae, technical report, support code, and release notes are freely available at

nomads.gfdl.noaa.gov/nomads/forms/mom4/CORE.html

This web page is supported for COREs by a collaboration between scientists at the National Center for Atmospheric Research (NCAR) and NOAA's Geophysical Fluid Dynamics Lab (GFDL).

The task of developing an atmospheric dataset suitable for global ocean-ice climate models is fraught with uncertainty and ambiguity. As argued in Section 2 and practiced in Section 10, one use of ocean-ice models is to assist in evaluating these datasets. Hence, the datasets must undergo regular reevaluation and updates to reflect newly acquired data as well as feedback from the modelling community. In particular, a version of NYF more appropriate for sea-ice forcing is presently under development.

4.2. Bulk formulae and salinity restoring

During early stages of this project, we originally thought that differences in algorithms and parameters used to compute bulk formulae would lead to trivial differences in ocean circulation relative to other model differences. This assumption proved to be wrong. A preliminary comparison between bulk formulae used in the Community Climate System Model (CCSM) and the GFDL model led to flux differences that were far too large to ignore when the goal is to run models with the same atmospheric forcing given the same ocean-ice state. In particular, the momentum stresses from atmospheric winds were larger with the GFDL formulation (based on Beljaars, 1994) and the latent heat fluxes were larger with the CCSM formulation (described in Large and Yeager (2004) and Large (2005)). The differences have been traced to differences in the neutral transfer coefficients (roughness lengths). This result highlights the use of ocean-ice models to compare simulations with different bulk formulae. These simulations can provide useful feedback onto the refinement of bulk formulae. Such, however, is not the purpose of the present study.

The Large and Yeager (2004) atmospheric state was developed using the CCSM bulk formulae. These formulae represent fits to observed data in both stable and unstable conditions spanning wind speeds from less than 1 m s^{-1} to more than 25 m s^{-1} . Given the close relation between the derived atmospheric state and the bulk formulae, we decided that all models in this study would employ the CCSM formulae, rather than each group using their own particular formulae.

Salinity or fresh water forcing was a point of debate amongst the participants in this study, largely due to difficulties raised in Section 3.4. The basic question is: how strongly should SSS be restored? Some simulations removed restoring under sea ice, whereas others retained restoring. Some ran with extremely weak restoring with the piston velocity of $50 \text{ m}/4 \text{ years}$, and some explored a range of restoring scenarios. We have more to say on this issue in Section 16. Quite generally, these issues highlight the utility of complementing ocean-ice simulations with fully coupled climate simulations, where ambiguities with salinity forcing are absent.

Choices made by each of the ocean and sea ice model for salinity restoring and salt/water normalisation are detailed in Section B3. They can be sum-

marised as follows:

- weak salinity restoring for NCAR-POP, FSU-HYCOM, and MPI;
- strong salinity restoring for GFDL-MOM, GFDL-HIM, and Kiel-ORCA;
- variable salinity restoring for KNMI-HYCOM
- salt/water normalisation for all, except FSU-HYCOM, Kiel-ORCA, and KNMI-MICOM.

4.3. Three proposed COREs

We propose three COREs, whose basic elements are outlined here.

- CORE-I: This experiment is aimed at investigations of the climatological mean ocean and sea ice states realized using the idealized repeating NYF of Large and Yeager (2004). Models should ideally be run to quasi-equilibrium of the deep circulation, which is on the order of many hundreds to thousands of years (England, 1995; Stouffer, 2004).
- CORE-II: This experiment is aimed at investigations of the forced response of the ocean and/or ocean hindcast. It therefore will employ a more recent version of the interannually varying dataset from Large and Yeager (2008), rather than the idealized repeating normal year. CORE-II may also facilitate more direct comparisons with observations of time dependent phenomena, and thus be of direct use for ocean reanalysis. It is critical to note that the utility of these experiments depends largely on the impact of initial conditions as well as model drift. These issues remain at the forefront of present research.
- CORE-III: This is a perturbation experiment involving ideas proposed by Gerdes et al. (2005a, 2006). Here enhanced fresh water flux enters the North Atlantic in response to increased meltwater runoff distributed around the Greenland coast. Response of the regional and global ocean and sea ice system on the decadal to centennial time scales is the focus of CORE-III. This experimental design is motivated by possible increases in Greenland meltwater that may occur due to anthropogenic global warming.

We focus in this paper on CORE-I. During the early stages of exploring CORE-I simulations, we hoped that 100 years would provide a sufficient time to expose general model behaviour and model differences. 100 years was the choice taken for the comparison of German ocean-ice models discussed in

Fritzsch et al. (2000) which used the forcing from Röske (2005). Unfortunately, 100 years proved insufficient for highlighting differences of overturning circulation behaviour. In particular, drifts in the water masses in some of the simulations caused either the overturning circulation to drastically weaken within 100 years, or to experience unrealistic oscillations after a few hundred years (Sections 15 and 16). Simulations of 500 years length exposed many of these issues, whereas 100 years was insufficient. Notably, even though many issues were exposed only after multiple-century integrations, there is no guarantee that 500 years is sufficient to sample the phase space of the models run with the CORE-I design. 500 years is therefore considered a pragmatic compromise amongst the participants in this study.

4.4. Differences in methods

Use of the Large and Yeager (2004) NYF dataset and bulk formulae with no temperature restoring for 500 year ocean-ice simulations is basically what defines CORE-I. This experimental design leaves open many details for each group to choose based on their judgement. Consequently, as shown in the Appendices, experimental design and model details followed by the groups differed in many aspects. For various reasons based on specifics of numerical algorithms, computational and human resources, and/or contrary scientific judgements, we were unable to remove all differences. Indeed, we did not put much effort at reducing these differences, as such would have sacrificed our ability to make progress towards a common experimental framework.

Certainly some differences in methods are expected with comparisons, and as such, can add to the strength of the project by exposing alternative approaches to the scrutiny of a larger group of scientists. Nonetheless, differences in model formulation and implementation of forcing add to the difficulty of uncovering mechanisms for simulation disagreements. For example, no two models used precisely the same grid resolution; some models used implied salt fluxes while others used real water fluxes (see Appendix B3 for details); and differences in ice albedo schemes were common. Such differences might be important for determining why, as shown later, the models exhibit varying behaviours of their simulated Atlantic overturning circulations.

4.5. Models in this study

The ocean and sea ice models employed in this study include the following (see Appendix A for details and references):

- NCAR-POP: This model is comprised of the ocean and sea ice components from the CCSM climate model using a zonal resolution of roughly one degree, with enhanced meridional resolution in the tropics. The ocean component uses geopotential vertical coordinates.
- FSU-HYCOM: This model is comprised of the HYCOM ocean model code within the CCSM framework used in the NCAR-POP simulations, with the same horizontal grid resolution, coupler, and sea ice model. The ocean component uses hybrid isopycnic-pressure coordinates, with pressure in the upper ocean mixed layer and isopycnic beneath.
- GFDL-MOM: This model is comprised of the ocean and sea ice components from the GFDL climate model using a zonal resolution of one degree, with enhanced meridional resolution in the tropics. The ocean component uses geopotential vertical coordinates.
- GFDL-HIM: This model replaced the geopotential MOM code with the isopycnal layered Hallberg Isopycnal Model (HIM), in which the vertical is discretized with potential density layers. The vertical and horizontal resolution is comparable to the GFDL-MOM simulation.
- KNMI-MICOM: This model is based on the MICOM isopycnal ocean model with zonal resolution of two degrees and enhanced meridional resolution in the tropics. The sea ice component is from Bentsen et al. (2004).
- MPI: This is the ocean and ice model components of the coupled climate model from the Max-Planck-Institute. The horizontal resolution gradually varies between 12km close to Greenland and 150km in the tropical Pacific. The ocean component uses geopotential vertical coordinates.
- Kiel-ORCA: This model is comprised of the NEMO modelling system, with the OPA 9 ocean model coupled to the LIM sea ice model with two degree zonal resolution, with enhanced meridional resolution in the tropics. The ocean component uses geopotential vertical coordinates.

All geopotential models, as well as GFDL-HIM, employ the Boussinesq approximation, in which volume, not mass, of a fluid parcel is conserved,

and thus steric effects are absent from the prognostic equations. In contrast, the KNMI-MICOM and FSU-HYCOM ocean codes are both non-Boussinesq.

4.6. Goals of the analysis

The following sections survey results from simulations run with the ocean-ice models listed above using the CORE-I forcing. A key purpose of this presentation is to be illustrative and provocative rather than thorough on all points. That is, the analysis fails to fully assess each model’s ability to remain faithful to Nature’s ocean-ice system. Furthermore, the analysis is insufficient to identify mechanisms for model differences. Nonetheless, we do provide descriptions of the gross simulation features, and in certain places we provide tentative conjectures and criticisms that may help to explain model biases.

5. Globally averaged ocean temperature and salinity

Amongst the most basic of model diagnostics is the globally averaged ocean temperature and salinity. Assuming no interior sources and sinks, the globally averaged ocean temperature and globally averaged ocean salinity are affected by surface fluxes, and by exchange of heat and salt with the sea ice. Sections D5 and D6 in the Appendix detail the various processes contributing to the globally averaged temperature and salinity.

5.1. Descriptive comments

It is unlikely that the prognostic model SST and SSS will match observations, so that longterm global heat and freshwater trends, arising from nonzero surface flux “imbalances”, can be expected. Nonetheless, given that the atmospheric state is prescribed over a fixed annual cycle, and assuming the ocean model algorithms are based on conservative numerical methods, we expect the globally averaged temperature and salinity to reach an equilibrium state after a transient phase. The time scale for equilibration can be centuries to millenia, depending on the rates of ocean ventilation and mixing (England, 1995; Stouffer, 2004). Figures 3 and 4 present time series for the globally averaged temperature and salinity in the CORE-I simulations. The results are quite distinct between the models,

with some reaching relatively stable results after a few hundred years, whereas others continue to exhibit drift after 500 years. Note also the nonzero model differences at the start, with these differences probably arising from details of the initial conditions for the ocean and sea ice models (see Section B1 for details of the initial conditions).

For global mean salinity, the NCAR-POP, MPI, and GFDL-MOM simulations show nearly stationary behaviour throughout the full integration. For NCAR-POP, this result suggests that the normalisation procedure used for global mean salt is serving to counteract any net salt input to the ocean from restoring (see discussion of salt/water normalisation in Section B3). Both MPI and GFDL-MOM employ water fluxes, and so the only transfer of salt into or out of the liquid ocean arises from the small amount exchanged with sea ice formation and melt. The stationary behaviour of global mean salinity indicates that potential drifts in global mean seawater volume are negligible, which is a result of the normalisation employed by each model to keep this drift small.

The Kiel-ORCA simulation shows a slow steady increase in global mean salinity. This model employs water fluxes and normalises these fluxes. However, the model includes relaxation to climatological temperature and salinity in the Mediterranean and Red Seas, which is a source of heat and salt. In a 100 year experiment without this relaxation, the global salinity increase is ten times smaller.

The GFDL-HIM, FSU-HYCOM, and KNMI-MICOM simulations each show a steady downward drift in global mean salinity, with KNMI-MICOM showing the most drift and GFDL-HIM the least. For the GFDL-HIM simulation, it is unclear what is the origin of the drift, given that it employs a normalisation to counteract any net input of water to the system. In contrast, the KNMI-MICOM and FSU-HYCOM simulations do not employ any normalisation, and this may explain the relatively large salinity drifts.

For global mean temperature, GFDL-MOM, MPI, and FSU-HYCOM show a general warming, with GFDL-MOM and MPI reaching a steady state around 4°C after about 300 years. In contrast, FSU-HYCOM shows a general warming for 300 years, then begins to cool, and continues to do so until the end of the 500 years. NCAR-POP and Kiel-ORCA show the least trend away from the initial conditions, with 500 year mean temperatures near the 3.6°C initial global mean temperature. Both appear to have reached a near steady state, though NCAR-

POP exhibits a slight cooling drift at year 500. The GFDL-HIM and KNMI-MICOM simulations both exhibit cooling, with the KNMI-MICOM simulation approaching global mean temperatures of 1.0°C , with little sign that this downward trend is slowing. In contrast, the GFDL-HIM simulation appears to be reaching an equilibrium at around 3.0°C .

6. Horizontally averaged temperature and salinity

More details about the temperature and salinity spin-up in the CORE-I simulations are provided in Figures 5 and 6. These figures show time series for the anomalous annual mean temperature and salinity as a function of depth, where the anomalies were created by taking the difference between the annual mean from the model and the annual mean from Conkright et al. (2002) and Steele et al. (2001). The near-surface and thermocline conditions show a rapid adjustment during the first 50-100 years, with comparatively small drifts thereafter. In contrast, deeper properties generally continue to drift throughout the simulation period.

For the temperature drifts, NCAR-POP shows an upper ocean warming, with waters below 1000m generally cooling. This cooling accelerates a bit towards the end of the simulation, consistent with the downward trend in global mean temperature seen in Figure 3. FSU-HYCOM shows an overall warming in the upper 500m, with little warming in the deeper oceans. The warming appears to be reversing as the 500 year mark is approached, which again is reflected in the cooling trend in global mean temperature starting around year 250. GFDL-MOM shows a general warming throughout the upper 2000m, with the exception of a cooling in the upper 100m. The deeper ocean shows very little trend. GFDL-HIM shows an upper ocean warming above 500m, and cooling throughout the deeper regions, extending to the abyss. The cooling dominates the full column, as revealed by the global mean time series in Figure 3. Kiel-ORCA shows the least drift of all the simulations above 2000m, with only a modest warming in the upper 1000m. The deeper ocean shows a small cooling trend. KNMI-MICOM, in contrast, shows the largest drift, with warming over the upper 100m, and strong cooling throughout the remaining ocean column. Finally, the MPI solution shows a slight cooling over the upper 100m, a strong warming beneath down to 2000m, and then a slight cool-

ing in the abyss.

For salinity drift, NCAR-POP shows a strong freshening in the upper 200m, counteracted by an increasing salinity down to around 1000m, and a slight freshening in the abyss. This characteristic drift pattern is also reflected in the MPI simulation, but with the MPI simulation showing a smaller drift magnitude. It is notable that these two simulations use the weak salinity restoring with a piston velocity of 50m/4years and normalisation. FSU-HYCOM also uses the weak salinity restoring, but without normalisation, and this simulation does not show the characteristic pattern of NCAR-POP and MPI. Kiel-ORCA does exhibit a similar pattern, though far more diffuse in the vertical and with a smaller amplitude. The GFDL-MOM and GFDL-HIM simulations show a strikingly similar drift pattern, with salty trend above roughly 1000m, and fresh trend beneath. Finally, the FSU-HYCOM and KNMI-MICOM simulations both show an overall freshening trend, with the FSU-HYCOM trend far smaller than KNMI-MICOM. Kiel-ORCA arguably has the smallest trend for all depths, though the global mean time series in Figure 4 indicates that the simulation has yet to settle down to a steady state to the degree seen in the NCAR-POP, MPI, and GFDL-MOM simulations. Again, both the FSU-HYCOM and KNMI-MICOM simulation do not perform a normalisation for the hydrology in the ocean-ice system (see Table 3), whereas all other simulations provide some means to ensure that the net salt or water entering the ocean-ice system remains within bound.

7. SST and SSS bias maps

Global maps of SST and SSS from the simulations are compared in Figures 7 and 8 to those from Conkright et al. (2002) for the World Ocean outside the Arctic, with Steele et al. (2001) used for the Arctic. Despite the strong negative feedback provided by the prescribed air-temperature, the models develop some regions with nontrivial difference patterns, and these are often found in multiple models. These results suggest that common modelling problems and/or deficiencies in the forcing may be responsible for the differences. It is notable that there are basin scale regions where Conkright et al. (2002) and Steele et al. (2001) SST tends to be greater than the Reynolds and Smith (1994) SST by 0.5°C to 1°C (e.g., North Atlantic subtropical gyre, Southern

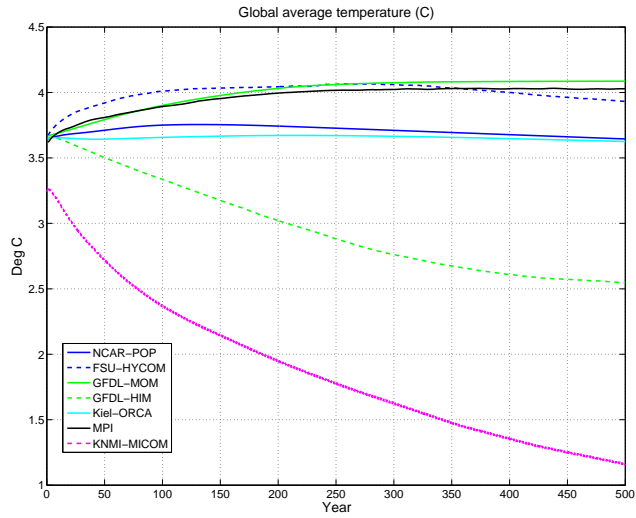


Fig. 3. Time series of the globally averaged annual mean liquid ocean temperature from the CORE-I simulations. This diagnostic is computed as $\langle \theta \rangle = \frac{\sum \theta dV}{\sum dV}$, with dV the grid cell volume, and the summation extending over the full ocean domain.

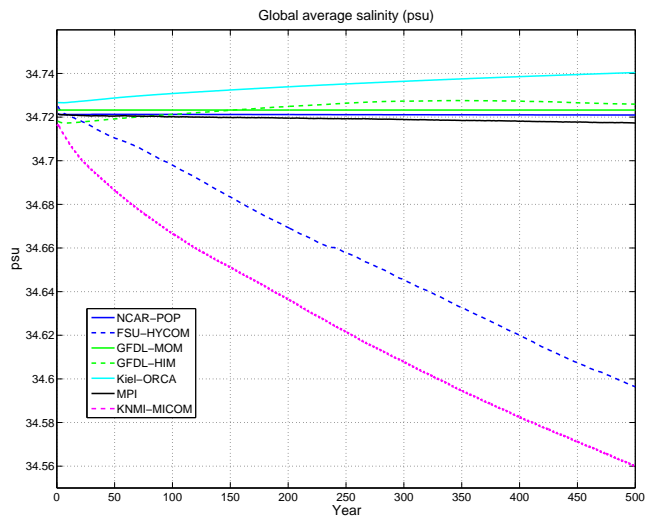


Fig. 4. Time series of the globally averaged annual mean ocean salinity from the CORE-I simulations. This diagnostic is computed as $\langle S \rangle = \frac{\sum S dV}{\sum dV}$, with dV the grid cell volume, and the summation extending over the full ocean domain.

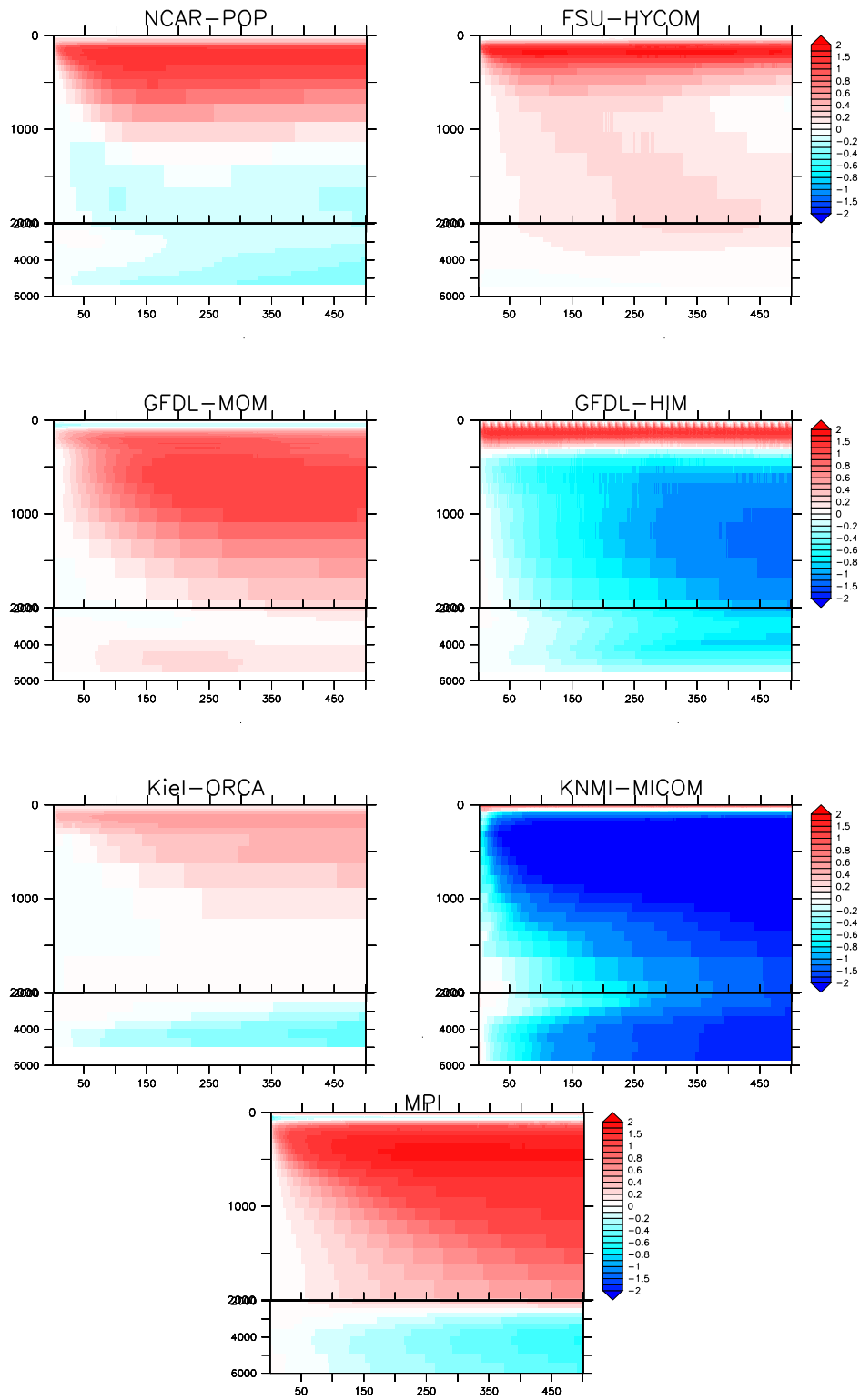


Fig. 5. Globally averaged drift of the annual mean temperature (degrees C) as a function of depth (metres on vertical axis) and time (years on horizontal axis). This drift is defined as $T_{\text{drift}}(z, t) = \left(\sum_{xy} dx dy dz (T_{\text{model}} - T_{\text{initial}}^{\text{ann}}) \right) / \left(\sum_{xy} dx dy dz \right)$, where $T_{\text{initial}}^{\text{ann}}$ is the annual mean from Conkright et al. (2002) and Steele et al. (2001), and \sum_{xy} is a horizontal sum. The upper 2000m is expanded relative to the deeper ocean, in order to highlight the generally larger surface drifts.

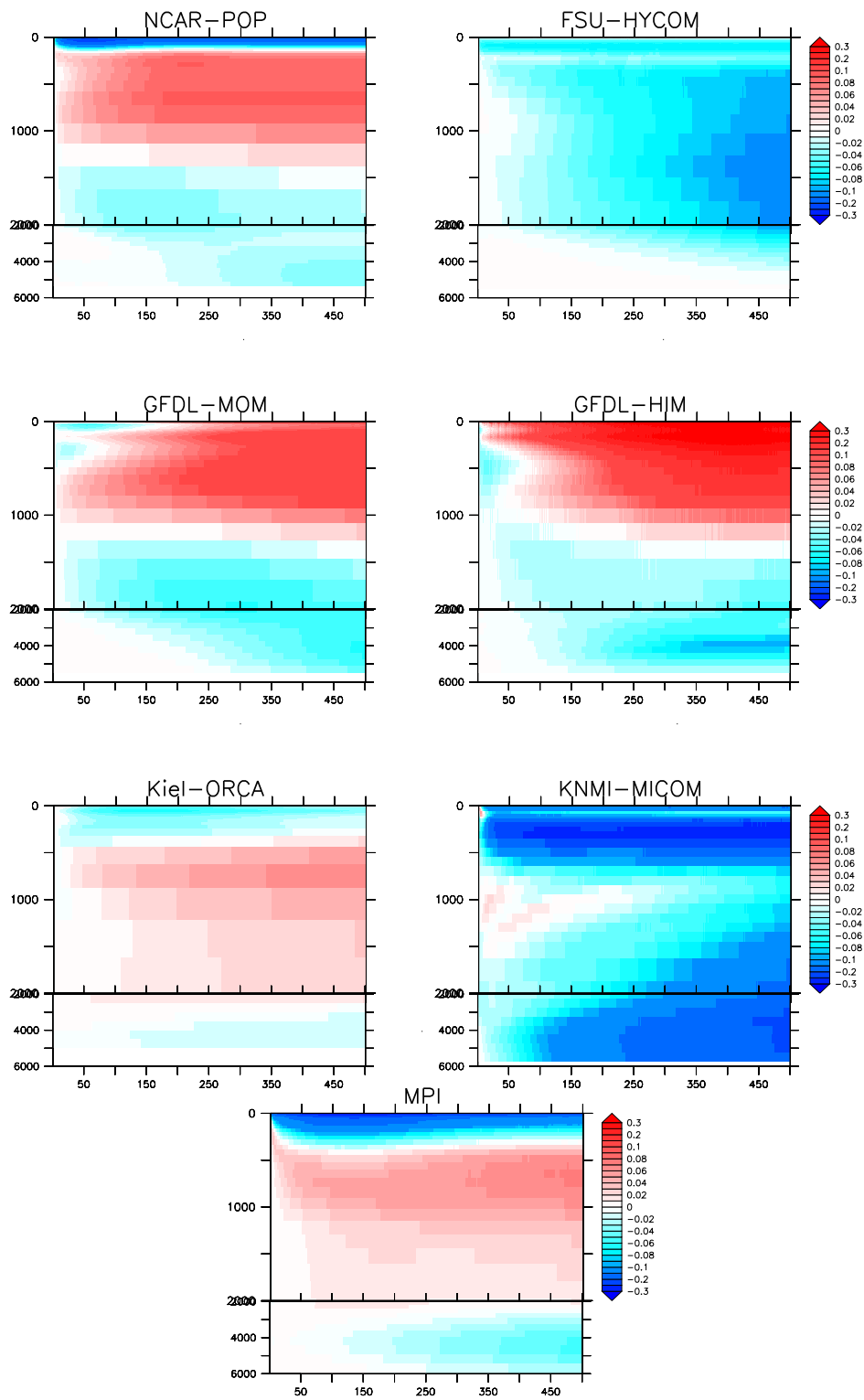


Fig. 6. Globally averaged drift of the annual mean salinity (psu) as a function of depth (metres on vertical axis) and time (years on horizontal axis). This drift is defined as $S_{\text{drift}}(z, t) = \left(\sum_{xy} dx dy dz (S_{\text{model}} - S_{\text{initial}}^{\text{ann}}) \right) / \left(\sum_{xy} dx dy dz \right)$, where $S_{\text{initial}}^{\text{ann}}$ is the annual mean from Conkright et al. (2002) and Steele et al. (2001), and \sum_{xy} is a horizontal sum. The upper 2000m is expanded relative to the deeper ocean, in order to highlight the generally larger surface drifts.

Ocean from $60^{\circ}E$ to Drake Passage) and colder by $0.5^{\circ}C$ to $2^{\circ}C$ (South Atlantic subtropical gyre, offshore the Indian Ocean sector of Antarctica). Therefore, only larger differences in Figure 7 are robust.

The SST biases reveal some anomaly patterns commonly found in models with the non-eddy permitting resolutions used here (see Table 1 in Appendix A for model details). In particular, large deviations are found along major frontal zones such as the western boundary currents in the North Atlantic and North Pacific. Biases in current structure generally lead to warm biases next to the adjacent continents, due to the spuriously northward extent of the warm boundary currents along the coasts, rather than the separation of the currents into the interior.

In the North Atlantic subpolar gyre region, all models show signs of difficulties maintaining the North Atlantic drift in the proper position, as well as problems navigating the northwest corner near Newfoundland. Problems with these currents leads to strong cold and warm biases, depending on details of the simulated currents.

All models exhibit a warm bias near the west coasts of the American and African continents. This bias is possibly due to poorly resolved coastal upwelling associated with the coarse grid resolution. There may also be a contribution from errors in the direction of near coastal winds. Note that due to the use of satellite radiation in Large and Yeager (2004), these warm biases are not due to problems with simulated clouds which may be present in re-analysis products.

In the Pacific, both the MPI and KNMI-MICOM simulations show a distinct cool pattern in the central and eastern tropical Pacific, with warming elsewhere in the basin. The cool patterns are noticeably larger than the other simulations. Indeed, all but FSU-HYCOM show a characteristic slight cooling in the central equatorial Pacific. FSU-HYCOM, in contrast, shows a slight warming extending from South America, gradually becoming a slight cooling in the far western equatorial Pacific.

As discussed in Section D6 in the Appendix, the ocean salinity is affected by the hydrological cycle associated with the prescribed atmospheric state; the use of a salinity or fresh water restoring term; and the presence of an overall normalisation to counteract potentials for drift. Details of the choices made by the various models for handling these issues differ somewhat, thus prompting us to expect there to be some differences in the evolution of global mean salinity (Section 5), horizontally

averaged salinity (Section 6), and SSS.

The NCAR-POP, FSU-HYCOM, and MPI simulations each use the very weak salinity restoring, with a piston velocity of $50m/4years$. For MPI, this weak restoring allows for the development of a rather salty surface bias in the Arctic, and generally fresh remainder of the World Ocean. NCAR-POP has a smaller SSS anomalies, especially in the tropical regions, though with noticeable fresh anomalies in the high northern latitudes. The FSU-HYCOM and Kiel-ORCA simulations both show strikingly small SSS anomalies. FSU-HYCOM uses the same very weak salinity restoring as NCAR-POP and MPI whereas Kiel-ORCA uses the relatively strong salinity restoring of $50m/300days$. For the FSU-HYCOM simulation, however, one should note that this model is undergoing a rather large fresh trend in the globally averaged time series of Figure 4. So it is unclear whether this small SSS anomaly pattern is stationary. The GFDL-MOM and GFDL-HIM simulations both show an overly salty SSS in most locations, with GFDL-HIM having significantly larger biases. Finally, the KNMI-MICOM simulation shows rather large biases across the northern latitudes, with the Southern Hemisphere maintaining modestly small biases.

8. Annual cycle at Ocean Weather Ship Echo

Analysis of the long-term adjustment behaviours seen in Sections 5-7 is complemented by an inspection of the annual cycle of near-surface thermal properties. In general, the temporal rate of change of ocean heat storage is balanced by the surface heat flux and the horizontal divergence of heat advection and diffusion in the ocean. As suggested by theory (Gill and Niiler, 1973) and confirmed by model simulation (Sarmiento, 1986) and observations (Hsiung et al., 1989), the seasonal cycle of heat content in the mid-latitudes is mainly balanced by the local surface flux (in contrast to the tropical oceans where heat content change is governed by transport divergence). A conspicuous deficiency of previous generation models was a much too weak seasonal heat storage in the mid-latitude regime (Sarmiento, 1986). Two critical factors for improving the amplitude of the summer heat gain were noted in model sensitivity studies: the surface heat flux formulation and the parameterisation of vertical mixing in the surface boundary layer (Böning and Herrmann, 1994). In this section, some

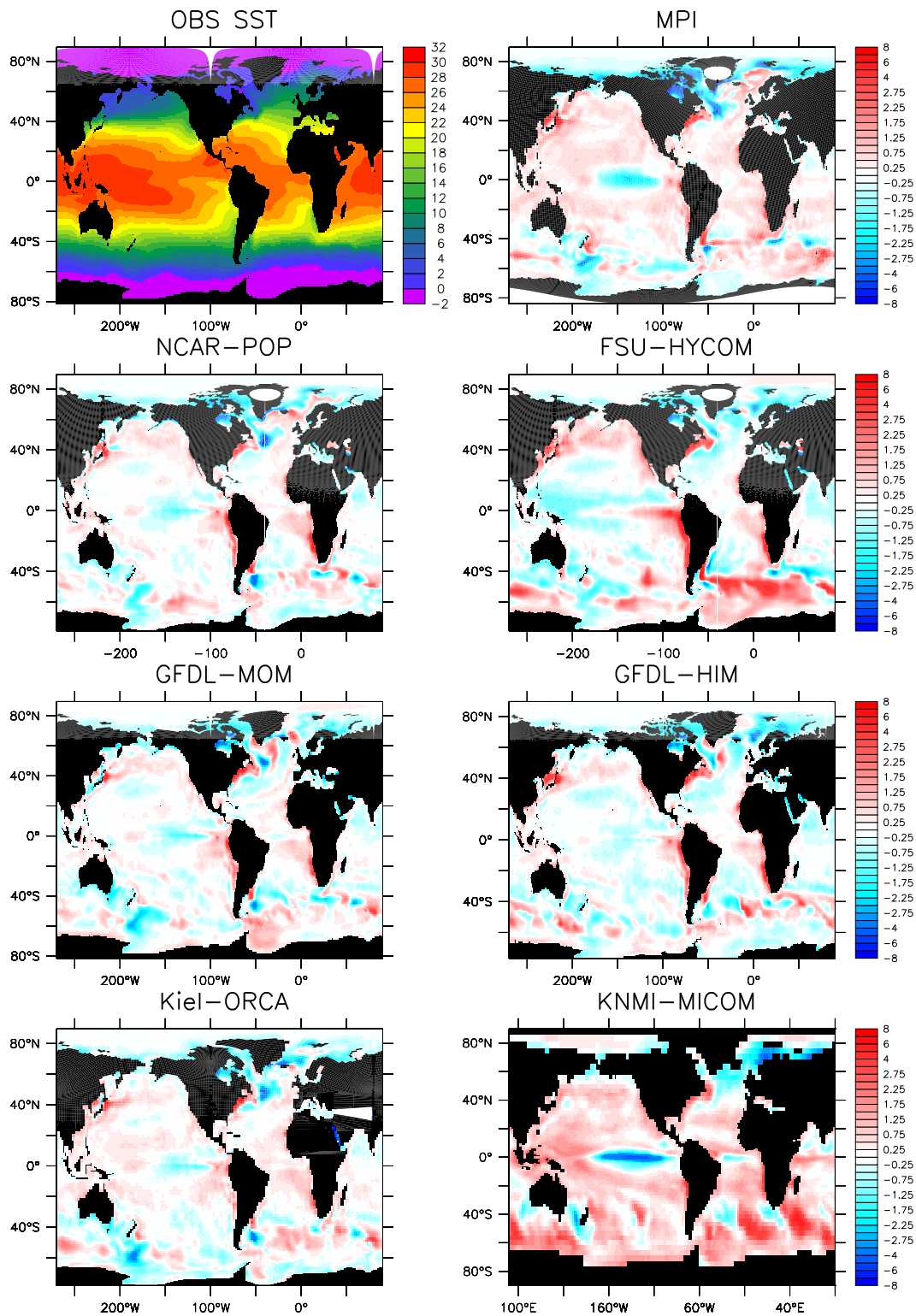


Fig. 7. Anomalous SST (degrees C) for years 491-500 from the simulations relative to the analysis (shown in the top middle panel) of Conkright et al. (2002) outside the Arctic, and Steele et al. (2001) in the Arctic.

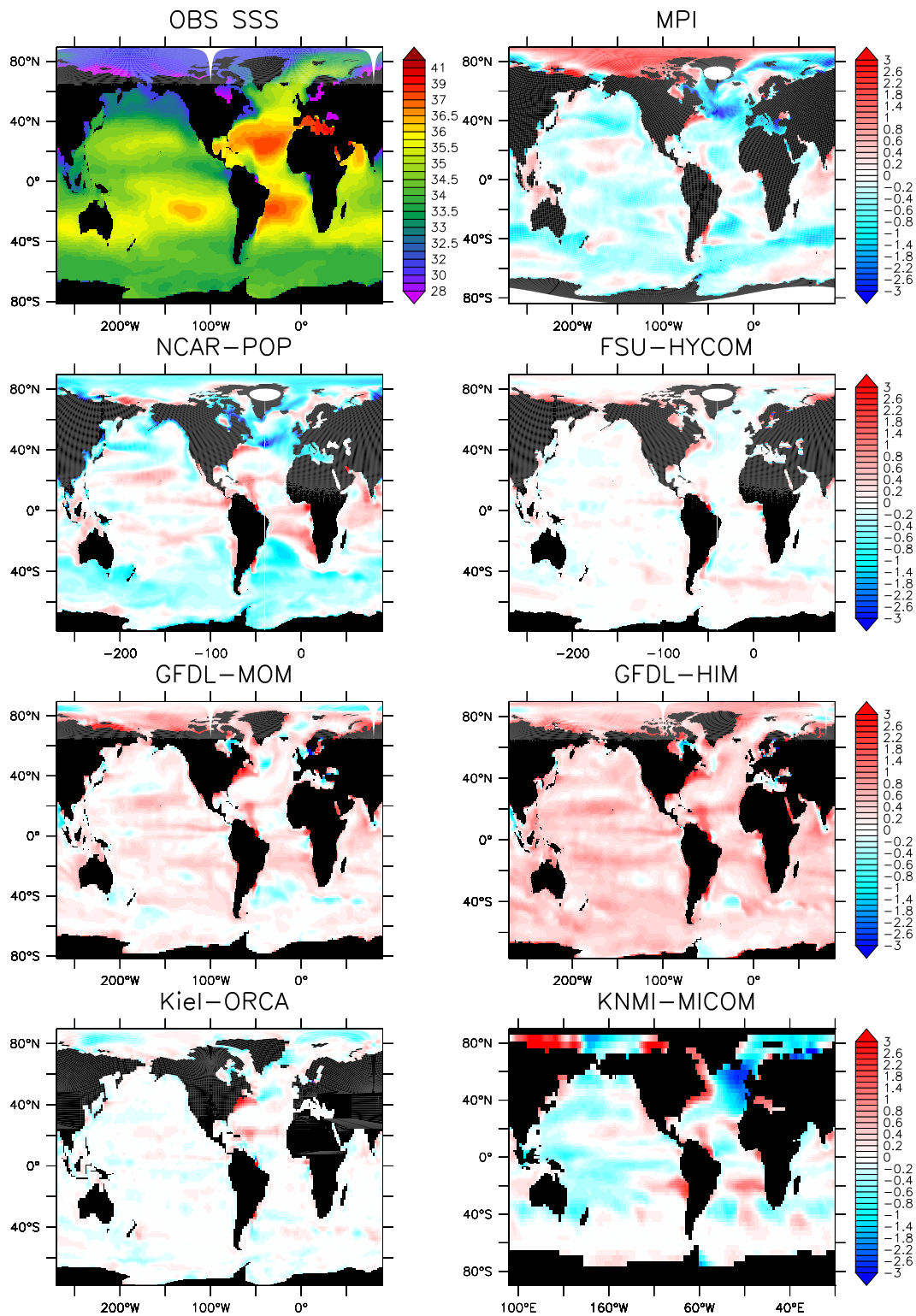


Fig. 8. Anomalous SSS (psu) for years 491-500 from the simulations relative to the analysis (shown in the top middle panel) of Conkright et al. (2002) outside the Arctic, and Steele et al. (2001) in the Arctic.

assessment regarding these factors is provided by comparing the simulations of heat content and SST changes with climatological data (Conkright et al. (2002); henceforth termed "Levitus") for a single site (Ocean Weather Ship (OWS) Echo at $35^{\circ}N$, $48^{\circ}W$) in the subtropical North Atlantic. In Figure 9, monthly values of these quantities are diagnosed, defining hysteresis loops characterizing the seasonal cycle for each model in relation to the data.

8.1. Descriptive comments

Temporal evolution of the hysteresis loop of seasonal heat storage at OWS Echo is exemplified for one experiment in Figure 9a. It confirms a near-surface equilibration period of roughly 100 years for the annual mean properties as suggested by the global mean time series in Figures 5 and 6. In contrast, the seasonal cycle of heat content and SST is equilibrated more rapidly, within the first decades of integration, with little change after year 50. Accordingly, due to the focus here on seasonal model behaviours for this diagnostic, depiction of the different model results in Figures 9b and 9c is provided for year 50. Given the possibility of remaining drifts of various degrees at this stage, the model-model differences in the annual mean properties have to be interpreted with some caution. It is interesting to note, however, that all models except GFDL-HIM have tended to drift towards a lower (on average, about 5%) heat content during this initial spin-up phase. The influence of a specific model choice is exemplified in Figure 9b, by contrasting the MPI-control run (MPI-A) using a weak SSS relaxation (piston velocity of 50 m/4 years) with MPI-B using stronger SSS restoring (50 m/300 days; see Table 3).

FSU-HYCOM, which shows a relatively large deviation from Levitus (Figure 9c), has not equilibrated yet at year 50. The heat content in this model presents both an oscillation and an increasing trend with time, differing from Kiel-ORCA shown in Figure 9a. In particular, the annual mean heat content in FSU-HYCOM increases from $4390^{\circ}C\text{-m}$ at year 50 to $4482^{\circ}C\text{-m}$ at year 400.

The first aspect to be noted concerning the simulation of the seasonal cycle is that the observed phase appears reasonably captured in all models, apart from slight differences in the month of the minimum heat content during the winter. However, there are some quantitative differences in the annual range of SST: in particular, with the exception of the MPI

cases, all models fail to reach the maximum summer temperatures, showing a cold bias during August-September of $1^{\circ}C$ in the z-models, and even more in the isopycnal models KNMI-MICOM and GFDL-HIM. The common weakness in the summer heat gain across a host of models using different mixed layer schemes, seems to suggest a problem with the surface heat flux, e.g., a bias in the solar radiation during this season.

8.2. Critical comments

The area encompassed by each model is a bit less than Conkright et al. (2002). The NCAR-POP and GFDL-MOM simulations use the same mixed layer scheme from Large et al. (1994) with closely analogous choices for the mixing options in the scheme. This agreement in mixed layer schemes may account for the agreement in model behaviour, in which both simulations show less heat content than seen in the observations. FSU-HYCOM also employs the mixed layer scheme from Large et al. (1994), but shows a shift in heat content. The reason for this shift is currently under investigation. Although the GFDL-HIM and KNMI-MICOM simulations are both isopycnal models, they exhibit a shift in heat content relative to one another, likely due to differences in the formulation of their respective bulk mixed layer schemes (see Section A4 for discussion of the mixed layer schemes in the two isopycnal models).

In general, the spread in model results may indicate that forcing may not be the only problem with the simulations.

9. Sea ice concentrations

High-latitude processes, including the distribution and strength of convective areas and the seasonal cycle of polar sea ice cover, are among the most challenging aspects of the climate system to accurately simulate. In particular, these aspects are very sensitive to the choice of atmospheric boundary conditions and model configurations, as emphasized by Proshutinsky et al. (2001) for the AOMIP project. A detailed examination of the parameter sensitivities encountered in the host of CORE experiments performed is beyond the scope of this paper. Hence, we restrict the following presentation to gross properties of the sea ice simulations.

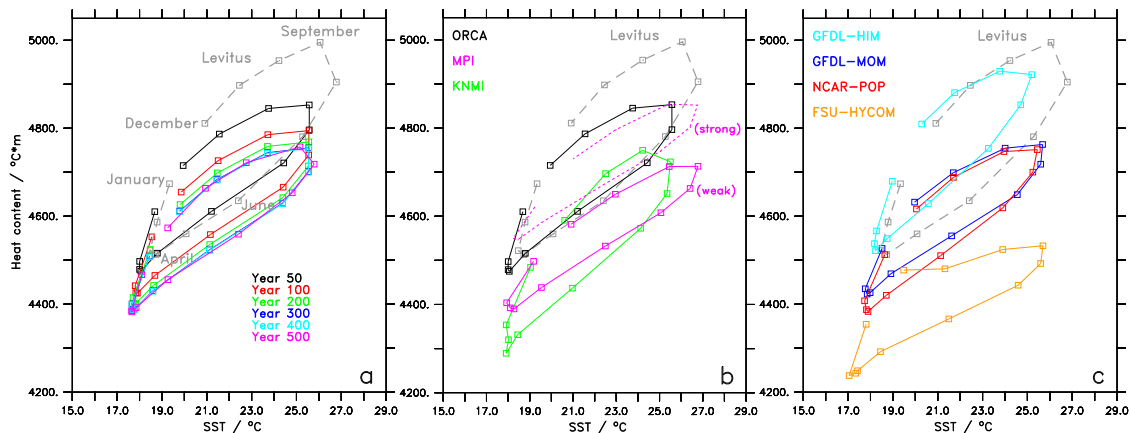


Fig. 9. Monthly values of the heat content (vertically integrated temperature) over the upper 250m versus sea surface temperature taken at Ocean Weathership Echo ($48^{\circ}W, 35^{\circ}N$). The Conkright et al. (2002) climatology (termed “Levitus” in this figure) is shown as the dashed curve in each panel. Panel (a) time evolution in Kiel-ORCA taken at selected years during the 500 year spin-up; Panels (b) and (c): comparison of various models at simulation year 50. Note that there are two contributions from MPI: the solid line is the control MPI-A simulation, which uses a piston velocity of 50m/4 years for salinity restoring; the dashed line is for the MPI-B, which uses the stronger piston velocity 50m/300 days.

Although both sea ice concentration (area of sea ice per grid cell area) and thickness are needed to determine the heat and water content of the ice, only the concentration is analysed, because more observations are available for comparison. Sea ice concentration reflects the effects of dynamic and thermodynamic processes, originating in the ocean-sea ice system and in the atmosphere. Thus, sea ice concentration is a good diagnostic of the models’ ability to reproduce near surface processes in high latitudes. Recently, sea ice concentration has been examined for AOMIP ocean-sea ice models by Johnson et al. (2007) and for coupled climate models by Zhang and Walsh (2006). Additional literature regarding the interannual and long-term variability of sea ice concentration and the quality of the observations can be found there. Recent model-model and model-observation comparisons for sea ice thickness and sea ice drift can be found in Gerdes and C. Köberle (2007) and Martin and Gerdes (2007), respectively.

Figure 10 shows the temporal evolution of the annual mean sea ice area in both hemispheres. The stability of the northern sea ice area, despite relatively large changes in the large scale oceanic flow (see for instance Figure 25), indicates that with prescribed atmospheric conditions, the oceanic heat transport into the Arctic has little impact on the sea ice properties. The pace of adjustment in integral sea ice measures is far less rapid in the Antarctic, with its huge seasonal cycle, than in the Arctic.

The KNMI-MICOM simulation has roughly half

the sea ice of other models in both hemispheres, and it displays a steady recovery trend in the Southern Hemisphere following a rapid adjustment to near zero ice following initialisation. For the other models, sea ice area and extent are more similar. Both the GFDL-HIM and MPI simulations show far more variability in Southern Hemisphere sea ice than the other models, with GFDL-HIM failing to reach a steady state.

Figure 11 maps the ice concentration in March (month of observed largest Northern Hemisphere sea ice coverage), and Figure 12 shows the same for September (month of observed largest Southern Hemisphere sea ice coverage). We compare simulated results to the satellite sea ice concentration climatology over years 1979-2004 compiled by Comiso (1999 (updated 2005)).

For the wintertime Northern Hemisphere (March), all models show a sizable coverage across the Arctic, with modest differences in the southern extent of the sea ice into the subpolar North Atlantic and North Pacific. Likewise, the wintertime Southern Hemisphere (September) simulations are reasonably close to the observations in the gross sense, with the exception of the KNMI-MICOM simulation, which shows a relatively small ice area. In general, the KNMI-MICOM simulation has the least amount of ice, nearly losing its summertime Arctic and summertime Antarctic ice areas. It is unclear what prompts the KNMI-MICOM simulation to produce such a small amount of sea ice in both hemispheres.

This result is consistent, however, with the broadly warm SSTs in the Southern Hemisphere (Figure 7), though the Northern Hemisphere SST shows a far smaller bias. The remaining models are able to maintain a nonzero southern summertime ice area, though generally a bit smaller than the observations, and with some variability amongst the simulations. Likewise, the summertime Northern Hemisphere simulations show a rather broad spread, with KNMI-MICOM and NCAR-POP showing less ice than observed, and MPI showing more sea ice than observations.

There are considerable differences in summertime Northern Hemisphere sea ice coverage from the observations, as determined from passive microwave sea ice concentration data (e.g., Johnson et al., 2007). Nonetheless, the pronounced summertime sea ice concentration differences among the models is significant. In general, all models show too little summertime sea ice in the Eurasian Arctic compared to the Comiso (1999 (updated 2005) data.

10. Tropical Pacific

Realistic ocean simulations in the Tropical Pacific are important for coupled ocean atmosphere simulations of El Niño Southern Oscillations (ENSO) (Latif et al., 1998). Integrity of the simulation is dependent especially on the wind stress and model’s ability to maintain a tight thermocline, with the latter dependent on vertical mixing (Meehl et al., 2001) as well as horizontal friction.

For comparison with observed hydrography and currents at the equator, we employ the isopycnal analysis of Johnson et al. (2002). It is based on measurements of zonal velocity, temperature, and salinity from 172 meridional sections taken mostly in the 1990s from $143^{\circ}E$ to $95^{\circ}W$ in the Tropical Pacific. This analysis preserves pycnocline structure across the equator where isopycnals and isotherms dome above the Equatorial Undercurrent (EUC) and slump below (Wyrтки and Kilonsky, 1984). This temperature structure results in a thermocline with tighter vertical gradient just off the equator than at the equator.

10.1. Descriptive comments

Time series for the zonal transport in the equatorial undercurrent (not shown) reveal that the spin-up for the equatorial current occurs within a

few decades. Temperature and zonal current remain qualitatively similar at year 50 to those towards the end of the simulation (not shown).

Figure 13 shows the annual mean equatorial upper ocean temperature from the observations of Johnson et al. (2002), as well as results from the model simulations averaged over years 491-500. The KNMI-MICOM simulation exhibits very cold water towards the bottom of the thermocline and very warm water in the upper ocean mixed layer in the west Pacific, with temperature range going beyond the scale of the other models. The FSU-HYCOM simulation shows far too deep and eastward penetration of the western warm pool, thus suppressing the eastern shoaling of the thermocline. The other models show very similar profiles for the thermocline that agree fairly well to the observations, with the exception of somewhat too uniform and deep western warm pool in the simulations.

We garner a complementary picture by focusing on the SST errors to the tropical Pacific. As noted in the global SST bias map of Figure 7, all models show too much warming off the west coast of South America, perhaps due to lack of resolution required for representation of coastal upwelling, and perhaps due to problems with the prescribed atmospheric state. Additionally, all models show overly cool waters in the central to eastern portion of the Pacific, with the MPI and KNMI-MICOM simulations especially cold. KNMI-MICOM is overly warm in the Maritime Continent regions of the west Pacific, thus creating a huge zonal SST difference relative to observations. The MPI and Kiel-ORCA simulations are also warm outside the cold tongue region, but less so than KNMI-MICOM. The GFDL-MOM, GFDL-HIM, and NCAR-POP simulations show similar error patterns with generally smaller amplitude than the remaining models.

Figure 14 shows the zonal current along the equator from Johnson et al. (2002). Its strength peaks just above 1 m s^{-1} . The previous class of global IPCC models had very sluggish currents, with speeds slower by factors of two or three. As shown in this figure, most of the models in the present study have simulated undercurrent speeds approaching, or exceeding, 1 m s^{-1} . The two exceptions are the MPI and FSU-HYCOM simulations, which show very weak undercurrents.

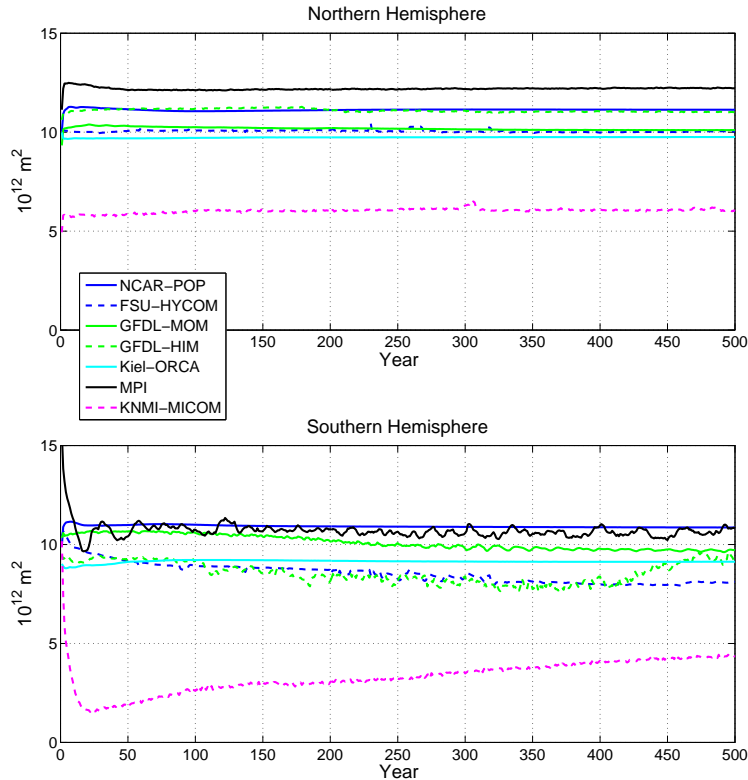


Fig. 10. Time series for the annual mean sea ice area (in units of 10^{12} m^2) for the Northern Hemisphere (top) and Southern Hemisphere (bottom). The hemispheric ice area depicts the area of a grid box multiplied by the ice concentration (area of ice cover per area of grid box), and summed over each hemisphere. The vertical axes are the same, thus facilitating a direct comparison between hemispheres.

10.2. Critical comments

The weak undercurrents in the MPI and FSU-HYCOM simulations may stem from differing model limitations. For MPI, the simulation is apparently handicapped by its coarse meridional resolution (coarser than 1.7° in the Tropical Pacific). While the core velocities are too low, the integrated transport of about 50Sv is in the range of observations. Moreover, initial results (not shown) from a new model version with 0.4° global resolution reveal maximum undercurrent velocities of well above 1 m s^{-1} .

For the FSU-HYCOM simulation, problems representing the tropical thermocline and mixed layer (Figure 13) likely contribute to the problems with the undercurrent. The source of these problems is not clear since (a) the equatorial undercurrent and associated thermocline are well represented during

the first 20 years of the simulation and (b) other long term simulations at comparable horizontal resolution do not exhibit this behavior (R. Bleck, personal communication). One should note that the magnitude of the viscosity coefficients chosen in FSU-HYCOM are quite large and may result in a weakening of the equatorial circulation. A solution to the equatorial problem in FSU-HYCOM remains under investigation.

All the models display too much near surface westward flow in the east. A likely contributor is the wind direction bias relative to QSCAT, which suggests that the wind stress should be more cross-equator, and hence less westward. Adjustments for this bias are made in version 2 of the CORE forcing (Large and Yeager, 2008).

The thermal stratification in the eastern Pacific is poor in all models (Figure 13), with too sharp, too shallow thermocline above 50m depth, and too thin

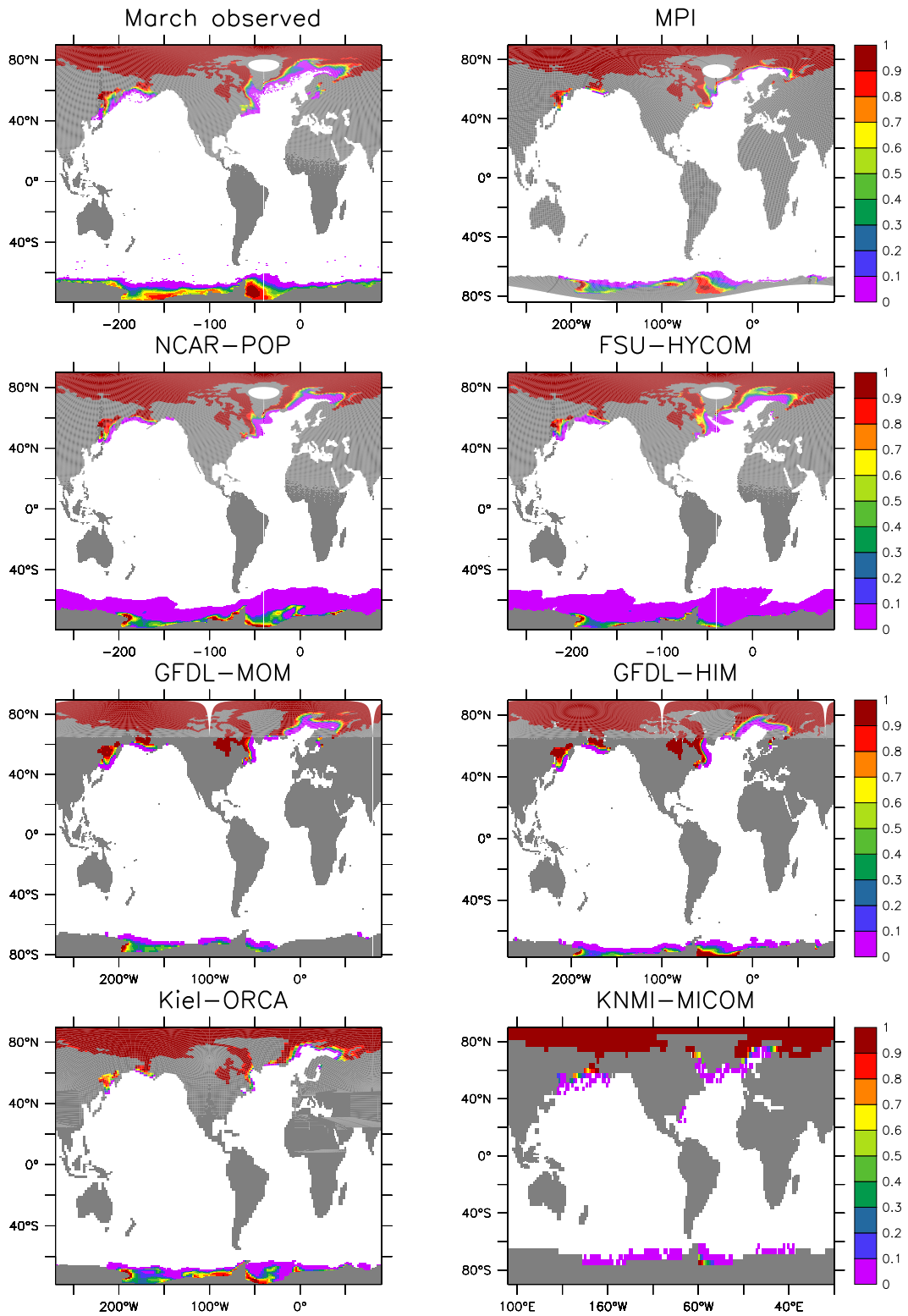


Fig. 11. Sea ice concentration (area of sea ice per grid cell area) in March (month of maximum observed Northern Hemisphere sea ice coverage), averaged over years 491-500 in the CORE-I simulations, as well as the observations taken from Comiso (1999 (updated 2005)).

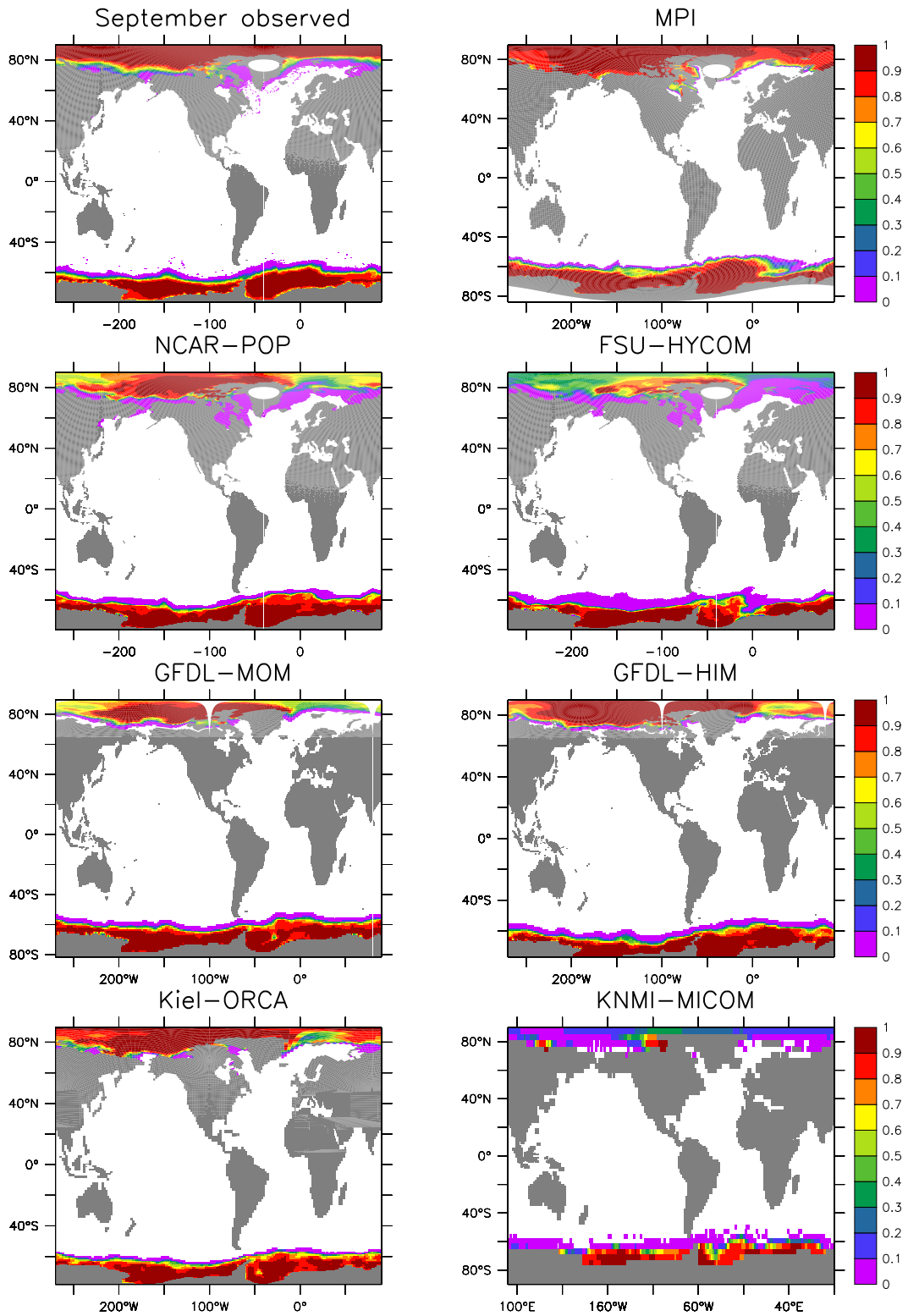


Fig. 12. Sea ice concentration (area sea ice per area grid cell) in September (month of the maximum observed Southern Hemisphere sea ice coverage), averaged over years 491-500 in the CORE-I simulations, as well as the observations taken from Comiso (1999 (updated 2005)).

layer between 12 and 14C, with the exception of the too thick layer in MPI. Problems with the eastern termination of the Equatorial Undercurrent are also evident in the velocity fields of Figure 14.

11. Mixed layer depths

The surface mixed layer is deepened by wind-driven mixing processes, Ekman-induced subduction, and convective overturn of gravitationally unstable water columns. It is the latter process that is particularly important in the formation of Subantarctic Mode Water (SAMW), North Atlantic Deep Water (NADW), and High Salinity Shelf Water around Antarctica. The mixed layer depth (MLD) attained in late winter is thus a crucial model diagnostic, as it reflects the depth of rapid overturn of surface waters, which is intimately related to water-mass formation and the regulation of global climate via the air-sea exchange of heat, freshwater, and CO₂.

In Section D7 of the Appendix, we detail issues involved with evaluating the mixed layer depth in an ocean model. In particular, we note that the models used in this study do not compute the mixed layer with the same algorithm. Hence, some of the differences can be explained by different methodology, though many differences are large enough to be robust.

11.1. Observed mixed layer depth

The observed estimate of MLD (Figure 15a) is derived from a long-term monthly mean climatology of $\theta - S$ (described below). This approach potentially underestimates the observed decade-long maximum MLD as it (1) ignores interannual variability in the convective mixed layer, (2) derives MLD from a somewhat horizontally smoothed $\theta - S$ field, and (3) can be based on under-sampled wintertime hydrography. However, alternate estimates, such as using raw Argo and WOCE data, do not provide a complete global coverage of mixed layer depth. Where data permitted, we compared the patterns of MLD obtained using raw and climatological $\theta - S$. In general, both yield similar patterns of MLD, only with shallower mixed layers in the climatologically-derived estimates in the far North Atlantic and Labrador Sea.

A critical MLD benchmark is the distribution of maximum depths of mixing attained during an an-

nual cycle. This depth influences the rate of uptake of anthropogenic CO₂ and the vigour of the ocean thermohaline circulation. It also marks the genesis of oceanic ventilation of thermocline (Stommel, 1979; Williams et al., 1995) and mode waters (McCartney, 1977). An estimate of the observed distribution of maximum MLD is shown in Figure 15a. This estimate is derived from the World Ocean Atlas 2005 (WOA05) data set (Locarnini et al., 2006; Antonov et al., 2006), based on monthly mean density. The WOA05 dataset includes Argo data recorded up until the end of 2005. The mixed layer depth h is calculated as the depth where the buoyancy difference ΔB between the surface and h is 0.0003m s^{-2} . Here, $\Delta B = g(1 - \sigma_0/\sigma_h)$, where σ is potential density. Data coverage is too sparse to derive the maximum MLD from anything shorter than a monthly coverage of global density, whereas in model simulations the MLD can be diagnosed at each model timestep. The potential errors arising from deriving the maximum MLD from a monthly climatology of $\theta - S$ were itemized above. Generally speaking, the fields shown in Figure 15a should be viewed as a large-scale estimate of the horizontal variations in global mixed layer depth, with highest values obtained in the formation regions for SAMW, NADW, and High Salinity Shelf Water around Antarctica.

11.2. Descriptive comments

In the observed analysis, deep winter mixed layers dominate the subpolar North Atlantic Ocean associated with NADW overturning. There is also a band of deep mixed layers or ‘pycnostads’ associated with SAMW, extending along much of the northern flank of the Antarctica Circumpolar Current (ACC) in the Indian and Pacific Oceans. Adjacent to Antarctica, particularly in the Weddell Sea, there is also evidence of deep mixed layers associated with the formation of High Salinity Shelf Water and, eventually, Antarctic Bottom Water. Only modest mixed layer depths are observed over much of the remainder of the global ocean (of the order of 50-100 metres, consistent with typical open ocean wind-driven mixed layer depth).

All models, except KNMI-MICOM, simulate a band of deep mixed layers associated with Subantarctic Mode Water (SAMW), although in GFDL-HIM these mixed layers are only seen in the Indian Ocean and south of Australia. In FSU-HYCOM and MPI, the SAMW pycnostads are

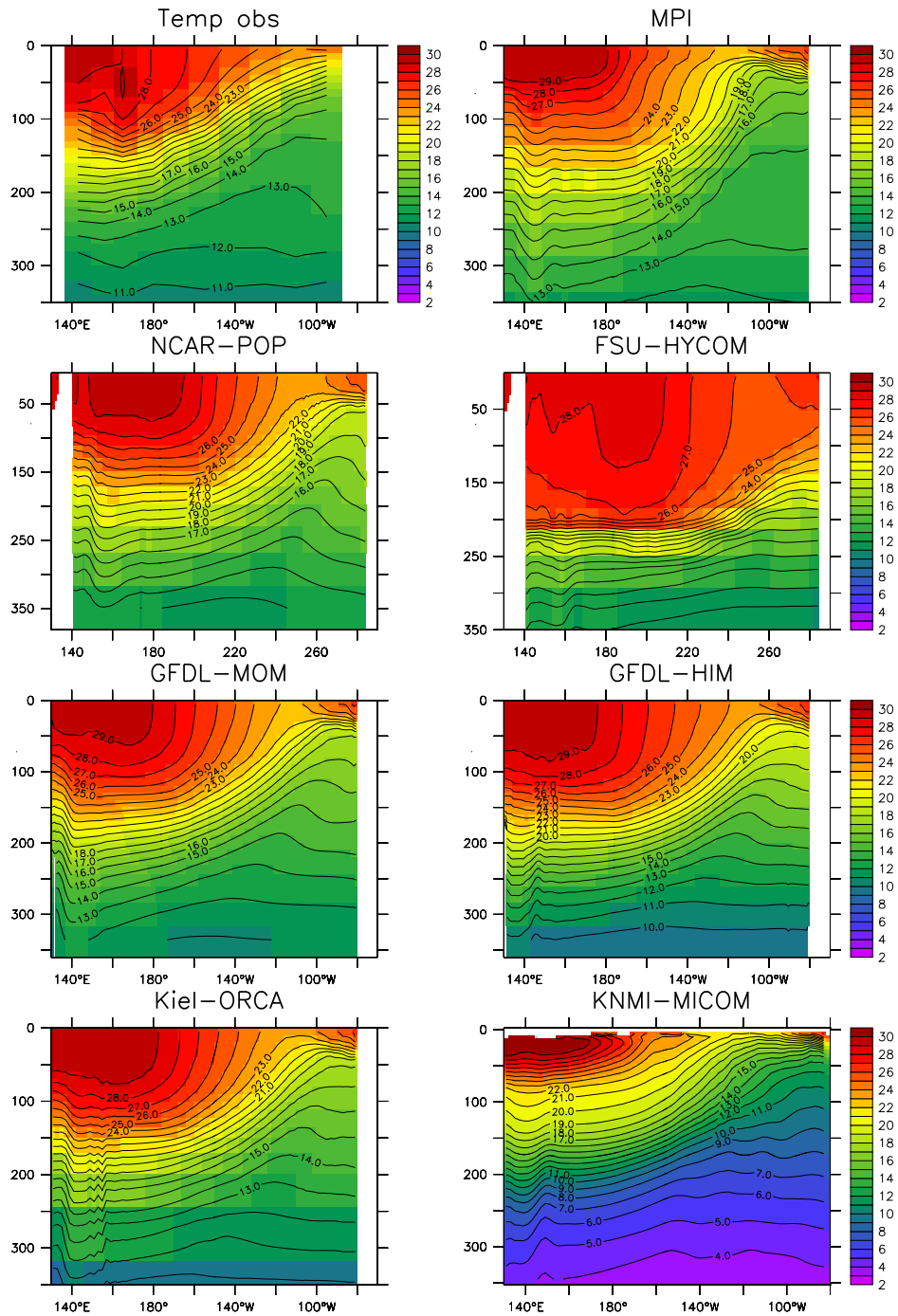


Fig. 13. The upper ocean temperature on the equator in the Pacific, with model results from a time mean over years 491-500 of the simulation. Top panel: Observed temperature from Johnson et al. (2002). Remaining panels are from the simulations. Note that the choppy KNMI-MICOM field near the surface arises from the remapping from density to depth space. Additional limitations of the submitted dataset preclude contouring this image.

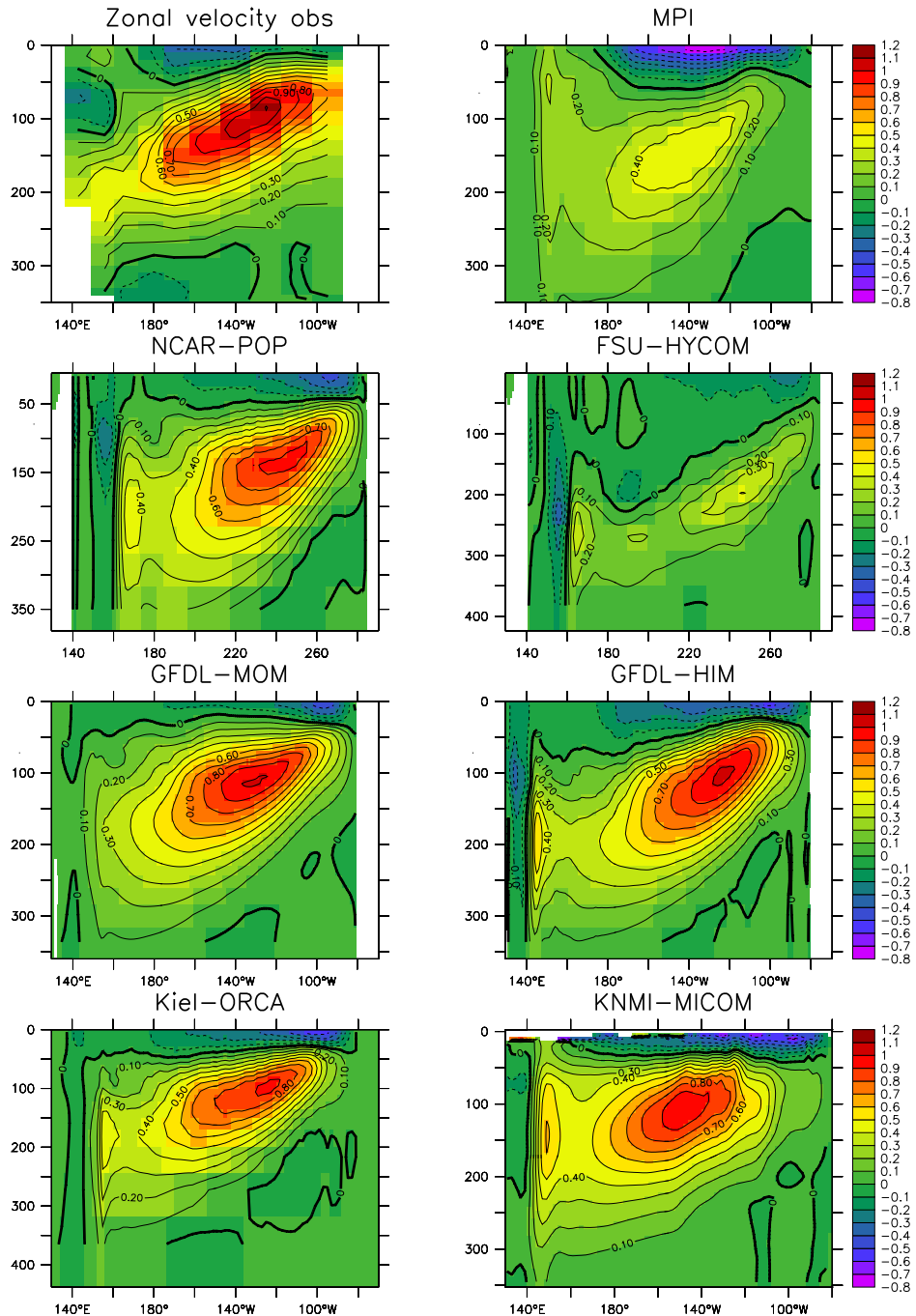


Fig. 14. The upper ocean zonal velocity component on the equator in the Pacific. The model results are time mean over years 491-500 of the simulations. Top panel: Observed zonal velocity component from Johnson et al. (2002).

generally too broad and deep. To the south, many of the models simulate very deep mixed layers in the Weddell gyre, in contrast to observations where MLD is generally shallow within the gyre, and deep at its southern edges. Deep mixed layers are limited to the Antarctic margin in the Ross Sea in the observed and most of the models, apart from the two GFDL simulations and MPI, wherein deep mixed layers are simulated within the polar gyres. In the North Atlantic all models capture a reasonably good mixed layer field, apart from KNMI-MICOM where the MLD seems chronically too shallow over all regions of water-mass formation. Overall, the NCAR-POP and Kiel-ORCA simulations yield the best general agreement with the observed estimate of maximum MLD.

11.3. *Critical comments*

The model-model comparison for MLD is complicated by the fact that each simulation uses a slightly different diagnostic method for calculating the depth of the mixed layer (see Section D7 in the Appendix). Ideally in a comparison study such as this, a method for diagnosing the surface mixed layer depth is agreed prior to experimentation, so that a meaningful model-model analysis can be made.

The two isopycnal models GFDL-HIM and KNMI-MICOM report the thickness of the layer with active surface wind and buoyancy induced turbulence, with markedly different fields for the two simulations. The KNMI-MICOM underestimates the MLD in all key areas of water-mass formation, leaving its ocean interior poorly ventilated from surface waters, and therefore too cold and fresh (Section 12). In contrast, GFDL-HIM has relatively deep Southern Ocean and North Atlantic mixed layer depths, creating a much more ventilated simulation than KNMI-MICOM.

Among the z-level models, significant differences in simulated MLD are also noted. The GFDL-MOM and NCAR-POP simulations employ the same mixed layer scheme from Large et al. (1994). Hence, there is likely another reason explaining their differences, with treatment of the neutral physics scheme (Gent and McWilliams, 1990) perhaps the main reason. As documented in Table 6 (Section D4), GFDL-MOM chooses a spatially dependent diffusivity which is a maximum of around $600 \text{ m}^2 \text{ s}^{-1}$ in the centre of the ACC, and reduces to nearly $100 - 200 \text{ m}^2 \text{ s}^{-1}$ outside the main region

of strong baroclinicity (Figure 14 of Griffies et al., 2005). NCAR-POP, in contrast, uses the larger value of $600 \text{ m}^2 \text{ s}^{-1}$ globally. Additionally, GFDL-MOM chooses to begin tapering the neutral fluxes with a neutral slope of $1/500$ (Gnanadesikan et al., 2007), whereas NCAR-POP extends their fluxes untapered until a slope of $3/10$ (Danabasoglu et al., 2007). Both the diffusivity and slope maximum differences provide a stronger tendency from Gent and McWilliams (1990) in NCAR-POP to relax the neutral slopes, thus shoaling the mixed layers, relative to GFDL-MOM. Recent tests with the GFDL-MOM coupled climate model indicate that the mixed layer in the Weddell Sea is significantly reduced by allowing for a larger maximum slope.

12. Zonal average potential temperature and salinity

One of the most widely assessed benchmarks of ocean model performance is the distribution of global ocean potential temperature and salinity ($\theta - S$) in the latitude and depth plane. There are several reasons that this diagnostic is popular. Foremost, the way heat and salt are distributed with latitude and depth is directly set by the global thermohaline and wind-driven circulation, reflecting the rate and properties of large-scale water mass ventilation. Water mass overturning rates are themselves intimately tied to important climate related quantities such as oceanic carbon uptake and the poleward transport of heat and freshwater. Whereas poleward property transports and the processes directly tied to water mass formation (e.g., convective overturning, mixing, and downslope gravity currents) are extremely difficult to measure directly, the global zonal mean climatology of oceanic $\theta - S$ is well constrained by measurements. This situation arises from the decades of hydrographic surveys, including those associated with the World Ocean Circulation Experiment (WOCE), as well as the slow overturning rates in the intermediate, mid depth and abyssal oceans. Modellers thus tend to rely on the traditional hydrographic parameters ($\theta - S$) to provide a means for assessment of the water-mass formation processes operating in global models. England and Maier-Reimer (2001) show that CFCs and radiocarbon can equally be used in this context.

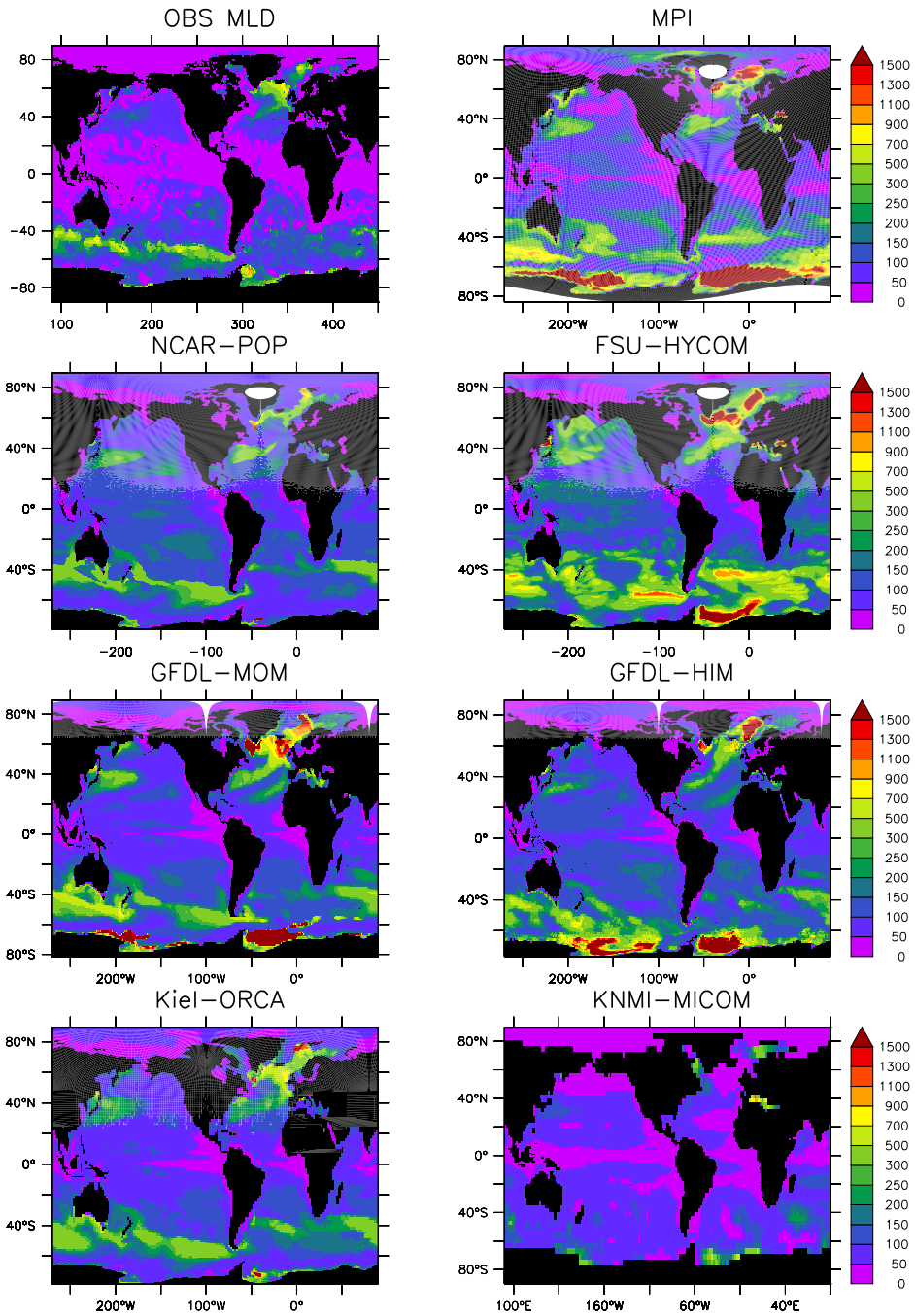


Fig. 15. This figure maps the maximum mixed layer depths during simulation years 491-500, as well as an approximation to that found in Nature. That is, for the simulations, at each horizontal grid point, the maximum mixed layer depth is found during years 491-500 for either the instantaneous mixed layer depth (NCAR-POP) or the monthly averaged mixed layer depth (remaining models). Definition of the mixed layer for the models and observations differ somewhat, with details provided in the text.

12.1. Descriptive comments

A comparison between the global zonal-mean anomalous θ and S in the CORE simulations is presented in Figures 16 and 17, respectively. The model minus observed fields are derived taking an average in the simulations during years 491-500. While 500 years is much shorter than the equilibration time of the mid depth oceans (e.g., England, 1995; Stouffer, 2004), it is sufficient to reveal significant model drift away from the observed mean $\theta - S$.

The NCAR-POP, GFDL-MOM, and MPI simulations are generally too warm and saline at 500-1000 metres depth, and contain warm deep anomalies in the North Atlantic. GFDL-HIM and FSU-HYCOM show notable warm anomalies in the subsurface tropical oceans. Below 1000m, GFDL-HIM and KNMI-MICOM exhibit warm anomalies in the deep water formation regions around 70°N, but significant cooling elsewhere. The Kiel-ORCA simulation shows weaker anomalous temperature patterns, whereas the KNMI-MICOM simulation is generally cold throughout the ocean interior, and very cold in the upper tropical and mid-latitude ocean, suggestive of overly vigorous upwelling into these regions. The zonal biases in salinity for NCAR-POP, GFDL-MOM, GFDL-HIM, and MPI show generally too much salt in the upper ocean and penetrating into the deep North Atlantic, and fresh anomalies below about 3000m depth. Notable errors in model θ , S are apparent in most models in the AAIW region. Overall, the Kiel-ORCA simulation shows the weakest bias in the interior $\theta - S$.

12.2. Critical comments

Errors in model simulations of global ocean $\theta - S$ result from at least one of the following problems:

- erroneous surface $\theta - S$,
- spurious rates of ocean overturning within the surface mixed layer,
- incorrect interior ocean circulation and tracer advection,
- unrealistic simulated interior mixing processes.

Errors in surface $\theta - S$ may be a result of incorrect air-sea heat and freshwater fluxes, errors in surface circulation and mixing, or a combination. Thus, the diagnosis of simulated subsurface $\theta - S$ against observations can be ambiguous: errors may be symptomatic of any number of problems in ocean model forcing, circulation and/or subgrid-scale physics.

Additionally, differences between models may also arise from the varying forms of surface salinity restoring. The results of this analysis suggest that coupled ocean-ice models should agree with the observed zonal average $\theta - S$ to within $\pm 1^\circ\text{C}$ and $\pm 2\text{psu}$ over the majority of the latitude-depth plane, with differences outside the ranges $\pm 2^\circ\text{C}$ and $\pm 4\text{psu}$ confined to only a few isolated regions.

The erroneous heat and salt content localized near 60-70°N in some of the simulations suggest overly rapid surface overturn in this region. An analysis of the GFDL-MOM simulated mixed layer depth confirms this (Figure 15), with a broad region of deep overturn just north of 60°N. The deep wintertime mixed layers in the North Atlantic may be related to erroneous surface $\theta - S$ and/or to interior mixing errors in regions of steeply sloped isopycnals. In contrast, the spuriously fresh salinities in northern regions of the Kiel-ORCA and KNMI-MICOM simulations are symptomatic of their relatively weak meridional overturning circulation (MOC) (Section 15) and shallow North Atlantic mixed layers (Figure 15).

Apart from differences in model versus observed $\theta - S$ originating in the North Atlantic, Figures 16 and 17 also reveal large-scale discrepancies in Southern Ocean water mass properties. All but the NCAR-POP and Kiel-ORCA simulations exhibit Antarctic Bottom Water (AABW) signatures that are too cold, and all appear too fresh, with NCAR-POP showing the least bias. Antarctic Intermediate Water (AAIW) is for the most part too warm and saline, with Kiel-ORCA and KNMI-MICOM showing somewhat smaller salt anomalies than the other simulations. These biases exist despite the fact that many of the models appear to capture reasonable rates for AABW overturn ($\sim 10\text{ Sv}$; Figure 24), and none exhibit excessive mixed layer depths near regions of Subantarctic Mode Water formation (Figure 15), a classical reason for warm saline AAIW in models (Danabasoglu et al., 1994; England, 1995). Danabasoglu and Marshall (2007) provide an updated discussion of sensitivity of AABW to mesoscale eddy parameterisations.

A possible cause for the cold and fresh AABW in the simulations is that the crude corrections of the NCEP surface air temperature surrounding Antarctica in the Large and Yeager (2004) dataset (see their Section 5.3) are insufficient. Another cause may be the inability of the models to simulate narrow bands of convection over the Antarctic continental shelf (see Figure 15). Some of the models (GFDL-MOM,

GFDL-HIM, and FSU-HYCOM) instead simulate large-scale regions of deep mixed layers in the central Weddell Gyre, in contrast to the observations (Figure 15a). This region is spuriously ice free during late summer, meaning fresher sea surface salinities than observed (Figure 8), and greater air sea heat loss on the annual mean. This thermohaline forcing causes deep open ocean convection in the region, spiking AABW with erroneously cold and fresh waters. In addition, for the Kiel-ORCA and KNMI-MICOM simulations, their weak NADW leaves the mid-depth Southern Ocean, and in particular Circumpolar Deep Water, too cold and fresh. The warm saline AAIW in contrast is possibly a result of incorrect subduction and/or isopycnal mixing rates in the region. However, given that this bias features in nearly all of the simulations, it suggests there to be biases in the surface forcing. We refer the reader to Russell et al. (2006) who provide a framework for understanding Southern Ocean simulations in fully coupled climate models.

13. Volume transport through Drake Passage

Vertically integrated volume transport of seawater through selected regions of the ocean provide modellers with an important benchmark to evaluate the integrity of simulated water masses and ocean currents, as well as the boundary forcing impacting the transport.³ In this section, we display results for the Drake Passage transport, which measures the zonal flow through the smallest latitudinal extent of the Antarctic Circumpolar Current (ACC). Drake Passage transport has been measured using various methods, with a low value around 100Sv from Orsi et al. (1995) and high value of 135Sv from Cunningham et al. (2003). T. Whitworth (1983) and Whitworth and Peterson (1985) give 134 ± 13 Sv, with at least some of the models respecting this value.

The Drake Passage transport is a strong function of the winds, baroclinicity across the ACC, and buoyancy forcing (e.g., Hallberg and Gnanadesikan,

³ Note that the MICOM and HYCOM codes used by KNMI-MICOM and FSU-HYCOM are based on the non-Boussinesq dynamics, in which mass transport is appropriate rather than volume transport. All other models are Boussinesq. To facilitate comparison between simulations, we chose to quote volume transport. Differences in transport resulting from considering the distinction between mass and volume are inconsequential in view of the magnitude of the inter-model differences.

2006, and references therein). It is thus very difficult to determine the causes of model inaccuracies in simulating this transport. Nonetheless, comparing the simulated transports between the models provides a useful diagnostic for this critical part of the ocean climate system (Russell et al., 2006).

13.1. Descriptive comments

Figure 18 shows the time series for annual mean Drake Passage transport from the CORE simulations. The transport in the KNMI-MICOM simulation rapidly dives toward a very small value, reaching about 25Sv after 500 years. The GFDL-HIM simulation also has a decreasing trend, though reaches a steady state of roughly 75Sv after 500 years. Each of the other models reach stronger steady states somewhat sooner. NCAR-POP, FSU-HYCOM, GFDL-MOM, and Kiel-ORCA have values between 135Sv-150Sv, which is on the high side of measurements. FSU-HYCOM is still showing a trend after 500 years. The MPI transport simulated by the experiment with weak restoring is relatively high and exhibits variability around a mean of roughly 180Sv. The corresponding transport from the strong-restoring experiment is 135Sv and quite stable. This comparison and results from GFDL-MOM and Kiel-ORCA using weak salinity restoring with piston velocity of 50m/4years are presented in Sections 16.1 and 16.2, and these two both show nonstationary behaviour. So it is only the NCAR-POP simulation that exhibits relatively stationary behaviour for the Drake Passage transport with the weak 50m/4years piston velocity.

13.2. Critical comments

Gent et al. (2001) discuss factors that set the mean Drake Passage volume transport in numerical ocean models, including those coupled to a sea-ice model. Russell et al. (2006) present an analysis of the ACC transport in coupled climate models. In general, there is a strong dependency (order 100Sv) on model parameters and topography. Together with a dependency on the thermohaline circulation off the Antarctic shelf, and hence on the sea-ice model, these are the probable contributors to the large spread in Figure 18. More consistent model results would be expected if these factors were made more similar across models (e.g., use the same topography and sea ice model). There is also a strong de-

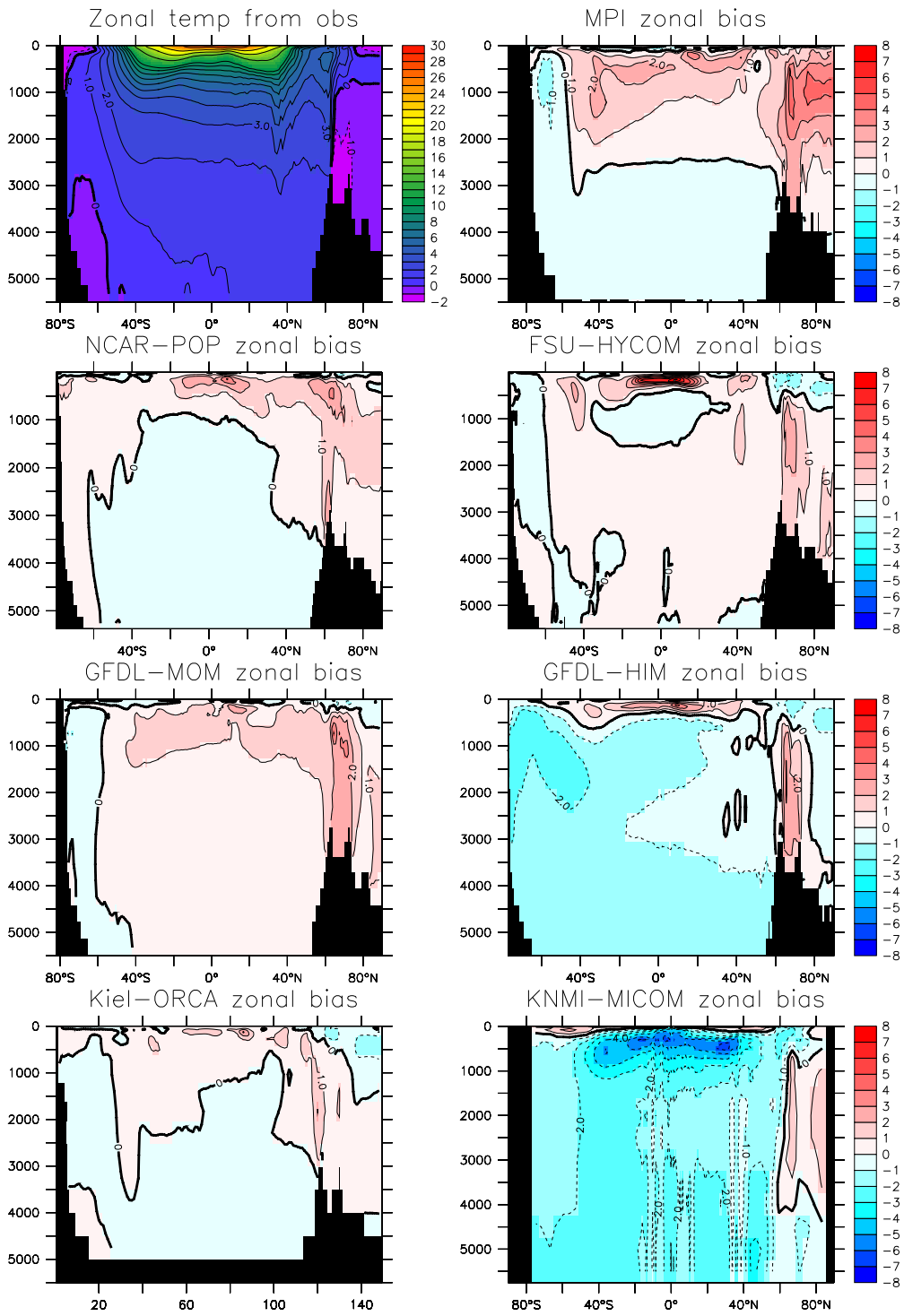


Fig. 16. Anomalous zonal mean decadal mean (years 491-500) potential temperature for the simulations relative to the analysis (shown in the top panel) of Conkright et al. (2002) for all but the Arctic, and Steele et al. (2001) for the Arctic.

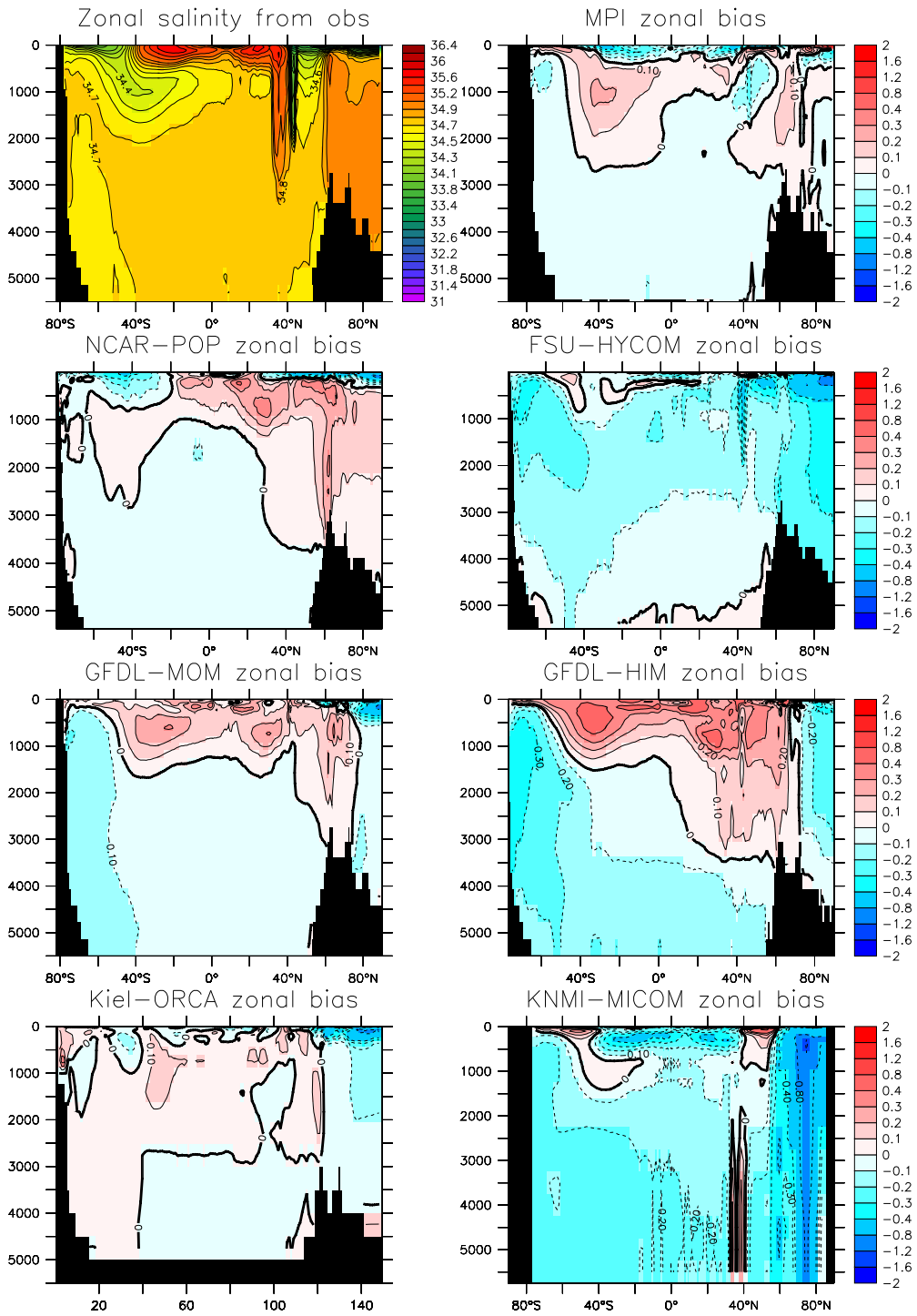


Fig. 17. Anomalous zonal mean decadal mean (years 491-500) salinity for the simulations relative to the analysis (shown in the top panel) of Conkright et al. (2002) for all but the Arctic, and Steele et al. (2001) for the Arctic.

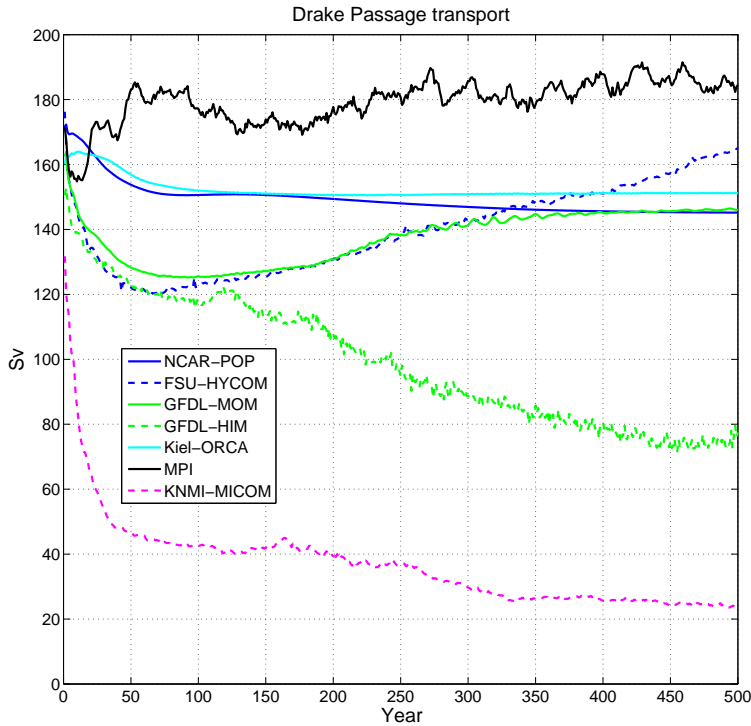


Fig. 18. Time series for the annual mean vertically integrated transport of seawater volume through the Drake Passage in the CORE-I simulations. Estimates of this transport from observations include $134 \pm 13\text{Sv}$ from T. Whitworth (1983), Whitworth and Peterson (1985), as well as more recent analysis from Cunningham et al. (2003).

pendency of the transport on the wind driven meridional Ekman transport, but with common wind forcing this aspect should be similar in all models.

14. Poleward heat transport

Poleward heat transport by the climate system is a response in the atmosphere and ocean to differential solar heating, with more warming in the tropics than the poles (Peixoto and Oort, 1992). In contrast to the atmosphere, ocean heat transport is greatly modified by meridional land-sea boundaries. It is critical for climate models to have a partitioning of poleward transport that is well distributed between the atmosphere and ocean in order for the model to maintain a relatively stable climate state under steady solar forcing (Weaver and Hughes, 1996; Gordon et al., 2000).

In Section D3 of the Appendix we detail the fundamental issues involved with computing the implied heat transport, based on surface ocean heat fluxes.

In this section, we discuss methods for determining implied heat transports from estimates of the surface turbulent and radiative heat fluxes. There are many uncertainties in this calculation. We then compare these results to the direct measurements of ocean heat transport from the CORE simulations.

14.1. Heat transports from observations and reanalysis

Determining which atmospheric product to use for computing the implied heat transport remains a subject of some discussion in the literature (e.g., Taylor, 2000). We present two results. The first is taken from Trenberth and Caron (2001), who estimate heat fluxes over a 50 month period starting in February 1985. The heat fluxes are computed as a residual between two much larger terms: top of atmosphere radiation and estimates of atmospheric heat transport divergence based on reanalysis data. Implied ocean heat transports are derived only after

forcing a global balance by assuming transports at $65^\circ N$ and $68^\circ S$, and adding excess heating to one or more of the Southern Ocean basins. The result from NCEP reanalysis, presented in Figure 19, is typically $0.5PW$ greater than derived from ERA-15. Hence, this implied transport is not a target model simulations should necessarily aim to reproduce closely.

The second implied heat transport is computed by running the normal year forcing (NYF) from Large and Yeager (2004) over the SST provided by Hurrell et al. (2006).⁴ Although ice-ocean heat fluxes are neglected, we provide further discussion of this second calculation using the NYF, because there are ambiguities, that do not arise when determining an implied heat transport from the interannually varying “observations.”

Notably, when running the NYF atmospheric state over the Hurrell et al. (2006) SSTs in years 1958-2000, the global ocean feels a net annual mean surface turbulent plus radiative heat flux shown in Figure 20. There is a very large range in the interannual global mean heat imbalance, although the 43-year mean imbalance is only $-0.55W m^{-2}$. This range in net heat flux is much larger for the normal year state variables than for interannually-varying state variables (see Large and Yeager (2004), Figure 31). Herein lies a major source of difficulty in computing an implied heat transport for the NYF. It arises because the natural correlations between SST and the overlying atmospheric state are lost when the NYF is put above the SST from any particular year.

The differences in net surface heat flux year to year are concentrated in tropical, western boundary current, and ice edge regions. It is in these regions where the loss of correlation between SST and the NYF atmospheric state are most serious. To generate implied heat transports, the global imbalances are subtracted uniformly over the ocean. This procedure presents a large (erroneous) redistribution of heat flux for years where the imbalance is large. Also, it is ambiguous in that confining the redistribution to the Southern Hemisphere, as in Trenberth and Caron (2001), would significantly change the result. Thus, it is unclear what annual cycle of SST is appropriate for pairing with the Normal Year for computing expected fluxes and the heat transport

⁴ These SSTs are based on a merge of an SST analyses of satellite and *in situ* data (Reynolds et al., 2002) starting in 1982, with historical SSTs reconstructed from ship observations beginning in 1871 (Rayner et al., 2003).

diagnostic. Perhaps the answer is an SST which generates a small (or slightly positive) global annual heating, which we assume to be the case in Nature for the last half of the 20th century due to anthropogenic global warming. A further criteria would be an implied transport distribution similar to the interannually varying forcing, because this would suggest a reasonable correlation between SST and the NYF atmosphere.

We present four variations on the net implied heat flux computed from the Large and Yeager (2004) NYF over the 43 years of Hurrell et al. (2006) SST. First, at each latitude a 43-year mean for the heat transport is computed. Although no single year realizes this mean implied heat transport, it provides a guide to the implied transport that may be expected from the ocean-ice simulations. Figure 21 shows this result. Then, instead of using the full 43 years, we segregate the 43 years according to the level of global annual heat imbalance:

- (a) years with imbalance less than $-1W m^{-2}$ (20 years);
- (b) years with imbalance greater than $1W m^{-2}$ (15 years);
- (c) years with imbalance between $-1W m^{-2}$ and $1W m^{-2}$ (8 years).

The resulting plots in Figure 21 show that case (c) agrees well with the interannually varying forcing result, case (b) has the best agreement with the Trenberth and Caron (2001) results in Figure 19, and case (a) generates the least agreement with both.

This analysis illustrates the extreme level of ambiguity associated with comparing CORE NYF implied heat transport with observations. Fortunately, model to model comparison is straightforward. Without doubt, reasonable alternative assumptions can lead to different results in what to expect from a model forced with NYF. In general, it is difficult to partition model and forcing error, but this is more of a problem with NYF, than when the forcing varies interannually.

14.2. Model results

Figure 22 shows the northward global heat transport computed from the ocean model fields averaged over simulation years 491-500. Note that MPI and KNMI-MICOM report the implied transports computed from the time integrated net radiative and turbulent heat flux. All other models report the heat transport computed from the ocean.

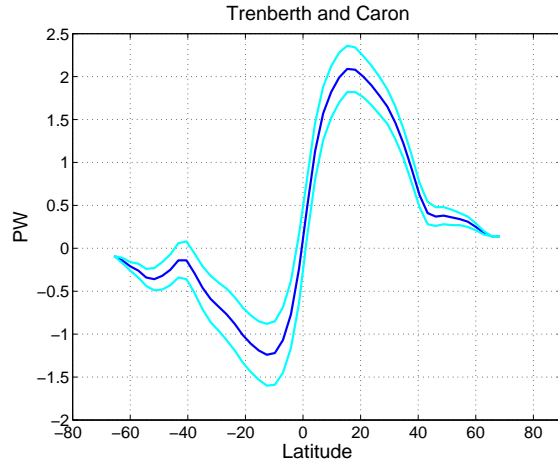


Fig. 19. Estimated implied meridional heat transport for the global ocean from Trenberth and Caron (2001). Their approach estimates heat fluxes from a combination of reanalysis and data products. Units are $\text{PW}=10^{15}$ Watts. The middle curve is their estimate, with the bounding curves providing the error bars.

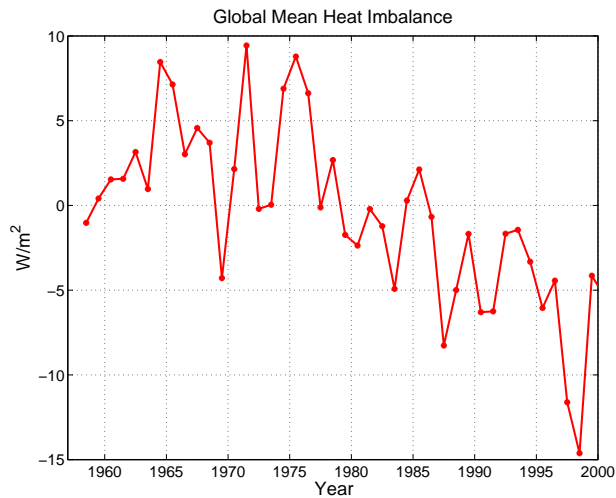


Fig. 20. Time series for the net radiative plus turbulent fluxes of heat into the ocean arising from the NYF atmosphere from Large and Yeager (2004) and the 43 years of Hurrell et al. (2006) SSTs. The nonzero net fluxes for each year represent a nontrivial imbalance of ocean heating. The time average for these fluxes over the 43 years is -0.55 W m^{-2} .

MPI shows the greatest transport in the subtropical regions of both hemispheres, with values in the north nearing 2PW and 1.5PW in the south. All models except KNMI-MICOM are generally in agreement with one another, each showing roughly 1.5PW maximum transport in the Northern Hemisphere, with GFDL-HIM simulation showing the most poleward transport in the Southern Hemisphere. The Kiel-ORCA simulation is somewhat weaker in the north, though with a sizable poleward transport in the south. Finally, the KNMI-MICOM

simulation is an outlier, with less than 1PW in the north and nearly 3PW poleward transport in the south. The reason for this large southward transport is unknown, but it is notable that this model has a huge drift in water mass properties (Sections 5 and 6), which likely affects the model's ability to produce a realistic heat transport.

The NCAR-POP, GFDL-MOM, and GFDL-HIM transport most closely resembles the NYF implied transport computed with years seeing an imbalance between -1 W m^{-2} and 1 W m^{-2} (see Figure 21).

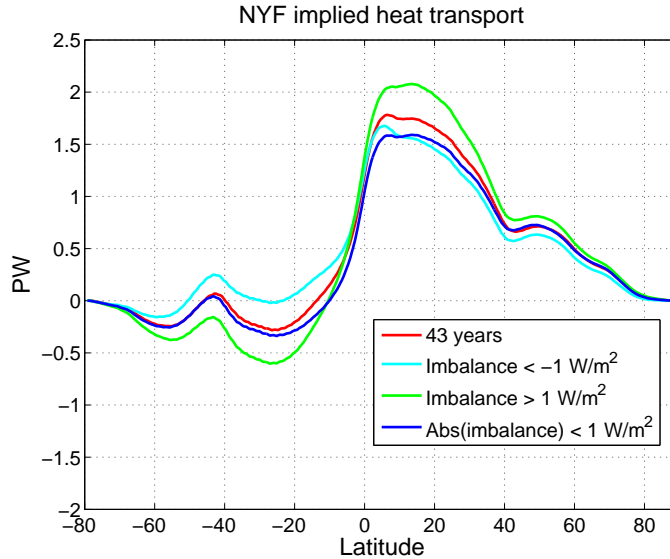


Fig. 21. Four examples of implied poleward heat transport available from the normal year forcing (NYF) from Large and Yeager (2004) when run over the SST provided by Hurrell et al. (2006). The four curves represent heat transport from the full 43 years of SST, as well as the following segregation: (a) years with imbalance less than -1 W m^{-2} (20 years); (b) years with imbalance greater than 1 W m^{-2} (15 years); (c) years with imbalance between -1 W m^{-2} and 1 W m^{-2} (8 years).

Tropical heating, as measured by the difference in northward heat transport between 20°N and 20°S , ranges from about 2.3 to 3.5 PW over the models. This range is bracketed by the values of 2.0 and 3.5 PW given, respectively, from the CORE interannually varying forcing of Large and Yeager (2004) and from Trenberth and Caron (2001). The fluxes from both imply that about 1 PW more of this tropical heating is lost to the north (1.7 to 2.2 PW) than to the south (0.3 to 1.3 PW). In this regard the models are inconsistent. Only NCAR-POP and GFDL-MOM are in good agreement, while GFDL-HIM and MPI only lose about 0.3 PW more to the north. More heat is lost to the south by Kiel-ORCA, FSU-HYCOM and KNMI-MICOM. All curves of Figure 22, except FSU-HYCOM and KNMI-MICOM, indicate some heating in the vicinity of 50°S . Except for GFDL-HIM, this heating is within 0.3 PW of balancing all the heat lost farther south.

15. Meridional overturning streamfunction

The meridional overturning streamfunction diagnoses the transport of volume or mass, and it is commonly used to summarise various features of the large scale circulation, particularly the effects from thermohaline forcing. In Section D2 in the Appendix, we detail methods for how this streamfunc-

tion is computed in the models. Here, we present results from the CORE-I simulations.

15.1. Descriptive comments

The thermohaline circulation is most active in the North Atlantic and the Southern Ocean. We thus find it convenient to examine the former in the Atlantic MOC streamfunction (Figure 23) and the latter in the global streamfunction (Figure 24), where each is time averaged over years 491-500 of the simulations. Additionally, Figure 25 shows the time series for the maximum of the North Atlantic streamfunction at 45°N beneath the wind driven Ekman layer. This time series exhibits the multi-centennial time scales for the spin-up of this circulation.

The structure of the North Atlantic streamfunctions (Figure 23) largely reflects the time series in Figure 25, with the GFDL-MOM simulation showing the most vigorous overturning streamfunction, and KNMI-MICOM the weakest. Note that the FSU-HYCOM figure was obtained by interpolating from its native hybrid pressure-density-terrain following vertical coordinates to depth, whereas GFDL-HIM and KNMI-MICOM use their native potential density as the vertical coordinate. The other models plot the streamfunction using their native depth coordinates.

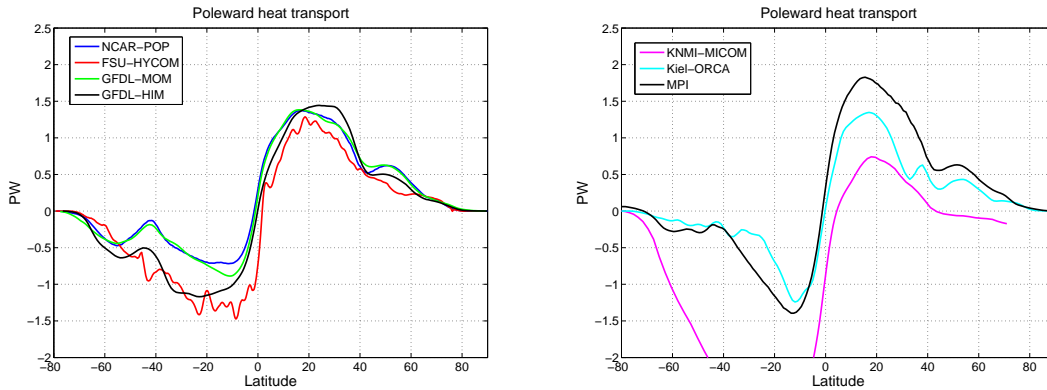


Fig. 22. Northward heat transport for the global ocean as determined by the ocean models. Results are decadal means from model years 491-500, and the values include both resolved advective and SGS transport contributions. Units are $\text{PW}=10^{15}$ Watts.

The global meridional streamfunction (Figure 24) reveals the following features:

- The strength of the vigorous tropical wind-driven cells;
- The Deacon Cell, which is a cell present in the Eulerian overturning streamfunction that is driven by the Southern Hemisphere subpolar westerlies around 60°S ;
- The AABW cell adjacent to Antarctica;
- An abyssal cell centred at around $20 - 40^\circ\text{S}$.

The models display wide-ranging values for these overturning cells. All z-level models as well as FSU-HYCOM exhibit reasonably realistic AABW cells of strength 5-10 Sv. In contrast, the isopycnal models show more vigorous overturning off Antarctica, particularly KNMI-MICOM, which exhibits substantial water-mass transformation at the high-latitude Southern Ocean, feeding an abyssal cell that ventilates much of the global ocean at mid-depth. Indeed, there is little evidence for northern sinking in this model, leaving its water-masses too cold and fresh almost everywhere (Figures 16 and 17).

The Deacon cell is nearly absent when plotted in density-latitude space (Webb, 1994; Karoly et al., 1997), as seen in the isopycnal models. In depth-latitude space, the Deacon Cell takes values in excess of 20Sv in each of the z-level models, as well as in FSU-HYCOM. In general, the depth-latitude Deacon Cell is reduced in magnitude by the Gent and McWilliams (1990) parameterisation, with equatorward meridional velocity counteracted by poleward parameterised eddy induced velocity (Hirst and McDougall, 1998). The strength of the reverse Deacon Cell arising from the GM streamfunction is set by parameters used in the SGS

parameterisation, such as the diffusivity and treatment of steep neutral slope regions (see Table 6 in Section D4), as well as the model distribution of density and layer thickness. A larger “GM-effect” acts to reduce, or nearly eliminate, the Deacon Cell (Danabasoglu et al., 1994), whereas a smaller GM-effect maintains the Deacon Cell at values such as those seen in the z-model simulations shown here.

Apart from KNMI-MICOM, the strength of the abyssal cells is in fairly good agreement with observed estimates, ranging between 6 and 20 Sv. Interestingly, the abyssal cell nearly vanishes in density coordinates in GFDL-HIM, indicating no substantial water-mass transformation associated with the abyssal cell in this isopycnal model.

15.2. Critical comments

A MOC with realistic transport strength and structure is important for maintaining a realistic ocean climate. For reasons detailed in Section 3, it is sometimes quite difficult to realize a stable overturning circulation, especially in ocean-ice models. The behaviour of the ocean-ice models in this study indeed reflects on this sensitivity, with some models “refusing” to stabilize at a circulation reflecting observations (e.g., for NADW, $\sim 15\text{Sv}$), whereas others appear to reach a stable value either with a very weak salinity restoring (NCAR-POP, FSU-HYCOM, and MPI), or stronger restoring (GFDL-MOM and Kiel-ORCA). It is notable that the two isopycnal models appear to have the most difficulty reaching a steady state, with the GFDL-HIM simulation showing large amplitude variations, whereas the KNMI-MICOM simulation settles into a very

weak, nearly absent, overturning circulation in the NADW cell.

Another shortcoming of the MOC analysis is the differing vertical coordinate systems used to plot the streamfunctions in the z-level and isopycnal models. In particular, the isopycnal models show very noisy results when projected onto the latitude-depth plane (also the case for the hybrid coordinate model FSU-HYCOM; see Figure 24b), and the z-models have the complement problem in latitude-density coordinates, unless diagnosing the advective fields at each model time step. Future comparisons should diagnose the MOC fields on the same vertical axis, and this should ideally be done in neutral density space, as then the MOC can be directly interpreted in terms of water-mass conversion rates across different density classes.

16. Surface freshwater forcing and MOC behaviour

The long-term spin-up behaviour of global ocean-ice models depends critically on surface freshwater forcing, particularly in the high-latitude areas important for setting properties of the deep and bottom water masses. As discussed in Section 3.3, salinity restoring is generally used to damp drifts in water mass properties, and in particular for the purpose of maintaining a stable MOC. During the development of CORE-I, some groups examined sensitivities to salinity/water restoring. Given that four of the seven groups were unable to stabilize the MOC without including a nontrivial restoring (i.e., piston velocity greater than the 50m/4years used by NCAR-POP, FSU-HYCOM, and MPI), we summarise results from experiments conducted with MPI, GFDL-MOM, and Kiel-ORCA which exhibit issues that may arise when ocean-ice models are run using the CORE-I design.

16.1. *Two experiments from MPI and GFDL-MOM*

Two experiments were run with the GFDL-MOM and MPI ocean-ice models, where the only difference is the strength of the salinity restoring.

- GFDL-A and MPI-A use a piston velocity of 50 m/4 years, which is the same as the NCAR-POP and FSU-HYCOM simulations. MPI-A is the standard experiment from MPI described in other sections of this paper.

- GFDL-B and MPI-B use the larger piston velocity of 50 m/300 days. GFDL-B is the standard GFDL-MOM experiment presented in other sections of this paper.

No salinity restoring occurs under sea ice in these experiments.

Figure 26 shows results for the Atlantic overturning streamfunction index and the Drake Passage transport. For the GFDL-A simulation, the North Atlantic exhibits a significant weakening of the MOC during the first 100 years, a gradual increase over the next 150 years. At year 250, both experiments show a reasonable MOC index, but then a series of growing amplitude multi-decadal oscillations start after year 250. Although we have not examined the variability mechanism, we conjecture that the oscillations are similar to the mixed boundary condition thermohaline oscillations discussed in such papers as Zhang et al. (1993), Greatbatch and Peterson (1996), and De Verdière and Huck (1999). This behaviour is distinct from the more stable GFDL-B experiment. Furthermore, the GFDL-CM2 climate model of Delworth et al. (2006), which uses the same ocean and sea ice components, exhibits multi-decadal variability (not shown), but this occurs about a stable mean state of roughly 20Sv, and does not exhibit the large drifts and huge amplitude oscillations seen in GFDL-A. We thus consider the very weakly restored experiment GFDL-A to be unsuitable for studying the mean ocean climate in GFDL-MOM. It is for this reason that all other results from the GFDL-MOM simulations focus exclusively on GFDL-B. Additionally, the transition to nonlinear oscillatory behaviour in GFDL-A, seen only after a few hundred years, provided motivation to run all the CORE-I simulations at least 500 years.

In contrast to the experience with GFDL-MOM, the two MPI simulations are more robust to changes in the salinity restoring. Note, however, that the Drake Passage transport in the weakly restored simulation MPI-A (the control MPI simulation) exhibits nontrivial temporal variability about a rather large mean. This contrasts to the behaviour in the stronger restoring case, in which the simulation is reflective of the NCAR-POP, GFDL-MOM, and Kiel-ORCA results shown in Figure 18. Nonetheless, the MPI-A and MPI-B simulations are far closer in behaviour than the analogous GFDL-MOM simulations. Because of the desire to maintain a very small salinity restoring, the MPI-A simulation was used as the control throughout the other sections of

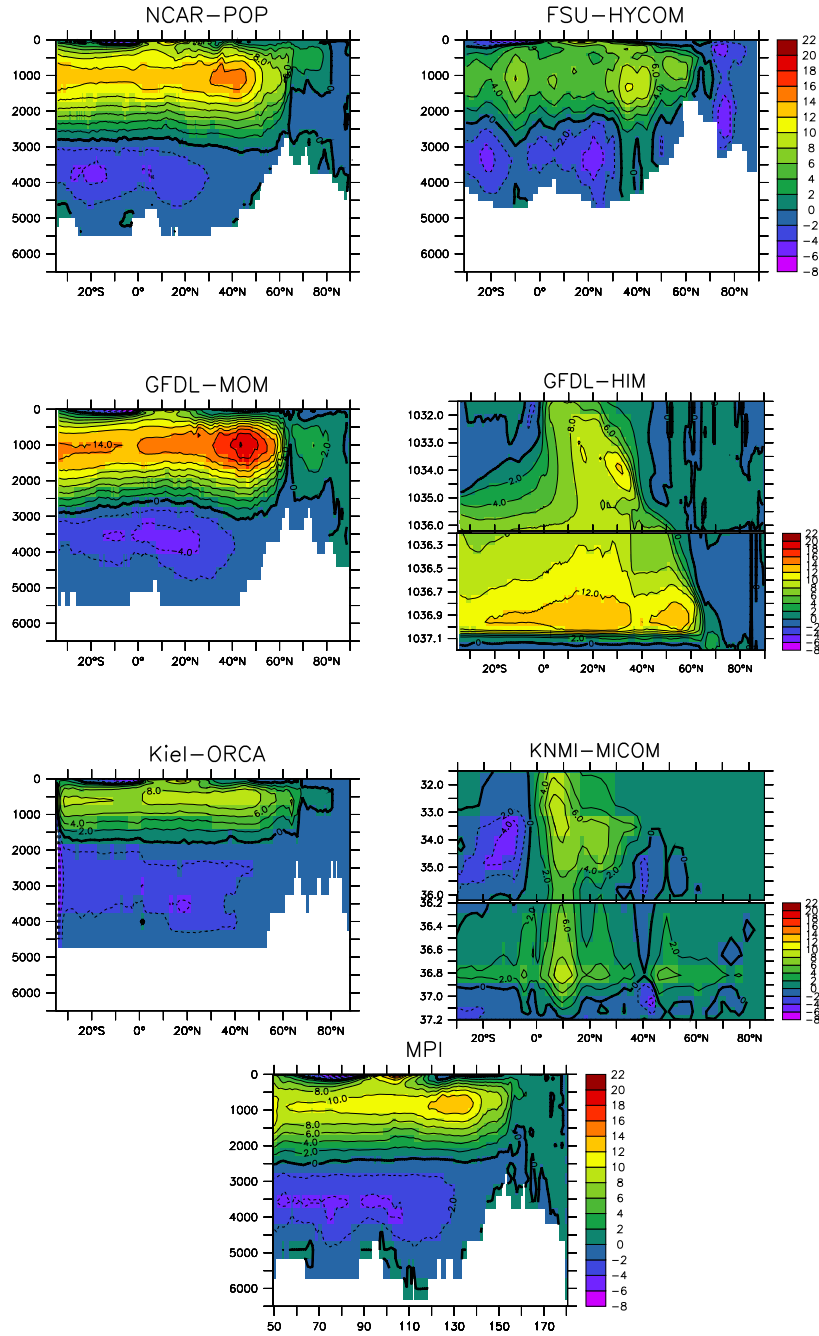


Fig. 23. The Atlantic basin meridional overturning streamfunction (in units of $Sv=10^6 \text{ m}^3 \text{ sec}^{-1}$), time averaged over years 491-500 in the CORE-I simulations. The FSU-HYCOM simulation has its results interpolated to depth based vertical coordinates, whereas GFDL-HIM and KNMI-MICOM retain the potential density coordinates referenced to 2000db ($1035 - \rho_{2000}$) (as in equation (30)). The remaining models plot the streamfunction with respect to their native depth coordinate (equation (29)). The depth plots extend to 6500m, which illustrates the differences in the deepest level represented by the different models, with MPI going the deepest. The density axes for GFDL-HIM and KNMI-MICOM are split into two regions to highlight the overturning cells in the high latitudes, as distinguished from the wind driven Ekman cells active in lower latitude lighter waters.

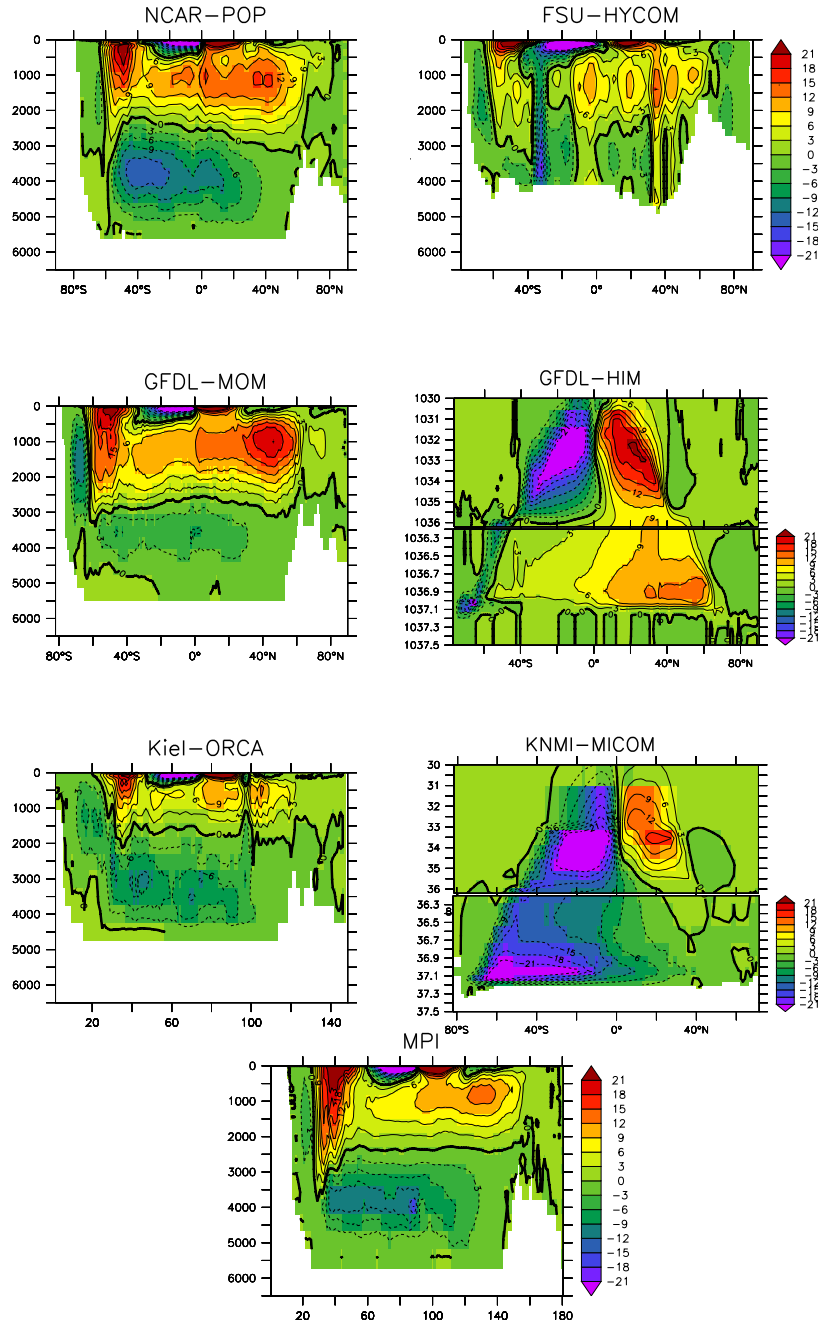


Fig. 24. The global meridional overturning streamfunction (in units of $Sv=10^6 \text{ m}^3 \text{ sec}^{-1}$), time averaged over years 491-500 in the CORE-I simulations. Note that the FSU-HYCOM simulation has its results interpolated to depth based vertical coordinates, whereas GFDL-HIM and KNMI-MICOM retain the potential density coordinates referenced to 2000db ($1035 - \rho_{2000}$) (as in equation (30)). The remaining models plot the streamfunction with respect to their native depth coordinate (equation (29)). The depth plots extend to 6500m, which illustrates the differences in the deepest level represented by the different models, with MPI going the deepest. The density axes for GFDL-HIM and KNMI-MICOM are split into two regions to highlight the overturning cells in the high latitudes, as distinguished from the wind driven Ekman cells active in lighter waters in the lower latitudes.

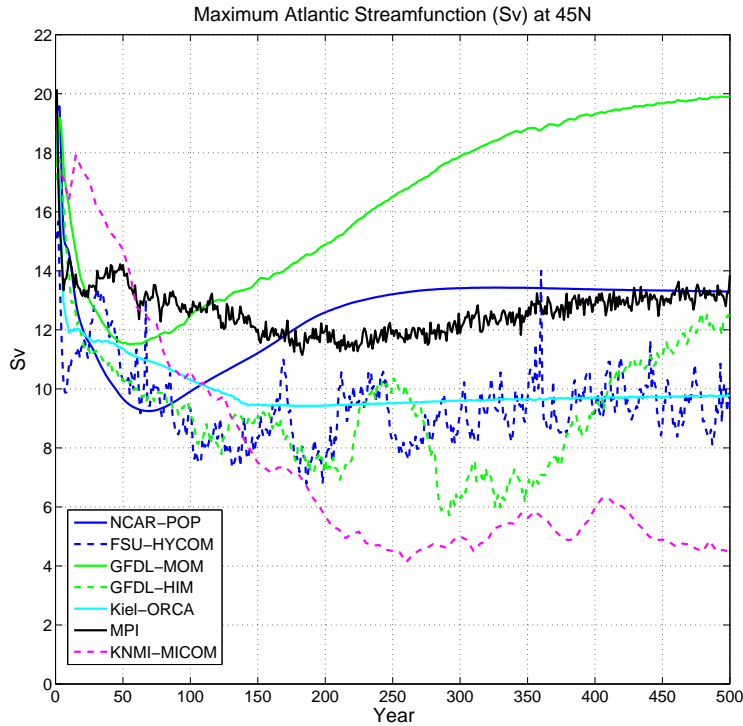


Fig. 25. Time series of the annual mean Atlantic meridional overturning streamfunction index (vertical axis) for model years 1-500 (horizontal axis) in units of $Sv=10^6 \text{ m}^3 \text{ sec}^{-1}$. The index is computed as the maximum Atlantic MOC streamfunction at $45^\circ N$ in the region beneath the wind driven Ekman layer. Note that the GFDL-MOM simulation was extended to 600 years to verify that it was reaching a steady state for the overturning.

this paper.

16.2. Kiel-ORCA experiments

Multiple experiments were run with the Kiel-ORCA configuration to examine sensitivity to surface freshwater forcing in the high-latitude areas of the Atlantic and Arctic. The experiments were broadly split into three classes: (a) no restoring; (b) "weak" restoring of surface salinity corresponding to a piston velocity of $50 \text{ m}/4 \text{ years}$; (c) "strong" restoring corresponding to $50 \text{ m}/300 \text{ days}$. Results are shown in Figure 27, with details of the restoring provided in the figure caption.

With zero or weak restoring, Kiel-ORCA exhibits a rapid decline of the North Atlantic MOC during the first 100-200 years, without any signs of multi-decadal oscillations seen in GFDL-A. Oscillations do appear, however, in the Drake passage transport (Figure 27b), with the oscillations emerging near the start of the simulations. Similar oscillations

were common in earlier versions of NCAR-POP, but damped by changing model parameters rather than increasing salinity restoring. Further oscillations are exhibited in the ORCA simulations in the northward transport of Antarctic Bottom Water (AABW) in the abyssal cell of the Southern Ocean (Figure 27c), with the maxima (minima) in abyssal cell transport closely corresponding to maxima (minima) in the ACC. Causal relations between the strength of bottom water formation and ACC-transport has been discussed by Timmermann et al. (2005) based on experimentation with a previous version of the Kiel-ORCA model.

None of the Kiel-ORCA experiments realized a MOC transport larger than $15 Sv$, with most showing much smaller values. Hydrographic properties of the overflow waters spilling across the Greenland-Iceland sill indicate that MOC evolution is closely tied to drift in the overflow density (Figure 27d), defined as the maximum density over the Denmark Strait sill before entrainment (see Latif et al., 2006,

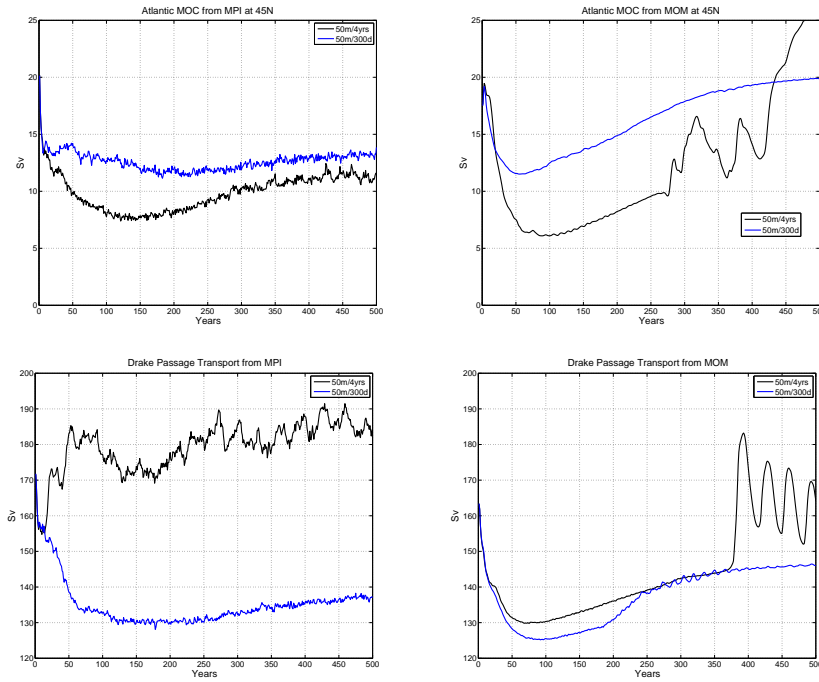


Fig. 26. Time series for the annual mean volume transports (in units of $Sv=10^6 \text{ m}^3 \text{ s}^{-1}$) from two MPI and two GFDL-MOM simulations. The black lines are from ocean-ice simulations with the salinity restoring converted to a water flux using a weak piston velocity of $50 \text{ m}/4 \text{ years}$ (this is the standard MPI simulation), and the blue lines are from ocean-ice simulations with larger piston velocity $50 \text{ m}/300 \text{ days}$ (this is the standard GFDL-MOM simulation). Shown are results for the maximum meridional-depth overturning streamfunction at $45^\circ N$, and the eastward transport in the Southern Ocean through the Drake Passage (the observational estimate from T. Whitworth, 1983; Whitworth and Peterson, 1985; Cunningham et al., 2003, is roughly $135 Sv$).

for similar analyses). Due to progressive freshening (not shown), all Kiel-ORCA cases drift from the observed climatological outflow density (of about 28.0). A drift in simulated Arctic Ocean salinities may be caused by:

- Excessive freshwater fluxes into the Arctic basin related, for example, to the prescribed precipitation or coastal run-off fields;
- An erroneous simulation of the freshwater export to the North Atlantic associated with, for example, the export of sea ice with the East Greenland Current;
- Too little import of saline waters with the North Atlantic Current.

As suggested by the increase in MOC (and overflow density) in the reduced-precipitation case ORCA-E, SSS restoring for the Kiel-ORCA simulations can only partly offset the high-latitude freshwater budget. Nonetheless, the small increase in the Atlantic MOC in ORCA-E relative to ORCA-B, despite reducing Arctic precipitation to the lower

limit of observational uncertainty, indicates that the strong, global retoring of ORCA-D is compensating for much more than precipitation error.

A particularly interesting aspect of Figure 27 is the rather tight relationship between mid-latitude MOC transport and overflow density, noted before by Latif et al. (2006). Apparently, neither changes in the restoring configuration nor in the prescribed precipitation field (ORCA-E) affect the relation between density and MOC. The main effect of these model choices is to determine the amplitude of the drift realized by a particular simulation.

16.3. A grid resolution hypothesis

Each of the many experiments conducted by the various models serve to highlight sensitivities to boundary conditions described in Section 3. So can one identify a model feature that predisposes it to retaining a nontrivial overturning with the CORE forcing using only a weak salinity restoring?

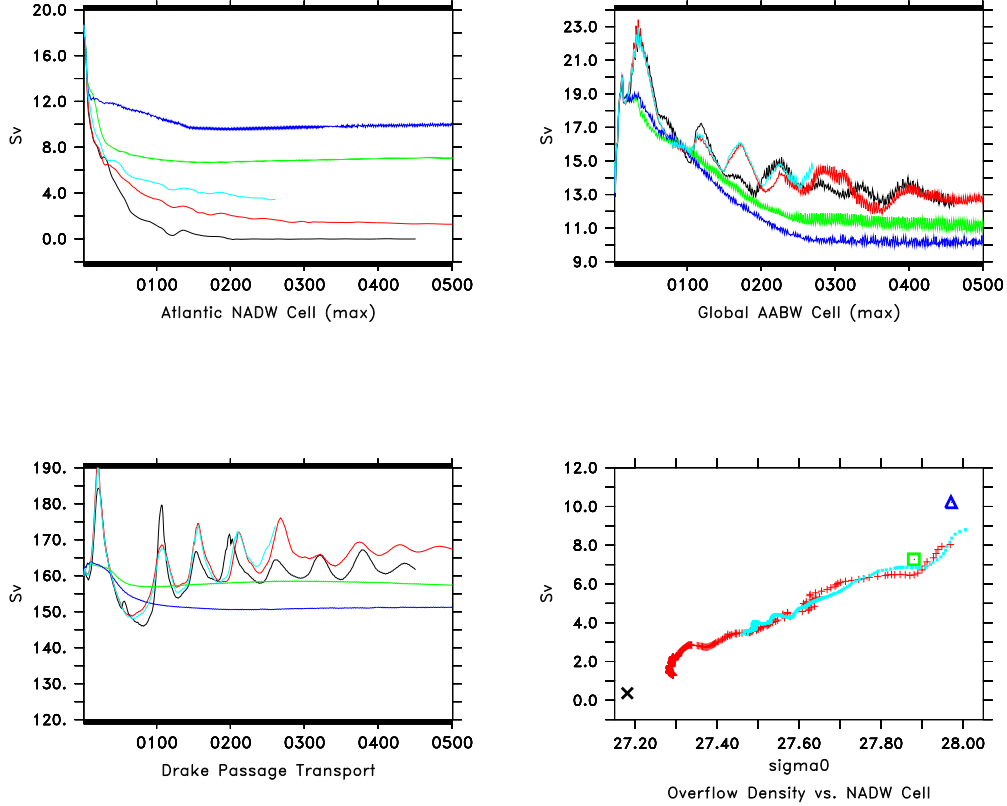


Fig. 27. Time series for the annual mean volume transports from a sequence of Kiel-ORCA experiments: ORCA-A (black), ORCA-B (red), ORCA-C (green), ORCA-D (dark blue; serving as the reference experiment), ORCA-E (light blue). These experiments differ by details of their salinity restoring: ORCA-A: no restoring; ORCA-B: weak restoring globally (no restoring under sea ice); ORCA-C: weak restoring, except for strong restoring in the (sub-)polar oceans (south of 45S, and north of 65N, also under sea ice); ORCA-D: strong restoring globally, including under sea ice. Note that over the Gulf Stream the restoring term is restrictive restoring, i.e., no freshening is allowed. This experiment represents the Kiel-ORCA reference case for all results shown in other sections. ORCA-E: restoring as in ORCA-B, but with reduced Arctic precipitation. Here, the net CORE precipitation is reduced over the Arctic and sub-Arctic oceans by 30%. This is roughly the scale to which the observational products are uncertain as to the magnitude of the net precipitation in the Arctic. Upper left: maximum MOC transport in the North Atlantic, Lower left: ACC transport through Drake passage, Upper right: maximum transport of the abyssal cell in the southern hemisphere; Lower right: scatter plot of yearly values (beginning at year 20) of annual mean MOC transport and overflow density (maximum density over the Denmark Strait sill before entrainment) for ORCA-B (red) and ORCA-E (light blue), and of the MOC/overflow values for the last decade of the 500-year integrations for ORCA-A (black cross), ORCA-C (green square), ORCA-D (blue triangle).

One hypothesis does present itself after considering characteristics of the seven model configurations. Namely, those models (NCAR-POP, FSU-HYCOM, and MPI) with fine horizontal grid resolution in the North Atlantic are reasonably stable with the weak piston velocity of 50m/4years. In particular, MPI has 12km grid resolution in the region just south of Greenland, and this resolution is far more refined than the models showing less stable behaviour.

As detailed in Section 16.2, a significant difficulty with the Kiel-ORCA simulations relates to the water budget in the deep water formation regions, which

again reflects on the processes identified in Section 3. We conjecture that these sensitivities are shared by all the ocean-ice models models in this study. However, we propose that those models able to transport more salt into these regions, such as via a more vigorous Gulf Stream, North Atlantic Current, and subpolar gyre are able to retain a stable overturning and thus to be less sensitive to details of the direct hydrological forcing in the high latitudes. This hypothesis is supported by the non-trivial overturning in the coupled climate simulations from the coarse version of CCSM Yeager et al. (2006).

This hypothesis suggests that ocean-ice models of finer high latitude resolution, especially those explicitly representing mesoscale eddies as well as a more detailed representation of the complex land-sea boundaries and associated boundary currents, will exhibit a *more* stable large-scale overturning circulation under the presence of fluctuations in the hydrological cycle. It will be of great interest to see whether future simulations with fine resolution models support this hypothesis.

17. Discussion and concluding remarks

Simulations with global ocean-ice models are more difficult to run in isolation than coupled atmosphere-land simulations. There are two main reasons for this distinction. First, the atmospheric fields needed to force the ocean-ice system are more numerous and less well known than the relatively well observed surface temperatures used to derive atmospheric fluxes over the ocean and specified sea-ice. Second, assuming that the atmosphere rapidly responds to a slowly varying SST, as in the Atmospheric Model Intercomparison Project (Gates, 1993), is a better assumption than the complement, which is assumed when running ocean-ice models with a prescribed atmospheric state. These points largely account for the lack of a community supported Ocean Model Intercomparison Project (OMIP) 15 years after the initiation of AMIP.

17.1. *Strategies for modelling the ocean-ice climate system*

There are two general conclusions one can draw when recognizing the difficulty running global ocean-ice models for climate purposes. First, one may choose to always run ocean-ice models coupled to a dynamical atmosphere and thus tolerate its inherent biases. That is, to jettison the notion of running ocean-ice models altogether, and thus to insist on only studying simulations from realistic climate models, where a prescribed atmosphere is unacceptable for the goals of, for example, climate change science. It is notable that this trend is being followed by scientists at GFDL, NCAR, and the Hadley Centre, where ocean-ice simulations are becoming less common than fully coupled climate simulations, even for purposes of developing ocean-ice models. Alternatively, one may consider realistic ocean-ice models coupled to atmospheric energy

balance models, such as discussed for the CORE-III experiments of Gerdes et al. (2005a, 2006). This approach requires far less overhead in atmospheric modelling expertise, and so is more available to groups focused just on ocean and sea ice modelling research. Either of these approaches which use a prognostic atmosphere precludes using the ocean-ice models for many of the purposes outlined in Section 2; most notably being an inability to perform reanalysis simulations aimed at interpreting the observed ocean record.

A second conclusion that can be drawn is to acknowledge the limitations inherent with ocean-ice models, and to thus use them for selected research studies within a hierarchy of numerical climate science tools, ideally with a fully coupled climate model as part of this hierarchy. This second conclusion is taken by many research groups, and it is used as motivation for our proposal of Coordinated Ocean-ice Reference Experiments (COREs) in hope of facilitating collaborative research and model development. That is, we aim with CORE to bring the ocean-ice models onto a common platform for experimental design and prescribed atmospheric state, from which remaining model differences can be used to ascribe reasons for diverging simulation behaviour.

Given our proposed CORE simulations, we illustrated results from seven ocean-ice models run with the CORE-I experimental design, which involves running the models for 500 years with the Normal Year Forcing from Large and Yeager (2004). These simulations allowed us to test the hypothesis that global ocean-ice models run under common forcing conditions will result in qualitatively similar simulations. This paper finds this hypothesis not to be valid in general, though the degree is very dependent on the particular diagnostic chosen, and the model sub-set. The nontrivial differences arising from the simulations are often not related to the vertical coordinate, which highlights the need to further examine the mechanisms accounting for the divergences.

17.2. *Concerning the merit of model comparisons*

A large part of this project was aimed at answering questions about the feasibility of bringing ocean-ice models onto a common design framework. A key reason for doing so is to ensure that differences between model simulations are not attributed to differ-

ences in atmospheric forcing. These questions have been answered in the affirmative by the existence of seven contributing groups having followed the CORE-I protocol. Questions regarding the merit of this exercise have also been raised. What is gained? Comparison projects can be fraught with difficulties in bringing the variety of model configurations into a reasonably controlled and understood setting. These difficulties make it tough to learn anything new regarding mechanisms for simulation differences, and this point can be used to criticise a model comparison project.

However, these difficulties are not an argument for the dis-utility of comparisons. Instead, the utility of such “come as you are” comparisons, with the present paper falling somewhat into this category, arise from the many new questions raised when simulations are compared and contrasted. Some questions relate to model fundamentals and practice, each of which are brought to an expanded “table” when modellers using different codes collaborate. This is a healthy situation, which is handicapped when groups conduct research in isolation. Many questions, such as the following, generally go unasked in the absence of such collaborations and associated comparisons.

- Why does the isopycnal model from KNMI-MICOM have so many problems matching the observations, relative to the other models? Is there something inherent in the chosen configuration, such as the choice of 16 layers rather than a number closer to the 50 chosen by GFDL-HIM? Or perhaps there are problems with certain fundamentals of the MICOM algorithm which have been remedied in GFDL-HIM. Namely, are problems conserving total heat and salt (see Appendix A5) manifesting in large drifts?
- What causes the relatively weak equatorial undercurrent in the FSU-HYCOM and MPI simulations? For MPI, this appears to be a result of insufficient resolution in the equatorial region. For the FSU-HYCOM simulation, an egregious problem is the transition from the tropical mixed layer, represented by pressure levels, to interior stratified ocean, represented by isopycnal layers. For all models, what is the role of the specific choice of friction?
- Is there a transition in horizontal grid resolution whereby refinements render a more stable overturning circulation compared to coarser simulations? If so, what features of the simulation, such as subpolar boundary currents and mesoscale

eddy features, are critical for maintaining a relatively stable North Atlantic simulation?

These, and other questions, motivate further focused study conducted in a more controlled setting that aim to elucidate mechanisms. Indeed, a project to be conducted through GFDL aims to run a subset of the models used here with the same grid, each using the CCSM sea ice and CCSM coupler.

In general, we propose that the preliminary experience with CORE-I documented in this paper provides a solid foundation for deeper study of many important questions about ocean-ice simulations as well as the mechanisms affecting climate means, variability, and stability. It also supports studies concerning the suitability of a prescribed atmospheric state for running ocean-ice models, and in particular the utility of the Large and Yeager (2004) dataset, both the normal year and the interannually varying versions.

17.3. COREs and coordination

CORE-I is one aspect of our envisioned coordinated suite of ocean-ice experiments. It focuses on the longer term ocean climate, with integration times sufficient for some models to reach a quasi-equilibrium. These simulations have direct relevance to the development of ocean-ice components of fully coupled climate models used for long term climate change simulations. Although fraught with the caveats detailed in Section 3, long term ocean-ice simulations from CORE-I render insights about the nature of quasi-equilibrium ocean-ice solutions. In particular, long term simulations provide a clear means to expose fundamental problems with both the model algorithms and the experimental design.

A complement task is realized with the envisioned CORE-II experiment, whereby models are run with the interannually varying atmospheric state from Large and Yeager (2004). These simulations are relevant for developing a mechanistic understanding of the observed ocean record compiled during the past few decades. Questions about how to initialize the models to minimise drift remain a key aspect of the CORE-II experiment, with this nontrivial question largely ignored in the present focus on the longer time scale CORE-I simulations. Finally, a third aspect of the ocean-ice climate system is addressed by considering the CORE-III experiment of Gerdes et al. (2005b), whereby the ocean is perturbed by a fresh water melt around Greenland,

meant to emulate the possible melt in a warming world. This experiment, which is similar in many regards to those conducted as part of the Coupled Model Intercomparison Project (see, for example, Hewitt et al., 2006), allows for an increased variety of ocean-ice participants, beyond those running fully coupled climate models. It also presents the ocean-ice system with an arguably more vigorous perturbation, which is both realistic and challenging to the model algorithms.

Each of the CORE-I, CORE-II, and CORE-III simulations benefit from coordination. This coordination assists by keeping groups more scientifically open. In particular, this openness prompts groups to rationally address aspects of their simulation which may otherwise be ignored or overlooked. Quite literally, peer-pressure provides a positive element in moving modellers towards a rational and robust platform. Coordination is also essential since the problems of ocean climate dynamics generally require teams of scientists approaching the questions from many complementary perspectives, and with differing tools. Such coordination generally is necessary for the climate modelling community to garner robust understandings of climate mechanisms, which ultimately increase our ability to rationally understand the past and to project into the future.

Acknowledgements

This paper was prompted by discussions occurring within the CLIVAR Working Group for Ocean Model Development (WGOMD) since the panel's first meeting in 2000. The WGOMD sponsored a Pilot-OMIP starting in 2001, with this project involving just a few models. The Pilot-OMIP was itself based on an earlier comparison of German ocean-ice models (Fritzsche et al., 2000). Both projects used the forcing dataset compiled by Röske (2005). The culmination of the Pilot-OMIP occurred at a CLIVAR workshop on ocean climate modelling held June 16-18, 2004 at the Geophysical Fluid Dynamics Laboratory in Princeton USA. At the workshop, it was decided that a research project should be conducted to flesh out a benchmark simulation prior to stepping all the way towards an OMIP. This conclusion then set the stage for COREs.

We thank David Bi, Rainer Bleck, Ge Peng, Frank Röske, and Ashwanth Srinivasan for critical support during various parts of this project. We thank Laurent Brodeau for documenting the effects of the code

error in the bulk formulae calculation described in Appendix C. Helene Banks, Riccardo Farnettic, and Ron Stouffer provided critical comments on earlier drafts of this paper. SMG thanks CSIRO Marine and Atmospheric Research in Australia for its hospitality during 2005, where this paper was initiated.

Appendix A: Contributing models

Table 1 summarises various aspects of the models used in this study, and Table 2 details the model release and the date when the simulations were run. These models are examples of high-end global ocean-ice configurations that are actively being used for climate research, either in the ocean-ice configurations employed in this paper, or as part of more complete climate or earth system modelling systems. Model details beyond those provided in this appendix can be found in the cited references.

A1. NCAR-POP

This is the ocean-ice component of the Community Climate System Model, CCSM3 (Collins et al., 2006). It consists of the Los Alamos Parallel Ocean Program (POP) documented by Smith and Gent (2004), which uses a z -coordinate for the vertical grid. This ocean is coupled to the Community Sea Ice Model (CSIM) that is documented by Briegleb et al. (2004). The resolution is roughly one degree (320 zonal grid cells) with refined meridional spacing in the equatorial region (384 meridional grid cells) and 40 unevenly spaced cells in the vertical. The north coordinate singularity is displaced over Greenland to avoid the traditional spherical coordinate singularity at the geographic North Pole. This displacement also has the added feature of refining the grid resolution in the Labrador Sea deep water formation region (Smith et al., 1995). It has been conjectured that this enhanced resolution in the deep water formation region is important for maintaining stability of the overturning circulation in the NCAR-POP CORE-I simulation. Systematic tests of this conjecture have only been performed at lower resolution Yeager et al. (2006).

The NCAR-POP ocean model component updates the tracer and baroclinic velocity fields using a leap-frog for the inviscid portion of the equations with a 3600s time step. The barotropic equations are updated implicitly in time using a leap-frog for the tendencies with 3600s time step (Dukowicz and

MODEL	VERTICAL (NUMBER)	ARAKAWA	HORIZONTAL	BAROTROPIC	BAROCLINIC/TRACER
NCAR-POP	z(40)	B	displaced(320x384)	implicit(3600s)	leap-frog (3600s)
FSU-HYCOM	hybrid(32) p, ρ_2, σ	C	displaced(320x384)	explicit(36s)	leap-frog (2160s)
GFDL-MOM	z(50)	B	tripolar(360x200)	explicit(80s)	staggered (7200s)
GFDL-HIM	ρ_2 (50)	C	tripolar(360x210)	explicit(60s)	pred-corrector (7200s+3600s)
KNMI-MICOM	ρ_2 (16)	C	tripolar(180x95)	explicit(72s)	leap-frog (1440s)
MPI	z(40)	C	displaced(256x220)	semi-implicit(4800s)	semi-implicit(4800s)
Kiel-ORCA	z(30)	C	tripolar(180x148)	implicit(5400s)	leap-frog (5400s)

Table 1

Summary of the ocean models used in this study. Indicated here are the model names; vertical coordinate and number of discrete vertical levels/layers; arrangement of variables on the horizontal grid according to the classification of Arakawa and Lamb (1977); orientation of the horizontal grid relative to the Arctic and the number of horizontal grid cells; use of explicit or implicit barotropic algorithm for computing the free surface and vertically integrated velocity, as well as the time step in seconds used for the barotropic equations; the time stepping method used for the inviscid portion of the baroclinic/tracer equations, as well as the time step in seconds. For the MPI ocean model, the semi-implicit method is implemented as a forward-backward scheme. For KNMI-MICOM and GFDL-HIM, the vertical is discretized according to layers of potential density referenced to 2000dbar. For HYCOM, the vertical is discretized according to pressure (upper ocean), potential density referenced to 2000dbar (interior ocean), and terrain following σ (bottom). All other models use geopotential vertical coordinates. For GFDL-HIM, tracers and diabatic processes are time stepped with 7200s, whereas adiabatic baroclinic dynamics are time stepped with 3600s.

MODEL	model release	dates for integration
NCAR-POP	POP1.4 and CSIM4	Mar 2004
FSU-HYCOM	HYCOM 2.2 and CSIM4	May-Aug 2007
GFDL-MOM	MOM4p0d	Feb-Apr 2005
GFDL-HIM	HIM-Fortran M-release	Feb-Mar 2007
KNMI-MICOM	MICOM 2.9	Jun 2006
MPI	mpiom-1.2.3	Oct-Dec 2006
Kiel-ORCA	NEMO 1.06	Feb-Apr 2006

Table 2

Details of the model release and the dates over which the simulations were run. This information specifies the code base used in the CORE-I simulations submitted by the various groups. Note the relatively wide range in calendar dates over which the simulations were run. This range reflects the effort needed by the various groups to develop the code base for running ocean-ice simulations, and the computational cost of running 500 year coupled ocean-ice simulations.

Smith, 1994). No water is transported across the ocean surface, thus necessitating salt fluxes rather than water fluxes.

Snow and ice albedos are important for computing the absorption of shortwave radiation in snow and sea ice system, and hence for setting the strength of the snow and ice albedo feedback. The NCAR-POP sea ice model employs an albedo following the approach documented in Briegleb et al. (2002). In particular, the NCAR-POP ice model distinguishes between visible (wavelengths $< 0.7\mu\text{m}$) and near-infrared (wavelengths $> 0.7\mu\text{m}$), since snow and ice spectral reflectivities are significantly higher in the visible band than the near-infrared band. The zenith angle dependence of snow and ice is ignored, and therefore the distinction between downwelling direct

and diffuse shortwave radiation is ignored. Snow and ice topography affect the scattering and transmission into the surface through shadowing effects and through variations in the angle of the surface from the horizon. These topography affects are also ignored. For further details of the NCAR-POP albedo, please refer to Briegleb et al. (2002). Although the ocean and sea ice components are the same as those used in the CCSM3.0 fully coupled climate model (e.g., Danabasoglu et al., 2006), there were some modifications made to the sea ice component parameters for use in the CORE simulations. The changes included slightly higher ice and snow albedos that are more in the middle of observed values.

The salinity of sea ice is 4ppt. This affords a nonzero salt flux between the ocean and sea ice as

ice melts and forms.

A2. FSU-HYCOM

FSU-HYCOM consists of the ocean-ice component to the Community Climate System Model (Collins et al., 2006) with HYCOM version 2.2 as the oceanic model. The horizontal grid and topography are identical to that of NCAR-POP. FSU-HYCOM is configured with 32 hybrid layers and each coordinate surface is assigned a reference isopycnal. The model continually checks whether or not grid points lie on their reference isopycnals and, if not, tries to move them vertically toward the latter (Bleck, 2002). However, the grid points are not allowed to migrate when this would lead to excessive crowding of coordinate surfaces. Thus, in the mixed layer or in shallow water, vertical grid points are geometrically constrained to remain at a fixed pressure while being allowed to join and follow their reference isopycnals over the adjacent deep ocean. Therefore, HYCOM behaves like a pressure coordinate model in the mixed layer or other unstratified regions, like an isopycnic coordinate model in stratified regions, and like a conventional terrain-following model in very shallow and/or unstratified oceanic regions (Chassignet et al., 2003, 2006).

The horizontal resolution of FSU-HYCOM is identical to NCAR-POP. The baroclinic and barotropic time steps are 2160s and 36s respectively. The model uses a non-slab K-profile parameterisation (KPP) mixed layer submodel (Large et al., 1994) and virtual salt flux at the ocean surface. The sea ice model employed by FSU-HYCOM is the same version of CSIM as used by NCAR-POP.

A3. GFDL-MOM

This is the ocean-ice component of the GFDL climate model (Delworth et al., 2006; Griffies et al., 2005; Gnanadesikan et al., 2006) which uses the z-coordinate ocean code MOM4 documented by Griffies et al. (2004) and Griffies (2004), and the sea ice model as discussed in Delworth et al. (2006). The model grid uses 360 cells in the zonal direction (one degree), 200 latitudinal cells (1/3 degree at the equator), and 50 vertical cells (22 in the upper 220m). The horizontal grid is tripolar as prescribed by Murray (1996) so that there is no coordinate singularity at the North Pole and grid cells in the Arctic are roughly of equal area. The model updates

the tracer and baroclinic velocity using a staggered scheme (Griffies, 2004; Griffies et al., 2005) with a 7200s time step for both the inviscid dynamics and dissipative physics, and the barotropic fields are updated explicitly with a predictor-corrector algorithm (Griffies et al., 2001; Griffies, 2004) using an 80s time step. The ocean model allows water to be transported across the ocean surface, thus does not employ surface salt fluxes.

As described in the Appendix to Delworth et al. (2006), the GFDL Sea Ice Simulator (SIS) employs an albedo following the CCSM approach documented in Briegleb et al. (2002) and described above, yet with some modifications. In particular, the GFDL model does not distinguish between visible and near infrared surface shortwave radiation, and the spectral albedos of Briegleb et al. (2002) are combined in a fixed ratio: 54% visible and 47% near infrared. The dry and wet albedos for snow are 0.80(dry)/0.68(wet) and for ice are 0.58(dry)/0.51(wet). Additionally, the Briegleb et al. (2002) scheme has been modified so that wet albedos are used within $10^{\circ}K$ of the melting temperature, rather than $1^{\circ}K$ of melting.

The salinity of sea ice is 5ppt. This affords a nonzero salt flux between the ocean and sea ice as ice melts and forms.

A4. GFDL-HIM

For GFDL-HIM, we replace the MOM code with the Hallberg Isopycnal Model (HIM). Details and references for HIM are provided by Hallberg and Gnanadesikan (2006). This model shares certain features with the more widely used Miami Isopycnal Model (MICOM) described by (Bleck et al., 1992), in that it is a C-grid isopycnal layer model using a split-explicit momentum equation solver. However, many of the numerical and physical algorithms differ, with detailed references provided in Hallberg and Gnanadesikan (2006). In particular, the mixed layer can be considered a “refined” bulk mixed layer which allows for velocity shears within the mixed layer (Hallberg, 2003). This shear introduces Ekman-driven mixed layer destabilization and shear-driven restratification. These processes are permitted in geopotential coordinate models, so long as the mixed layer has a suitable vertical resolution; they are absent, however, in the mixed layer scheme used in the MICOM code used by KNMI-MICOM (Section A5).

The GFDL-HIM model is configured with 360 cells in the zonal direction (one degree resolution), 210 latitudinal cells (1/3 degree at the equator), and 50 vertical potential density layers, with potential density referenced to 2000dbar. As for the GFDL-MOM simulation, the horizontal grid is tripolar as prescribed by Murray (1996). The model updates the tracer and diabatic processes using a predictor corrector scheme with a 7200s time step, and 3600s for the inviscid dynamics; the barotropic fields are updated explicitly with a predictor-corrector algorithm using an 60s time step. The ocean model allows water to be transported across the ocean surface, and so does not employ surface salt fluxes.

A5. KNMI-MICOM

The version of the ocean-ice model for this manuscript used at KNMI (Royal Netherlands Meteorological Institute) is part of the *SpeedO* coupled modelling framework described by Hazeleger et al. (2003). The ocean component is a global implementation of the Miami Isopycnal Model (MICOM) described by Bleck et al. (1992). The model uses an isopycnal vertical coordinate below the mixed layer. A bulk mixed layer scheme is used for the upper ocean with varying coefficients according to Gaspar et al. (1990). The model is coupled to an ice model that has been used by Bentsen et al. (2004). Notably, this model does not include frazil ice. The convection algorithm and decabbling algorithm of Bentsen et al. (2004) have been implemented. Thermobaric effects are included in the ocean model. No water is transported across the ocean surface, thus necessitating salt fluxes rather than real water fluxes.

The ocean-ice model has a curvilinear C-grid with 3 poles, one in Antarctica, one in Canada and one in Siberia (tripolar grid as described in Sun and Bleck (2006)). The zonal resolution is two degrees (180 grid cells), with meridional resolution two degrees outside the tropics and refined to one-half degree in the tropics (95 grid cells). The vertical resolution has 16 isopycnal layers, with potential density referenced to 2000dbar. The baroclinic and tracer time steps are 1440s, and the barotropic time step is 72s.

As seen by the various model metrics, the KNMI-MICOM simulation is an outlier, failing to match the physical integrity of the other simulations for nearly all examined diagnostics. Some fundamental problems with the MICOM version 2.9 algorithms

are mentioned here as possible reasons for the simulation difficulties.

- Thermobaricity and cabbeling:
 - Thermobaricity and cabbeling cause checkerboard-like instabilities when the code is configured for refined vertical resolutions at high densities. These instabilities were suppressed by reducing the vertical resolution down to the 16 used here. Doing so, unfortunately, then led to poor representation of deep ocean processes. Note that Hallberg (2005) documents this problem and provides a partial fix. The Hallberg (2005) fix is used in GFDL-HIM simulations, as required since GFDL-HIM uses a much finer vertical layer spacing with 50 layers.
 - The decabbling routine uses a very diffusive first order upstream advection, which can create huge levels of spurious diapycnal mixing. In particular, there were large temperature and salinity drifts between the deep and upper ocean due to cabbeling.
 - Thermobaricity corrections are not globally defined. Instead, they use a region-dependent *a priori* temperature and salinity profile to calculate the thermobaricity corrections. There were large hydrography drifts within layers between adjacent regions with different thermobaricity algorithms, most noticeably between the Southern Ocean and Atlantic.
- MICOM uses the MPDATA scheme of Smolarkiewicz (1984) for the horizontal advection of tracers. MICOM advects density and spiciness rather than the traditional potential temperature and salinity. MPDATA, however, does not conserve tracer content. To maintain constant global tracer amounts, MICOM applies a global correction. This correction adds extra horizontal and vertical diffusion between the layers, and it can exchange tracers between disconnected parts of a layer.
- Another contribution to the highly diffusive simulation comes from the Robert-Asselin filter (e.g., Haltiner and Williams, 1980) used to damp the splitting mode in the leap frog time-stepping scheme.

A6. MPI

The Max-Planck-Institute ocean model (MPIOM) is the ocean-sea ice component of the Max-Planck-Institute climate model (Roeckner et al., 2006;

Jungclauss et al., 2006). MPIOM is a primitive equation model (C-Grid, z-coordinates, free surface) with the hydrostatic and Boussinesq assumptions. It includes an embedded dynamic/thermodynamic sea ice model with a viscous-plastic rheology following Hibler (1979) and a bottom boundary layer scheme for the flow across steep topography. A model description can be found at Marsland et al. (2003).

The model configuration applied here has 40 vertical levels, with 20 in the upper 600m and the bottom topography is approximated by partial cells. The horizontal resolution of the bipolar grid with poles over Greenland and Antarctica gradually varies between a minimum of 12km close to Greenland and 190km in the tropical Pacific. The model time stepping for tracers and velocity is formulated semi-implicitly with a time step of 4800s. Sea ice has a salinity of 5psu. MPI’s sea ice is initialised with thickness and concentration depended on the initial sea surface temperature.

A7. Kiel-ORCA

This is the ocean-ice component of the IPSL climate model (Institut Pierre-Simon Laplace) in Paris. It uses the NEMO framework (Nucleus for European Modelling of the Ocean) version 1.06, with the OPA9 z-coordinate ocean code (Madec et al., 1999) and LIM2 sea ice (Timmermann et al., 2005).

The ocean model has 30 vertical levels (16 levels in the upper 200m) and a quasi-isotropic global tripolar grid, with two degrees resolution in the zonal direction and a meridional refinement at the equator to 0.5° ($180 \times 149 \times 30$ grid cells). Topography is represented with partial cells (Barnier et al., 2006). The model updates the tracer and baroclinic velocity using a leap-frog time step of 5400s for the inviscid dynamics and a forward step for dissipative physics. The barotropic equations are updated implicitly in time according to the algorithm of Roullet and Madec (2000) (using same time steps as the baroclinic). The ocean model allows water to be transported across the ocean surface, thus does not employ surface salt fluxes.

Appendix B: CORE-I experimental design

We summarise here the experimental design. There are notable details which differ between the

models. Even in a cordial research collaboration such as this, we found it difficult to remove all differences, either because of incompatibilities between model algorithms, or scientific disagreements. It is conjectured that differences identified below are less important than the decision to use the same forcing, including the same bulk formulae, both based on Large and Yeager (2004). This conjecture, however, remains unproven without further sensitivity experiments which are beyond the scope of this paper.

B1. Initial conditions

We document here the methods used by the various modelling centres for initialising their experiments. Ocean initial conditions are thought to be of minor consequence on the centennial to multi-centennial time scale of CORE-I. Shorter integrations, however, are influenced by the initial conditions, especially in deep water formation regions and regions of sea ice cover. For sea ice, in the absence of restoring of surface salinity, the initial sea ice permanently changes the total salt content in the ocean. In particular, if starting with zero sea ice, sea ice formation will increase ocean salinity due to the conversion of sea water from liquid to ice. As long as the sea ice does not fully melt again, this salinity bias will persist. The presence of a salinity restoring will dampen this bias.

Ocean models are initialised with zero velocity. The initial temperature and salinity fields are taken from the January temperature and salinity profiles of Steele et al. (2001). The exception is that KNMI-MICOM used initial conditions based on Levitus et al. (1994). Temperature and salinity fields are interpolated to the respective ocean model grid.

- Ice models are initialised in the following manners.
- NCAR-POP: Zero initial ice concentration and zero ice volume.
 - FSU-HYCOM: Zero initial ice concentration and zero ice volume.
 - GFDL-HIM: Sea ice is taken from an earlier spinup.
 - GFDL-MOM: Sea ice is taken from an earlier spinup.
 - KNMI-MICOM: Zero initial ice concentration and zero ice volume.
 - Kiel-ORCA: Initial sea ice has a uniform thickness of 1m in the Southern Hemisphere for grid boxes below freezing with a lead fraction of 10%, and 0m in the Northern Hemisphere.

- MPI: Sea ice is initialised for SST values between -0.5°C and the freezing point of sea water, with linearly increasing 0 to 4m thickness and a 100% concentration.

B2. Fluxes from Large and Yeager (2004)

The normal year forcing for heat, moisture, and momentum are provided by the Large and Yeager (2004) compilation of solar radiation, atmospheric state, and river runoff. The atmospheric state as well as the model’s prognostic SST and surface currents determine the turbulent momentum (wind stress), turbulent heat (sensible and latent), and turbulent moisture (evaporation) fluxes. These fluxes are computed at each ocean-ice coupling time step using the CCSM bulk formulae described in Large and Yeager (2004) and ???. Further details on the forcing are provided in Appendix C.

B3. Salinity/water forcing

We provide here some details of the water or salinity forcing used in the models. Again, the salinity or water restoring is not physical, but it has been found necessary to prevent uncontrolled drifts in salinity as a response to inaccurate precipitation. For models using the salt flux boundary condition (NCAR-POP, FSU-HYCOM, and KNMI-MICOM), fresh water from precipitation, evaporation, and river runoff is converted to an implied salt flux via ⁵

$$F_{(\text{implied})}^{(\text{salt})} = q_w S_o, \quad (1)$$

where S_o is a global reference salinity. The NCAR-POP and FSU-HYCOM simulations set $S_o = 34.7\text{psu}$. Salinity restoring is applied to the ocean according to the restoring salt flux

$$F_{(\text{restore})}^{(\text{salt})} = V_{\text{piston}} (S^{(\text{data})} - SSS) \quad (2)$$

with a nonnegative piston velocity V_{piston} . Hence, when the prognostic surface salinity SSS is larger than monthly surface salinities $S^{(\text{data})}$, a negative restoring salt flux $F_{(\text{restore})}^{(\text{salt})}$ is added to the top ocean model grid cell.

The salinities $S^{(\text{data})}$ are available at <http://nomads.gfdl.noaa.gov/nomads/forms/mom4/CORRE.html>.

This field is discussed in Section 4.7 of Large and Yeager (2004). Briefly, it is derived from the

⁵ Some authors refer to this salt flux as a *virtual* flux, since it is not present in Nature (Huang, 1993).

Conkright et al. (2002) and Steele et al. (2001) monthly mean data. We averaged the top two vertical levels to remove overly fresh biases from river input. Because of processing errors in the Foxe Basin, levels 1-4 were reset to level 5, and Hudson Bay April level 1 was reset to level 2. Near the Antarctic, we use the Doney and Hecht (2002) recipe, which prescribes the following.

- Where the bathymetric depth is less than or equal to 300m, the column-max salinity replaces the near-surface salinity.
- Where the depth is greater than this depth but less than or equal to 1000m, the column-max salinity is blended with the near-surface salinity.
- Where the bathymetric depth is greater still, the near-surface salinity is used, without any enhancement.

The column-max salinity is derived from a search through all 12 months, and not just from a search of the one particular month in question.

No restoring is applied under sea ice in the KNMI-MICOM simulation, but note that grid cells may have partial ice cover such that some restoring does take place over those cells. The NCAR-POP simulation used a salinity piston velocity of 50m/4years in all regions of the ocean (including under sea ice), with the exception of enclosed marginal seas. The KNMI-MICOM simulation used a regionally dependent piston velocity: zero piston velocity in the region 35°N - 65°N , 275°E and 0°E (North Atlantic); 50m/50days north of 75°N ; 50m/30days south of 50°S ; and 50m/1500days for the remainder. Gradients between the different zones were linearly interpolated in a 10 degree strip. The NCAR-POP simulation removes the global mean restoring every time step, and multiplies precipitation by a factor based on the global salinity change over the previous year such that

$$\text{net salt input} = \sum F_{(\text{implied})}^{(\text{salt})} dx dy \quad (3)$$

is damped towards zero. This procedure compensates for the initial growth of sea-ice. KNMI-MICOM and FSU-HYCOM do not apply a normalisation, and such may account for the fresh drift seen in the salinity bias time series (Figure 6).

The GFDL-HIM, GFDL-MOM, MPI, and Kiel-ORCA models transport water across the ocean surface, and thus add fresh water directly to the volume conservation equation (i.e., the free surface height evolution). Correspondingly, the restoring salt flux in equation (2) is converted to a fresh water flux q_w^{restore} via

$$q_w^{\text{restore}} = -F_{(\text{restore})}^{(\text{salt})}/SSS. \quad (4)$$

Hence, water is added ($q_w^{\text{restore}} > 0$) where the model's prognostic SSS is greater than $S^{(\text{data})}$, and subtracted ($q_w^{\text{restore}} < 0$) where $SSS < S^{(\text{data})}$. For the Kiel-ORCA simulations, no extra water normalisation is applied, and this may account for the fresh drift seen in the salinity bias time series (Figure 6). In the GFDL and MPI simulations, a global normalisation is applied at each time step so that over the ocean surface,

$$\text{net water input to ocean} = \sum (P - E + R + q_w^{\text{restore}}) \quad (5)$$

is set to zero.⁶ This normalisation keeps the ocean volume constant yet has the unphysical effect of adding a water flux nonlocally, even to a region that may have zero salinity bias. Nonetheless, it serves to avoid longterm drifts in the salinity and sea level.

Choices made for the models are summarised in Table 3.

Appendix C: Forcing the ocean-ice models

The ocean and sea ice systems are forced by air-sea and air-ice fluxes, respectively, and coupled through ice-ocean fluxes. In a coupled ocean-ice model, the latter are internally computed according to the specific ice-model. The purpose of this appendix is to highlight details of computing the atmospheric forcing, which have proven important in setting up the CORE simulations. Although it focuses on air-sea fluxes, many of the issues also pertain to air-ice fluxes, and the similarities and differences in flux calculations over the two surfaces are detailed in Large (2005).

C1. Air-sea fluxes

This section follows closely Section 2 from Large and Yeager (2008). The ocean is forced by fluxes of freshwater, F_s , heat, Q_{surf} , and momentum, τ . In general, contributions come both from the open ocean fraction, f_o , and the sea-ice fraction, $1 - f_o$:

$$\tau = f_o \tau_{as} + (1 - f_o) \tau_{io} \quad (6)$$

$$F_s = f_o F_{as} + (1 - f_o) F_{io} + R \quad (7)$$

$$Q_{surf} = f_o Q_{as} + (1 - f_o) Q_{io}, \quad (8)$$

where the subscript *as* refers to air-sea fluxes, *io* refers to ice-ocean, and R is the continental runoff. The air-sea heat and freshwater fluxes have components :

$$Q_{as} = Q_S + Q_L + Q_E + Q_H \quad (9)$$

$$F_{as} = P + E, \quad (10)$$

where all fluxes are positive when momentum, water, or heat enter the ocean.

The shortwave solar radiation, Q_S , with wavelengths between 0.3μ and 3μ , is computed from the solar insolation Q_I incident on the ocean surface, minus that which is reflected from the ocean surface

$$Q_S = Q_I (1 - \alpha) \quad (11)$$

with the albedo $\alpha \approx 0.065$ appropriate for the ocean (Payne, 1972). The net surface longwave solar

$$Q_L = Q_A - \sigma SST^4 \quad (12)$$

results from a downwelling longwave flux Q_A from the atmosphere, minus the blackbody radiation from the ocean back to the atmosphere with $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ }^\circ\text{K}^{-4}$ the Stefan-Boltzmann constant.

Bulk formulae parameterise the turbulent fluxes (wind stress τ , evaporation E , latent heat Q_E , and sensible heat Q_H) in terms of the near surface atmospheric state (wind \mathbf{U} , potential temperature θ , specific humidity q , and density ρ) and surface ocean state (SST and surface ocean current \mathbf{U}_o).⁷ The air at the ocean surface is assumed to be saturated at a specific humidity $q_s(SST)$ that is a function of the SST. The vapor pressure over seawater is about 2% less than that over fresh water (see equation (5) in Large and Yeager, 2004). This difference is not negligible compared to the 20% subsaturation of marine air that drives evaporation. Consequently, the effect is included in the $q_s(SST)$ function used by all models in the present comparison. A different function of surface temperature is used over sea-ice.

All models in this paper use the same bulk formulae and atmospheric state in a two step process. First, the meteorological data are referenced to a

⁶ The GFDL-MOM simulations incorrectly included the sea ice melt as part of the normalisation. It should not be included, since it is only the net water in the ocean plus ice systems that should be normalised to have a net zero sum.

⁷ The SST and ocean current are taken from the surface ocean model grid cell values. This is an appropriate choice, since the bulk formulae presented here are developed from fits to bulk properties of the upper ocean, rather than skin properties.

EXPERIMENT	SALT/WATER FLUX	V_{piston} (m d ⁻¹)	RESTORE REGION	UNDER ICE	NORMAL HYDROLOGY
NCAR-POP	salt	50/(4*365)	global	yes	yes
FSU-HYCOM	salt	50/(4*365)	global	yes	no
GFDL-MOM-A	water	50/(4*365)	global	no	yes
GFDL-MOM-B (control)	water	50/(300)	global	no	yes
GFDL-HIM	water	50/(300)	global	yes	yes
KNMI-MICOM	salt	regional	regional	no	no
MPI-A (control)	water	50/(4*365)	global	fraction	yes
MPI-B	water	50/(300)	global	fraction	yes
Kiel-ORCA	water	50/(300)	regional	yes	yes

Table 3

Summary of choices made for the surface water/salt forcing in the various experiments. All experiments were run for 500 years. Two experiments are mentioned for GFDL-MOM and MPI, corresponding to the perturbation experiments described in Section 16. The control experiments for both models are noted, as these are the settings used for all other sections. Column two indicates whether the experiment used salt fluxes or water fluxes for hydrological forcing. Column three gives the piston velocity in metres per day, which determines the strength of the restoring used for surface salinity/water forcing. Column four notes the region over which the salinity/water restoring fluxes are set. The Kiel-ORCA simulation used regionally varying restoring as discussed in Figure 27. The KNMI-MICOM simulations used the following regionally varying piston velocity: zero piston velocity in the region 35°N-65°N, 275°E and 0°E (North Atlantic); 50m/50days north of 75°N; 50m/30days south of 50°S; and 50m/1500days for the remainder. Gradients between the different zones were linearly interpolated in a 10 degree strip. Column five indicates whether restoring was applied under sea ice. The “fraction” entry for MPI indicates the SSS restoring term was weighted by $(1 - \text{ice concentration})$, so that no restoring was applied with 100% ice. Column six refers to whether the experiment normalised the precipitation minus evaporation plus runoff in some manner to reduce drift.

common reference height. The bulk transfer coefficients are formulated at 10m height and neutral stability, so the second step is to shift them to the atmospheric stability and the reference height. This step is often overlooked when forcing ocean and ocean-ice models, where sometimes data from different heights (e.g. 10m winds and 2m temperature and humidity from reanalyses) are used directly with the 10m bulk formulae. For accuracy, it is essential to use a common height, and the height shifting and flux algorithms should be consistent, because they are closely related. The ocean state is provided by the individual ocean models.

A reference height of $\delta = 10\text{m}$ was chosen for CORE-I, so that there would be no need for any height shift of either the wind data, or the transfer coefficients. The shift of air temperature and humidity to 10m is done off-line using observed SST, so that the CORE atmospheric state is given at 10m (Table 4), where the dimensionless atmospheric stability is denoted as ζ_{10} . The turbulent fluxes are then given by

$$\tau = \rho C_D(\zeta_{10}) |\Delta\mathbf{U}| \Delta\mathbf{U} \quad (13)$$

$$E = \rho C_E(\zeta_{10}) [q - q_s(SST)] |\Delta\mathbf{U}| \quad (14)$$

$$Q_E = \Lambda_v E \quad (15)$$

$$Q_H = \rho c_p C_H(\zeta_{10}) (\theta - SST) |\Delta\mathbf{U}|, \quad (16)$$

where $c_p \approx 1000.5 \text{ J/kg}$ is the specific heat capacity of air, $\Lambda_v = 2.5 \times 10^6 \text{ J/kg}$ is the latent heat of vaporization for water, and $\Delta\mathbf{U} = \mathbf{U} - \mathbf{U}_o$ is the difference between the atmospheric winds and ocean current (following Pacanowski (1987)).⁸

Large and Yeager (2004) detail the methodology for shifting the neutral 10m transfer coefficients for drag, C_D , sensible heat transfer, C_H , and evaporation, C_E , to the stability of the coefficients in these equations, where

$$\zeta_{10} = \frac{\delta \kappa g \sqrt{\rho}}{|\tau|^{3/2}} \left(\frac{Q_H}{c_p \theta} + \frac{Q_E}{\Lambda_v (q + .608^{-1})} \right), \quad (17)$$

In this relation, $\delta = 10\text{m}$ is the atmospheric reference height, $\kappa = 0.4$ is the von Karman constant, g is the gravitational acceleration, and

$$\frac{M_{\text{dry air}}}{M_{\text{water}}} - 1 = .608, \quad (18)$$

with $M_{\text{dry air}} = 28.97$ the molecular weight of dry air, and $M_{\text{water}} = 18.015$ the molecular weight of fresh water. The stability shift is accomplished according to

⁸ Pacanowski (1987) noted that it is important to take into account surface ocean currents in the computation of wind stress, especially where ocean currents reach large values, such as in the equatorial Pacific. Models in this study all follow this approach for computing the wind stress.

$$C_D(\zeta_{10}) = C_D \left(1 - \frac{\sqrt{C_D}}{\kappa} \psi_m(\zeta_{10}) \right)^{-2} \quad (19)$$

$$C_H(\zeta_{10}) = C_H \sqrt{\frac{C_D(\zeta_{10})}{C_D}} \left(1 - \frac{C_H}{\kappa \sqrt{C_D}} \psi_h(\zeta_{10}) \right)^{-1} \quad (20)$$

$$C_E(\zeta_{10}) = C_E \sqrt{\frac{C_D(\zeta_{10})}{C_D}} \left(1 - \frac{C_E}{\kappa \sqrt{C_D}} \psi_h(\zeta_{10}) \right)^{-1}, \quad (21)$$

where ψ_m and ψ_h are empirical dimensionless momentum and heat flux profiles which are functions of stability with the following behaviour

$$\psi_h = -5 \zeta_{10} \quad \text{with } \zeta_{10} \geq 0 \text{ (stable)} \quad (22)$$

$$\psi_m = -5 \zeta_{10} \quad \text{with } \zeta_{10} \geq 0 \text{ (stable)} \quad (23)$$

$$\psi_h > \psi_m > 0 \quad \text{with } \zeta_{10} < 0 \text{ (unstable)}. \quad (24)$$

Unfortunately, only the NCAR-POP and FSU-HYCOM fluxes were computed with the correct stability shift. The code used by the other models computed incorrect transfer coefficients at ζ_{10} for both heat, C_H^* , and moisture, C_E^* , where the fractional error is given by

$$\frac{C_H^*}{C_H(\zeta_{10})} = \frac{1 - \frac{\sqrt{C_D}}{\kappa} \psi_m(\zeta_{10})}{1 - \frac{C_H}{\kappa \sqrt{C_D}} \psi_h(\zeta_{10})} \quad (25)$$

$$\frac{C_E^*}{C_E(\zeta_{10})} = \frac{1 - \frac{\sqrt{C_D}}{\kappa} \psi_m(\zeta_{10})}{1 - \frac{C_E}{\kappa \sqrt{C_D}} \psi_h(\zeta_{10})}. \quad (26)$$

Most of the ocean is cooled by negative Q_H and Q_E , which makes ζ_{10} negative, the atmosphere unstable and ψ_h more positive than ψ_m . Therefore, the fractional error is usually greater than unity, and largest in regions where high SST and low wind speeds combine to make a significantly unstable atmosphere, with ψ_h typically twice ψ_m . The heat flux error is greatest where these regions coincide with large latent heat flux loss, with the error increasing the loss. For example, using observed SST, the net heat flux into the ocean is reduced by 8 to 10 W m^{-2} over much of the tropical Indian and western Pacific oceans, over the western boundary currents with the exception of the Brazil current, and off western Australia (Laurent Brodeau, personal communication, 2007). The expected SST response would be for NCAR-POP and FSU-HYCOM to be warmer in these regions by at most 0.3°C . Elsewhere, the effect is smaller and poleward of about 40° degrees latitude, the heat flux reduction is typically less than 3 W m^{-2} .

C2. The Large and Yeager (2004) dataset

Table 4 summarises the Large and Yeager (2004) dataset used to compute the normal year forcing for heat, moisture, and momentum. This data includes monthly mean precipitation, daily short and long solar radiation, six hourly atmospheric meteorological state at 10m (temperature, humidity, and vector wind), and annual mean river runoff. The atmospheric state as well as the model's prognostic SST and surface currents determine the turbulent momentum (wind stress), turbulent heat (sensible and latent), and turbulent moisture (evaporation) fluxes. These fluxes are computed at each ocean-ice coupling time step using the CCSM bulk formulae described in Large and Yeager (2004).

C3. River runoff

The river runoff from Large and Yeager (2004) has a single time step as it represents annual mean runoff. This data has been spread out from the river mouths in a manner used in CCSM climate simulations. This approach is thought to account for some unresolved mixing that occurs at river mouths in Nature. A remapping scheme is provided at the GFDL web site quoted in Section 4.1. This scheme maps the river data onto a model grid of differing resolution, so long as the new grid is logically rectangular (as all grids used in this study).

Modellers are encouraged to test different runoff specifications, such as with a seasonal cycle as in (Röske, 2005). If doing so, we recommend that a correction be made to keep the total annual flux of runoff similar to the value in the Large and Yeager (2004) dataset in order to facilitate comparisons.

C4. Normalisation of moisture fluxes

As detailed in Section B3, some models apply a normalisation to maintain a balance in the globally integrated moisture flux (precipitation minus evaporation plus river runoff plus restoring) equal to zero. This normalisation reduce long term drifts. We summarise the choices made for the simulations in Table 3.

FIELD	AVAILABLE TIMES	USED TO COMPUTE
runoff	annual mean	river runoff into ocean
precipitation	12 months	precipitation
downwelling shortwave	365 days	ocean shortwave heating
downwelling longwave	365 days	ocean longwave heating
10m air temp	six hourly for 365 days	sensible heating (using air temp, SST, and bulk formulae)
10m humidity	six hourly for 365 days	evaporation and latent heating (assuming saturation over ocean water)
10m vector winds	six hourly for 365 days	wind stress (given vector winds, surface currents, and bulk formulae)
10m vector winds	six hourly for 365 days	sensible and latent heating, and evaporation (along with SST)
sea level pressure	six hourly for 365 days	atmospheric loading on ocean and surface atmospheric density

Table 4

Summary of the forcing used in the CORE-I simulations as provided by Large and Yeager (2004) fields and CCSM bulk formulae. Column 1 is the field provided by their dataset, column 2 is the temporal amount of data provided, column 3 is the quantity computed for use in the ocean-ice model forcing. Note that vector winds are listed twice due to their use in computing a large number of surface fluxes.

C5. Use of sea level pressure

The sea level pressure is used to compute the air density (see equation (24) of Large and Yeager, 2004). It also presents a load on the ocean-ice models (the *inverse barometer* effect), which drives barotropic motion. However, all ocean-ice models in this study discard this pressure loading, thus considering the atmosphere to be massless for purposes of ocean-ice dynamics.

C6. High frequency meteorological data

It is desirable to use high frequency (6 hourly) meteorological data as provided by Large and Yeager (2004). A one month run of an AMIP model was used at GFDL to explore the flux errors associated with averaged meteorological inputs. With daily winds, temperatures, and humidities, latent heat fluxes are under estimated broadly over the winter storm track band by some tens of $W m^{-2}$. There was also a smaller underestimate located in the summer storm track band. Experiments that refined the temporal resolution of the flux inputs individually showed that high frequency winds are most important for reducing the error, but temperature and humidity frequency also contribute. When all inputs are given at 6 hourly frequency, the global RMS error is about $1 W m^{-2}$ versus near $8 W m^{-2}$ for daily inputs.

C7. Treatment of saltwater vapor pressure

Models here use the same treatment of saltwater vapor pressure. The vapor pressure over seawater is about 2% less than that over fresh water (see equation (5) in Large and Yeager, 2004). This difference is not negligible compared to the 20% subsaturation of marine air that drives evaporation. Consequently, the effect should be included in all models participating in a comparison.

C8. Shortwave penetration

The evolution of SST depends critically on the absorption of solar shortwave radiation and demands that this flux be separated from the other heat fluxes (Rosati and Miyakoda, 1988). How this radiation is attenuated into the ocean differs in the models. As described in Appendix B to Danabasoglu et al. (2006), shortwave absorption in NCAR-POP and FSU-HYCOM is based on monthly and spatially varying SEAWIFS chlorophyll climatology following Ohlmann (2003). Similarly, GFDL-MOM and GFDL-HIM prescribe the chlorophyll climatology of Sweeney et al. (2005) and attenuates vertically according to Morel and Antoine (1994). KNMI-MICOM places all of its shortwave radiation into the bulk mixed layer. MPI and Kiel-ORCA use an e-folding dependency for the shortwave penetration as documented in Paulson and Simpson (1977), with clear water (type I in Jerlov (1968) classification). We summarise these choices in Table 5.

EXPERIMENT	WATER TYPE FOR SHORTWAVE
NCAR-POP	chlorophyll monthly climatology
FSU-HYCOM	chlorophyll monthly climatology
GFDL-MOM	chlorophyll monthly climatology
GFDL-HIM	chlorophyll monthly climatology
KNMI-MICOM	shortwave in mixed layer
MPI	uniform Jerlov water type I
Kiel-ORCA	uniform Jerlov water type I

Table 5

Table indicating the manner that shortwave radiation is attenuated as it penetrates into the ocean.

Appendix D: Diagnostics used to evaluate the simulations

The purpose of this section is to summarise the diagnostics used to evaluate the model simulations. Some diagnostics used in this paper are compared to observations, and so represent *metrics* used to gauge the physical integrity of the simulations. Others are juxtaposed between the different simulations to highlight various features of the models. Following a list of diagnostics, we present some fundamentals useful for the computation of certain diagnostics.

All models in this study employ a generalized orthogonal grid in the horizontal. As noted in Appendix A, four of the models employ tripolar grids (GFDL-MOM, GFDL-HIM, KNMI-MICOM, and Kiel-ORCA), whereas FSU-HYCOM, NCAR-POP and MPI use a bipolar grid with the North Pole displaced over Greenland. Generalized horizontal coordinates are commonly chosen for global models as they remove the spherical coordinate singularity from the Arctic Ocean, and thus allow for improved simulation integrity in this region. Unfortunately, given the relative immaturity of software plotting packages, non-spherical grids lead to difficulties directly comparing the simulations (e.g., making difference maps) in regions where the grid metric factors do not agree, such as the Arctic. We do not attempt here to remedy this shortcoming, but instead choose to allow each model to plot results on their own native grid.

D1. List of diagnostics

To facilitate contributions from research groups employing the CORE-I experimental design, we summarise the diagnostics used in this paper.

- Figures 3 and 4: Time series of the globally aver-

aged annual mean temperature and salinity over the 500 year simulations. The initial condition for the simulation is taken from Conkright et al. (2002) for the World Ocean outside the Arctic, and Steele et al. (2001) for the Arctic.

- Figures 5 and 6: Globally averaged drift of the annual mean temperature and salinity as a function of depth and time over the 500 year simulations. The initial condition for the simulation is taken from Conkright et al. (2002) for the World Ocean outside the Arctic, and Steele et al. (2001) for the Arctic.
- Figure 7: Anomalous SST for years 491-500 from the simulations relative to the analysis of Conkright et al. (2002) outside the Arctic, and Steele et al. (2001) in the Arctic.
- Figure 8: Anomalous SSS for years 491-500 from the simulations relative to the analysis of Conkright et al. (2002) outside the Arctic, and Steele et al. (2001) in the Arctic.
- Figure 9: Monthly values of the heat content (vertically integrated temperature) over the upper 250m versus sea surface temperature taken at Ocean Weathership Echo ($48^{\circ}W, 35^{\circ}N$). Results from simulated year 50 are compared.
- Figure 10: 500 year time series for the annual mean sea ice area in the Northern Hemisphere and Southern Hemisphere.
- Figures 11 and 12: Maps of the sea ice concentration (area sea ice per grid cell area) in March and September averaged over years 491-500.
- Figure 13: Time mean of year 491-500 temperature on the equator in the Pacific.
- Figure 14: Time mean of year 491-500 zonal current on the equator in the Pacific.
- Figure 15: Maximum mixed layer depth within years 491-500.
- Figures 16 and 17: Anomalous zonal mean decadal mean (years 491-500) potential temperature and

salinity for the simulations relative to the analysis of Conkright et al. (2002) for all but the Arctic, and Steele et al. (2001) for the Arctic.

- Figure 22: Northward heat transport directly computed in the models for the global ocean averaged over years 491-500.
- Figure 18: 500 year time series for the annual mean vertically integrated transport of seawater volume through the Drake Passage
- Figure 25: 500 year time series of the annual mean Atlantic meridional overturning streamfunction index. The index is computed as the maximum Atlantic MOC streamfunction at $45^\circ N$ in the region beneath the wind driven Ekman layer.
- Figure 23: The Atlantic basin meridional overturning streamfunction, time averaged over years 491-500. The vertical coordinate can be chosen according to the vertical coordinate of the model.
- Figure 24: The global meridional overturning streamfunction, time averaged over years 491-500. The vertical coordinate can be chosen according to the vertical coordinate of the model.

D2. Meridional overturning streamfunction

This section supplements the results from Section 15, where we presented the meridional overturning streamfunctions for the Atlantic and World Oceans. The meridional overturning streamfunction diagnoses the transport of volume poleward

$$\Psi(y, z) = - \int dx \int_{-H}^z dz' (v + v^*). \quad (27)$$

It is commonly used to summarise various features of the large scale circulation, particularly effects from thermohaline forcing. In equation (27), the vertical limits extend from the ocean bottom at $z = -H(x, y)$ to a depth z , and the zonal integral extends over the globe or within a closed ocean basin. We include a possibly nonzero divergence-free sub-grid scale (SGS) velocity v^* . For the geopotential coordinate models used in this study, v^* is determined by the parameterisation of Gent et al. (1995), where

$$v^* = -\partial_z (\kappa S_y), \quad (28)$$

with $S_y = -\rho_{,y}/\rho_{,z}$ the meridional slope of the neutral direction, and $\kappa > 0$ a diffusivity (with units of squared length per time). The overturning streamfunction is thus given by the SGS plus Eulerian terms

$$\Psi(y, z) = \int dx \kappa S_y - \int dx \int_{-H}^z dz' v, \quad (29)$$

where we set $\kappa S_y = 0$ at the ocean bottom. The contribution from the Gent et al. (1995) parameterisation to both volume and tracer transport is most prominent in the Southern Ocean, where the basin is large and the neutral slopes generally steepen towards the pole. In other basins, such as the North Atlantic, its contribution is modest to negligible, depending on the chosen diffusivity.

For the isopycnal models employed by GFDL-HIM and KNMI-MICOM, the meridional overturning streamfunction is computed with the vertical integral proceeding from the ocean bottom upwards to a surface of constant potential density, so that

$$\Psi(y, \rho) = - \int dx \int_{-H}^{\rho} d\rho' v (\partial z / \partial \rho'), \quad (30)$$

where $dz = d\rho' (\partial z / \partial \rho')$ is the thickness between isopycnal layers. Since the product $v (\partial z / \partial \rho')$ is non-linear, it is important for isopycnal models to retain this product over each time step when forming the time average streamfunction. Otherwise, the resulting streamfunction could suffer from aliasing problems.

Beneath the ocean surface, the no-normal flow condition on the velocity field implies that the overturning streamfunction is a constant along land-sea boundaries, and choice of the reference streamfunction value as chosen above leads to a vanishing streamfunction on these boundaries. In a model run without water flux forcing, the streamfunction vanishes along the surface as well. In contrast, fresh water input means that the streamfunction is generally nonzero at the ocean surface.

D3. Poleward heat transport

For the ocean-ice system, the prescribed atmospheric state used in CORE implies a poleward heat transport, given a sea surface temperature and a few assumptions. We summarise the formulation in order to highlight assumptions. For this purpose, start with the vertically integrated temperature budget for a Boussinesq ocean

$$\partial_t \left(\int_{-H}^{\eta} dz \theta \right) = -\nabla \cdot \left(\int_{-H}^{\eta} dz (\mathbf{u} \theta + \mathbf{F}) \right) + Q_{\text{heat}} / (\rho_o C_p). \quad (31)$$

The left hand side represents transient storage of temperature in a column of ocean fluid from the surface at $z = \eta$ to the bottom at $z = -H$. This storage is affected by the convergence of heat through the vertically integrated resolved horizontal flow, $\mathbf{u}\theta$, and parameterised SGS processes, \mathbf{F} . For each of the models in this study, the SGS parameterisations include the neutral diffusion scheme from Redi (1982) and the eddy induced advection from Gent and McWilliams (1990).

The storage term in equation (31) is also affected by heat transfer across the ocean surface

$$Q_{\text{heat}} = Q_{\text{surf}} + Q_{\text{geo}} + \rho_o C_p \sum_{P,E,R} q_w \theta_w. \quad (32)$$

Q_{surf} is the surface heat flux (positive into the ocean) arising from interactions between the ocean and atmosphere, and ocean and sea ice

$$Q_{\text{surf}} = Q_{\text{air-sea}} + Q_{\text{ice-sea}}. \quad (33)$$

These fluxes account for shortwave and longwave solar radiation, and turbulent fluxes from sensible and latent heating. Additional ice-sea fluxes arise from the interaction between liquid seawater and frozen sea ice. For example, sea ice forms as liquid seawater is cooled below its freezing point. The cooling that would otherwise contribute to a sub-freezing liquid seawater is generally transferred to the sea ice model as frazil is formed. This transfer represents a sign definite heating of the liquid ocean, and it thus appears as part of the net heating in equation (32). Conversely, when frozen precipitation or calving land glaciers enter the liquid ocean, the latent heat of fusion required to melt the frozen water acts as a cooling for the liquid seawater.

The solid earth presents a heat source to the ocean which is included in the geothermal heating term. This heating is typically set to zero, but is retained by the Kiel-ORCA simulation, with a uniform value of 0.086 W m^{-2} applied throughout the World Ocean.

The final contribution to the heat flux in equation (32) arises from the transfer across the ocean surface of water with a temperature θ_w . This transfer occurs with a volume per area per time precipitation ($P > 0$ for water entering ocean), evaporation ($E > 0$ for water leaving ocean), and river runoff ($R > 0$ for water entering ocean). This water heat flux is absent in virtual salt flux simulations (see Section B3 and Table 3), since these models do not exchange water with other parts of the climate system. For the models in this study using real water fluxes, θ_w for $P, E,$

and R is equated to the SST of the grid cell participating in the water transfer.⁹ For precipitation and evaporation, the term $\theta_w (P - E)$ generally provides a global net cooling of the ocean, since evaporation transfers water away from the ocean at a temperature typically higher than precipitation adds water. This net heat loss must be compensated by ocean heat transfer.

If the ocean is in a steady state (say over annual or decadal time scales), then heating over these time scales will balance the divergence of heat throughout an ocean column. Now extend this balance to a horizontal domain, in which the zonal boundaries are either periodic or no-flow, and the southern boundary at $y = y_s$ is no-flow (as for the Antarctic continent). Use of the Divergence Theorem indicates that meridional transport through the northern boundary balances heat input through the ocean surface

$$\rho_o C_p \int dx \int_{-H}^{\eta} dz (v\theta + F^y) = \int_{y_s}^y dy \int dx Q_{\text{heat}}. \quad (34)$$

That is, in a steady state, the area integrated heat flux into the ocean over the region south of a given latitude (right hand side) is balanced by meridional transport of heat within the ocean at the specific latitude (left hand side). For example, if there is a net surface flux of heat out of the ocean south of a chosen latitude (e.g., cold air over warmer water), then maintenance of a steady state requires that this surface heat loss is balanced by a transport of heat within the ocean towards the south across this latitude. A meridional ocean heat transport is thus *implied* by the net surface forcing. This result allows us to estimate the meridional ocean heat transport through knowledge of just the surface fluxes.

As described by Large and Yeager (2004), knowledge of the atmospheric state and the sea surface temperature provide sufficient information to compute surface turbulent air-sea heat fluxes by using the bulk formulae. The radiative and turbulent heating then provide $Q_{\text{air-sea}}$. The water transport term $(P - E + R)\theta_w$ is generally subdominant to Q_{surf} . So it is common to estimate the implied ocean heat

⁹ This choice is suggested for precipitation in state-of-the-art coupled climate models, since the condensate in atmospheric models generally has zero heat capacity. Likewise, many river models do not carry the temperature of river water. There is a non-conservation of heat in the simulated global climate system due to these limitations of the atmosphere and river components in coupled climate models.

transport just with $Q_{\text{air-sea}}$. Hence, estimates of the ocean heat transport implied from $Q_{\text{air-sea}}$ are useful if the following assumptions hold:

- ocean heat storage is small over the time scales of interest (i.e., ocean has reached a steady state);
- heat transport due to the movement of water across the ocean surface is ignored;
- heat fluxes between ice-sea are ignored;
- bottom heat fluxes (e.g., geothermal fluxes) are ignored.

These assumptions are made for the implied transports for the Normal Year Forcing shown in Figure 21. Note that the missing $Q_{\text{ice-sea}}$ fluxes generally give more cooling toward the poles, implying more poleward transport (steeper negative slopes) in both hemispheres than shown in Figure 21.

D4. Comments on diagnosing heat and volume transport

As the ocean heat transport is a nonlinear product of terms, it is desirable to compute the transport in a simulation at each time step using the model numerics, and maintain a time average online during the simulation. In table 6, we note how the transport was computed by the models. For the overturning streamfunction (see Section D2), the streamfunction is a nonlinear product when the thickness of a layer can change, as for a nonlinear free surface z-model (GFDL-MOM) and for an isopycnal or hybrid model (GFDL-HIM, FSU-HYCOM, KNMI-MICOM).

The problem with using the mean velocity times the mean thickness for the overturning streamfunction, or mean velocity times the mean thickness times mean temperature for the heat transport, is that there tend to be very large correlations between thickness and the other two fields (especially near-surface). So working with time mean fields would be a very poor approximation to the true fluxes.

In an isopycnal model, one complication is that the near-surface layers are non-isopycnal, and hence do not trivially map into the discretized density space. GFDL-HIM uses the layer density over a short interval (an hour to a few days) to partition the fluxes in these variable-density layers into the discrete density space matching the target densities for the interior layers. The total transport is exactly what the model uses, but there is a slight dependence on the frequency with which the fluxes are recast into pure density coordinates, and in the choice of the lateral mapping of densities onto velocity points.

Because for efficiency these are not calculated every time-step, GFDL-HIM cannot simply use the ratio of the buoyancy flux to the volume flux to determine the buoyancy at which the fluxes occur. With the 2-hour frequency used in the GFDL-HIM CORE runs, it is expected that these uncertainties are negligible.

Finally, the transport arising from SGS mesoscale parameterisation via neutral diffusion (Redi, 1982) and eddy induced advection (Gent and McWilliams, 1990) play a role in establishing the heat transport and meridional overturning streamfunction, especially in the Southern Ocean. Table 6 thus provides details regarding how the models chose the parameters setting the strength of these schemes.

D5. Budget for volume averaged temperature

Integrating the heat budget (31) over the full extent of the World Ocean leads to an evolution equation for the globally integrated potential temperature

$$\rho_o C_p \partial_t \left(\sum \theta dV \right) = \sum Q_{\text{heat}} dA. \quad (35)$$

In this equation, dV is the grid cell volume, dA is the horizontal area of a cell. Locally, the surface turbulent and radiative fluxes tend to dominate heating from water transfer by one or two orders of magnitude, with frazil yet another order of magnitude smaller. The turbulent and radiative terms, as well as the water term, have both positive and negative contributions, whereas frazil always represents a heating of the liquid ocean. Hence, the global sum of these terms can be comparable and thus are important to retain in a budget analysis.

Note that models using real water fluxes have a seawater volume that is generally nonconstant, given the ability to exchange water between other components of the climate system. These models, however, generally implement a normalisation term similar to the salinity normalisation (Section B3 and Table 3), thus minimising the drift. Absent this normalisation, drifts in volume averaged temperature and salinity may result from drifts in seawater volume.

One means of checking whether the model results are sensible is to consider a rough order of magnitude calculation to determine the temperature trend that one may expect, given a net heat flux. Letting $\langle \theta \rangle = \sum \theta dV / \sum dV$, equation (35) indicates that

EXPERIMENT	MOC TIME AVERAGE	HEAT TIME AVERAGE	GM	NEUTRAL	MAX SLOPE
NCAR-POP	instant	instant	600	600	3/10
FSU-HYCOM	daily	daily	$0.05 (\Delta)^3$	$0.05 (\Delta)^3$	–
GFDL-MOM	instant	instant	$100 \leq \kappa \leq 600$	600	1/500
GFDL-HIM	instant	instant	600	600	–
KNMI-MICOM	instant	monthly	0.005Δ	0.005Δ	–
MPI	instant	instant	0.0025Δ	0.0025Δ	$\Delta z / \Delta$
Kiel-ORCA	instant	instant	$100 \leq \kappa \leq 3000$	1000*	1/100

Table 6

Table indicating the manner that the models diagnose the decadal mean heat transport and meridional overturning streamfunction. “Instantaneous” implies that the fields are computed using each time step of the model simulation, then time averaged. “Daily” means that daily mean fields were used, with nonlinear products computed from the dailies. The same holds for “monthly.” The models include SGS contributions from Redi (1982) neutral diffusion and eddy induced advection from Gent and McWilliams (1990). These schemes play an important role in setting the overturning streamfunction and the heat transport, especially in the Southern Ocean. Hence, we include in this table some information about the implementation of these SGS schemes. The GM diffusivity and neutral diffusivity are given in units of $\text{m}^2 \text{sec}^{-1}$. The maximum neutral slope is noted for the geopotential models, where tapering of the neutral physics fluxes is applied for regions of greater neutral slopes. For the isopycnal models, there is no maximum slope parameter. The following provides more details for certain models.

FSU-HYCOM: Interface height smoothing (corresponding to Gent and McWilliams (1990)) is applied through a biharmonic operator, with a mixing coefficient determined by the grid spacing Δ (in metres) times a constant velocity scale 0.05 m s^{-1} . For regions where the FSU-HYCOM simulation has coordinate surfaces aligned with constant pressure (mostly in the upper ocean mixed layer), Gent and McWilliams (1990) is not implemented, and lateral diffusion is oriented along pressure surfaces rather than rotated to neutral directions.

GFDL-MOM: A range of GM diffusivity is used with value determined according to the vertically integrated baroclinicity (Griffies et al., 2005).

KNMI-MICOM: This model reports implied transports from the time integrated net heat flux. It uses a grid scale dependent diffusivity applied with a Laplacian operator, using a velocity scale of 0.005 m s^{-1} .

MPI: Monthly mean vertical velocities are used to diagnose the MOC as a way to include sea level in the diagnostic. The heat transport is computed from the implied transport taken from the time integrated net surface heat flux. It uses a grid scale dependent diffusivity applied with a Laplacian operator, using a velocity scale of 0.0025 m s^{-1} . The maximum neutral slope is set by the grid ratio $\Delta z / \Delta$, with Δ the appropriate horizontal grid spacing.

Kiel-ORCA: GM diffusivity is set according to the first baroclinic Rossby radius, and its neutral diffusivity is scaled according to the grid spacing, with $1000 \text{ m}^2 \text{ sec}^{-1}$ the nominal value.

the global mean liquid ocean temperature evolves according to

$$(V \rho_o C_p) \partial_t \langle \theta \rangle = \sum Q_{\text{heat}} dA, \quad (36)$$

where V is the liquid ocean volume. Typical values for the World Ocean yield $V \rho_o C_p \approx 5.4 \times 10^{24} \text{ J}/^\circ\text{C}$ and $\sum dA = 3.6 \times 10^{14} \text{ m}^2$, leading to the decadal scale temperature trend

$$\frac{\Delta \langle \theta \rangle}{\text{decade}} \approx 0.02 Q_{\text{heat}}. \quad (37)$$

For example, with a 1 W m^{-2} net heating of the ocean over the course of a decade, we expect a temperature trend of roughly 0.02°C per decade, or 0.2°C per century.

D6. Budget for volume averaged salinity

As a supplement to the results in Section 5, we note that the globally integrated salinity in the liq-

uid ocean evolves according to

$$\partial_t \left(\sum S dV \right) = \sum Q_{\text{salt}} dA, \quad (38)$$

where

$$Q_{\text{salt}} = Q_{\text{ice}} + v_{\text{piston}} (SSS - S_{\text{obs}}) + S_o (P - E + R) + \text{normal} \quad (39)$$

is the salt flux. Here, Q_{ice} is the flux of salt passing between the liquid ocean and sea ice, and v_{piston} is the piston velocity setting the strength of the sea surface salinity (SSS) restoring to the observed values S_{obs} . The term $S_o (P - E + R)$ arises from the conversion of water flux to a salt flux, with S_o a global constant. The normalisation term is used to minimise the salt drift, which would generally occur in its absence given that the globally integrated restoring term need not be zero. The restoring term, the virtual salt flux term, and the normalisation term are absent in models that use a real water flux. Instead, these models convert the restoring salt

fluxes into an implied water flux, they insert water directly to the model, thus changing its volume, and they introduce a normalisation for the surface water forcing. Further details of this budget, as well as the choices made by the various models, are provided in Section B3 and Table 3.

D7. Mixed layer depths

The purpose of this section is to supplement the mixed layer diagnostics presented in Section 11. We start by noting that the diagnosis of observed and model mixed layer depth is ambiguous. The mixed layer has near-zero gradients of $\theta - S$ and density, as well as tracers such as CFCs. So most techniques to estimate the MLD rely on either a threshold gradient or a threshold change in one of these quantities, normally in potential temperature θ or density (see, for example Lorabacher et al., 2006; De Boyer Montegut et al., 2004; Monterey and Levitus, 1997). Relying solely on θ has the advantage of good observational data coverage, but this approach neglects salinity stratification associated with barrier layers (see e.g., Sprintall and Tomczak, 1992). In contrast, relying solely on density overlooks density-compensating changes in $\theta - S$, exaggerating the mixed layer depth. Given these caveats, with water-mass formation regulated by upper ocean stratification and surface water overturn, an assessment of model MLD is extremely important.

There are two general classes of definitions for the diagnosed mixed layer used by the models in this study. The first is based on meeting a buoyancy criteria, where the vertical buoyancy derivative exceeds some critical value (as is used in the observed estimate of Figure 15a). The second definition equates the mixed layer depth to the depth over which vigorous turbulent mixing from surface wind and buoyancy forcing occurs. The following provides more details of this diagnostic as used in each model. Note that once a mixed layer depth is determined within a particular model, the models all then report the maximum mixed layer depth attained for the years 491-500. In all models this is based on the maximum monthly mean value attained, except in NCAR-POP, wherein the maximum instantaneous MLD is reported. These depths are presented in Figure 15.

- NCAR-POP (buoyancy criteria): The mixed layer depth at a particular horizontal point is the shallowest depth where the interpolated verti-

cal derivative of buoyancy equals a local critical buoyancy gradient (Large et al., 1997). This critical gradient is just the maximum of all the discrete gradients given by surface minus model level buoyancy, divided by level depth.

- FSU-HYCOM (buoyancy criteria): HYCOM specifies an equivalent temperature jump of $\Delta\theta = 0.1^\circ\text{C}$, and an associated density jump $\Delta\sigma = -\alpha\Delta\theta$, with α the local value for $\partial\rho/\partial\theta$. Given this jump in density, the algorithm then computes the potential density difference between two adjacent layers starting from the ocean surface. Whenever this adjacent density difference is greater than the pre-defined density jump $\Delta\sigma$, an interpolation is found to determine the depth where the difference equals to the jump, which then defines the mixed layer depth.
- GFDL-MOM (buoyancy criteria): The mixed layer depths are based on an interpolation to find the first depth where the difference in buoyancy relative to the surface is greater than 0.0003m s^{-2} . This approach follows that described in Levitus (1982), and amounts to that used in NCAR-POP, yet with a constant critical buoyancy difference used for all depth levels.
- GFDL-HIM (turbulent mixing criteria): The mixed layer depths are defined as the thickness of the layer with vigorous turbulence, defined as the depth over which energy input or lost by wind and buoyancy forcing is dissipated.
- KNMI-MICOM (turbulent mixing criteria): The mixed layer depths are defined as the thickness of the top layer (bulk mixed layer) which is not isopycnic.
- MPI (buoyancy criteria): Mixed layer depth is determined as the depth at which the density difference from the sea surface is 0.125 sigma units, following Monterey and Levitus (1997).
- Kiel-ORCA (buoyancy criteria): Mixed layer depth is determined as the depth at which the density difference from the sea surface is 0.1 sigma units.

Appendix E: List of Acronyms

- AABW = Antarctic Bottom Water
- AAIW = Antarctic Intermediate Water
- ACC = Antarctic Circumpolar Current
- AMIP = Atmospheric Model Intercomparison Project
- CCSM = Community Climate System Model

- CFC = Chlorofluorocarbon
- CLIVAR= Climate Variability and predictability project
- CM = Climate Model
- CORE = Coordinated Ocean-ice Reference Experiment
- EUC = Equatorial Undercurrent
- ENSO = El Niño Southern Oscillation
- FSU = Florida State University
- GFDL = Geophysical Fluid Dynamics Laboratory
- GM = Gent-McWilliams parameterisation (Gent and McWilliams, 1990)
- HIM = Hallberg Isopycnal Model
- HYCOM = HYbrid Coordinate Ocean Model
- IPSL = Institut Pierre Simon Laplace
- KNMI = Royal Netherlands Meteorological Institute
- KPP = K-Profile Parameterisation
- MICOM = Miami Isopycnal Coordinate Ocean Model
- MLD = Mixed Layer Depth
- MOC = Meridional Overturning Circulation
- MOM = Modular Ocean Model
- MPI = Max Planck Institute
- NADW = North Atlantic Deep Water
- NCAR = National Center for Atmospheric Research
- NEMO = Nucleus for European Modelling of the Ocean
- NOAA = National Oceanic and Atmospheric Administration
- NYF = Normal Year Forcing from Large and Yeager (2004)
- ORCA = Ocean model configuration of the NEMO ocean code
- OMIP = Ocean Model Intercomparison Project
- OPA = Source code for NEMO
- OWS = Ocean Weather Ship
- P-E+R = Precipitation minus Evaporation plus River runoff
- POP = Parallel Ocean Program
- SAMW = Subantarctic Mode Water
- SSS = Sea Surface Salinity
- SST = Sea Surface Temperature
- WGOMD = Working Group for Ocean Model Development
- WOA = World Ocean Atlas
- WOCE = World Ocean Circulation Experiment

References

- Antonov, J., Locarnini, R., Boyer, T., Mishonov, A., Garcia, H., 2006. World Ocean Atlas 2005, volume 2: Salinity. U.S. Government Printing Office 62, NOAA Atlas NESDIS, Washington, D.C., 182 pp.
- Arakawa, A., Lamb, V. R., 1977. The UCLA general circulation model. In: Chang, J. (Ed.), *Methods in Computational Physics: General Circulation Models of the Atmosphere*. Vol. 17. Academic Press, pp. 174–265.
- Barnier, B., 1998. Forcing the ocean. In: Chassignet, E. P., Verron, J. (Eds.), *Ocean Modeling and Parameterization*. Vol. 516 of NATO ASI Mathematical and Physical Sciences Series. Kluwer, pp. 45–80.
- Barnier, B., Madec, G., Penduff, T., Molines, J., Treguier, A., Sommer, J. L., Beckmann, A., Bistoch, A., Böning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., Cuevas, B. D., 2006. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy permitting resolution. *Ocean Dynamics* 56, 543–567.
- Barnier, B., Siefridt, L., Marchesiello, P., 1995. Thermal forcing for a global ocean circulation model using a three-year climatology of ECMWF analyses. *Journal of Marine Research* 6, 363–380.
- Beljaars, A., 1994. The parameterization of surface fluxes in large-scale models under free convection. *Quarterly Journal of the Royal Meteorological Society* 121, 255–270.
- Bentsen, M., Drange, H., Furevik, T., Zhou, T., 2004. Simulated variability of the Atlantic meridional overturning circulation. *Climate Dynamics* 22, 701–720.
- Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modelling* 4, 55–88.
- Bleck, R., Rooth, C., Hu, D., Smith, L. T., 1992. Salinity-driven thermocline transients in a wind and thermohaline forced isopycnic coordinate model of the North Atlantic. *Journal of Physical Oceanography* 22, 1486–1505.
- Böning, C. W., Herrmann, P., 1994. Annual cycle of poleward heat transport in the oceans: results from high-resolution modeling of the Equatorial North Atlantic. *Journal of Physical Oceanography* 24, 91–107.
- Briegleb, B., Bitz, C., Hunke, E., Lipscomb,

- W., Holland, M., Schramm, J., Moritz, R., 2004. Scientific description of the sea ice component in the Community Climate System Model, Version 3. NCAR Technical Report No. NCAR/TN-463+STR available online at <http://www.cesm.ucar.edu>.
- Briegleb, B., Hunke, E., Bitz, C., Lipscomb, W., 2002. Description of the Community Climate System Model Version 2 Sea Ice Model. NCAR Technical Report available online at <http://www.cesm.ucar.edu>, 60pp.
- Bryan, F., 1986. High latitude salinity effects and interhemispheric thermohaline circulations. *Nature* 323, 301–304.
- Chassignet, E., Hurlburt, H., Smedstad, O., Halliwell, G., Wallcraft, A., Metzger, E., Blanton, B., Lozano, C., Rao, D., Hogan, P., Srinivasan, A., 2006. Generalized vertical coordinates for eddy-resolving global and coastal forecasting. *Oceanography* 19, 20–31.
- Chassignet, E., Smith, L., Halliwell, G., Bleck, R., 2003. North Atlantic simulation with the HYbrid Coordinate Ocean model (HYCOM): Impact of the vertical coordinate choice, reference density, and thermobaricity. *Journal of Physical Oceanography* 33, 2504–2526.
- Collins, W., Bitz, C., Blackmon, M., Bonan, G., Bretherton, C., Carton, J., Chang, P., Doney, S., Hack, J., Henderson, T., Kiehl, J., Large, W., McKenna, D., Santer, B., Smith, R., 2006. The community climate system model: CCSM3. *Journal of Climate* 19, 2122–2143.
- Comiso, J., 1999 (updated 2005). Bootstrap sea ice concentrations for NIMBUS-7 SMMR and DMSP SSM/I, June to September 2001. National Snow and Ice Data Center Digital Media.
- Conkright, M., Antonov, J., Baranova, O., Boyer, T., Garcia, H., Gelfeld, F., Johnson, D., Locarnini, R., Murphy, P., O'Brien, T., Smolyar, I., Stephens, C., 2002. World Ocean Database 2001, Volume 1: Introduction. NOAA Atlas NESDIS 42, U.S. Government Printing Office 13, NOAA, Washington, D.C., 167 pp.
- Cox, M., Bryan, K., 1984. A numerical model of the ventilated thermocline. *Journal of Physical Oceanography* 14, 674–687.
- Cunningham, S., Alderson, S., King, B., Brandon, M., 2003. Transport and variability of the Antarctic Circumpolar Current in Drake Passage. *Journal of Geophysical Research* 108, Art. 8084.
- Danabasoglu, G., Ferrari, R., McWilliams, J., 2007. Sensitivity of an ocean general circulation model to a parameterization of near-surface eddy fluxes. *Journal of Climate* submitted.
- Danabasoglu, G., Large, W. G., Tribbia, J. J., Gent, P. R., Briegleb, B. P., McWilliams, J. C., 2006. Diurnal coupling in the tropical oceans of CCSM3. *Journal of Climate* 19, 2347–2365.
- Danabasoglu, G., Marshall, J., 2007. Effects of vertical variations of thickness diffusivity in an ocean general circulation model. *Ocean Modelling* 18, 122–141.
- Danabasoglu, G., McWilliams, J., Gent, P., 1994. The role of mesoscale tracer transports in the global ocean circulation. *Science* 264, 1123–1126.
- De Boyer Montegut, C., Madec, G., Fischer, A., Lazar, A., Iudicone, D., 2004. Mixed layer depth over the global ocean: An examination of profile data and a profile based climatology. *Journal of Geophysical Research* 109, doi:10.1029/2004JC002378.
- De Verdière, C., Huck, T., 1999. Baroclinic instability: A wavemaker for oceanic interdecadal variability. *Journal of Physical Oceanography* 29, 865–892.
- Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., Cooke, W. F., Dixon, K. W., Dunne, J., Dunne, K. A., Durachta, J. W., Findell, K. L., Ginoux, P., Gnanadesikan, A., Gordon, C., Griffies, S. M., Gudgel, R., Harrison, M. J., Held, I. M., Hemler, R. S., Horowitz, L. W., Klein, S. A., Knutson, T. R., Kushner, P. J., Langenhorst, A. L., Lee, H.-C., Lin, S., Lu, L., Malyshev, S. L., Milly, P., Ramaswamy, V., Russell, J., Schwarzkopf, M. D., Shevliakova, E., Sirutis, J., Spelman, M., Stern, W. F., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, F., Zhang, R., 2006. GFDL's CM2 global coupled climate models - Part 1: Formulation and simulation characteristics. *Journal of Climate* 19, 643–674.
- Doney, S. C., Hecht, M. W., 2002. Antarctic bottom water formation and deep water chlorofluorocarbon distributions in a global ocean climate model. *Journal of Physical Oceanography* 32, 1642–1666.
- Dukowicz, J. K., Smith, R. D., 1994. Implicit free-surface method for the Bryan-Cox-Semtner ocean model. *Journal of Geophysical Research* 99, 7991–8014.
- England, M. H., 1995. The age of water and ventilation timescales in a global ocean model. *Journal of Physical Oceanography* 25, 2756–2777.
- England, M. H., Maier-Reimer, E., 2001. Using chemical tracers to assess ocean models. *Reviews of Geophysics* 39, 29–70.

- Fritzsche, B., Gerdes, R., Hiller, W., Latif, M., Legutke, S., Maier-Reimer, E., Olbers, D., Röske, F., 2000. The Ocean Model Intercomparison Project (OMIP). Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, and Max-Planck-Institute for Meteorology, Hamburg, 139 pages.
- Gaspar, P., Gregoris, Y., Lefevre, J., 1990. A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: Tests at station Papa and long-term upper ocean study site. *Journal of Geophysical Research* 95, 16179–16193.
- Gates, W., 1993. AMIP: The Atmosphere Model Intercomparison Project. *BAMS* 73, 1962–1970.
- Gent, P., Large, W., Bryan, F., 2001. What sets the mean transport through drake passage? *Journal of Geophysical Research* 106, 2693.
- Gent, P. R., McWilliams, J. C., 1990. Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography* 20, 150–155.
- Gent, P. R., Willebrand, J., McDougall, T. J., McWilliams, J. C., 1995. Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography* 25, 463–474.
- Gerdes, R., C. Köberle, 2007. Comparison of Arctic sea ice thickness variability in IPCC Climate of the 20th Century experiments and in ocean–sea ice hindcasts. *Journal of Geophysical Research* 112 (C04S13), doi:10.1029/2006JC003616.
- Gerdes, R., Griffies, S., Hurlin, W., 2005a. Reaction of the oceanic circulation to increased melt water flux from greenland - a test case for ocean general circulation models. *Clivar Exchanges* 10 (28–31).
- Gerdes, R., Hurka, J., Karcher, M., Kauker, F., Köberle, C., 2005b. Simulated history of convection in the Greenland and Labrador Seas 1948–2001. In: monograph, A. (Ed.), *The Nordic seas: An integrated perspective*. Vol. 158. American Geophysical Union, Washington, DC, pp. 221–238.
- Gerdes, R., Hurlin, W., Griffies, S., 2006. Sensitivity of a global ocean model to increased run-off from greenland. *Ocean Modelling* 12, 416–435.
- Gill, A. E., Niiler, P., 1973. The theory of the seasonal variability in the ocean. *Deep-Sea Research* 20 (9), 57–64.
- Gnanadesikan, A., Dixon, K. W., Griffies, S. M., Balaji, V., Beesley, J. A., Cooke, W. F., Delworth, T. L., Gerdes, R., Harrison, M. J., Held, I. M., Hurlin, W. J., Lee, H.-C., Liang, Z., Nong, G., Pacanowski, R. C., Rosati, A., Russell, J., Samuels, B. L., Song, S. M., Spelman, M. J., Stouffer, R. J., Sweeney, C. O., Vecchi, G., Winton, M., Wittenberg, A. T., Zeng, F., Zhang, R., 2006. GFDL’s CM2 global coupled climate models-Part 2: The baseline ocean simulation. *Journal of Climate* 19, 675–697.
- Gnanadesikan, A., Griffies, S., Samuels, B., 2007. Effects in a climate model of slope tapering in neutral physics schemes. *Ocean Modelling* 17, 1–16.
- Gordon, C., Cooper, C., Senior, C. A., Banks, H., J. M. Gregory, T. C. J., Mitchell, J. F. B., Wood, R. A., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics* 16, 147–168.
- Greatbatch, R., Peterson, K., 1996. Interdecadal variability and oceanic thermohaline adjustment. *Journal of Geophysical Research* 101, 20467–20482.
- Griffies, S. M., 2004. *Fundamentals of ocean climate models*. Princeton University Press, Princeton, USA, 518+xxxiv pages.
- Griffies, S. M., Gnanadesikan, A., Dixon, K. W., Dunne, J. P., Gerdes, R., Harrison, M. J., Rosati, A., Russell, J., Samuels, B. L., Spelman, M. J., Winton, M., Zhang, R., 2005. Formulation of an ocean model for global climate simulations. *Ocean Science* 1, 45–79.
- Griffies, S. M., Harrison, M. J., Pacanowski, R. C., Rosati, A., 2004. *A Technical Guide to MOM4*. NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, USA, 337 pp.
- Griffies, S. M., Pacanowski, R., Schmidt, M., Balaji, V., 2001. Tracer conservation with an explicit free surface method for z -coordinate ocean models. *Monthly Weather Review* 129, 1081–1098.
- Hallberg, R., 2005. A thermobaric instability in Lagrangian vertical coordinate ocean models. *Ocean Modelling* 8, 227–300.
- Hallberg, R., Gnanadesikan, A., 2006. On the role of eddies in determining the structure and response of the wind-driven southern hemisphere overturning: Results from the Modeling Eddies in the Southern Ocean (MESO) project. *Journal of Physical Oceanography* 36, 2232–2252.
- Hallberg, R. W., 2003. The suitability of large-scale ocean models for adapting parameterizations of boundary mixing and a description of a refined bulk mixed layer model. In: Müller, P., Garrett, C. (Eds.), *Near-Boundary Processes and Their Parameterization*. Proceedings of the 13th ‘Aha Hu-

- liko'a Hawaiian Winter Workshop. University of Hawaii at Manoa, pp. 187–203.
- Haltiner, G. T., Williams, R. T., 1980. *Numerical Prediction and Dynamic Meteorology*. John Wiley and Sons, New York, USA.
- Haney, R. L., 1971. Surface thermal boundary conditions for ocean circulation models. *Journal of Physical Oceanography* 1, 241–248.
- Hazeleger, W., Severijns, C., Haarsma, R., Selten, F., Sterl, A., 2003. Speedo - model description and validation of a flexible coupled model for climate studies. KNMI-Technical Report R257, 1–38.
- Hellerman, S., Rosenstein, M., 1983. Normal monthly wind stress over the World Ocean with error estimates. *Journal of Physical Oceanography* 13, 1093–1104.
- Hewitt, C., Broccoli, A., Crucifix, M., Gregory, J., Mitchell, J., Stouffer, R., 2006. The effect of a large freshwater perturbation on the glacial North Atlantic Ocean using a coupled general circulation model. *Journal of Climate* 19, 4436–4447.
- Hibler, W., 1979. A dynamic thermodynamic sea ice model. *Journal of Physical Oceanography* 9, 815–846.
- Hirst, A., McDougall, T., 1998. Meridional overturning and diapycnal transport in a z-coordinate ocean model including eddy-induced advection. *Journal of Physical Oceanography* 28, 1205–1223.
- Holloway, G., Dupont, F., Golubeva, E., Hakkinen, S., Hunke, E., Jin, M., Karcher, M., Kauker, F., Maltrud, M., Morales-Maqueda, M., Maslowski, W., Platov, G., Stark, D., Steele, M., Suzuki, T., Wang, J., Zhang, J., 2007. Water properties and circulation in Arctic Ocean models. *Journal of Geophysical Research* 112, doi:10.1029/2006JC003642.
- Hsiung, J., Newell, R. E., Houghtby, T., 1989. The annual cycle of oceanic heat storage and oceanic meridional heat transport. *Quarterly Journal of the Royal Meteorological Society* 115 (485), 1–28.
- Huang, R. X., 1993. Real freshwater flux as a natural boundary condition for the salinity balance and thermohaline circulation forced by evaporation and precipitation. *Journal of Physical Oceanography* 23, 2428–2446.
- Hurrell, J., Hack, J., Shea, D., Caron, J., Rosinski, J., 2006. A new sea surface temperature and sea ice boundary data set for the Community Atmosphere Model. *Journal of Climate* in prep.
- Jayne, S., Marotzke, J., 1999. A destabilizing thermohaline circulation-atmosphere-sea ice feedback. *Journal of Climate* 12, 642–651.
- Jerlov, N. G., 1968. *Optical Oceanography*. Elsevier.
- Johnson, G. C., Sloyan, B. M., Kessler, W. S., McTaggart, K. E., 2002. Direct measurements of upper ocean currents and water properties across the tropical Pacific during the 1990's. *Progress in Oceanography* 52, 31–36.
- Johnson, M., Gaffigan, S., Hunke, E., Gerdes, R., 2007. A comparison of Arctic Ocean sea ice concentration among the coordinated AOMIP model experiments. *Journal of Geophysical Research* 112 (C04S11), doi:10.1029/2006JC003690.
- Jungclaus, J., Keenlyside, N., Botzet, M., Haak, H., Luo, J.-J., Latif, M., Marotzke, J., Mikolajewicz, U., Roeckner, E., 2006. Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. *Journal of Climate* 19, 3952–3972.
- Karoly, D., McIntosh, P., Berrisford, P., McDougall, T. J., Hirst, A., 1997. Similarities of the Deacon cell in the Southern Ocean and Ferrel cells in the atmosphere. *Quarterly Journal of the Royal Meteorological Society* 123, 519–526.
- Killworth, P. D., Smeed, D., Nurser, A., 2000. The effects on ocean models of relaxation toward observations at the surface. *Journal of Physical Oceanography* 30, 160–174.
- Large, W., Yeager, S., 2004. Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies. NCAR Technical Note: NCAR/TN-460+STR. CGD Division of the National Center for Atmospheric Research.
- Large, W. B., 2005. Surface fluxes for practitioners of global ocean data assimilation. In: Chassignet, E. P., Verron, J. (Eds.), *Ocean Weather Forecasting: an Integrated View of Oceanography*. Vol. 577. Springer, pp. 229–270.
- Large, W. G., Danabasoglu, G., Doney, S. C., McWilliams, J. C., 1997. Sensitivity to surface forcing and boundary layer mixing in a global ocean model: annual-mean climatology. *Journal of Physical Oceanography* 27, 2418–2447.
- Large, W. G., McWilliams, J. C., Doney, S. C., 1994. Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics* 32, 363–403.
- Large, W. G., Yeager, S., 2008. Global means and seasonal to decadal variability of air-sea fluxes. *Journal of Climate* submitted.
- Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., O'Brien, J., Rosati, A., Schneider, E., 1998. A review of the predictability and prediction of ENSO. *Journal of Geophysical*

- Research 103 (C7), 14375–14393.
- Latif, M., Böning, C., Willebrand, J., Biastoch, A., Dengg, J., Keenlyside, N., Madec, G., Schweckendiek, U., 2006. Is the thermohaline circulation changing? *Journal of Climate*, 4631–4637.
- Levitus, S., 1982. *Climatological atlas of the world ocean*. U.S. Government Printing Office 13, NOAA, Washington, D.C., 163 pp.
- Levitus, S., Burgett, R., Boyer, T., 1994. *World Ocean Atlas 1994*. vol 3: Salinity; vol. 4: Temperature. NOAA Atlas NESDIS 3 and 4, NOAA, U.S. Dept. of Commerce, Washington, D.C.
- Locarnini, R., Mishonov, A., Antonov, J., Boyer, T., Garcia, H., 2006. *World Ocean Atlas 2005*, volume 1: Temperature. U.S. Government Printing Office 61, NOAA Atlas NESDIS, Washington, D.C., 182 pp.
- Lohmann, G., Gerdes, R., 1998. Sea ice effects on the sensitivity of the thermohaline circulation in simplified atmosphere-ocean-sea ice models. *Journal of Climate* 11, 2789–2803.
- Lohmann, G., Gerdes, R., Chen, D., 1996. Sensitivity of the thermohaline circulation in coupled oceanic gcm-atmospheric ebm experiments. *Climate Dynamics* 12, 403–416.
- Lorbacher, K., Dommenges, D., Niiler, P., Köhl, A., 2006. Ocean mixed layer depth: A subsurface proxy of ocean-atmosphere variability. *Journal of Geophysical Research* 111-C07010, doi:10.1029/2003JC002157.
- Madec, G., Delecluse, P., Imbard, M., Lévy, C., 1999. OPA 8.1 ocean general circulation model reference manual. In: *Note du Pôle de modélisation*. Paris, France, p. 91 pp.
- Maltrud, M., McClean, J., 2005. An eddy resolving global $1/10^\circ$ ocean simulation. *Ocean Modelling* 8, 31–54.
- Marsh, R., de Cuevas, B., Coward, A., Nurser, A., Josey, S., 2005. Water mass transformation in the North Atlantic over 1985-2002 simulated in an eddy-permitting model. *Ocean Science* 1, 127–144.
- Marshall, J., Schott, F., 1999. Open-ocean convection: observations, theory, and models. *Reviews of Geophysics* 37, 1–64.
- Marsland, S., Haak, H., Jungclaus, J., Latif, M., Röske, F., 2003. The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Modelling* 5, 91–127.
- Martin, T., Gerdes, R., 2007. Sea ice drift variability in Arctic Ocean Model Intercomparison Project models and observations. *Journal of Geophysical Research* 112 (C04S10), doi:10.1029/2006JC003617.
- McCartney, M., 1977. Subantarctic mode water. In: Angel, M. (Ed.), *A Voyage of Discovery: George Deacon 70th Anniversary Volume*. Supplement to *Deep-Sea Research*. Pergamon Press, Oxford, pp. 103–119.
- Meehl, G., Gent, P. R., Arblaster, J., Otto-Bliesner, B., Brady, E., Craig, A., 2001. Factors that affect the amplitude of El Niño in global coupled climate models. *Climate Dynamics* 17, 515–526.
- Monterey, G., Levitus, S., 1997. Climatological cycle of mixed layer depth in the world ocean. U.S. government printing office, NOAA NESDIS, Washington, D.C., 5 pp.
- Morel, A., Antoine, D., 1994. Heating rate within the upper ocean in relation to its bio-optical state. *Journal of Physical Oceanography* 24, 1652–1665.
- Murray, R. J., 1996. Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics* 126, 251–273.
- Ohlmann, J., 2003. Ocean radiant heating in climate models. *JOC* 16, 1337–1351.
- Orsi, A., Whitworth, T., Nowlin, W., 1995. On the meridional extent and fronts of the antarctic circumpolar current. *Deep-Sea Research* 42A, 641–673.
- Pacanowski, R. C., 1987. Effect of equatorial currents on surface stress. *Journal of Physical Oceanography* 17, 833–838.
- Paulson, E. A., Simpson, J. J., 1977. Irradiance measurements in the upper ocean. *Journal of Physical Oceanography* 7, 952–956.
- Payne, R., 1972. Albedo of the surface ocean. *Journal of Atmospheric Sciences* 29, 959–970.
- Peixoto, J. P., Oort, A. H., 1992. *Physics of Climate*. American Institute of Physics, 520 + xxxix pp.
- Proshutinsky, A., Steele, M., Zhang, J., Holloway, G., Steiner, N., Häkkinen, S., Holland, D., Gerdes, R., Köberle, C., Karcher, M., Johnson, M., Maslowski, W., Zhang, Y., Hibler, W., Wang, J., 2001. The Arctic Ocean Model Intercomparison Project (AOMIP). *EOS* 82, 637–644.
- Rahmstorf, S., Marotzke, J., Willebrand, J., 1996. Stability of the thermohaline circulation. In: Krauss, W. (Ed.), *The warm water sphere of the North Atlantic Ocean*. Borntraeger, pp. 129–157.
- Rahmstorf, S., Willebrand, J., 1995. The role of temperature feedback in stabilizing the thermohaline circulation. *Journal of Physical Oceanography* 25, 787–805.
- Rayner, N., Parker, D., Horton, E., Folland, C.,

- Alexander, L., Rowell, D., Kent, E., Kaplan, A., 2003. Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research* 108(D14), 4407, DOI:10.1029/2002JD002670.
- Redi, M. H., 1982. Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography* 12, 1154–1158.
- Reynolds, R. W., Rayner, N., Smith, T. M., Stokes, D., Wang, W., 2002. An improved *in situ* and satellite SST analysis for climate. *Journal of Climate* 15, 1609–1625.
- Reynolds, R. W., Smith, T. M., 1994. Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate* 7 (6), 929–948.
- Rivin, I., Tziperman, E., 1997. On the sensitivity of air-sea fluxes to SST perturbations. *Journal of Climate* 11, 2431–2446.
- Robinson, I., 2005. Satellite oceanography. In: Chassignet, E. P., Verron, J. (Eds.), *GODEA Summer School*. Springer/Kluwer.
- Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornbluh, L., Manzini, E., Schlese, U., Schulzweida, U., 2006. Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *Journal of Climate* 19, 3771–3791.
- Rosati, A., Miyakoda, K., 1988. A general circulation model for upper ocean simulation. *Journal of Physical Oceanography* 18, 1601–1626.
- Röske, F., 2005. A global heat and freshwater forcing dataset for ocean models. *Ocean Modelling* in press.
- Roullet, G., Madec, G., 2000. Salt conservation, free surface, and varying volume. a new formulation for ogcms. *JGR* 105, 23927–23947.
- Russell, J., Stouffer, R., Dixon, K., 2006. Intercomparison of the southern ocean circulations in the ipcc coupled model control simulations. *Journal of Climate* 19, 4560–4575.
- Sarmiento, J. L., 1986. On the North and Tropical Atlantic heat balance. *Journal of Geophysical Research* 91 (C10), 11677–11690.
- Schweckendiek, U., Willebrand, J., 2005. Mechanisms affecting the overturning response in global warming simulations 18, 4925–4936.
- Smith, R., Gent, P., 2004. Reference Manual for the Parallel Ocean Program (POP). Los Alamos Technical Report No. LAUR-02-2484.
- Smith, R. D., Kortas, S., Meltz, B., 1995. Curvilinear coordinates for global ocean models. Los Alamos preprint LA-UR-95-1146.
- Smolarkiewicz, P., 1984. A fully multidimensional positive definite advection transport algorithm with small implicit diffusion. *Journal of Computational Physics* 54, 325–362.
- Sprintall, J., Tomczak, M., 1992. Evidence of the barrier layer in the surface layer of the tropics. *Journal of Geophysical Research* 97, 7305–7316.
- Steele, M., Morfley, R., Ermold, W., 2001. PHC: A global ocean hydrography with a high-quality Arctic Ocean. *Journal of Climate* 14, 2079–2087.
- Stommel, H., 1961. Thermohaline convection with two stable regimes of flow. *Tellus* 13, 224–228.
- Stommel, H., 1979. Determination of water mass properties of water pumped down from the ekman layer to the geostrophic flow below. *Proceedings of the National Academy of Science, USA* 76, 3051–3055.
- Stouffer, R. J., 2004. Time scales of climate response. *Journal of Climate* 17, 209–217.
- Sun, S., Bleck, R., 2006. Multi-century simulations with the coupled GISS-HYCOM climate model: control experiments. *Climate Dynamics* 26, 407–428.
- Sweeney, C., Gnanadesikan, A., Griffies, S. M., Harrison, M., Rosati, A., Samuels, B., 2005. Impacts of shortwave penetration depth on large-scale ocean circulation and heat transport. *Journal of Physical Oceanography* 35, 1103–1119.
- T. Whitworth, 1983. Monitoring the transport of the Antarctic Circumpolar Current at Drake Passage. *Journal of Physical Oceanography* 13, 2045–2057.
- Taylor, P., 2000. Final Report of the Joint WCRP/SCOR Working Group on Air-Sea Fluxes: Intercomparison and validation of ocean-atmosphere energy flux fields. WCRP-112, WMO/TD-No.1036. World Climate Research Programme, p. 303pp.
- Timmermann, R., Goosse, H., Madec, G., Fichefet, T., Ethé, C., Duliére, V., 2005. On the representation of high latitude processes in the ORCA-LIM global coupled sea ice-ocean model. *Ocean Modelling* 8, 175–201.
- Trenberth, K., Caron, J., 2001. Estimates of meridional atmosphere and ocean heat transports. *Journal of Climate* 14, 3433–3443.
- Weaver, A. J., Hughes, T. M., 1996. On the incompatibility of ocean and atmosphere models and the need for flux adjustments. *Climate Dynamics* 13, 141–170.
- Webb, D. J., 1994. The Deacon cell and the other meridional cells in the Southern Ocean. *Journal*

- of Physical Oceanography 24, 429–442.
- Whitworth, T., Peterson, R., 1985. Volume transport of the antarctic circumpolar current from bottom pressure measurements. *Journal of Physical Oceanography* 15, 810–816.
- Williams, R., Spall, M., Marshall, J., 1995. Does Stommel’s mixed layer ‘demon’ work? *Journal of Physical Oceanography* 25, 3089–3102.
- Wyrski, K., Kilonsky, B., 1984. Mean water and current structure during the Hawaii-to-Tahiti shuttle experiment. *Journal of Physical Oceanography* 14, 242–254.
- Yeager, S. G., Shields, C. A., Large, W. G., Hack, J. J., 2006. The low-resolution CCSM3. *Journal of Climate* 19, 2545–2566.
- Zhang, S., Greatbatch, R., Lin, C., 1993. A re-examination of the polar halocline catastrophe and implications for coupled ocean-atmosphere models. *Journal of Physical Oceanography* 23, 287–299.
- Zhang, X., Walsh, J., 2006. Toward a seasonally ice-covered Arctic Ocean: Scenarios from the IPCC AR4 model simulations. *Journal of Climate* 19, 1730–1747.