This manuscript was prepared for Ocean Science (Discussions)
with version 7.2 of the egu LaTeX class egu.cls.
Date: July 6, 2007

OS-LOGO

A Proposal for Coordinated Ocean-ice Reference Experiments (COREs)

Stephen M. Griffies[1], Arne Biastoch[2], Claus Böning[2], Frank Bryan[3], Eric Chassignet[4], Matthew England[5], Rüdiger Gerdes[6], Helmuth Haak[7], Robert W. Hallberg[1], Wilco Hazeleger[8], Johann Jungclaus[7], William G. Large[3], Gurvan Madec[9], Bonita L. Samuels[1], Markus Scheinert[2], Camiel A. Severijns[8], Harper L. Simmons[10], Anne Marie Treguier[11], Mike Winton[1], Stephen Yeager[3], and Jianjun Yin[4]

[1]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, USA
[2]Leibniz IfM-GEOMAR, Kiel, Germany
[3]National Center for Atmospheric Research, Boulder, USA
[4]Center For Ocean-Atmosphere Prediction Studies, Florida State University, Tallahasse, USA
[5]University of New South Wales, Sydney, Australia
[6]Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven, Germany
[7]Max-Planck-Institut für Meteorologie, Hamburg, Germany
[8]Royal Netherlands Meteorological Institute (KNMI), de Bilt, Netherlands
[9]Laboratoire Dynamique d'Oceanographie et de Climatologie, Paris, France
[10]International Arctic Research Center, University of Alaska, USA
[11]Labortorie de Physique de Oceans, IFREMER, Plouzane, France

Abstract. Coordinated Ocean-ice Reference Experiments (COREs) are proposed as a tool to explore the behaviour of global ocean-ice models under common surface boundary forcing. We highlight issues arising when designing coupled global ocean and sea ice experiments, such as difficulties formulating a consistent forcing dataset and experimental protocol. Particular focus is given to the hydrological forcing, with details key to realizing simulations with stable overturning circulations.

The dataset of Large and Yeager (2004) was developed for coupled ocean and sea ice models, and we found it to be suitable for our purposes. Simulations with this dataset 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 forcing produce qualitatively similar simulations. This hypothesis holds reasonably well (with some exceptions) for upper ocean tropical behaviour, but is less valid when examining deeper properties, especially in the high latitudes.

1 Introduction

Simulations with coupled ocean-ice models are commonly 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 differences in forcing methods can lead to large deviations in circulation behaviour and sensitivities. Such difficulties create practical barriers in comparing simulations from different modelling groups.

Correspondence to: Stephen.Griffies@noaa.gov

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. Notably, our proposal for COREs does not resolve problems related to forcing global ocean-ice models. Instead, we believe that proposing a set of interesting and feasible experiments, exposing experimental design details, and illustrating the behaviour of a suite of models run within the CORE framework, 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, where simulations are generally accumulated for analyses by a broad community. Intercomparison 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 embarking upon an analogous global ocean-ice model intercomparison project (i.e., an OMIP), 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 describing 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 the rudimentary methods for analyzing these simulations. This analysis provides guidance for others who employ the CORE experimental design. It furthermore allows for an examination of the hypothesis that similarly forced models produce similar simulations. This hypothesis is shown here to be largely valid for tropical upper ocean circulations, where most of the models (with some exceptions) exhibit similar strengths of the undercurrent as well as tropical temperature structure. However, the hypothesis generally fails when considering deep water properties and transports, with a rather wide spread found in model results. 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. We nonetheless present conjectures at places in the manuscript which may indicate reasons for certain diverging behaviours.

This paper can be roughly split into three parts, with the first documenting the state of the art in ocean-ice coupled modelling. Section 2, we highlight some uses of ocean-ice models, thus further arguing for the relevance of a reference experimental design. In Section 3, we review 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. Section 5 presents the spin-up behaviour of the simulations as reflected in the horizontally averaged temperature and salinity biases for the 500 years. This section also presents the biases in surface temperature for the final decade of the simulations. Section 6 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 7 considers the sea ice spin-up in both ice area and extent, and presents maps for the ice cover in March and September. Section 8 focuses attention onto the Tropical Pacific ocean simulation, which is critical for studies of El Niño Southern Oscillations (ENSO). Section 9 presents the zonal averaged anomalies of temperature and salinity during the final decade of the 500 year simulations. Section 10 presents the global poleward heat transport. Section 11 exhibits the time series for volume transport through the Drake Passage. Sections 12 and 13 then consider the meridional overturning circulation (MOC), with emphasis on its sensitivity to hydrological forcing. Section 14 closes the main portion of the paper with discussion and conclusions.

The third part of this paper is presented in four appendices that detail various aspects of the models used in this study, the experimental protocol, the methods use to force the models, and the metrics used to evaluate the simulations. Much of this material is often omitted in ocean-ice 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 climate 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 insolating effects of ice cover relative to open ocean. Quite generally, due to the importance of sea ice in ocean circulation, and vice-versa, realistic modelling studies of global ocean climate must include an interactive sea ice model coupled to the ocean.

Coupled ocean-ice models form an important sub-group in the climate system. They are often developed together prior to coupling to other components such as 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, the ocean-ice model places 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 and computationally expensive. Thus, a piecemeal 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 Arctic Ocean Model Intercomparison Project (AOMIP) providing one example with significant scientific impact (Proshutinsky et al., 2001). 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 sensitivites to perturbations, can prove distinct from ocean-ice models. 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 are frequently used for fruitful scientific research and model development purposes. We summarize here a few uses which 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.
- Ensembles of ocean-ice models can be run with a broader suite of algorithms and parameterizations than climate models. Such flexibility 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 boundary datasets 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 dataset, and which show similar biases, suggest that there are problems with the forcing datasets. In these ways, models can provide feedback onto the development of datasets (e.g., Large and Yeager, 2006).
- 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).
- 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 comparing simulations from different ocean-ice models using the same surface boundary 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-ice models

A coupled ocean-ice model requires momentum, thermal and hydrological exchanges with the atmosphere to drive the simulated ocean and ice fields. Respectively, these exchanges take the form of stress from atmospheric winds, of radiative and turbulent fluxes of heat, and of precipitation, continental runoff and evaporation. The latter 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 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). This approach requires no atmospheric information. It also limits the errors that can be realized in the simulated surface ocean properties. However, fluxes are non-zero only when SST and/or SSS differ

[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 volume averaged value, and so do not precisely reflect the surface values measured, say, from a satellite. See Robinson (2005) for more discussion.

from the observations. Furthermore, there is no direct link between the thermal and hydrological forcing. Nonetheless, these fluxes provide a strong negative feedback that limits the errors that can be realized in the simulated surface 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 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, for realistic simulations, 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.

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 transports 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 huge error bars (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 numerically 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 vleocities 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 formulaes for the ocean-atmosphere heat flux.

This feedback between the SSTs and the atmospheric state provides a nontrivial space-time dependent damping of SSTs which acts to reduce model drift. 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, and 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 valid, because of errors in both the ocean-ice models, and errors in the atmospheric datasets used to force them. The latter represent only an approximation to Nature, and the uncertainties can be large. Thus, even perfect ocean-ice models are exposed to limitations inherent in the forcing. 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.

We now expose two problems that arise when running ocean-ice models with a prescribed atmospheric state. The problems are intimately related, but we describe them here as separate mechanisms for clarity.

3.4.1 Salinity fluxes and mixed boundary conditions

The first problem relates to anticipated errors in the surface fluxes for salinity or fresh water, especially precipitation. These 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, such salinity restoring has no physical basis. It is thus desirable physically to use at most a weak restoring, which allows increased and typically more realistic, variability in the surface salinity and deep circulation, and that can be regarded as a correction to the precipitation.

In Nature the restoring timescale for SSS is very much longer than the effective SST restoring timescale, and the thermohaline fluxes move into a regime commonly known as *mixed boundary conditions* (Bryan, 1987). Stommel (1961) showed that ideal thermohaline systems forced with mixed boundary conditions admit multiple equilibria. The ability of any ocean or ocean-ice model to represent the oceanic adjustment in the more complete climate system with coupling feedback to the atmosphere, 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 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 column can become prone to freshwater pooling at the surface, which could provide a positve 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, overly strong vertical mixing and convection. The net effect on the simulated thermohaline circulation would be a weaking if the ice edge were overall too extensive, and a strengthening if too contracted. We illustrate these situations in Figure 1.

Other important feedback processes involving sea ice that affect sensitivity of the thermohaline circulation are discussed in Lohmann et al. (1998) and Jayne and Marotzke (1999). 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.

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 from 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 establising benchmark simulations, thus

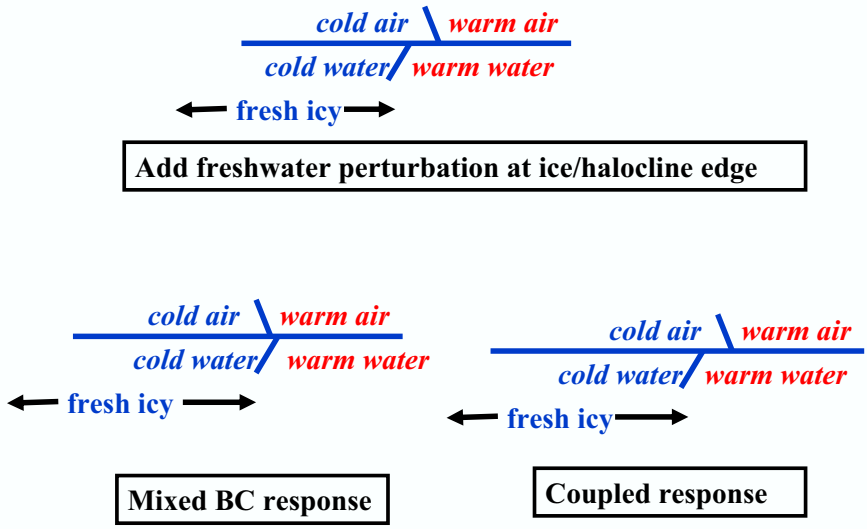

Fig. 1. This schematic illustrates the different responses of the ocean-ice system to changes in surface boundary fluxes. Top panel: Consider a cold-air outbreak (coming from the left and moving to the right in this schematic), such as occurs with synoptic activity over sub-polar regions. This outbreak 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.

facilitating comparisons and further refinements to the flux data sets and experimental design.

The standard practice we propose is termed a *Coordinated Ocean-ice Reference Experiment* (CORE). There are three COREs 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 and evolving 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 developed for the requirements of the Community Climate System Model (CCSM-POP) ocean-ice components (Collins 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 and periodically updated (Large and Yeager, 2006), 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 (2006)) 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. 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 forcing. The normal year has been constructed to retain synoptic variability (i.e., atmospheric storms), with a seamless transition from 31 December to 1 January.

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

- monthly varying precipitation (12 time steps per year);
- daily varying shortwave and longwave radiative fluxes (365 time steps per year, and so no diurnal cycle and no leap years);
- six-hourly varying meteorological fields: 10m air temperature, humidity, zonal wind, meridional wind, and sea level pressure (4 x 365 x 43 time steps per year with 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 by a collaboration between scientists at the National Center for Atmospheric Research (NCAR) and NOAA's Geophysical Fluid Dynamics Lab (GFDL) in support of COREs.

The task of developing a forcing dataset suitable for global ocean-ice climate models is fraught with uncertainty and ambiguity. As argued in Section 2, 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.

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. However, 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 forcing given the same ocean 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).

The Large and Yeager (2004) fields were 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 $1\,\mathrm{m\,s^{-1}}$ to more than $25\,\mathrm{m\,s^{-1}}$. We therefore 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 3.3. The basic question is: how strong should the damping be? Some simulations removed restoring under sea ice, whereas others retained restoring. Some ran with extremely weak restoring with the piston velocity of $50\,\mathrm{m}/4$ years, and some explored a range of restoring scenarios. We have more to say on this issue in Section 13.

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 (order hundreds to thousands of years).
- CORE-II: This experiment is aimed at investigations of the forced response of the ocean and/or ocean hindcast. It therefore employs the interannual varying dataset of Large and Yeager (2004). Other elements of the integration are the same as CORE-I.
- CORE-III: This is a perturbation experiment involving ideas proposed by Gerdes et al. (2005a, 2006). Here, enhanced fresh water 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 12 and 13). 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 Regarding differences in methods

Use of the Large and Yeager (2004) 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 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 virtual salt fluxes while others used real water fluxes; and differences in ice albedo schemes were common. These points may be critical for determining why, as shown later, some models maintained a nontrivial Atlantic overturning even with very weak salinity restoring (piston velocity of 50m/4years), some steadily weakened without a stronger restoring, and some failed to retain a sizable overturning even with strong salinity restoring. These differences could have been reduced with a more restrictive experimental design. However, such would have precluded participation from a large fraction of the seven groups in this paper. We thus chose a "come as you are" approach for the comparisons presented here.

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):

- CCSM-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.
- CCSM-HYCOM: This model is comprised of the HYCOM ocean model code within the CCSM framework used in the CCSM-POP simulations. That is, CCSM-HYCOM is the same as CCSM-POP, with the only modification being the ocean code.
- 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 rather than geopotential levels. The vertical and horizontal resolution is comparable to the GFDL-MOM simulation.
- KNMI: This model is comprised of ocean and sea ice components with zonal resolution of two degrees and enhanced meridional resolution in the tropics. The ocean component uses isopycnal vertical coordinates based on the Miami Isopycnic Ocean Model (MICOM).
- MPI: This is the ocean and ice model components of the coupled climate model from the Max-Planck-Institute. The ocean component uses geopotential vertical coordinates and has 40 vertical levels. The horizontal resolution gradually varies between a minimum of 12km close to Greenland and 150km in the tropical Pacific.
- ORCA: This model is comprised of the NEMO modelling system, whith 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. The MICOM code used by KNMI, and CCSM-HYCOM, are both non-Boussinesq, and so they include steric effects.

4.6 Goals of the analysis

The remaining sections of this paper survey results from simulations run with various ocean-ice models 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, it is insufficient to identify mechanisms for model differences. Nonetheless, we do highlight gross features that differ. Additionaly, through extra experiments performed by three of

the groups, we illustrate how the simulated circulations are sensitive to model and forcing details. In particular, these simulations highlight difficulties that some of the groups had in determining a suitable surface salinity forcing.

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, and ORCA), whereas CCSM-HYCOM, CCSM-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.

To facilitate producing difference maps, and to clearly compare model fields, participants in the 4th Assessment Report of the Intergovenmental Panel on Climate Change are required to map ocean model output to the same common spherical grid, according to the specifications described at the Project for Climate Model Diagnosis and Intercomparison (PCMDI) website

$$\texttt{http://www-pcmdi.llnl.gov/ipcc/about\_ipcc.php}$$

Although potential exists for us to employ the mapping software from PCMDI, we chose not to add this extra level of analysis burden on the participants in this study. Hence, difference maps are restricted to differences to the respective model's initial conditions. Each group is responsible for computing these differences for their individual model. Notably, we do not consider the present study to be a formal model comparison project involving participants outside of those involved with running the models. Rather, this study represents a prototype of what may later become more widespread and organized. Hence, differences in gridded model output do not compromise these limited goals.

5 Spin-up behavior and surface properties

Spin-up of the CORE-I simulations is illustrated in Figures 2 and 3, which show time series for the volume weighted anomalous temperature and salinity as a function of depth. 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 continue to drift throughout the simulation period.[2]

For the temperature drifts, there no two models that exhibit the same qualitative behaviour throughout the water column. The CCSM-POP, GFDL-MOM, and MPI simulations each show surface cooling and upper-ocean warming, but the GFDL-MOM shows slight warming in the abyss whereas CCSM-POP and MPI show cooling. The ORCA and GFDL-HIM simulations both show warming in the upper ocean and cooling throughout the depths beneath 500m, with the magnitude of these drifts smaller in ORCA than GFDL-HIM. The KNMI simulation shows the largest drift overall with a strong cooling in the mid-ocean region along with a strong warming in the upper ocean similar to GFDL-HIM.

For salinity, CCSM-POP and MPI show a near surface freshening, with MPI's extending to 500m whereas CCSM-POP's remains just in the uppper levels. Both the GFDL-MOM and GFDL-HIM simulations evolve towards salty upper ocean and fresh deeper ocean, with GFDL-MOM exhibiting smaller magnitudes. The KNMI simulation is uniformly fresh throughout the water column for nearly the full 500 years. As shown in Table 1 in Appendix A, this model does not choose to perform a normalization for the hydrology in the ocean-ice system, whereas all but the ORCA simulation provide some means to ensure that the net salt or water entering the ocean-ice system remains within bound. 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 unless some form of normalization is applied. Such is a limitation of the ocean-ice system absent an interactive land or atmospheric component.

Global maps of SST from the simulations are compared in Figure 4 to those from Conkright et al. (2002) for the World Ocean outside the Arctic, with Steele et al. (2001) used for the Arctic. All models show the typical anomaly patterns expected for the resolutions of models used here, with largest deviations concentrated along major frontal zones such as the western boundary currents and North Atlantic current. This bias leads to a warm bias on the east coast of USA, and cool bias in the subpolar Atlantic. The SSTs are additionally too warm in the middle latitude regions on the west of continents, possibly due to poorly resolved coastal upwelling. In the Pacific, both the MPI and KNMI simulations show distinct cool pattern in the central and eastern tropical Pacific, with warming elsewhere in the basin. We consider more in this region in Section 8.

6 Annual cycle at Ocean Weather Ship ECHO

An example of the near-surface annual cycle is given in Figure 5. Here, we diagnose the heat content and temperature for Ocean Weather Ship ECHO in the subtropical North Atlantic. Apart from elucidating the adjustment behaviour of

[2] It is instructive to compare the behaviour of the water mass properties with the transport indices in Figure 18 discussed in Section 13.1. Even if a simulation appears dynamically adjusted after 100 years, as deduced from the transport time series, anomalies in simulated transports continue adding to drifts in the deep temperature and salinity fields.

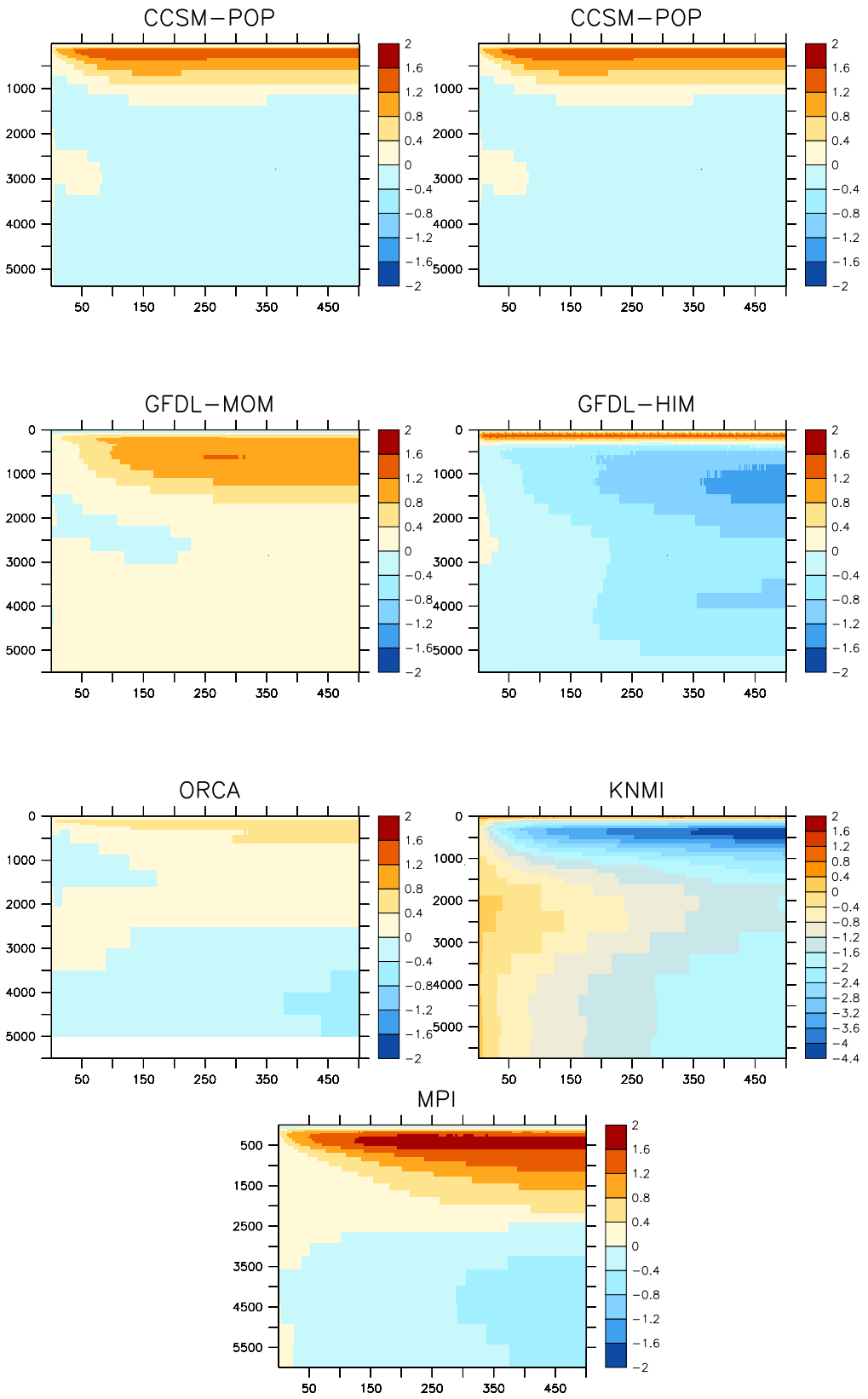

Fig. 2. Globally averaged drift of the annual mean temperature as a function of depth (vertical axis in metres) and time (years on horizontal axis). This drift is defined as $T_{\text{drift}}(z,t) = \left(\sum_{xy} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,(T_{\text{model}} - T_{\text{initial}})\right) / \left(\sum_{xy} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\right)$. Note the larger negative range for the KNMI plot and the larger contour intervals. All other plots have the same range and contour interval.

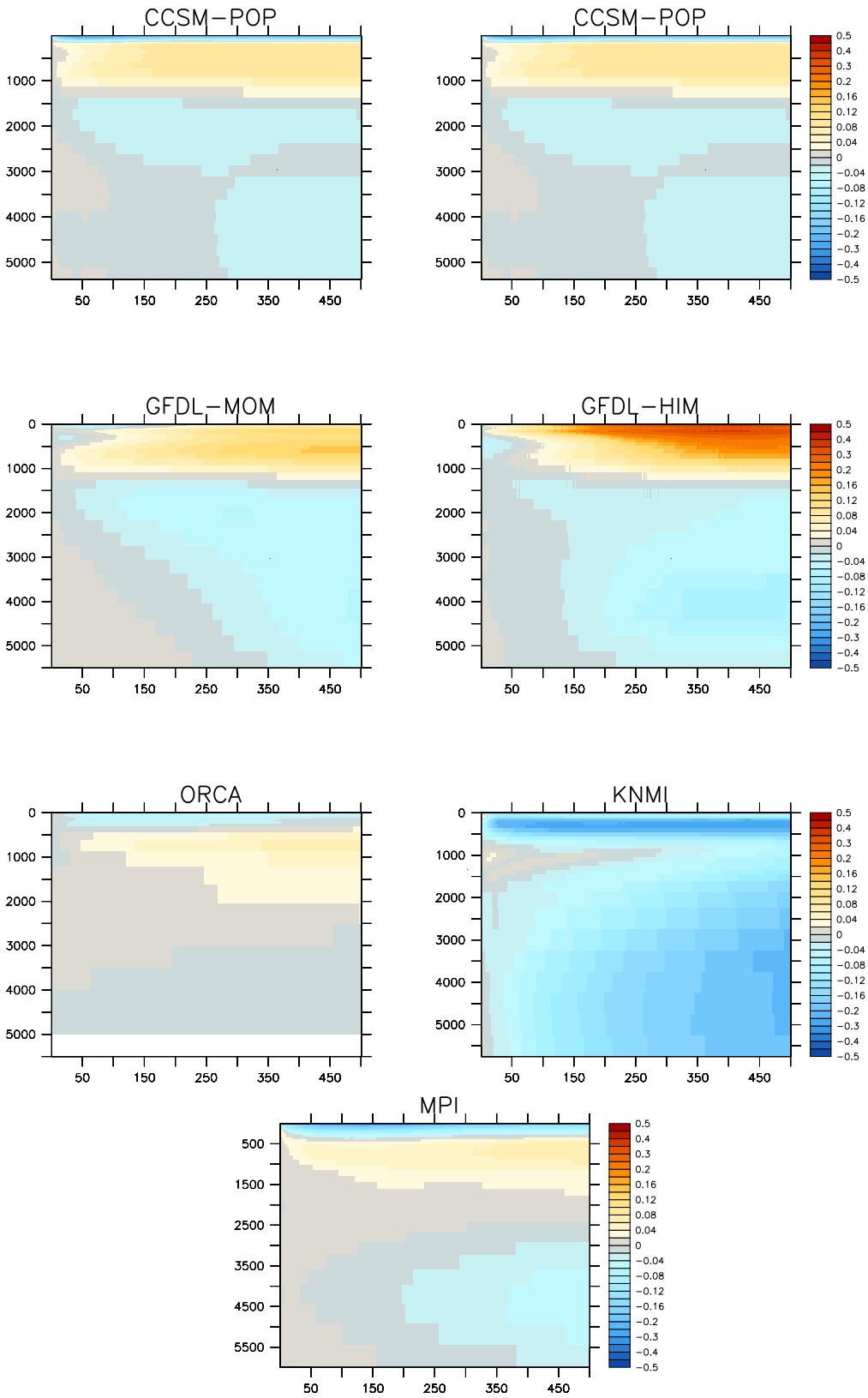

Fig. 3. Globally averaged drift of the annual mean salinity as a function of depth and time (years on horizontal axis). This drift is defined as $s_{\text{drift}}(z,t) = \left(\sum_{xy} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, (s_{\text{model}} - s_{\text{initial}}\right) / \left(\sum_{xy} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z\right)$.

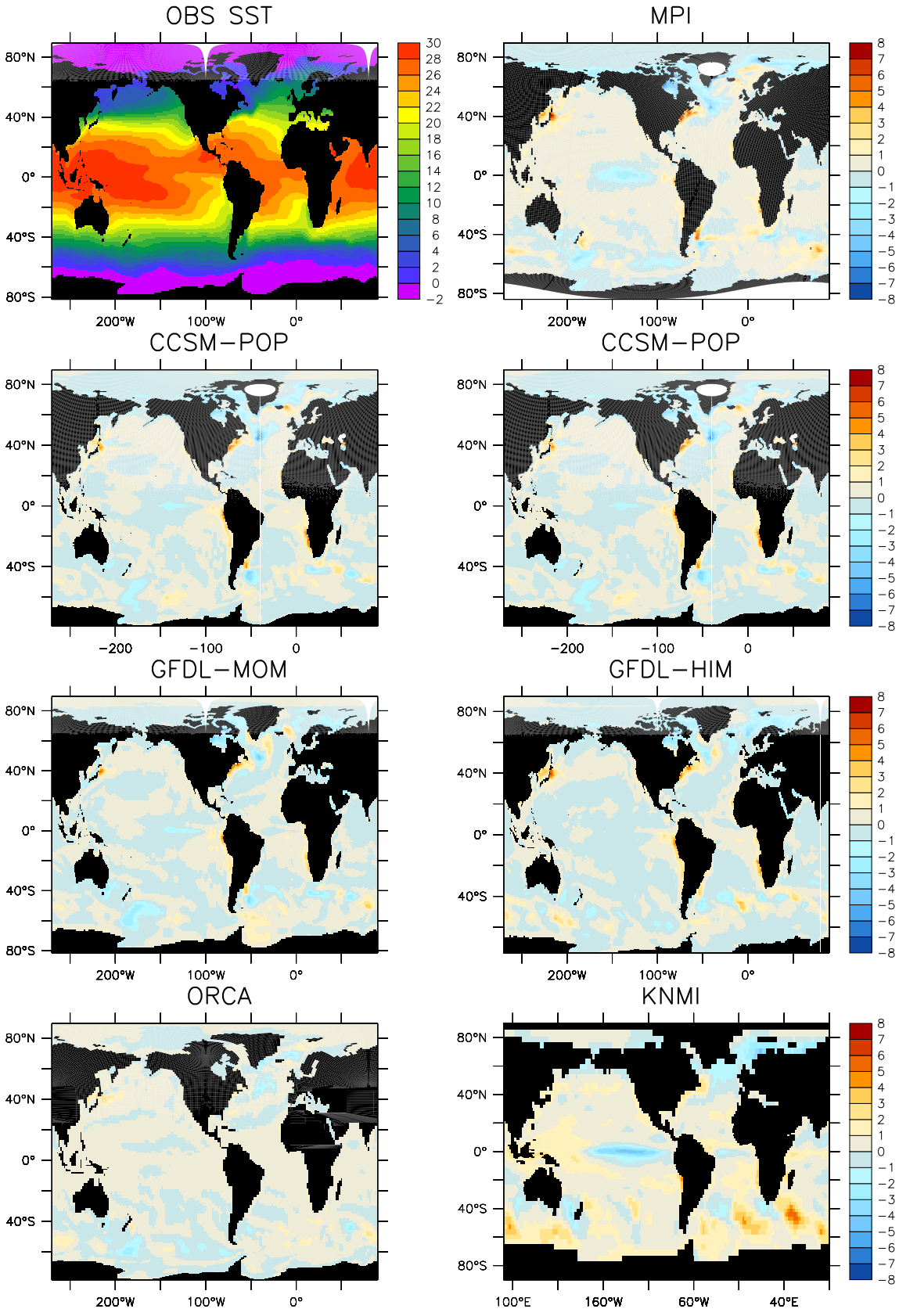

Fig. 4. Anomalous SST for years 491-500 from the simulations relative to the analysis (shown in the top middle panel) of Conkright et al. (2002) outside the Arctic, and Steele et al. (2001) in the Arctic.

the simulations, this figure also provides a useful diagnostic to assess the ability of bulk forcing (in conjunction with different boundary layer formulations) to realistically simulate an important aspect of the ocean's role in climate.

Temporal evolution of the hysteresis loop exemplified for one experiment in Figure 5a confirms the near-surface equilibration time of 50-100 years as seen in the global mean time series in Figures 2 and 3. Figure 5b compares the results of various model simulations at year 50. Each model possesses a reasonable seasonal cycle, though with some shift relative to Levitus. Furthermore, the area encompassed by each model is a bit less than Conkright et al. (2002), with ORCA containing systematically more heat content, whereas the CCSM-POP and GFDL-MOM simulations are slightly less. The CCSM-POP and GFDL-MOM simulations use the same mixed layer scheme from Large et al. (1994), thus possibly accounting for their close agreement. Although the GFDL-HIM and KNMI simulations are both isopycnal models, they exhibit a shift in heat content, likely due to basic 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).

7 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.

As shown by the temporal evolution of the annual mean sea ice area in Figure 6, 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 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 initialization. For the other models, sea ice area and extent are more similar. Both the GFDL-HIM and ORCA simulations show far more variability in Southern Hemisphere sea ice than the other models, with GFDL-HIM possibly not yet reaching a steady state.

Figure 7 maps the ice concentration (sea ice area per grid box area) in March (month of largest Northern Hemisphere sea ice coverage), and Figure 8 shows the same for September (month of 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).

The models exhibit a much larger cycle in the Southern Hemisphere, which reflects the huge seasonal cycle in the observations. The KNMI simulation generally shows the least amount of ice, nearly loosing its summertime Arctic ice cover and showing rather low wintertime Southern Hemisphere ice concentration.

8 Tropical Pacific

Realistic ocean simulations in the Tropical Pacific are important for maintaining proper coupled ocean atmosphere simulation of El Niño Southern Oscillations (ENSO) (Latif et al., 1998), which is critical for simulations of climate change. 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 $143°E$ to $95°W$ 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 (Wyrtki and Kilonsky, 1984). This temperature structure results in a thermocline with tighter vertical gradient just off the equator than at the equator.

Time series for the zonal transport in the equatorial undercurrent (not shown)[3] 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 9 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. Note the somewhat cold water found in ORCA and GFDL-HIM, and the very cold water in KNMI, found towards the bottom of the thermocline. The KNMI simulation also exhibits very warm water in the upper ocean mixed layer in the west Pacific, going beyond the scale of the other models. The remaining 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 4, 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.

[3] Figure 18 in Section 13.1 shows the GFDL-MOM time series.

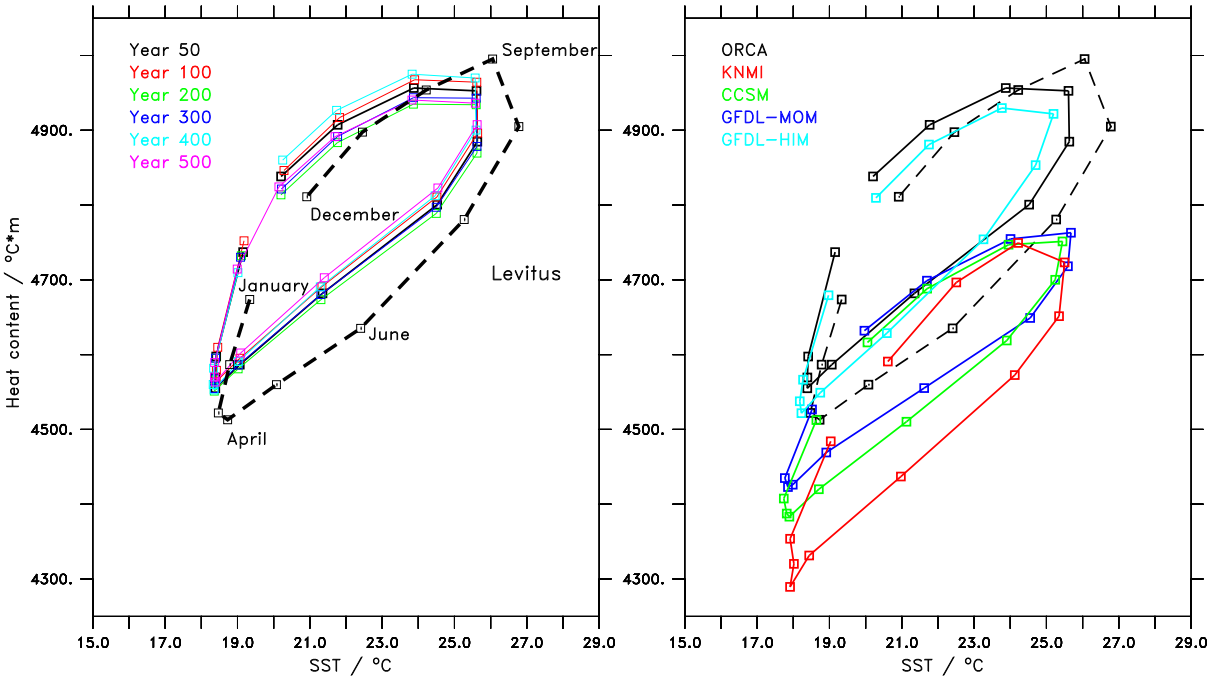

Fig. 5. Monthly values of the heat content (vertically integrated temperature) over the upper 250m versus sea surface temperature taken at Ocean Weathership ECHO ($48^\circ W, 35^\circ N$). The climatology of Conkright et al. (2002) is shown as the dashed curve in both figures. (a) time evolution in ORCA taken at selected years during the 500 year spin-up, (b) comparison of various models at year 50: ORCA(black), GFDL-MOM(dark blue), GFDL-HIM(light blue), CCSM-POP(green), KNMI(red).

Additionally, all models show overly cool waters in the central to eastern portion of the Pacific, with the MPI and KNMI simulations especially cold. KNMI 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 ORCA simulations are also warm outside the cold tonge region, but less so than KNMI. The GFDL-MOM, GFDL-HIM, and CCSM-POP simulations show similar error patterns with generally smaller amplitude than the remaining models.

Figure 10 shows the zonal current along the equator from Johnson et al. (2002). Its strength peaks a bit above $1\,\mathrm{m\,s^{-1}}$. The previous class of IPCC models had very sluggish currents, with speeds slower by factors of two or three. As shown in this figure, each model in the present study has simulated undercurrent speeds approaching, or exceeding, $1\,\mathrm{m\,s^{-1}}$. The only exception is the MPI simulation, which shows a relatively weak undercurrent. This model is perhaps overly dissipated by its choice for horizontal friction, or limited by its relatively coarse meridional resolution at the equator (coarser than 1°), with both handicapping the ability to simulate a strong undercurrent.

9 Zonal average potential temperature and salinity

One of the most widely assessed benchmarks of ocean model performance is the distribution of global 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 relatively 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. Modelers 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.

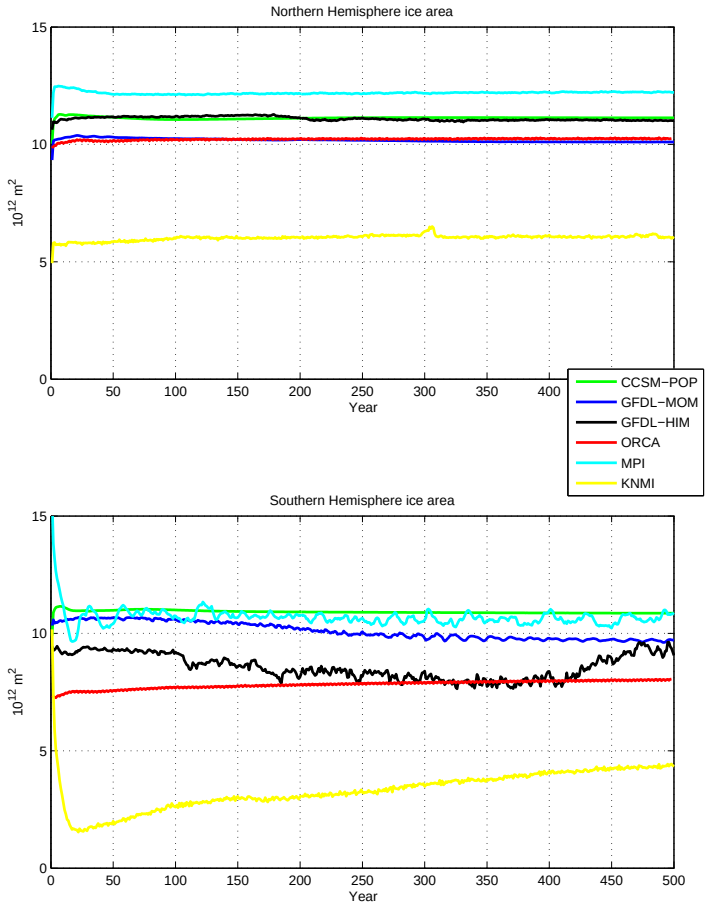

Fig. 6. Time series for the annual mean sea ice area (in units of 10^{12} m^2) for the Northern Hemisphere (top row) and Southern Hemisphere (bottom row), with CCSM-POP(green), GFDL-MOM(blue), GFDL-HIM(black), ORCA(red), MPI(cyan), and KNMI(yellow). The hemispheric ice area depicts the area of a grid box multipled 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.

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,
- unrealistic interior mixing processes in the model.

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 physics.

A comparison between the global zonal-mean anomalous θ and S in the CORE simulations is presented in Figures 11 and 12, respectively. The model minus observed fields are derived taking an average in the simulations during years 491-500. While this time scale 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 CCSM-POP, GFDL-MOM, and MPI simulations are too warm and salty in the upper ocean, and contain deep warm anomalies in the North Atlantic. GFDL-HIM also shows some warm anomalies in the surface ocean and warm anomalies in the deep water formation regions around 70^{o}N, but with a less broad pattern of warm anomalies. The ORCA simulation shows even less anomalous temperature, whereas

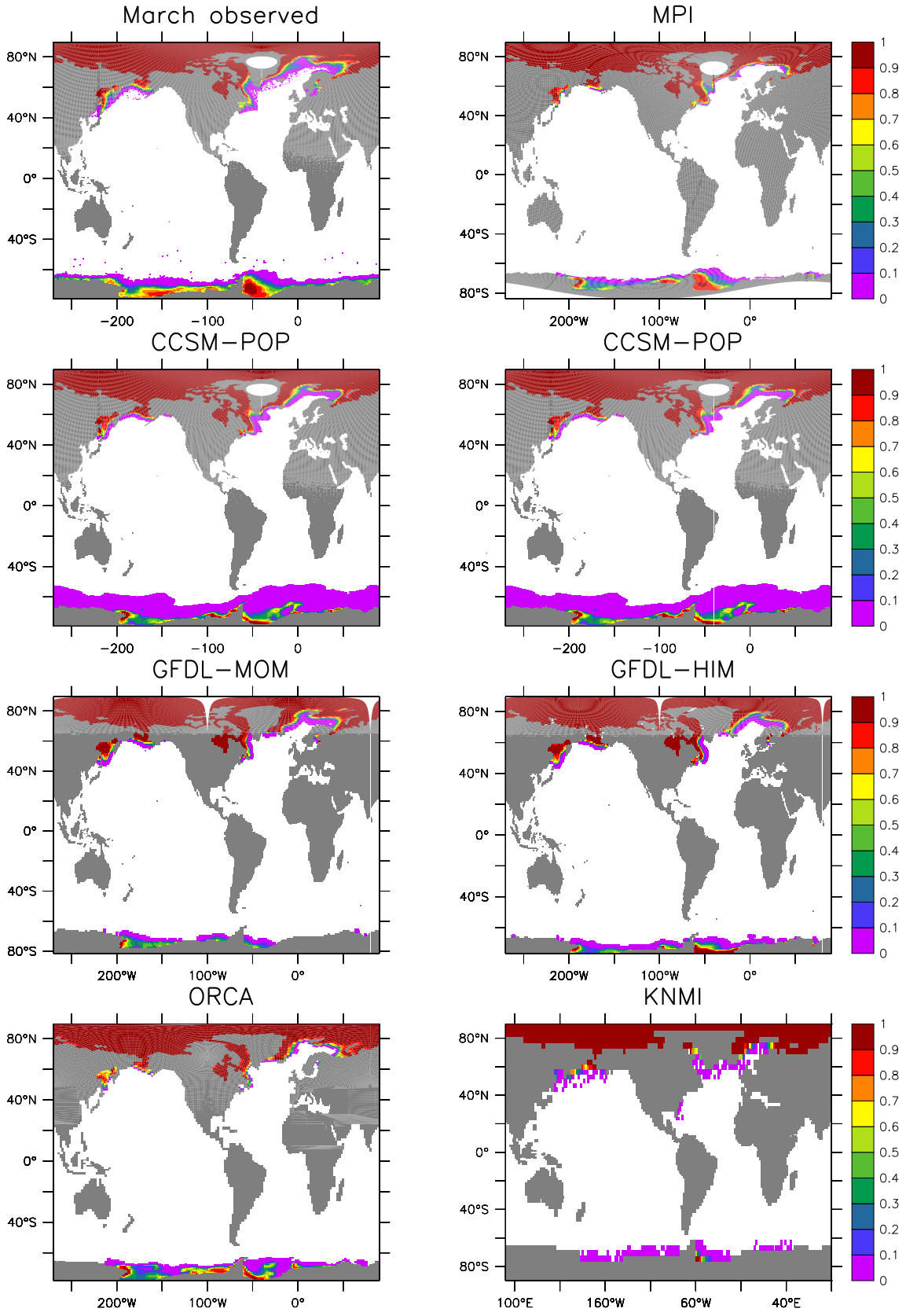

Fig. 7. Sea ice concentration (area of sea ice per grid cell area) in March (month of maximum observed Northern Hemisphere sea ice coverage), averaged over years 491-500 in the CORE-I simulations, as well as the observations taken from Comiso (1999 (updated 2005). ORCA still shows ice extent. CCSM-HYCOM will replace one of CCSM-POP.

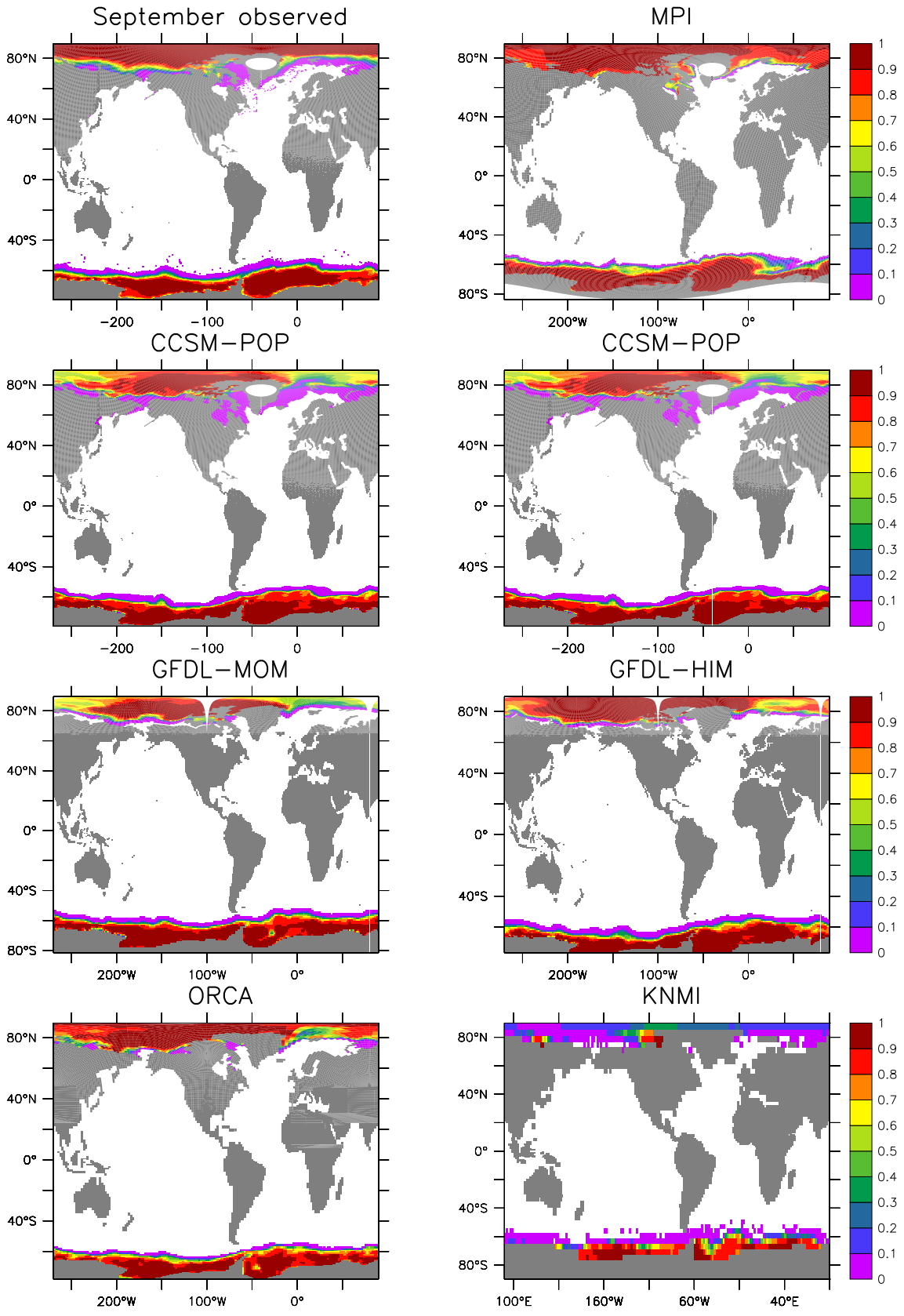

Fig. 8. Sea ice concentration (area sea ice per area grid cell) in September (month of the maximum observed Southern Hemisphere sea ice coverage), averaged over years 491-500 in the CORE-I simulations, as well as the observations taken from Comiso (1999 (updated 2005). ORCA still shows ice extent. CCSM-HYCOM will replace one of CCSM-POP.

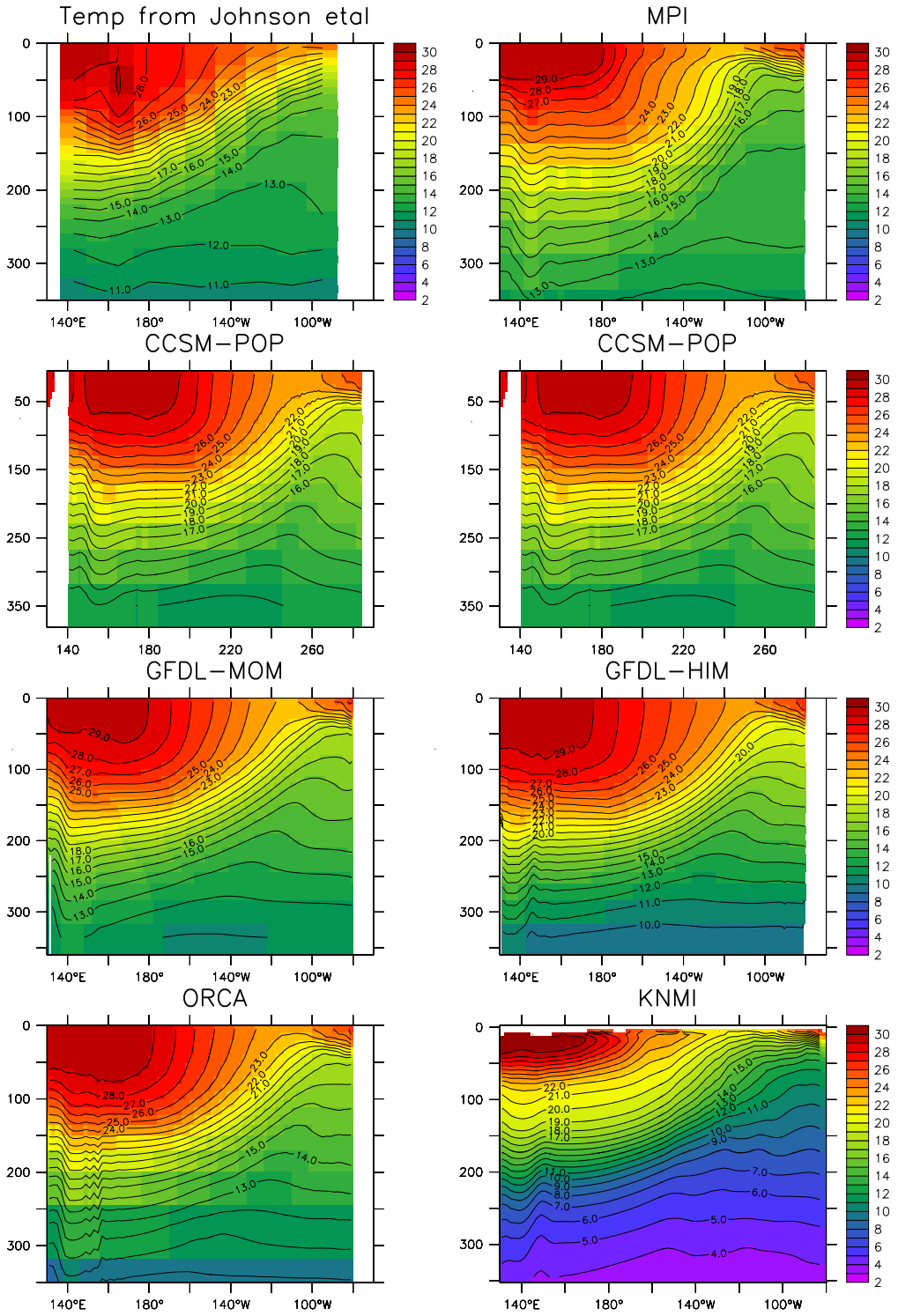

Fig. 9. This figure displays the upper ocean temperature on the equator in the Pacific, with model results from a time mean over years 491-500 of the simulation. Top panel: Observed temperature from Johnson et al. (2002). Remaining panels are from the simulations. Note that the choppy KNMI field near the surface arises from the remapping from density to depth space. Additional limitations of the submitted dataset preclude contouring this image.

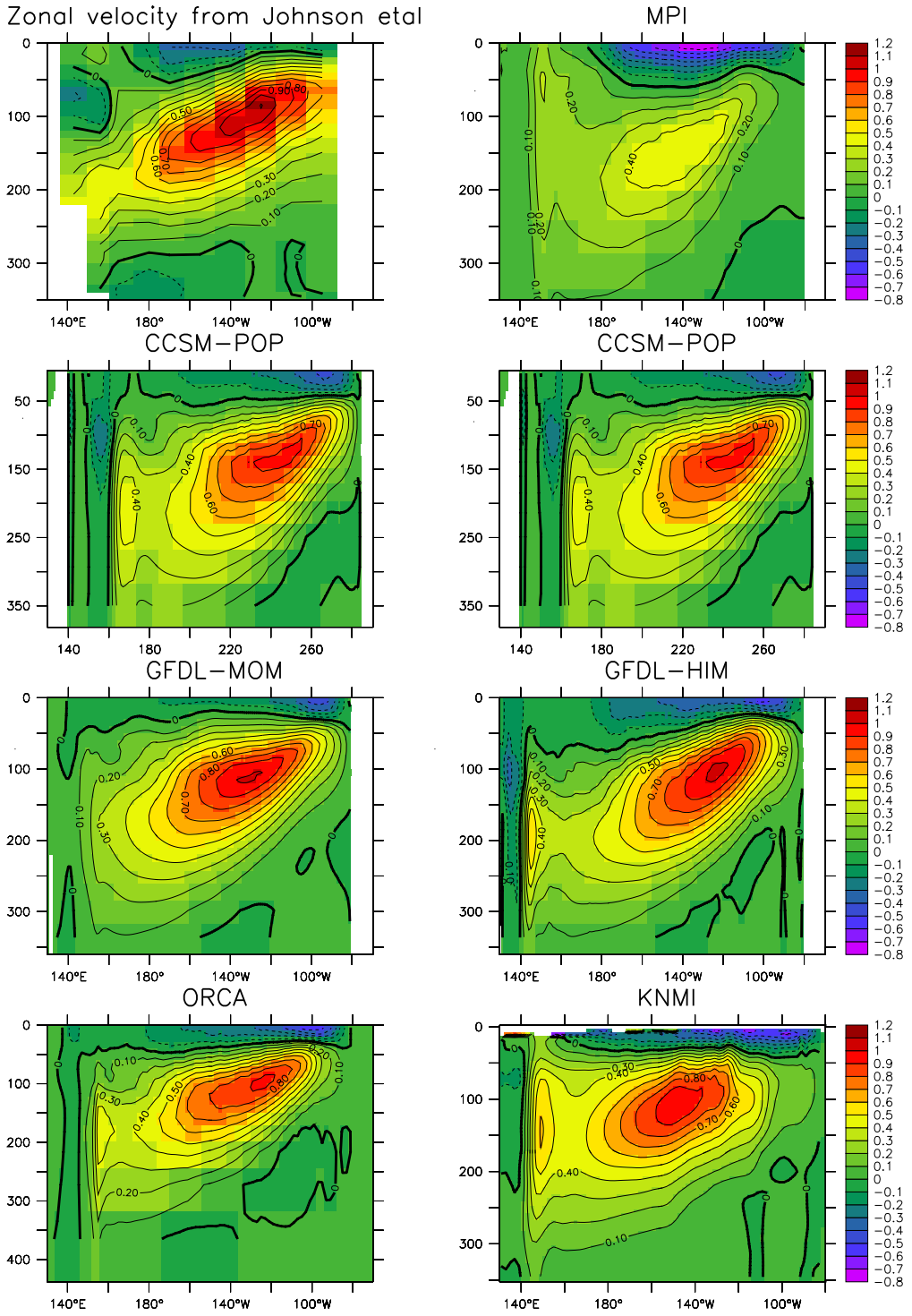

Fig. 10. This figure displays the upper ocean zonal velocity component on the equator in the Pacific. The model results are time mean over years 491-500 of the simulations. Top panel: Observed temperature from Johnson et al. (2002). Note the CCSM-POP uses cm/sec to measure velocity, with all other models using metre/sec.

the KNMI simulation is generally cold throughout the ocean interior, very cold in the upper tropical and mid-latitude ocean, and a bit too warm in the far north. The zonal biases in the salinity for CCSM-POP, GFDL-MOM, GFDL-HIM, and MPI show somewhat too much salt in the upper ocean and penetrating into the deep North Atlantic, with the MPI simulation showing the least bias. ORCA and KNMI are more broadly fresh throughout the ocean, with KNMI especially fresh in the deep northern regions.

The erroneous heat and salt content localized near 60-70^{o}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 (not shown), with a broad region of deep overturn just north of 60^{o}N (Gnanadesikan et al., 2006). This deep mixed layer occurs despite the GFDL-MOM meridional overturning rates, described in Section 12, being only modestly too vigorous relative to observations. 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 ORCA and KNMI simulations is symptomatic of its rather weak meridional overturning circulation (MOC) (Section 12). Curiously, ORCA also shows weak anomalously warm θ in this region, despite its slow overturning. This bias must arise from localized errors in SST in regions where the model forms late winter pycnostads.

Apart from differences in model versus observed $\theta - S$ originating in the North Atlantic, Figures 11 and 12 also reveal large scale discrepancies in Southern Ocean water mass properties. All but the CCSM-POP simulation exhibit Antarctic Bottom Water (AABW) signatures that are too cold, and all appear too fresh, with CCSM-POP showing the least bias. Antarctic Intermediate Water (AAIW) is for the most part too warm and saline, with ORCA and KNMI showing somewhat smaller salt anomalies than the other simulations. These biases exist despite the fact that most models appear to capture reasonable rates for AABW overturn ($\sim$ 10 Sv), and none exhibit excessive mixed layer depths near regions of Subantarctic Mode Water formation (a classical reason for warm saline AAIW in models reference???). A likely cause for the cold and fresh AABW in the simulations is that the models simulate a region of large scale deep mixed layer in the central Weddell Gyre, which contrasts with observations (reference???). This region is spuriously ice free during late summer, meaning fresher sea surface salinities (need to check that this is so) 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 ORCA and KNMI simulations, their weak NADW leaves the middle depth Southern Ocean, and in particular Circumpolar Deep Water, too cold and fresh. The warm saline AAIW in contrast is possibily 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, which are furthere evidenced by the very weak poleward implied heat transport in the Southern Hemisphere exhibited in Figure 13.

- I need to fill in a recent best guess range for observed NADW formation paragraph 3 in lieu of NN-MM Sv. Any suggestions welcome. Anne Marie suggests Speer and Lumkin, or Ganachoud, or Talley?
- Arne, The AABW overturning curve shown in Fig. 12 appears to be the strength of what I'd call the abyssal cell centred near 20S. Is that right? My plot of ORCA-D had only just over 10 Sv in the global AABW cell. This affects paragraph 4.
- Implicit in the above is some discussion of MLDs in the respective models, which I gleaned from Fig sent by Arne and from the Gnanadesikan, Griffies and Samuels paper (Fig. 6). It may be worth adding a simple zonal-mean plot of MLD in observed vs. the models (a la Gnanadesikan et al. Fig 6c).
- Last paragraph: I'm guessing about the SSS properties as these were not shown in the core draft.pdf I had access to (only SST).
- Last paragraph again, I'm unsure why the model's AAIW is too warm/saline. Mixed layers look pretty good in the region. It is perhaps model-obs differences in isopycnal slopes (and therefore downward mixing rates). Any suggestion here welcomed.

10 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 coordinated 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).

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

$$\partial_t \left(\int \mathrm{d}x\, \mathrm{d}z\, \theta \right) = -\partial_y \left(\int \mathrm{d}x\, \mathrm{d}z\, (v\, \theta + F^y) \right) \quad (1)$$

$$+ \int \mathrm{d}x\, (Q^{\mathrm{surf}}/(\rho_o\, C_p) + q_{\mathrm{w}}\, T_{\mathrm{w}}). \quad (2)$$

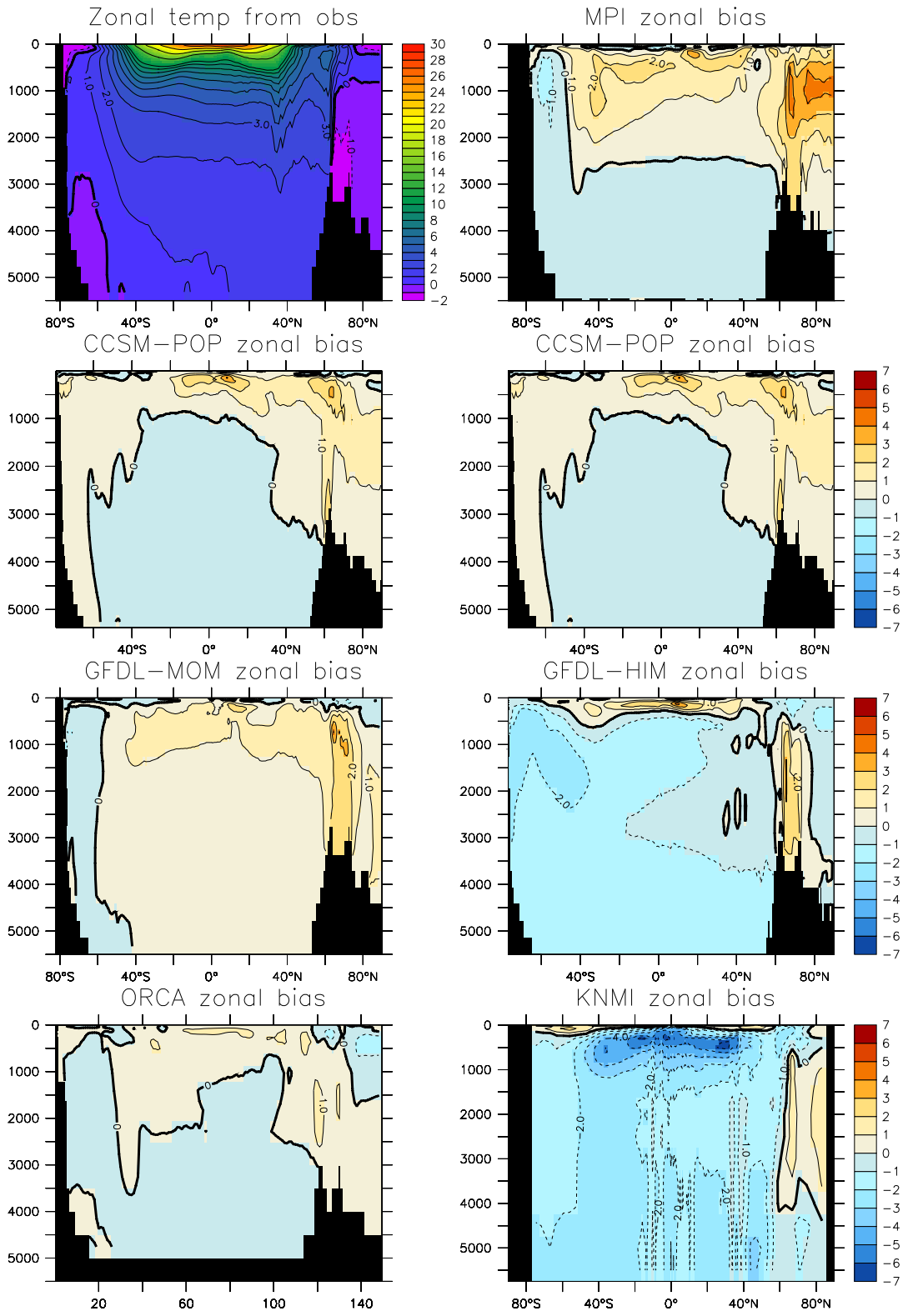

Fig. 11. Anomalous zonal mean decadal mean (years 491-500) potential temperature for the simulations relative to the analysis (shown in the top panel) of Conkright et al. (2002) for all but the Arctic, and Steele et al. (2001) for the Arctic.

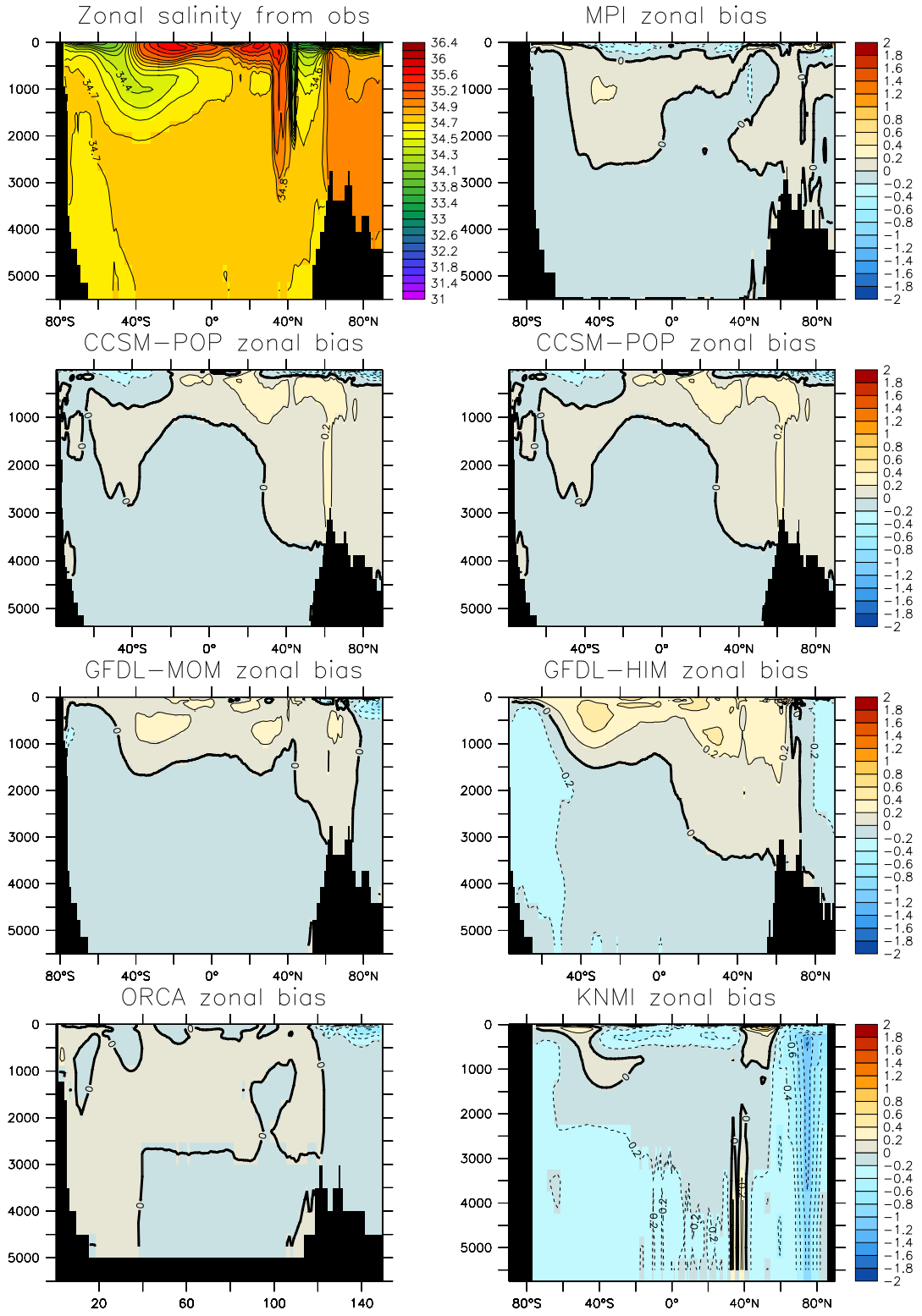

Fig. 12. Anomalous zonal mean decadal mean (years 491-500) salinity for the simulations relative to the analysis (shown in the top panel) of Conkright et al. (2002) for all but the Arctic, and Steele et al. (2001) for the Arctic.

The left hand side represents transient storage of heat in the ocean. On the right hand side, $v\,\theta$ is the advective temperature transport due to the resolved model velocity field, and F^y is the meridional flux arising from SGS processes, such as diffusion and the eddy-induced advective transport prescribed by Gent et al. (1995). Q^{surf} is the surface heat flux (positive into the ocean) arising from shortwave and longwave solar radiation, and turbulent fluxes from sensible and latent heating. The fresh water term $q_{\text{w}}\,\theta_{\text{w}}$ accounts for the transport of water across the ocean surface occuring at a temperature θ_{w} with a volume per area per time q_{w} ($q_{\text{w}} > 0$ for water entering the ocean) . The freshwater term is absent in the CCSM-POP and KNMI simulations due to their use of salt rather than water fluxes. We have ignored in this formulation any of the generally minor contributions from interior and bottom heat sources.

Assuming a steady state, and dropping the surface fresh water term, yields a balance between surface heating and the meridional divergence of meridional ocean heat transport

$$\int \mathrm{d}x\, Q^{\text{surf}} = \partial_y \left(\rho_o\, C_p \int \mathrm{d}x\, \mathrm{d}z\, (v\,\theta + F^y) \right). \quad (3)$$

Assuming the meridional ocean transport vanishes at the southern boundary at $y = y_s$ (along the Antarctic continent) renders the steady state balance for any particular latitude

$$\int_{y=y_s}^{y} \mathrm{d}y \int \mathrm{d}x\, Q^{\text{surf}} = \rho_o\, C_p \int \mathrm{d}x\, \mathrm{d}z\, (v\,\theta + F^y). \quad (4)$$

This relation states that the area integrated heat flux into the ocean over the region south of any given latitude is balanced by a meridional transport of heat within the ocean at the specific latitude. 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 warm water), then this ocean heat loss is balanced by a transport of heat within the ocean towards the south at this latitude. As described by Large and Yeager (2004), knowledge of the atmospheric state and the sea surface temperature provide sufficient information to compute surface heat fluxes by using the bulk formulae. This calculation leads to an *implied* ocean heat transport (left hand side of equation (4)). Given the implied heat transport, we can determine the meridional ocean heat transport at a particular latitude (right hand side). Again, determination of the ocean heat transport from the implied heat transport is valid if the following assumptions hold: (1) ocean heat storage is small (i.e., steady state); (2) heat transport due to the movement of water across the ocean surface is ignored; (3) bottom heat fluxes (e.g., geothermal fluxes) are ignored. These assumptions are reasonable, especially since surface air-sea heat fluxes are generally dominant in much of the ocean, even though there remain large uncertainties in these fluxes. Hence, it is common to compare implied heat transports from various atmospheric data products to direct measurements of ocean heat transports.

Determining which atmospheric product to use for computing the implied heat transport remains a subject of some discussion (e.g., Taylor, 2000). We present two results. First, we compute the implied northward heat transport by running the normal year forcing (NYF) from Large and Yeager (2004) over the Hurrell et al. (2006) SST from 1958-2000 and computing heat fluxes using the bulk formulae from Large and Yeager (2004). This range of SST provides some sense for the range available from the model simulations. At each latitude, a 43-year mean 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. The second result is taken from Trenberth and Caron (2001), who estimate heat fluxes from a combination of reanalysis and data products and use these fluxes to compute an implied heat transport.

Figure 13 shows these implied northward global heat transport, as well as the ocean heat transport directly computed from the ocean model fields averaged over simulation years 491-500. Notably, the Large and Yeager (2004) implied heat transport shows the least amount of poleward transport in the middle latitude Southern Hemisphere, with Trenberth and Caron (2001) and the ocean-ice models showing somewhat more transport, comparable to the northward transport in the Northern Hemisphere. For the models, 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. The CCSM-POP, GFDL-MOM, and GFDL-HIM simulations 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 ORCA simulation is somewhat weaker in the north, though with a sizable poleward transport in the south. Finally, the KNMI simulation is a clear outlier, with less than 1PW in the north and nearly 3PW poleward transport in the south.

11 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.[4] 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). The transport is a strong function of the winds, baroclinicity across the ACC, and buoyancy forcing (e.g., Hallberg and Gnanade-

[4] Each of the models in this study employs the Boussinesq approximation, and so the volume transport is considered, rather than more fundamental mass transport appropriate for a non-Boussinesq model.

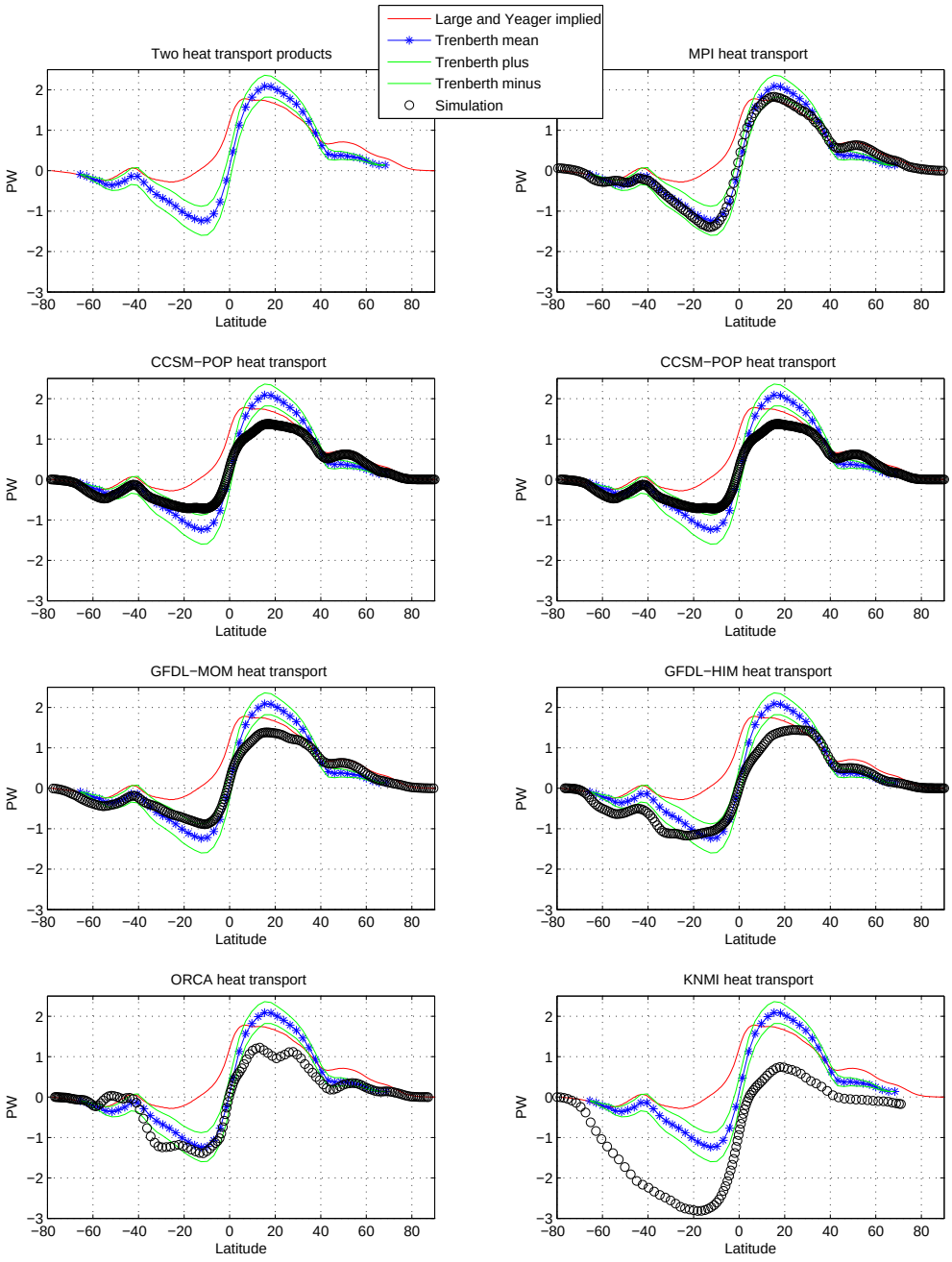

Fig. 13. Northward heat tranport for the global ocean as determined by the ocean models (circles). Results are decadal means from model years 491-500, and the values include both resolved advective and SGS transport contributions. Also shown is the implied northward heat tranport derived from the atmospheric state variables and bulk formulae of the Normal Year Forcing (NYF) of Large and Yeager (2004) used in CORE-I simulations (line), and the SSTs from Hurrell et al. (2006). 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). At each latitude, a 43-year mean is computed. Additionally, we exhibit the implied transports (stars) from Trenberth and Caron (2001) along with their range. This estimate is based on heat fluxes from a combination of reanalysis and data products. Units are PW=10^{15}Watts for each transport.

sikan, 2006, and references therein). 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).

Figure 14 shows the time series for annual mean Drake Passage transport from the CORE simulations. The transport in the KNMI 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. CCSM-POP, GFDL-MOM, and ORCA converge to values between 135Sv-150Sv, which is on the high side of measurements. The MPI transport is relatively high and exhibits nontrivial variability around a mean of roughly 180Sv. Results from GFDL-MOM and ORCA using weak salinity restoring with piston velocity of 50m/4years are presented in Sections 13.1 and 13.2, and these two both show nonstationary behaviour. So it is only the CCSM-POP simulation that exhibits relatively stationary behaviour for the Drake Passage transport with the weak 50m/4years piston velocity.

12 Meridional overturning streamfunction

The meridional overturning streamfunction diagnoses the transport of volume poleward

$$\Psi(y,z) = -\int \mathrm{d}x \int_{-H}^{z} \mathrm{d}z' \, (v + v^*). \tag{5}$$

It is commonly used to summarize various features of the large scale circulation, particularly effects from thermohaline forcing. In equation (5), 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 subgrid scale (SGS) velocity v^*. For the geopotential coordinate models used in this study, v^* is determined by the parameterization of Gent et al. (1995), where

$$v^* = -\partial_z \, (\kappa \, S_y), \tag{6}$$

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

$$\Psi(y,z) = \int \mathrm{d}x \, \kappa \, S_y - \int \mathrm{d}x \int_{-H}^{z} \mathrm{d}z' \, v, \tag{7}$$

where we set $\kappa \, S_y = 0$ at the ocean bottom. The contribution from the Gent et al. (1995) parameterization 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, 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

$$\Psi(y,\rho) = -\int \mathrm{d}x \int_{-H}^{\rho} \mathrm{d}\rho' \, v \, (\partial z/\partial \rho'), \tag{8}$$

where $\mathrm{d}z = \mathrm{d}\rho' \, (\partial z/\partial \rho')$ is the thickness between isopycnal layers.

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 (e.g., CCSM-POP and KNMI, which use salt fluxes), the streamfunction vanishes along the surface as well. In contrast, for the GFDL-MOM, GFDL-HIM, MPI, and ORCA simulations, fresh water input means that the streamfunction is generally nonzero at the ocean surface.

As the thermohaline circulation is most active in the North Atlantic basin, we examine the Atlantic MOC streamfunction. Figure 15 shows the time series for the maximum of the streamfunction at $45^\circ N$ beneath the wind driven Ekman layer. This time series exhibits the multi-centennial time scales for the spin-up of this circulation. Figure 16 shows the Atlantic meridional overturning streamfunction for the simulations averaged over years 491-500. The structure of these streamfunctions largely reflects on the time series in Figure 15, with the GFDL-MOM simulation showing the most vigorous overturning streamfunction, and KNMI the weakest.

A nontrivial MOC 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 ($\sim$15Sv), whereas others appear to reach a stable value either with a very weak salinity restoring (CCSM-POP and MPI), or somewhat stronger restoring (GFDL-MOM and 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 simulation settles into a very weak overturning circulation.

13 Salinity forcing and MOC behaviour

As seen in the bias time series of Figures 2 and 3, water masses in the deep ocean take many centuries to equilibrate

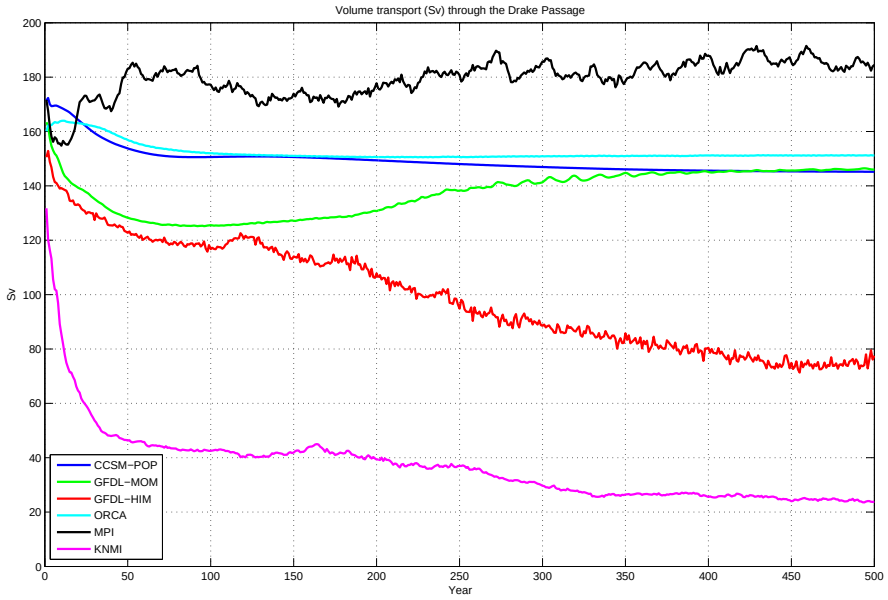

Fig. 14. Time series for the annual mean vertically integrated transport of seawater volume through the Drake Passage in the CORE-I simulations.

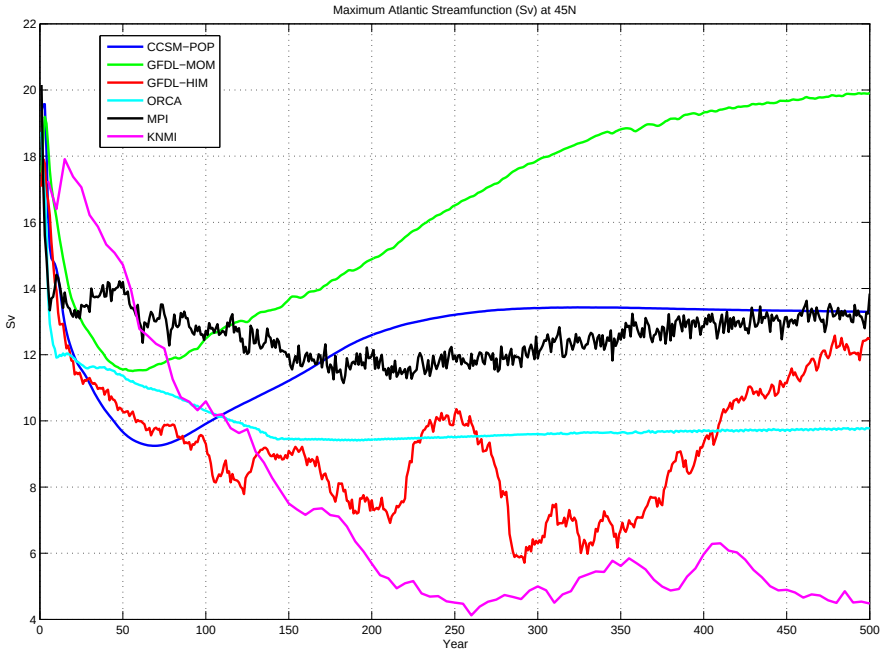

Fig. 15. Time series of the annual mean Atlantic meridional overturning streamfunction index (vertical axis) for model years 1-500 (horizontal axis) in units of Sv=10^6 m^3 sec^{-1}. The index is computed as the maximum Atlantic MOC streamfunction at $45^\circ N$ in the region beneath the wind driven 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.

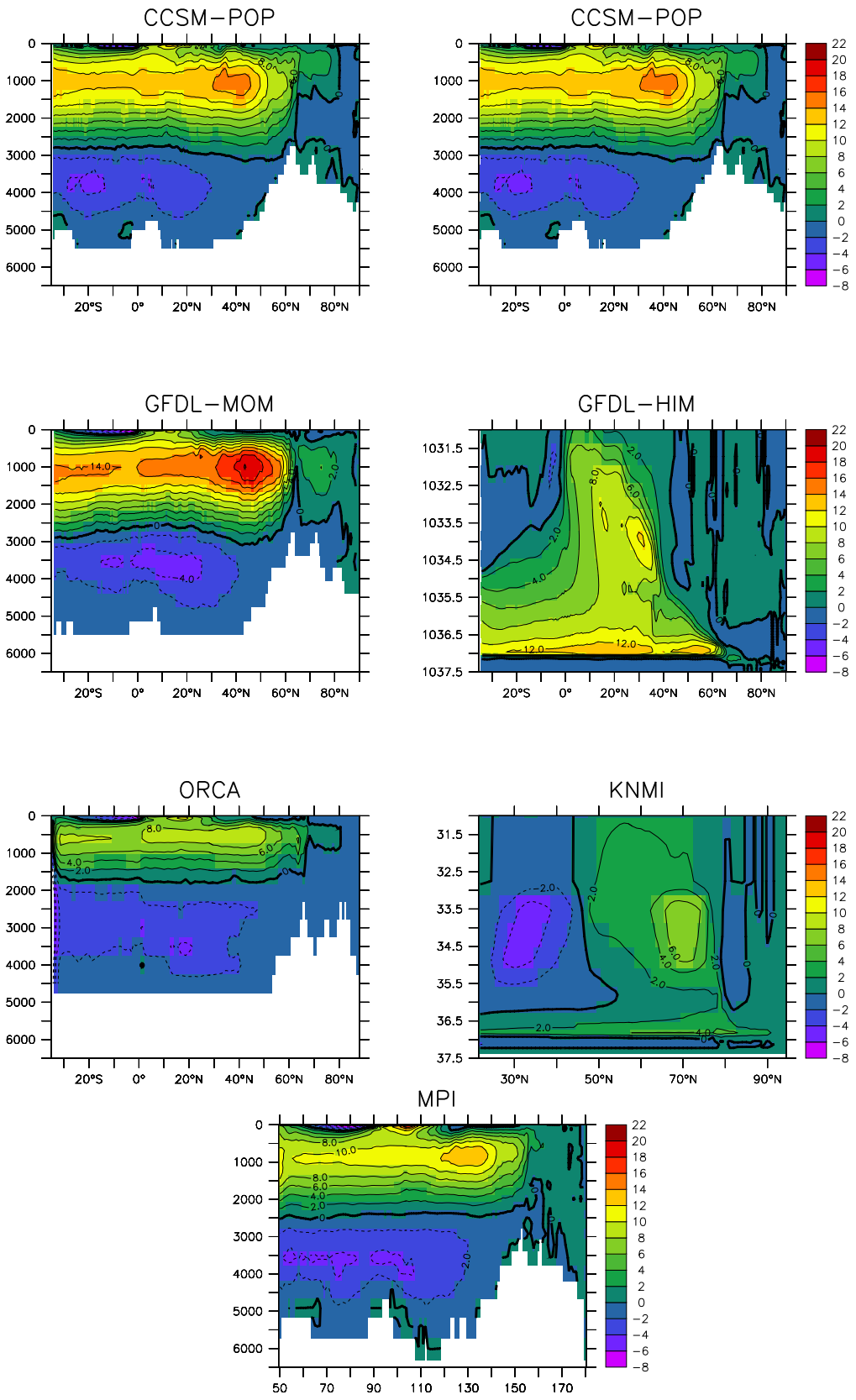

Fig. 16. The Atlantic basin meridional overturning streamfunction (in units of Sv=10^6 m^3 sec^{-1}), time averaged over years 491-500 in the CORE-I simulations. Note that the KNMI and GFDL-HIM simulations are plotted with respect to potential density referenced to 2000db ($1035 - \rho_{2000}$) as the vertical axis (equation (8)), whereas the other models plot the streamfunction with respect to depth (equation (7)). The depth models are plotted to 6500m, which illustrates the differences in the deepest level represented by the different models, with MPI going

in global ocean-ice simulations. Details of the hydrological forcing and ocean-ice interactions can be critical for the realism of the equilibrated deep water properties. 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, participating groups pursued many different experiments to test model sensitivities to salinity restoring. Each group varied their choice for salinity restoring, with the standard runs from CCSM-POP and MPI maintaining a globally weak forcing with piston velocity of 50m/4years; the standard runs from GFDL-MOM, GFDL-HIM, and ORCA strengthening the restoring to 50m/300days; and KNMI using stronger restoring in certain high latitude regions. Section B3 in the Appendix provides details of what the models chose.

The key aim of the sensitivity experiments was to determine a regime where the models exhibit a quasi-stable MOC for the multi-century scale simulations, and to identify geographical regions where the salinity forcing is most critical to maintaining MOC stability. Given that four of the seven groups were unable to stabilize the MOC without including a nontrivial restoring (i.e., piston velocity greater than 50m/4years), and noting that other groups not included in this study experienced similar difficulties, we summarize here select results from experiments conducted with GFDL-MOM, MPI, and ORCA in order to highlight difficulties. This discussion serves the reader interested in surveying issues which may arise when other ocean-ice models are run using the CORE-I design.

13.1 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 uses a piston velocity of $50\,\mathrm{m}/4$ years, which is the same as the two CCSM 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 $50\,\mathrm{m}/300$ days. GFDL-B is the standard GFDL-MOM experiment presented in other sections of this paper.

No salinity restoring occurs under sea ice in either of these experiments. However, tests using GFDL-MOM with salinity restoring globally (including under sea ice) did not qualitatively alter the following results.

We first present results from MPI in Figure 17 for the Atlantic overturning streamfunction index and the Drake Passage transport. In general, the MPI simulations are more robust to changes in the salinity restoring than the GFDL-MOM and ORCA simulations, described in the following. That is, both of the MPI simulations retain a nontrivial circulation in the Atlantic and ACC. Note, however, that the Drake Passage transport in the weakly restored simulation (the standard 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 CCSM-POP, GFDL-MOM, and ORCA results shown in Figure 14.

We next discuss the same two experiments run with the GFDL-MOM code. Figure 18 exhibits the time series for selected volume transports in the GFDL-A and GFDL-B simulations. The most striking feature is the large deviations found in GFDL-A. The North Atlantic exhibits a significant weakening of the MOC during the first 100 years, a gradual increase over the next 150 years, then a series of growing amplitude multi-decadal oscillations starting after year 250. Vertically integrated transports in other sectors of the model also show distinct behaviour for GFDL-A. In particular, note the inter-decadal oscillations in the Drake, Indonesian, and Equatorial Undercurrent transports starting after year 350. Although we have not examined the variability mechanism in GFDL-A, 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 CM2.1 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.

In conclusion, examination of the GFDL simulations indicate that the weakly restored case GFDL-A is far removed from the behaviour of the GFDL CM2.1 coupled model, as GFDL-A exhibits unphysically large and apparently unstable multi-decadal oscillations. 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.

13.2 ORCA experiments

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. We elucidate this dependence by a sequence of simulations with ORCA. Here, the surface salinity restoring is varied globally as well as in selected regions of the polar oceans. An additional experiment probed the influence of the uncertain freshwater input fields, by reducing the net CORE precipation over the Arctic and sub-Arctic oceans by 30%. In the following, we denote three different settings for the salinity restoring parameter:

- No restoring,

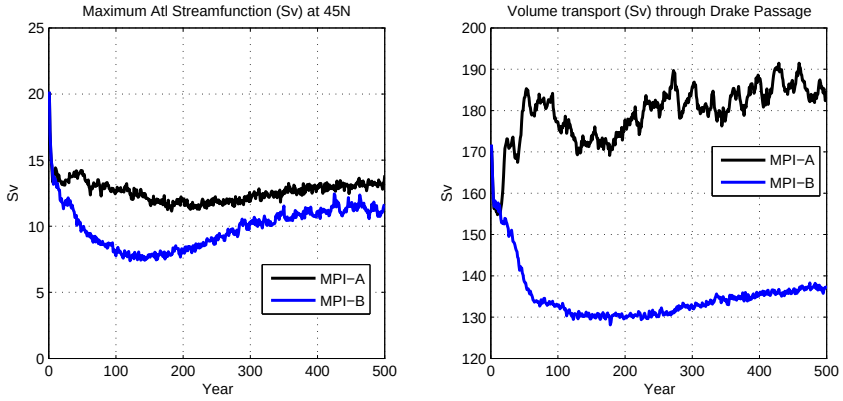

Fig. 17. Time series for the annual mean volume transports (in units of Sv=10^6 m^3 s^{-1}) from two MPI CORE-I simulations. The black line is from the MPI-A ocean-ice simulation with the salinity restoring converted to a water flux using a weak piston velocity of $50\,\mathrm{m}/4$ years (this is the standard MPI simulation discussed in other sections of this paper), and the blue line is from MPI-B ocean-ice model with larger piston velocity $50\,\mathrm{m}/300$ days. Top panel: Maximum meridional-depth overturning streamfunction at $45^\circ N$. Bottom panel: Eastward transport in the Southern Ocean through the Drake Passage.

– "Weak" restoring of surface salinity corresponding to a piston velocity of $50\,\mathrm{m}/4$ years,

– "Strong" restoring corresponding to $50\,\mathrm{m}/300$days.

We show results from the following experiments, all integrated for 500 years:

– 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 ORCA reference case for all results shown in other sections.

– ORCA-E: restoring as in ORCA-B, but with reduced Arctic precipation.

As for the GFDL-simulations, there are some strong differences in the time series of selected annual mean transports between these cases (see Figure 19). In both the "no" and the "weak" restoring cases there is a strong decline of the North Atlantic MOC during the first 100-200 years without, however, the occurrence of multi-decadal oscillations seen in GFDL-A. Such oscillations are, however, a main characteristic of the Antarctic Circumpolar Current (ACC) transport through Drake passage in these experiments (Figure 19b). In contrast to GFDL-A, these oscillations emerge already during the first decades of the model spin-up, with gradually decreasing amplitude during the course of the integrations. Similar oscillations as in the ACC transport are also exhibited by the northward transport of Antarctic Bottom Water (AABW) in the Southern Ocean (Figure 19c), with the maxima (minima) in AABW transport closely corresponding to maxima (minima) in the ACC. A causal relationship between the strength of bottom water formation and ACC-transport has been discussed by Timmermannn et al. (2005) based on experimentation with a previous version of the ORCA model. In particular, they noted that model drifts toward too strong AABW-cells and ACC-transports during the first decades of a model spin-up could be related to an increase in surface salinities near Antarctica. The ensuing excessive open ocean convection and dense columns of homogeneous water in the region of seasonal ice coverage resulted in excessive northward transports of bottom water, and a strongly enhanced baroclinic component of the ACC.

A stronger damping of sea surface salinity anomalies affects both the oscillating behaviour and strength of the AABW/ACC-system, as well as the drift in the North Atlantic MOC. An interesting result of the sequence is, however, the apparent inability of ORCA to maintain a realistic MOC transport (i.e. of order 15-20Sv) under the given CORE forcing. An inspection of the evolution of the hydrographic properties of the overflow waters spilling across the Greenland-Iceland sill indicates that the MOC evolution is closely tied to drift in the overflow density (Figure 19d). As a consequence of a progressing freshening (not shown), all model cases drift away from the observed climatological outflow density (of about 28.0). A drift in simulated Arctic Ocean salinities may be caused by various mechanisms, including

– Excessive freshwater fluxes into the Arctic basin related, for example, to the prescribed precipitation or

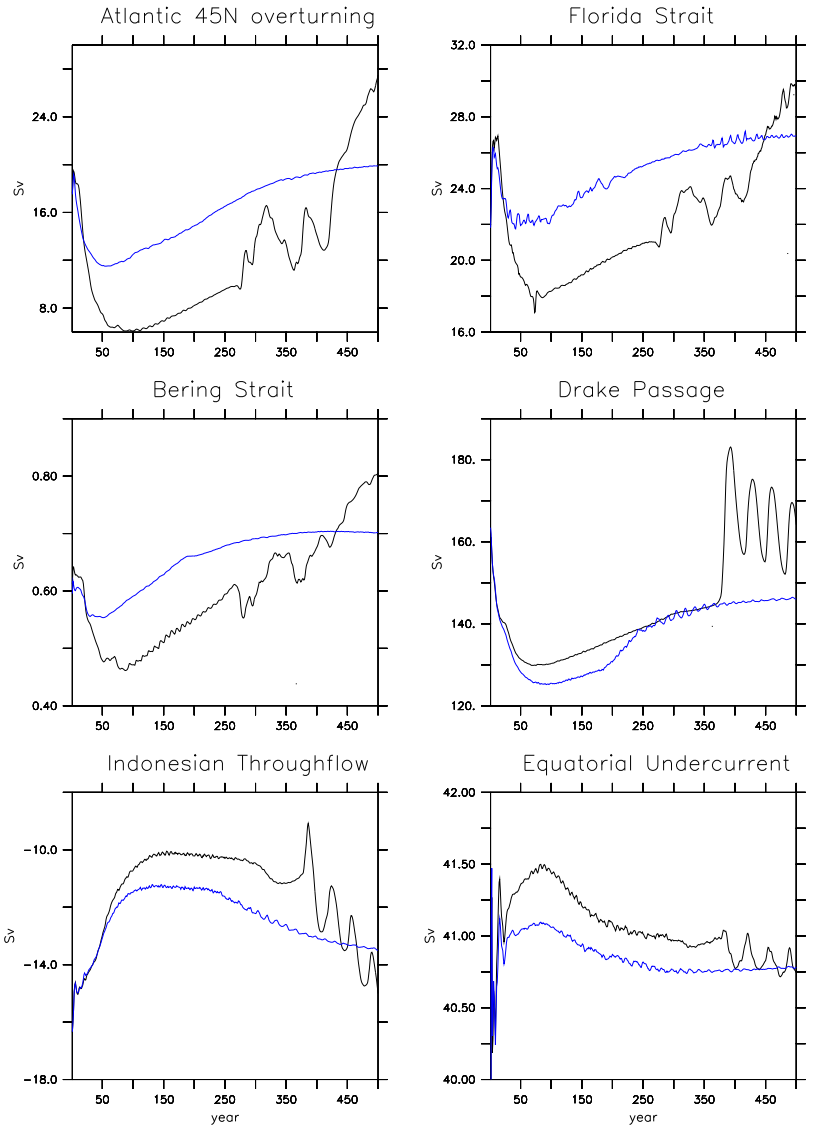

Fig. 18. Time series for the annual mean volume transports (in units of Sv=10^6 m^3 s^{-1}) for selected geographic regions from two GFDL CORE-I simulations. The black line is from the GFDL-A ocean-ice simulation with the salinity restoring converted to a water flux using a weak piston velocity of $50\,\mathrm{m}/4$ years, and the blue line is from GFDL-B ocean-ice model with larger piston velocity $50\,\mathrm{m}/300$ days. Top left: Maximum meridional-depth overturning streamfunction at $45^\circ N$. Top right: Northward transport in the Florida Current between Florida and Cuba, defined as $(81.5^\circ W, 23^\circ N - 27^\circ N)$ (observational estimates from Leaman et al., 1987, are between 29Sv-35Sv). Middle left: Northward transport through the Bering Strait, defined as $(167^\circ W - 171^\circ W, 65^\circ N)$ (an observational estimate from Roach et al., 1995, is roughly 0.8Sv). Middle right: Eastward transport in the Southern Ocean through the Drake Passage (an observational estimate from Cunningham et al., 2003, is roughly 135Sv). Lower left: Transport through the Indonesian Islands, defined here as $(100^\circ E - 140^\circ E, 6^\circ S)$ (an observational estimate from Gordon et al., 2003, is roughly 10Sv). Lower right: Eastward transport within the equatorial undercurrent at $(155^\circ W, 3^\circ S - 3^\circ N, 0 - 350\mathrm{m})$ (observational estimates from Lukas and Firing, 1984; Sloyan et al., 2003, range between 24Sv-36Sv).

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, a restoring of sea surface salinity can only partly remedy the effect of possible errors in the high-latitude freshwater budget. A particularly interesting aspect of Figure 19 is the rather tight relationship between mid-latitude MOC transport and overflow density, noted before in various model simulations reported by Latif et al. (2006). Apparently, neither changes in the restoring configuration nor in the prescribed precipitation field (ORCA-E) appear to affect the relation between density and MOC. The main effect of variations in these model choices is to determine the amplitude of the drift realized by a particular simulation.

13.3 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? To pursue this question, in addition to investigating details of the hydrological forcing, many experiments with GFDL-MOM explored modifications to the physical parameters, mostly related to the mesoscale eddy parameterizations and horizontal friction. No significant modification to the results shown already were found. However, one hypothesis does present itself after considering characteristics of the seven model configurations. Namely, those two models (CCSM-POP and MPI) with the finest horizontal grid resolution in the North Atlantic are 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 unstable behaviour.

As detailed in Section 13.2, a significant difficulty with the 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 suggests that ocean-ice models of finer high latitude resolution, such as those explicitly representing mesoscale eddies and those possessing a more detailed representation of the complex land-sea boundaries, will exhibit a *more* stable large-scale overturning circulation under the presence of fluctuations in the hydrological cycle.

14 Discussion and concluding remarks

Simulations with global ocean-ice models are perhaps the most difficult of the component models to run in isolation from the other components of the climate system. In particular, coupled atmosphere-land simulations, given an SST dataset, is more straightforward than running ocean-ice simulations with a prescribed atmospheric state. There are two main reasons for this distinction. First, fluxes used to force the ocean-ice system are less well known than the relatively well observed SST used to derive atmospheric fluxes over the ocean. 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.

There are two general conclusions one can draw from 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. 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 to the goals of, for example, climate change science. It is notable that this trend is being followed by scientists at GFDL, 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 ic modelling research.

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. This second conclusion is taken by many research groups, thus motivating our proposal for Coordinate Ocean-ice Reference Experiments (COREs) in hopes of facilitating collaborative research and model development. That is, we aim with CORE to bring the ocean-ice models onto a common platform for surface forcing, from which remaining model differences can be the main reason for diverging simulation behaviour.

Given our proposed CORE simulations, we illustrated results from seven ocean-ice models run with the CORE-I pro-

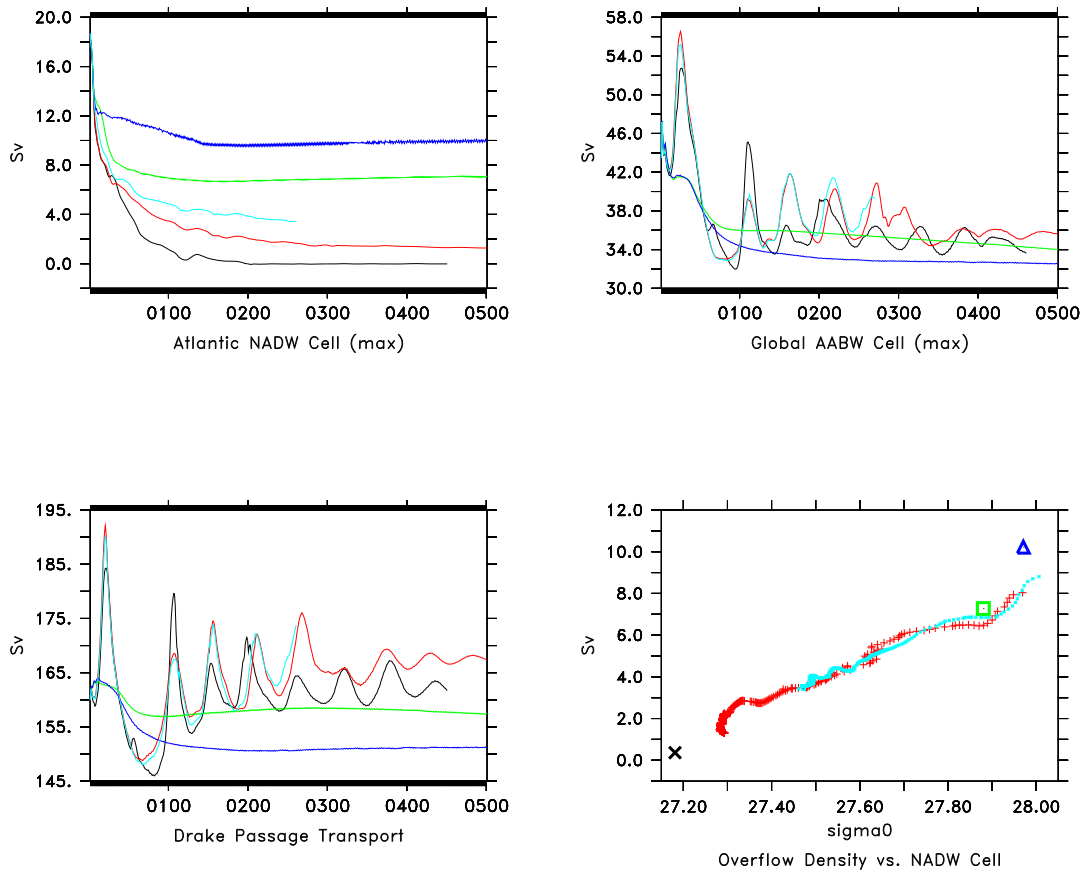

Fig. 19. Time series for the annual mean volume transports from a sequence of ORCA experiments: ORCA-A (black), ORCA-B (red), ORCA-C (green), ORCA-D (dark blue; serving as the reference experiment), ORCA-E (light blue): (a) maximum MOC transport in the North Atlantic, (b) ACC transport through Drake passage, (c) maximum transport of the AABW-cell in the southern hemisphere; (d) scatter plot of yearly values (beginning at year 20) of annual mean MOC transport and overflow density 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 ORCAS-A (black cross), ORCA-C (green square), ORCA-D (blue triangle).

tocol, 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 hypothesis held reasonably well (with some exceptions) for upper ocean tropical metrics, but was far less valid when examining deeper properties and transports, especially in the high latitudes. The nontrivial differences arising from the simulations highlight the need to further examine the mechanisms accounting for the divergence.

A large part of this project was aimed at answering questions about the feasibility of bringing ocean-ice models onto a common design framework. These questions have been answered in the affirmative by the existence seven contributing groups having followed the CORE-I protocol. Questions regarding the merit 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 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 in this category, arise from new questions raised when simulations are compared and contrasted. Indeed, many questions generaly go unasked in the absence of comparisons. These questions then motivate further focused study conducted in a more controlled setting that aim to elucidate mechanisms. In this manner, we propose that the preliminary experience with CORE-I documented in this paper provide a solid foundation for deeper study of MOC stability in ocean-ice simulations. It also supports studies concerning the suitability of a prescribed atmospheric state for running ocean-ice models for meeting the goals of ocean-ice modelling, and in particular the utility of the Large and Yeager (2004) dataset. In this regard, CORE

has great merit and should be the for further global ocean-ice model comparison projects.

Acknowledgements. This paper was prompted by discussions occurring at a CLIVAR workshop on ocean climate modelling held June 16-18, 2004 at the Geophysical Fluid Dynamics Laboratory in Princeton USA. The workshop was organized by the CLIVAR Working Group for Ocean Model Development (WGOMD). The WGOMD has facilitated a Pilot-OMIP since 2001. This project involved a handful of models and was itself based on an earlier comparison of German ocean-ice models (Fritzsch et al., 2000). Both of these earlier projects used the forcing dataset compiled by Röske (2005). Experience with the German project, the Pilot-OMIP, the Röske (2005) dataset, as well as discussions at the GFDL workshop were all critical for development of COREs.

We thank David Bi, 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. SMG wishes to thank CSIRO Marine and Atmospheric Research in Australia for its hospitality during 2005, where part of this paper was written.

Appendix A: Contributing models

Table 1 summarizes 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 CCSM-POP

This is the ocean-ice component to the Community Climate System Model (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. The ocean component is coupled to the sea ice model of 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 45 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 CCSM-POP CORE-I simulation. No systematic tests have been run to examine this conjecture.

The CCSM-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, and a 7200s forward time step for the dissipative physical processes. The barotropic equations are updated implicitly in time using a leap-frog for the tendencies with 3600s time step (Dukowicz and 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 feeback. The CCSM-POP sea ice model employs an albedo following the approach documented in Briegleb et al. (2002). In particular, the CCSM-POP ice model distinguishes between visible (wavelengths $< 0.7\mu$m) and near-infrared (wavelengths $> 0.7\mu$m), since snow and ice spectral reflectivities are significnatly 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 downwelleing 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 CCSM-POP albedo, please refer to Briegleb et al. (2002).

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 CCSM-HYCOM

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 fol-

MODEL	VERTICAL (NUMBER)	ARAKAWA	HORIZONTAL	BAROTROPIC	BAROCLINIC/TRACER
CCSM-HYCOM	hybrid ρ, p, σ	C	displaced(???)	explicit(???)	???
CCSM-POP	z(40)	B	displaced(320x384)	implicit(3600s)	leap-frog (3600s)
GFDL-HIM	$\rho_2(50)$	C	tripolar(360x210)	explicit(60s)	pred-corrector (7200s+3600s)
GFDL-MOM	z(50)	B	tripolar(360x200)	explicit(80s)	staggered (7200s)
KNMI	$\rho_2(16)$	C	tripolar(180x95)	explicit(72s)	leap-frog (1440s)
MPI	z(40)	C	displaced(256x220)	semi-implicit(4800s)	semi-implicit(4800s)
ORCA	z(31)	C	tripolar(182x149)	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; arrangment 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 and GFDL-HIM, the vertical is discretized according to layers of potential density referenced to 2000dbar. 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
CCSM-POP	POP1.4 and CSIM4	March 2004
CCSM-HYCOM	???	Jun-Aug 2007
GFDL-MOM	MOM4p0d	Feb-Apr 2005
GFDL-HIM	HIM-Fortran M-release	Feb-Mar 2007
KNMI	MICOM 2.9	June 2006
MPI	mpiom-1.2.3	Oct-Dec 2006
ORCA	???	???

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.

lowing 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 (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

The version of the ocean-ice model 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). Furthermore, the convection algorithm and decabbeling 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.

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; Jungclaus 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 level, with 20 in the upper 600m. The horizontal resolution gradually varies between a minimum of 12km close to Greenland and 150km 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 initialized with thickness and concentration depended on the initial sea surface temperture. The MPI CORE experiments use the sea ice albedo parameterisation suggested by Large and Yeager (2004).

A7 ORCA

This is a configuration of Version 9 of the OPA z-coordinate ocean model of Madec et al. (1999) coupled to the ice model of Fichefet and Maqueda (1999). The ocean model component has 31 vertical levels and the horizontal resolution is slightly coarser than 2° (182 x 149 x 31 grid cells), refined in the tropics to 0.5°. The horizontal grid is tripolar according to the methods of Madec and Imbard (1996). The model updates the tracer and baroclinic velocity using a leap-frog time step of 5400s for the time tendency. The barotropic equations are updated implicitly in time with a leap-frog 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 removing the need to employ salt fluxes. The ocean and sea ice models are part of the French climate modelling efforts in support of anthropogenic climate change studies at the Institut Pierre-Simon Laplace in Paris.

Appendix B: CORE-I experimental design

We summarize 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 initializing 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 fresh 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 initialized with zero velocity. The initial temperature and salinity fields are taken from the January temperature and salinity profiles of Conkright et al. (2002) for all basins except the Arctic, and Steele et al. (2001) for the Arctic. The exception is that KNMI used initial conditions based on Levitus et al. (1994). Temperature and salinity fields are interpolated to the respective ocean model grid.

Ice models are initialized in the following manners.

– CCSM-POP: 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: Zero initial ice concentration and zero ice volume.

– 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 initialized 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). 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. For models using the salt flux boundary condition (CCSM-POP and KNMI), fresh water from precipitation, evaporation, and river runoff is converted to an implied salt flux via

$$F^{(\text{salt})}_{(\text{implied})} = q_{\text{w}}\, S_o, \tag{9}$$

where S_o is a global reference salinity taken at

$$S_o = 35\text{psu}. \tag{10}$$

Likewise, salinity restoring is applied to the ocean according to the restoring salt flux

$$F^{(\text{salt})}_{(\text{restore})} = V_{\text{piston}}\,(S^{(\text{data})} - SSS) \tag{11}$$

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 $F^{(\text{salt})}_{(\text{restore})}$ 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/CORE.html.

They are derived from the 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 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 simulation, but note that grid cells may have partial ice cover such that some restoring does take place over those cells. The CCSM-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 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 CCSM-POP simulation applies a global normalization so that over the course of a year,

$$\text{net salt input} = \sum \left(F^{(\text{salt})}_{(\text{implied})} + F^{(\text{salt})}_{(\text{restore})}\right) \mathrm{d}x\, \mathrm{d}y \tag{12}$$

is damped towards zero, based on the value of this global sum from the previous model year. KNMI did not apply a normalization, and such may account for the fresh drift seen in the salinity bias time series (Figure 3).

The GFDL-HIM, GFDL-MOM, MPI, and 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 (11) is converted to a fresh water flux $q^{\text{restore}}_{\text{w}}$ via

$$q^{\text{restore}}_{\text{w}} = -F^{(\text{salt})}_{(\text{restore})}/SSS. \tag{13}$$

Hence, water is added ($q^{\text{restore}}_{\text{w}} > 0$) where the model's prognostic SSS is greater than $S^{(\text{data})}$, and subtracted ($q^{\text{restore}}_{\text{w}} < 0$)

where $SSS < S^{(\text{data})}$. For the ORCA simulations, no extra water normalization is applied, and this may account for the fresh drift seen in the salinity bias time series (Figure 3). In the GFDL and MPI simulations, a global normalization is applied at each time step so that over the ocean surface,

$$\text{net water input to ocean} = \sum (P - E + R + q_{\text{w}}^{\text{restore}}) \quad (14)$$

is set to zero.[5] Note that this normalization has the unphysical effect of adding a water flux nonlocally, even to a region that may have zero salinity bias. Nonetheless, it is necessary to avoid longterm drifts in the salinity and sea level.

Choices made for the various models are summarized in Table 3.

Appendix C: Forcing the ocean-ice models

The purpose of this appendix is to highlight details of the forcing which have proven important in setting up the CORE simulations.

C1 Air-sea fluxes

This section follows closely Section 2 from Large and Yeager (2006).

The ocean and sea ice system is forced by fluxes of freshwater, F_s, heat, Q_s, and momentum, $\boldsymbol{\tau}$. Partitioning these fluxes between the ocean, with area fraction f_o, and sea ice, with fraction $1 - f_o$, leads to

$$\boldsymbol{\tau} = f_o\, \boldsymbol{\tau}_{as} + (1 - f_o)\, \boldsymbol{\tau}_{io} \quad (15)$$
$$F_s = f_o\, F_{as} + (1 - f_o)\, F_{io} + R \quad (16)$$
$$Q_s = f_o\, Q_{as} + (1 - f_o)\, Q_{io}, \quad (17)$$

where the subscript as refers to air-sea fluxes, io refers to ice-ocean, and R is the river runoff. As with Large and Yeager (2004) and Large and Yeager (2006), we focus here on air-sea fluxes, in which

$$Q_{as} = Q_S + Q_L + Q_E + Q_H \quad (18)$$
$$F_{as} = P + E, \quad (19)$$

where all fluxes are positive when momentum, water, or heat enter the ocean. Solar radiation is comprised of the shortwave term, Q_S, with wavelengths between 0.3μ and 3μ, and longwave Q_L with wavelengths up to 50μ. Bulk formulae parameterize the turbulent fluxes (wind stress $\boldsymbol{\tau}$, evaportation E, latent heat Q_E, and sensible heat Q_S) 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 $\mathbf{U}_o$)

$$\boldsymbol{\tau} = \rho\, C_D\, |\Delta \mathbf{U}| \Delta \mathbf{U} \quad (20)$$
$$E = \rho\, C_E\, [q - q_s(SST)]\, |\Delta \mathbf{U}| \quad (21)$$
$$Q_E = \Lambda_v\, E \quad (22)$$
$$Q_H = \rho\, c_p\, C_H\, (\theta - SST)\, |\Delta \mathbf{U}|. \quad (23)$$

In these equations, $\Delta \mathbf{U} = \mathbf{U} - \mathbf{U}_o$ is the difference between the atmospheric winds and ocean current (following Pacanowski (1987)).[6] The heat specific capacity of air has a value $c_p \approx 1000.5$ J/kg, and $\Lambda_v = 2.5 \times 10^6$J/kg is the latent heat of vaporization for water. The surface air is assumed to be saturated using a specific humidity $q_s(SST)$ that is a function of the SST. Large and Yeager (2004) detail the methodology for computing the neutral 10m transfer coefficients for drag, C_D, sensible heat transfer, C_H, and evaporation, C_E, as well as the saturated specific humidity $q_s(SST)$.

The shortwave radiative flux is computed from the solar insolation Q_I incident on the ocean surface, minus that which is reflected from the ocean surface

$$Q_S = Q_I\, (1 - \alpha) \quad (24)$$

with the albedo $\alpha \approx 0.07$ appropriate for the ocean (Payne, 1972). The longwave radiation into the ocean

$$Q_L = Q_A - \sigma\, SST^4 \quad (25)$$

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

In the following sections of this Appendix, we comment more on these fluxes, as well as the datasets used for their computation.

C2 The Large and Yeager (2004) dataset

Table 4 summarizes 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).

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

[6] Pacanowski (1987) noted that it is important to take into account surface ocean currents in the computation of wind stress, es-

EXPERIMENT	SALT/WATER FLUX	$V_{\text{piston}}(\text{m d}^{-1})$	restore region	normalized hydrology	LENGTH (Y)
CCSM-POP	salt	50/(4*365)	global	yes	500
GFDL-MOM-A	water	50/(4*365)	global	yes	500
GFDL-MOM-B (control)	water	50/(300)	global	yes	500
GFDL-HIM	water	50/(300)	global	yes	500
KNMI	salt	regional	regional	no	500
ORCA	water	50/(300)	global	no	500
MPI-A (control)	water	50/(4*365)	global	yes	500
MPI-B	water	50/(300)	global	yes	500

Table 3. Summary of choices made for the surface water/salt forcing in the various experiments. Note that more than one experiment is mentioned for GFDL-MOM, MPI, and ORCA, corresponding to the perturbation experiments described in Section 13. The control experiments for both models are noted, as these are the settings used for all other sections. Column 2 indicates whether the experiment used salt fluxes or water fluxes for hydrological forcing; column 3 gives the piston velocity in metres per day which determines the strength of the restoring used for surface salinity/water forcing; column 4 notes the region over which the salinity/water restoring fluxes are set. The symbol $*$ for the ORCA simulation refers to coefficients that regionally vary according to the discussion in Section 13.2. The "regional" entries for the KNMI simulations refer to the 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 5 refers to whether the experiment normalized the precipitation minus evaporation plus runoff in some manner to reduce drift; and column 6 gives the length of the simulation.

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)
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.

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).

Modelers 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.

pecially where ocean currents reach large values, such as in the equatorial Pacific. As shown in Table 5, the models in this study all use this form for computing the wind stress.

C4 Normalization of moisture fluxes

To keep the globally integrated moisture flux (precipitation minus evaporation plus river runoff plus restoring) equal to zero, thus reducing long term drifts, some models apply a normalization, whereas others do not. We summarize the choices made for the simulations in Table 3.

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, as indicated in Table 5, 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 Referencing the meteorological data

All models in this paper use the same bulk formulae and atmospheric state in a two step process detailed in Large and Yeager (2004) and Large (2005). First, the meteorological data are referenced to a 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. A reference height of 10m 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 from 2m to 10m is done off-line using observed SST, so that the CORE forcing is given as the 10m wind, potential temperature and humidity in Table 4. 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 with the 10m bulk formulae. For accuracy in the turbulent flux calculations, it is essential that data and coefficients be referenced to the same height. The height shifting and flux algorithms are closely related and should be performed in a consistent manner.

With 10m meteorological data used in the flux calculations of equations (20)–(23), the transfer coefficents C_D, C_H and C_E need to be 10m coefficients at 10m atmospheric stability, represented by δ/L, where $\delta = 10$m and L is the Monin-Obukhov depth. The Monin-Obukhov depth is defined by the turbulent fluxes

$$L^{-1} = \kappa g \sqrt{\rho} |\boldsymbol{\tau}|^{-3/2} \left(\frac{Q_H}{C_p\, \theta\, (1 + .608\, q)} + \frac{Q_E}{\Lambda_v\, (q + .608^{-1})} \right), \tag{26}$$

where $\kappa = 0.4$ is the von Karman constant and g is the gravitational acceleration. The transfer coefficients are computed from the respective $\delta = 10$m neutral coefficients $C^N_{D\delta}$, $C^N_{H\delta}$, and $C^N_{E\delta}$, according to

$$C_D = \frac{C^N_{D\delta}}{\left(1 - \kappa^{-1} \sqrt{C^N_{D\delta}}\, \psi_m(\delta/L)\right)^2} \tag{27}$$

$$C_H = \sqrt{C_D / C^N_{D\delta}} \left(\frac{C^N_{H\delta}}{1 - \kappa^{-1} \sqrt{C^H_{D\delta}}\, \psi_h(\delta/L)} \right) \tag{28}$$

$$C_E = \sqrt{C_D / C^N_{D\delta}} \left(\frac{C^N_{H\delta}}{1 - \kappa^{-1} \sqrt{C^H_{D\delta}}\, \psi_h(\delta/L)} \right). \tag{29}$$

Here, ψ_m and ψ_h are empirical functions of stability defined by

$$\psi_h(\delta/L > 0) = -5\, \delta/L \qquad \text{stable} \tag{30}$$

$$\psi_m(\delta/L > 0) = -5\, \delta/L \qquad \text{stable} \tag{31}$$

$$\psi_h(0) = 0 \qquad \text{neutral stability} \tag{32}$$

$$\psi_m(0) = 0 \qquad \text{neutral stability} \tag{33}$$

$$\psi_h(\delta/L < 0) > \psi_m(\delta/L < 0) > 0 \qquad \text{unstable.} \tag{34}$$

Unfortunately, only the CCSM-HYCOM and CCSM-POP fluxes were computed with the correct stablity shift. An error in the Fortran code used by the other models gave incorrect heat and moisture coefficients C^*_H and C^*_E, where the erroneous coefficients are given by

$$C^*_H / C_H = \frac{1 - \kappa^{-1} \sqrt{C^N_{D\delta}}\, \psi_m(\delta/L)}{1 - \kappa^{-1} \sqrt{C^H_{D\delta}}\, \psi_h(\delta/L)} \tag{35}$$

$$C^*_E / C_E = \frac{1 - \kappa^{-1} \sqrt{C^N_{D\delta}}\, \psi_m(\delta/L)}{1 - \kappa^{-1} \sqrt{C^E_{D\delta}}\, \psi_h(\delta/L)} \tag{36}$$

Most of the ocean is cooled by negative Q_H and Q_E, which makes the Monin-Obukhov depth L negative, the atmosphere unstable, and thus both ψ_h and ψ_m positive. The code error is largest in regions where high SST and/or low wind speeds make a signifcantly unstable atmosphere, with ψ_h typically twice ψ_m, such that the error ratios (35) and (36) become greater than unity. The largest heat flux error is found in these highly unstable regions and areas of large latent heat flux loss. For example, using observed SST, the net heat flux into the ocean is reduced by $8 - 10$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). Elsewhere, the effect is smaller. Poleward of about 40 degress latitude the heat flux reduction is typically less than 3W m^{-2}. The expected SST signal is for CCSM to be warmer by at most $0.3^\circ C$.

C7 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.

C8 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.

C9 Shortwave penetration

The evolution of SST depends critically on the aborption of solar shortwave radiation and demands that this flux be separated from the other heat fluxes (Rosati and Miyakoda, 1988). How this radiation is attentuated into the ocean differs in the models. As described in Appendix B to Danabasoglu et al. (2006), shortwave absorption in CCSM 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 places all of its shortwave raditation into the bulk mixed layer. MPI and 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 summarize these choices in Table 5.

Appendix D: Metrics used to evaluate the simulations

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

- Figures 2 and 3: Globally averaged drift of the annual mean temperature and salinity as a function of depth and time over the 500 years 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 4: 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 5: Monthly values of the heat content (vertically integrated temperature) over the upper 250m versus sea surface temperature taken at Ocean Weathership ECHO ($48°W, 35°N$). Results from simulated year 50 are compared.
- Figure 6: 500 year time series for the annual mean sea ice area in the Northern Hemisphere and Southern Hemisphere.
- Figures 7 and 8: Maps of the sea ice concentration (area sea ice per grid cell area) in March and September averaged over years 491-500.
- Figure 9: Time mean of year 491-500 temperature on the equator in the Pacific.
- Figure 10: Time mean of year 491-500 zonal current on the equator in the Pacific.
- Figures 11 and 12: 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 13: Northward heat tranport directly computed in the models for the global ocean averaged over years 491-500.
- Figure 14: 500 year time series for the annual mean vertically integrated transport of seawater volume through the Drake Passage
- Figure 15: 500 year time series of the annual mean Atlantic meridional overturning streamfunction index. The index is computed as the maximum Atlantic MOC streamfunction at $45°N$ in the region beneath the wind driven Ekman layer.
- Figure 16: 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.

References

Arakawa, A. and Lamb, V. R.: The UCLA general circulation model, in Methods in Computational Physics: General Circulation Models of the Atmosphere, edited by J. Chang, vol. 17, pp. 174–265, Academic Press, 1977.

Barnier, B.: Forcing the ocean, in Ocean Modeling and Parameterization, edited by E. P. Chassignet and J. Verron, vol. 516 of *NATO ASI Mathematical and Physical Sciences Series*, pp. 45–80, Kluwer, 1998.

Barnier, B., Siefridt, L., and Marchesiello, P.: Thermal forcing for a global ocean circulation model using a three-year climatology of ECMWF analyses., Journal of Marine Research, 6, 363–380, 1995.

Beljaars, A.: The parameterization of surface fluxes in large-scale models under free convection, Quarterly Journal of the Royal Meteorological Society, 121, 255–270, 1994.

Bentsen, M., Drange, H., Furevik, T., and Zhou, T.: Simulated variability of the Atlantic meridional overturning circulation, Climate Dynamics, 22, 701–720, 2004.

Bleck, R., Rooth, C., Hu, D., and Smith, L. T.: Salinity-driven thermocline transients in a wind and thermohaline forced isopycnic coordinate model of the North Atlantic, Journal of Physical Oceanography, 22, 1486–1505, 1992.

Briegleb, B., Hunke, E., Bitz, C., , and Lipscomb, W.: Description of the Community Climate System Model Version 2 Sea Ice Model, NCAR Technical Report, available online at http://www.ccsm.ucar.edu, 60pp., 2002.

EXPERIMENT	MOMENTUM STRESS	SEA LEVEL PRESSURE	WATER TYPE FOR SHORTWAVE
CCSM-POP	$\rho_a \, C_D \, \Delta \mathbf{U} \, \lvert \Delta \mathbf{U} \rvert$	dynamically inactive	chlorophyll climatology
GFDL	$\rho_a \, C_D \, \Delta \mathbf{U} \, \lvert \Delta \mathbf{U} \rvert$	dynamically inactive	chlorophyll climatology
KNMI	$\rho_a \, C_D \, \Delta \mathbf{U} \, \lvert \Delta \mathbf{U} \rvert$	dynamically inactive	shortwave in mixed layer
ORCA	$\rho_a \, C_D \, \Delta \mathbf{U} \, \lvert \Delta \mathbf{U} \rvert$	dynamically inactive	uniform Jerlov water type I
MPI	$\rho_a \, C_D \, \Delta \mathbf{U} \, \lvert \Delta \mathbf{U} \rvert$	dynamically inactive	uniform Jerlov water type I

Table 5. Some more details of the choices used for forcing the various experiments. Column 2 indicates the form taken to compute momentum stress on the ocean surface, where $\Delta \mathbf{u}$ refers to the difference between ocean currents and atmospheric winds; column 3 notes whether seal level pressure, provided by the Large and Yeager (2004) dataset, dynamically forces the ocean barotropic mode or not; column 4 refers to the general manner whereby shortwave radiation is attenuated as it penetrates into the ocean.

Briegleb, B., Bitz, C., Hunke, E., Lipscomb, W., Holland, M., Schramm, J., and Moritz, R.: 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.ccsm.ucar.edu, 2004.

Bryan, F.: Parameter sensitivity of primitive equation ocean general circulation models, Journal of Physical Oceanography, 17, 970–985, 1987.

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., and Smith, R.: The Community Climate System Model: CCSM3, Journal of Climate, 19, 2122–2143, 2006.

Comiso, J.: Bootstrap sea ice concentrations for NIMBUS-7 SMMR and DMSP SSM/I, June to September 2001, National Snow and Ice Data Center, Digital Media, 1999 (updated 2005).

Conkright, M., Antonov, J., Baranova, O., Boyer, T., Garcia, H., Gelfeld, F., Johnson, D., Locarnini, R., Murphy, P., O'Brien, T., Smolyar, I., and Stephens, C.: World Ocean Database 2001, Volume 1: Introduction, NOAA Atlas NESD'is 42, U.S. Government Printing Office 13, NOAA, Washington, D.C., 167 pp., 2002.

Cox, M. and Bryan, K.: A numerical model of the ventilated thermocline, Journal of Physical Oceanography, 14, 674–687, 1984.

Cunningham, S., Alderson, S., King, B., and Brandon, M.: Transport and variability of the Antarctic Circumpolar Current in Drake Passage, Journal of Geophysical Research, 108, Art. 8084, 2003.

Danabasoglu, G., Large, W. G., Tribbia, J. J., Gent, P. R., Briegleb, B. P., and McWilliams, J. C.: Diurnal Coupling in the Tropical Oceans of CCSM3, Journal of Climate, 19, 2347–2365, 2006.

de Verdière, C. and Huck, T.: Baraoclinic instability: A wavemaker for oceanic interdecadal variability, Journal of Physical Oceanography, 29, 865–892, 1999.

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., and Zhang, R.: GFDL's CM2 Global Coupled Climate Models - Part 1: Formulation and Simulation Characteristics, Journal of Climate, 19, 643–674, 2006.

Doney, S. C. and Hecht, M. W.: Antarctic Bottom Water formation and deep water chlorofluorocarbon distributions in a global ocean climate model, Journal of Physical Oceanography, 32, 1642–1666, 2002.

Dukowicz, J. K. and Smith, R. D.: Implicit free-surface method for the Bryan-Cox-Semtner ocean model., Journal of Geophysical Research, 99, 7991–8014, 1994.

England, M. H.: The age of water and ventilation timescales in a global ocean model, Journal of Physical Oceanography, 25, 2756–2777, 1995.

England, M. H. and Maier-Reimer, E.: Using chemical tracers to assess ocean models, Reviews of Geophysics, 39, 29–70, 2001.

Fichefet, T. and Maqueda, M. M.: Modelling the influence of snow accumulation and snow-ice formation in the seasonal cycle of the Antarctic sea ice cover, CD, 15, 251–268, 1999.

Fritzsch, B., Gerdes, R., Hiller, W., Latif, M., Legutke, S., Maier-Reimer, E., Olbers, D., and Röske, F.: The Ocean Model Intercomparison Project (OMIP), Alfred Wegener Institude for Polar and Marine Research, Bemerhaven, and Max-Planck-Institute for Meteorology, Hamburg, p. 139 pages, 2000.

Gaspar, P., Gregoris, Y., and Lefevre, J.: 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, 16 179–16 193, 1990.

Gates, W.: AMIP: The Atmosphere Model Intercomparison Project, BAMS, 73, 1962–1970, 1993.

Gent, P. R., Willebrand, J., McDougall, T. J., and McWilliams, J. C.: Parameterizing eddy-induced tracer transports in ocean circulation models., Journal of Physical Oceanography, 25, 463–474, 1995.

Gerdes, R., Griffies, S., and Hurlin, W.: Reaction of the oceanic circulation to increased melt water flux from Greenland - a test case for ocean general circulation models, Clivar Exchanges, 10, 2005a.

Gerdes, R., Hurka, J., Larcher, M., Kauker, F., and Koeberle, C.: Simulated history of convection in the Greenland and Labrador Seas 1948-2001, in Climate Variability of the Nordic Seas, edited by B. C. for Climate Research, pp. 221–238, American Geophysical Union, Bergen, Norway, 2005b.

Gerdes, R., Hurlin, W., and Griffies, S.: Sensitivity of a global ocean model to increased run-off from Greenland, Ocean Modelling, 12, 416–435, 2006.

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., and Zhang,

R.: GFDL's CM2 Global Coupled Climate Models-Part 2: The Baseline Ocean Simulation, Journal of Climate, 19, 675–697, 2006.

Gordon, A., Susanto, R., and Vranes, K.: Cool Indonesian throughflow as a consequence of restricted surface layer flow, Nature, 425, 824–828, 2003.

Greatbatch, R. and Peterson, K.: Interdecadal variability and oceanic thermohaline adjustment, Journal of Geophysical Research, 101, 20 467–20 482, 1996.

Griffies, S. M.: Fundamentals of ocean climate models, Princeton University Press, Princeton, USA, 518+xxxiv pages, 2004.

Griffies, S. M., Pacanowski, R., Schmidt, R., and Balaji, V.: Tracer Conservation with an Explicit Free Surface Method for z-coordinate Ocean Models, Monthly Weather Review, 129, 1081–1098, 2001.

Griffies, S. M., Harrison, M. J., Pacanowski, R. C., and Rosati, A.: A Technical Guide to MOM4, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, USA, 337 pp, 2004.

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., and Zhang, R.: Formulation of an ocean model for global climate simulations, Ocean Science, 1, 45–79, 2005.

Hallberg, R. and Gnanadesikan, A.: 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, 2006.

Hallberg, R. W.: The suitability of large-scale ocean models for adapting parameterizations of boundary mixing and a description of a refined bulk mixed layer model, in Near-Boundary Processes and Their Parameterization, edited by P. Müller and C. Garrett, Proceedings of the 13th 'Aha Huliko'a Hawaiian Winter Workshop, pp. 187–203, University of Hawaii at Manoa, 2003.

Haney, R. L.: Surface Thermal Boundary Conditions for Ocean Circulation Models, Journal of Physical Oceanography, 1, 241–248, 1971.

Hazeleger, W., Severijns, C., Haarsma, R., Selten, F., and Sterl, A.: SpeedO - model description and validation of a flexible coupled model for climate studies, KNMI-Technical Report, R257, 1–38, 2003.

Hibler, W.: A dynamic thermodynamic sea ice model, Journal of Physical Oceanography, 9, 815–846, 1979.

Hurrell, J., Hack, J., Shea, D., Caron, J., and Rosinski, J.: A new sea surface temperature and sea ice boundary data set for the Community Atmosphere Model, Journal of Climate, in prep, 2006.

Jayne, S. and Marotzke, J.: A destabilizing thermohaline circulation-atmosphere-sea ice feedback, Journal of Climate, 12, 642–651, 1999.

Jerlov, N. G.: Optical Oceanography, Elsevier, 1968.

Johnson, G. C., Sloyan, B. M., Kessler, W. S., and McTaggart, K. E.: Direct measurements of upper ocean currents and water properties across the tropical Pacific during the 1990's, Progress in Oceanography, 52, 31–36, 2002.

Jungclaus, J., Keenlyside, N., Botzet, M., Haak, H., Luo, J.-J., Latif, M., Marotzke, J., Mikolajewicz, U., and Roeckner, E.: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM, Journal of Climate, 19, 3952–3972, 2006.

Killworth, P. D., Smeed, D., and Nurser, A.: The effects on ocean models of relaxation toward observations at the surface, Journal of Physical Oceanography, 30, 160–174, 2000.

Large, W. and Yeager, S.: 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, 2004.

Large, W. B.: Surface fluxes for practioners of global ocean data assimilation, in Ocean Weather Forecasting: an Integrated View of Oceanography, edited by E. P. Chassignet and J. Verron, vol. 577, pp. 229–270, Springer, 2005.

Large, W. G. and Yeager, S.: Global means and seasonal to decadal variability of air-sea fluxes, Journal of Climate, submitted, 2006.

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

Large, W. G., Danabasoglu, G., Doney, S. C., and McWilliams, J. C.: Sensitivity to surface forcing and boundary layer mixing in a global ocean model: annual-mean climatology, Journal of Physical Oceanography, 27, 2418–2447, 1997.

Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., O'Brien, J., Rosati, A., and Schneider, E.: A review of the predictability and prediction of ENSO, Journal of Geophysical Research, 103, 14 375–14 393, 1998.

Latif, M., Böning, C., Willebrand, J., Biastoch, A., Dengg, J., Keenlyside, N., Madec, G., and Schweckendiek, U.: Is the thermohaline circulation changing?, Journal of Climate, pp. 4631–4637, 2006.

Leaman, K., Molinari, R., and Vertes, P.: Structure and variability of the Florida Current at 27N: April 1982-July 1984, Journal of Physical Oceanography, 17, 565–583, 1987.

Levitus, S.: Climatological atlas of the world ocean, U.S. Government Printing Office 13, NOAA, Washington, D.C., 163 pp., 1982.

Levitus, S., Burgett, R., and Boyer, T.: 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., 1994.

Lohmann, G., Gerdes, R., and Chen, D.: Sensitivity of the thermohaline circulation in coupled oceanic GCM–atmospheric EBM experiments, Climate Dynamics, 12, 403–416, 1996.

Lohmann, G., Gerdes, R., and Chen, D.: Sea ice effects on the sensitivity of the thermohaline circulation in simplified atmosphere-ocean-sea ice models, Journal of Climate, 11, 2789–2803, 1998.

Lukas, R. and Firing, E.: The geostrophic balance of the Pacific equatorial undercurrent, Deep-Sea Research, 31, 61–66, 1984.

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

Madec, G., Delecluse, P., Imbard, M., and Lévy, C.: OPA 8.1 Ocean General Circulation Model reference manual, in Note du Pôle de modélisation, p. 91 pp, Paris, France, 1999.

Marshall, J. and Schott, F.: Open-ocean convection: observations, theory, and models, Reviews of Geophysics, 37, 1–64, 1999.

Marsland, S., Haak, H., Jungclaus, J., Latif, M., and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, Ocean Modelling, 5, 91–127, 2003.

Meehl, G., Gent, P. R., Arblaster, J., Otto-Bliesner, B., Brady, E., and Craig, A.: Factors that affect the amplitude of El Niño in global coupled climate models, Climate Dynamics, 17, 515–526, 2001.

Morel, A. and Antoine, D.: Heating rate within the upper ccean in

relation to its bio-optical state, Journal of Physical Oceanography, 24, 1652–1665, 1994.

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

Ohlmann, J.: Ocean Radiant Heating in Climate Models, JOC, 16, 1337–1351, 2003.

Orsi, A., Whitworth, T., and Nowlin, W.: On the meridional extent and fronts of the Antarctic Circumpolar Current, Deep-Sea Research, 42A, 641–673, 1995.

Pacanowski, R. C.: Effect of equatorial currents on surface stress, Journal of Physical Oceanography, 17, 833–838, 1987.

Paulson, E. A. and Simpson, J. J.: Irradiance measurements in the upper ocean., Journal of Physical Oceanography, 7, 952–956, 1977.

Payne, R.: Albedo of the surface ocean, Journal of Atmospheric Sciences, 29, 959–970, 1972.

Peixoto, J. P. and Oort, A. H.: Physics of Climate, American Institute of Physics, 520 + xxxix pp, 1992.

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., and Wang, J.: The Arctic Ocean Model Intercomparison Project (AOMIP), EOS, 82, 637–644, 2001.

Rahmstorf, S. and Willebrand, J.: The Role of Temperature Feedback in Stabilizing the Thermohaline Circulation, Journal of Physical Oceanography, 25, 787–805, 1995.

Rahmstorf, S., Marotzke, J., and Willebrand, J.: Stability of the thermohaline circulation, in The Warmwatersphere of the North Atlantic, edited by W. Krauss, pp. 129–157, Borntraeger, 1996.

Rayner, N., Parker, D., Horton, E., Folland, C., Alexander, L., Rowell, D., Kent, E., and Kaplan, A.: Global Analyses of SST, Sea Ice and Night Marine Air Temperature Since the Late Nineteenth Century, Journal of Geophysical Research, 108(D14), 4407,DOI:10.1029/2002JD002 670, 2003.

Reynolds, R. W., Rayner, N., Smith, T. M., Stokes, D., and Wang, W.: An improved *in situ* and satellite SST analysis for climate, Journal of Climate, 15, 1609–1625, 2002.

Rivin, I. and Tziperman, E.: On the sensitivity of air-sea fluxes to SST perturbations, Journal of Climate, 11, 2431–2446, 1997.

Roach, A., Aagard, K., Pease, C., Salo, S., Weingartner, T., Pavlov, V., and Kulakov, M.: Direct measurements of transport and water properties through Bering Strait, Journal of Geophysical Research, 100, 18 443–18 457, 1995.

Robinson, I.: Satellite Oceanography, in GODEA Summer School, edited by E. P. Chassignet and J. Verron, Springer/Kluwer, 2005.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model, Journal of Climate, 19, 3771–3791, 2006.

Rosati, A. and Miyakoda, K.: A general circulation model for upper ocean simulation, Journal of Physical Oceanography, 18, 1601–1626, 1988.

Röske, F.: A global heat and freshwater forcing dataset for ocean models, Ocean Modelling, in press, 2005.

Roullet, G. and Madec, G.: Salt conservation, free surface, and varying volume. A new formulation for OGCMs, JGR, 105, 23 927–23 947, 2000.

Sloyan, B., Johnson, G., and Kessler, W.: The Pacific Cold Tongue: A pathway for interhemispheric exchange, Journal of Physical Oceanography, 33, 1027–1043, 2003.

Smith, R. and Gent, P.: Reference Manual for the Parallel Ocean Program (POP), Los Alamos Technical Report No. LAUR-02-2484, 2004.

Smith, R. D., Kortas, S., and Meltz, B.: Curvilinear coordinates for global ocean models, Los Alamos preprint, LA-UR-95-1146, 1995.

Steele, M., Morfley, R., and Ermold, W.: PHC: A global ocean hydrography with a high-quality Arctic Ocean, JOC, 14, 2079–2087, 2001.

Stommel, H.: Thermohaline convection with two stable regies of flow, Tellus, 13, 224–228, 1961.

Stouffer, R. J.: Time scales of climate response, Journal of Climate, 17, 209–217, 2004.

Sun, S. and Bleck, R.: Multi-century simulations with the coupled GISS-HYCOM climate model: control experiments, Climate Dynamics, 26, 407–428, 2006.

Sweeney, C., Gnanadesikan, A., Griffies, S. M., Harrison, M., Rosati, A., and Samuels, B.: Impacts of shortwave penetration depth on large-scale ocean circulation and heat transport, Journal of Physical Oceanography, 35, 1103–1119, 2005.

Taylor, P.: 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, p. 303pp, World Climate Research Programme, 2000.

Timmermannn, R., Goosse, H., Madec, G., Fichefet, T., Ethe, C., and Duliére, V.: On the representation of high latitude processes in the ORCA-LIM global coupled sea ice-ocean model, Ocean Modelling, 8, submitted, 2005.

Trenberth, K. and Caron, J.: Estimates of meridional atmosphere and ocean heat transports, Journal of Climate, 14, 3433–3443, 2001.

Weaver, A. J. and Hughes, T. M.: On the incompatibility of ocean and atmosphere models and the need for flux adjustments, Climate Dynamics, 13, 141–170, 1996.

Wyrtki, K. and Kilonsky, B.: Mean water and current structure during the Hawaii-to-Tahiti shuttle experiment, Journal of Physical Oceanography, 14, 242–254, 1984.

Zhang, S., Greatbatch, R., and Lin, C.: A re-examination of the polar halocline catastrophe and implications for coupled ocean-atmosphere models, Journal of Physical Oceanography, 23, 287–299, 1993.