Ocean Modelling
January 2016, Volume 97 Pages 65-90
https://doi.org/10.1016/j.ocemod.2015.11.007
https://archimer.ifremer.fr/doc/00310/42164/

North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part II: Inter-annual to decadal variability

Danabasoglu Gokhan [1,*], Yeager Steve G. [1], Kim Who M. [2], Behrens Erik [3], Bentsen Mats [4], Bi Daohua [5], Biastoch Arne [3,9], Bleck Rainer [7,8], Boening Claus [3], Bozec Alexandra, Canuto Vittorio M. [8], Cassou Christophe [10], Chassignet Eric [9], Coward Andrew C. [11], Danilov Sergey [12], Diansky Nikolay [13], Drange Helge [14,15], Farneti Riccardo [16], Fernandez Elodie [10,17], Fogli Pier Giuseppe [18], Forget Gael [19], Fujii Yosuke [20], Griffies Stephen M. [21], Gusev Anatoly [13], Heimbach Patrick [19], Howard Armando [8,22], Ilicak Mehmet [4], Jung Thomas [12], Karspeck Alicia R. [1], Kelley Maxwell [8], Large William G. [1], Leboissetier Anthony [8], Lu Jianhua [9], Madec Gurvan [23], Marsland Simon J. [5,6], Masina Simona [18,24], Navarra Antonio [18,24], Nurser A. J. George [11], Pirani Anna [25], Romanou Anastasia [8,26], Salas Y Melia David [27], Samuels Bonita L. [21], Scheinert Markus [3], Sidorenko Dmitry [12], Sun Shan [7], Treguier Anne-Marie [28], Tsujino Hiroyuki [20], Uotila Petteri [5,6,29], Valcke Sophie [10], Voldoire Aurore [27], Wang Qiang [12], Yashayaev Igor [30]

[1] NCAR, Boulder, CO 80301 USA.
[2] Texas A&M Univ, College Stn, TX USA.
[3] Helmholtz Ctr Ocean Res, GEOMAR, Kiel, Germany.
[4] Bjerknes Ctr Climate Res, Uni Res Climate, Bergen, Norway.
[5] CSIRO, Ctr Australian Weather & Climate Res, Melbourne, Vic, Australia.
[6] CSIRO, Bur Meteorol, Melbourne, Vic, Australia.
[7] NOAA Earth Syst Res Lab, Boulder, CO USA.
[8] NASA Goddard Inst Space Studies GISS, New York, NY USA.
[9] Florida State Univ, Ctr Ocean Atmospher Predict Studies COAPS, Tallahassee, FL 32306 USA.
[10] CERFACS, Toulouse, France.
[11] NOCS, Southampton, Hants, England.
[12] Alfred Wegener Inst Polar & Marine Res AWI, Bremerhaven, Germany.
[13] Russian Acad Sci, Inst Numer Math, Moscow, Russia.
[14] Univ Bergen, Inst Geophys, Bergen, Norway.
[15] Bjerknes Ctr Climate Res, Bergen, Norway.
[16] Abdus Salaam Int Ctr Theoret Phys, Trieste, Italy.
[17] Mercator Ocean, Toulouse, France.
[18] Ctr Euromediterraneo Sui Cambiamenti Climatici CM, Bologna, Italy.
[19] MIT, Cambridge, MA 02139 USA.
[20] Japan Meteorol Agcy, MRI, Tsukuba, Ibaraki, Japan.
[21] NOAA Geophys Fluid Dynam Lab GFDL, Princeton, NJ USA.
[22] CUNY Medgar Evers Coll, Brooklyn, NY 11225 USA.
[23] CNRS IRD UPMC, IPSL LOCEAN, Paris, France.
[24] INGV, Bologna, Italy.
[25] Abdus Salaam Int Ctr Theoret Phys, Int CLIVAR Project Off, Trieste, Italy.
[26] Columbia Univ, New York, NY USA.
[27] Ctr Natl Rech Meteorol CNRM GAME, Toulouse, France.
[28] IUEM, CNRS Ifremer IRD UBO, UMR 6523, Lab Phys Oceans, Plouzane, France.

[29] Finnish Meteorol Inst, FIN-00101 Helsinki, Finland.
[30] Fisheries & Oceans Canada, Bedford Inst Oceanog, Dartmouth, NS B2Y 4A2, Canada.

* Corresponding author : Gokhan Danabasoglu, email address : gokhan@ucar.edu

---

Abstract :

Simulated inter-annual to decadal variability and trends in the North Atlantic for the 1958-2007 period from twenty global ocean - sea-ice coupled models are presented. These simulations are performed as contributions to the second phase of the Coordinated Ocean-ice Reference Experiments (CORE-II). The study is Part II of our companion paper (Danabasoglu et al., 2014) which documented the mean states in the North Atlantic from the same models. A major focus of the present study is the representation of Atlantic meridional overturning circulation (AMOC) variability in the participating models. Relationships between AMOC variability and those of some other related variables, such as subpolar mixed layer depths, the North Atlantic Oscillation (NAO), and the Labrador Sea upper-ocean hydrographic properties, are also investigated. In general, AMOC variability shows three distinct stages. During the first stage that lasts until the mid-to late-1970s, AMOC is relatively steady, remaining lower than its long-term (1958-2007) mean. Thereafter, AMOC intensifies with maximum transports achieved in the mid-to late-1990s. This enhancement is then followed by a weakening trend until the end of our integration period. This sequence of low frequency AMOC variability is consistent with previous studies. Regarding strengthening of AMOC between about the mid-1970s and the mid-1990s, our results support a previously identified variability mechanism where AMOC intensification is connected to increased deep water formation in the subpolar North Atlantic, driven by NAO-related surface fluxes. The simulations tend to show general agreement in their temporal representations of, for example, AMOC, sea surface temperature (SST), and subpolar mixed layer depth variabilities. In particular, the observed variability of the North Atlantic SSTs is captured well by all models. These findings indicate that simulated variability and trends are primarily dictated by the atmospheric datasets which include the influence of ocean dynamics from nature superimposed onto anthropogenic effects. Despite these general agreements, there are many differences among the model solutions, particularly in the spatial structures of variability patterns. For example, the location of the maximum AMOC variability differs among the models between Northern and Southern Hemispheres.

Highlights

► Inter-annual to decadal variability in AMOC from CORE-II simulations is presented. ► AMOC variability shows three stages, with maximum transports in mid- to late-1990s. ► North Atlantic temporal variability features are in good agreement among simulations. ► Such agreements suggest variability is dictated by the atmospheric data sets. ► Simulations differ in spatial structures of variability due to ocean dynamics.

Keywords : Global ocean - sea-ice modelling, Ocean model comparisons, Atmospheric forcing, Inter-annual to decadal variability and mechanisms, Atlantic meridional overturning circulation variability, Variability in the North Atlantic

1. Introduction

This study presents an analysis of the simulated inter-annual to decadal variability and trends in the North Atlantic Ocean for the 1958−2007 period from a set of simulations participating in the second phase of the Coordinated Ocean-ice Reference Experiments (CORE-II). It is Part II of our companion paper, Danabasoglu et al. (2014) (hereafter DY14), where the mean states in the Atlantic basin from these simulations are documented to provide a baseline for the present variability analysis.

Our primary focus is again on the Atlantic meridional overturning circulation (AMOC), but here we investigate representation of its inter-annual to decadal variability and trends in the participating models. As stated in DY14, AMOC is presumed to play a major role in decadal and longer time scale climate variability and in prediction of the earth's future climate on these time scales through its heat and salt transports and its impacts on sea surface temperatures (SSTs) and sea level. Due to lack of long and continuous AMOC observations, the main support for such an important role for AMOC in influencing the earth's climate comes from coupled general circulation model (CGCM) simulations. In long control simulations with CGCMs, usually for pre-industrial conditions run without either changes in radiative forcings or inclusion of anthropogenic forcings, AMOC intrinsic variability is rather rich with a variety of time scales, e.g., inter-annual, decadal, centennial. Furthermore, such low frequency AMOC anomalies tend to precede the basin scale SST anomalies in the Atlantic Ocean, thus suggesting a driving role for AMOC in models (e.g., Delworth et al., 1993; Danabasoglu, 2008; Kwon and Frankignoul, 2012; Delworth and Zeng, 2012; Danabasoglu et al., 2012). Hence, the basin scale, low frequency variability (40−70 year period) of the observed SSTs in the Atlantic Ocean is assumed to be linked to AMOC fluctuations. This basin scale SST variability is usually referred to as the Atlantic Multidecadal Variability (AMV) or Atlantic Multidecadal Oscillation. AMV represents an index of detrended, observed (North) Atlantic SST variability estimated from instrumental records and proxy data (Schlesinger and Ramankutty, 1994; Kushnir, 1994; Delworth and Mann, 2000). We also note that some studies suggest that variability of AMOC and upper-ocean temperatures may be potentially predictable on decadal time scales (e.g., Griffies and Bryan, 1997; Pohlmann et al., 2004; Msadek et al., 2010; Branstator and Teng, 2010), thus making appropriate initialization of the AMOC state for decadal prediction experiments an important endeavor.

For studies of AMOC variability and its mechanisms and prediction, CGCMs are an essential tool. However, their fidelity remains a serious concern, and a fundamental understanding of the mechanisms of simulated AMOC variability remains elusive (see Liu (2012) and Srokosz et al. (2012) for recent reviews). For example, the magnitude and dominant time scales of AMOC variability and its mechanisms can differ substantially from one model to another (see above references), from one version of a model to another (Danabasoglu, 2008; Danabasoglu et al., 2012), and, in some cases, even from one time segment of a model simulation to another (Kwon and Frankignoul, 2012, 2014). Some oceanic subgrid scale parameterizations are shown to affect the variability of AMOC as well, e.g., magnitude of vertical diffusivity coefficients (Farneti and Vallis, 2011); representation of the Nordic Sea overflows (Yeager and Danabasoglu, 2012) and of meso- and submesoscale eddies (Danabasoglu et al., 2012). In addition, various aspects of AMOC variability are sensitive to both the atmosphere and ocean model resolutions (Bryan et al., 2006). Given these significant model sensitivities and many unanswered questions, there is a critical need for improving our understanding of the mechanisms and assessing the fidelity and robustness of simulated AMOC variability against limited available observations.

The CORE-II hindcast experiments provide a common framework to address some of these issues. Specifically, they can be used to investigate AMOC variability and its mechanisms on seasonal, inter-annual, and decadal time scales and to understand and separate forced variability from natural variability – the latter in combination with (coupled) control experiments that exclude external and anthropogenic effects. Additionally, robustness of variability mechanisms across models can be evaluated. Continuous, observationally-based estimates of AMOC are available only starting in early 2004 through the Rapid Climate Change transbasin observing array installed along 26.5°N (RAPID; Cunningham et al., 2007). The CORE-II hindcasts – along with the reanalysis products – can provide complementary information on AMOC for the pre-RAPID era. Unfortunately, for our current work, the overlap period between the RAPID estimates and the model simulations is rather short, i.e., April 2004 through December 2007, making our annual-mean comparisons rather crude. Nevertheless, the solutions from the CORE-II hindcasts can be compared against other available observations in their representations of certain climate events, such as the mid-1990s warming of the subpolar North Atlantic. Identified variability mechanisms or their drivers associated with such events are expected to provide insight on AMOC variability in general, even though the CORE-II simulations cannot directly address intrinsic inter-annual to multi-decadal AMOC variability because the forcing data sets include external and anthropogenic effects. We note that several individual model studies, using the CORE-II protocol, have already demonstrated many realistic features of mean and variability in the North Atlantic in CORE-II hindcasts, including an investigation of the AMOC variability mechanisms associated with the mid-1990s warming of the subpolar North Atlantic (e.g., Yeager et al., 2012; Yeager and Danabasoglu, 2014; Gusev and Diansky, 2014).

Use of such hindcast simulations to investigate variability in the North Atlantic, particularly of the AMOC, is not new (e.g., Häkkinen, 1999; Eden and Willebrand, 2001; Bentsen et al., 2004; Beismann and Barnier, 2004; Böning et al., 2006; Biastoch et al., 2008; Deshayes and Frankignoul, 2008; Lohmann et al., 2009b; Brodeau et al., 2010; Robson et al., 2012). These studies employ various historical atmospheric datasets, e.g., National Centers for Environmental Prediction − National Center for Atmospheric Research (NCEP/NCAR) reanalysis (Kalnay et al., 1996), European Center for Medium-range Weather Forecasting (ECMWF) ERA-40 reanalysis (Uppala et al., 2005), or a combination of other datasets, to force regional Atlantic basin or global ocean models. They – along with the CORE-II hindcast studies mentioned in the previous paragraph – show that AMOC variability on inter-annual to decadal time scales is connected to surface buoyancy fluxes and wind stress associated with the North Atlantic Oscillation (NAO). A particularly robust feature of these and other studies is the strengthening of AMOC during the last few decades of the twentieth century. Specifically, the persistent positive NAO (NAO+) that occurred between the early 1970s and the mid-1990s is credited with enhanced deep water formation (DWF) and associated deepening of mixed layers in the subpolar North Atlantic, particularly in the Labrador Sea (LS) region. This in turn results in increased AMOC and northward heat transports that have been identified as the major contributors to the mid-1990s subpolar North Atlantic warming (e.g., Robson et al., 2012; Yeager et al., 2012). We note that this AMOC variability mechanism suggesting a prominent role for the NAO is very similar to the AMOC intrinsic variability mechanisms found in many CGCM control simulations (e.g., Dong and Sutton, 2005; Teng et al., 2011; Danabasoglu et al., 2012).

In the present study, our primary goal is to provide an evaluation of how participating models represent trends and variability in AMOC and in some other fields on inter-annual to decadal time scales under the common CORE-II forcing, with a focus on the North Atlantic. With the variability mechanism described above providing a background, other goals include i) an investigation of robust aspects of AMOC variability in these coarse resolution models in the presence of mean state differences discussed in DY14 and ii) an exploration of relationships between AMOC variability and those of some other fields such as NAO, mixed layer depths (MLDs), and the LS upper-ocean temperature, salinity, and density.

The paper is organized as follows. In section 2, we briefly summarize the CORE-

II framework, analysis methods, and participating models, including two additional contributions (labeled as FSU2 and GISS2) to those used in DY14. We document the variabilities in AMOC; North Atlantic SSTs; North Atlantic MLDs; upper-ocean central LS hydrographic properties; and subpolar gyre (SPG) circulation and SPG sea surface height (SSH) in sections 3 through 7. We then present the relationships between AMOC variability and i) those of meridional heat transport (MHT) in section 8 and ii) those of LS MLD, SPG circulation, SPG SSH, and NAO in section 9. The last section, i.e., section 10, has a summary and our conclusions. We provide short summaries of FSU2 and GISS2 along with a note on their vertical coordinate choices and a brief evaluation of their mean states in the North Atlantic in Appendix A. Appendix B details the departures from the CORE-II protocol that occurred in nearly half of the participating models. Finally, a list of major acronyms is included in Appendix C.

2. CORE-II framework, models, and analysis methods

The CORE-II experiments represent ocean – sea-ice hindcast simulations forced with the inter-annually varying atmospheric datasets over the 60-year period from 1948 to 2007. These forcing datasets were developed by Large and Yeager (2004, 2009). The CORE-II protocol requests that the simulations are integrated for no less than five repeat cycles of the 60-year forcing. There is no restoring term applied to SSTs. However, a form of surface salinity restoring may be used to prevent unbounded local salinity trends. Details of the CORE-II protocol are given in Griffies et al. (2012) and DY14.

Our present study includes two additional contributions to those used in DY14, thus bringing the total number of participating models to twenty. Both of the new participants, labeled as FSU2 and GISS2, are based on the HYbrid Coordinate Ocean

Model (HYCOM). The FSU simulation in DY14 uses an earlier HYCOM version which advects density and salinity, thus does not conserve heat. In contrast, FSU2 employs a formulation that advects temperature and salinity, conserving heat. GISS2 also uses this latter formulation and represents an updated version of the model described in Sun and Bleck (2006). Summaries of FSU2 and GISS2 model descriptions are provided in Appendix A.1 and Appendix A.2, respectively. For the descriptions of other models and their surface salinity restoring details, we refer to the Appendices in DY14. We use the same model naming convention in the present study as in DY14. For completeness and reference purposes, an updated list of the participating groups along with their model names and resolutions is reproduced in Table 1.

After the publications of DY14 and Griffies et al. (2014), it came to our attention that about half of the participating models did depart from the CORE-II protocol recommendations. These departures, detailed in Appendix B, include use of different bulk formulae, modifications of the Large and Yeager (2009) bulk formulae, and changes in the forcing datasets.

The 60-year repeat forcing cycle introduces an unphysical jump in the forcing from 2007 back to 1948 with the ocean state in 1948 identical to that of the end state of the forcing cycle. This approach impacts the solutions during the early years of the forcing period. Our analysis here uses only the 1958−2007 period from the fifth cycle of the simulations to partially avoid any adverse effects of this artificial jump in forcing. We employ standard correlation, regression, and empirical orthogonal function (EOF) analysis methods. The principal component (PC) time series are normalized to have unit variance. Thus, the EOF spatial pattern magnitudes correspond to one standard deviation changes in the PC time series. Unless otherwise noted, the time series are based on annual-mean data. In most of our analysis, we choose not to detrend the time series, because our interests include low-frequency, e.g., decadal, variability and trends. As discussed in DY14, about half of the models reach a practical AMOC equilibrium state as measured by small root-mean-square differences and high correlations of their AMOC time series between the fourth and fifth forcing cycles. However, remaining models, i.e., AWI, FSU, GFDL-MOM, ICTP, INMOM, and KIEL, as well as the two new contributions FSU2 and GISS2, do not fully obtain such an equilibrium state and show ongoing drifts in their AMOCs (see Figs. 1 and 2 of DY14), likely impacting magnitudes of some of our calculated trends. The time series are decomposed into their high- and low-frequency contents, using a Butterworth filter with a somewhat arbitrary cutoff period of 7 years. In some of the figures with time series, we also include the time series for the multi model mean, denoted as MMM. The MMM time series do not include MRI-A – the only contribution with data assimilation. The solutions from this MRI-A simulation are also provided to the Karspeck et al. (2015) study where a comparison of AMOC mean, variability, and trends from six data assimilation products is presented.

The statistical significance of various lead−lag correlations is examined using a Monte Carlo approach called a *parametric bootstrap*. In this approach, we assume that the annual average statistical properties of the variables being considered (e.g., AMOC and MLD) can be modeled as a first-order auto regressive process (AR1), with variance and damping coefficient estimated from the model time series (without low-pass filtering). Consistent with a standard t-test for evaluating the significance of correlation coefficients, we test the null-hypothesis that the two time series are independent at all lags, but that sampling error may lead to a non-zero correlation. We build empirical distributions for each lag with which to evaluate this null-hypothesis using 2000 samples formed in the following way: two independent time series of length 50 years are generated from the AR1 process and the anomaly correlation coefficient is computed for each lag after low-pass filtering. This approach will naturally account for changes in the degrees-of-freedom associated with the lag, the autocorrelation in the model, and the low-pass filtering. Obtained correlations that fall above (below) 97.5% (2.5%) of the samples from the empirical distribution at each lag are considered significant (i.e., statistically unlikely to have resulted from two uncorrelated time series) at the 95% confidence level.

As in DY14, we use the total AMOC transports in our analysis, i.e., the sum of the Eulerian-mean, mesoscale eddy, and submesoscale eddy contributions, if the latter two are available. Except INMOM, all models include a variant of the Gent and McWilliams (1990) parameterization (GM90) to represent the advective effects of the mesoscale eddies. Only four models (ACCESS, GFDL-GOLD, GFDL-MOM, and NCAR) employ a submesoscale eddy parameterization (Fox-Kemper et al., 2011) that contributes to the total transport. We note that in BERGEN the same submesoscale eddy parameterization is used only to modify the turbulent kinetic energy budget of the mixed layer model and it does not contribute to the total transport. Because we are primarily interested in large-scale sub-thermocline (below 500 m) characteristics of AMOC and the impacts of both the mesoscale and submesoscale eddies are largely confined to the upper few hundred meters in the North Atlantic, missing subgrid-scale contributions from some models is not expected to affect our findings. For convenience, we refer to total AMOC simply as AMOC in the rest of this paper.

Furthermore, we primarily use the representation of AMOC in depth−latitude space in our analysis. While this is the most common depiction and use of AMOC, an alternative is AMOC in density−latitude space – which we also consider, though briefly. As discussed in Kwon and Frankignoul (2014), the depth-space AMOC tends to stress sinking (deep water formation) across isopycnals. In contrast, the density-space AMOC is better at highlighting water mass transformations and, perhaps, at exposing the impacts of upper-ocean subpolar gyre in the North Atlantic. Zhang

(2010) also argues that the density-space AMOC better represents the meridional coherency of AMOC variability. Given that the information provided by either representation will likely be model dependent, both representations may be used to provide complementary analysis for detailed variability mechanism studies (see Kwon and Frankignoul, 2014).

3. AMOC variability

We start with the AMOC maximum transport time series at 26.5° and 45°N shown in Figs. 1 and 2, respectively. The time series are based on AMOC obtained in depth − latitude space. They are anomalies from the respective 50-year (1958−2007) means for each model: these means are given in parentheses next to the model labels in each figure and they are also listed in Table 2. The MMM time series are included in the figures. These two latitudes are chosen to represent low- and mid-latitude AMOC variability, respectively. The 26.5°N time series additionally permit a comparison of models' AMOC variability to that of the RAPID based estimates during a short overlap period.

Focusing on decadal and longer time scales at both latitudes, AMOC variability, in general, can be characterized in three stages. During the first stage that lasts until the mid- to late-1970s, AMOC is relatively steady, usually remaining weaker than its long-term mean. Thereafter, AMOC intensifies with maximum transports achieved in the mid- to late-1990s. This intensification is then followed by a weakening trend that continues until the end of our integration period. Maximum transports appear to occur earlier and the weakening trend appears to be more pronounced at 45°N than at 26.5°N. Unfortunately, there are no long-term continuous observations to verify this *general* AMOC behavior in our CORE-II simulations. However, many modeling studies discussed in section 1 corroborate the AMOC variability depicted in Figs. 1 and 2 (e.g., Häkkinen, 1999; Eden and Willebrand, 2001; Bentsen et al., 2004; Beismann and Barnier, 2004; Böning et al., 2006; Deshayes and Frankignoul, 2008; Lohmann et al., 2009b; Brodeau et al., 2010; Robson et al., 2012). Similar trend behavior is also seen in some reanalysis products (Pohlmann et al., 2013).

There are, however, exceptions to the above generalizations. For example, CMCC, FSU, MIT, MRI-A, and NOCS show either very weak or no noticeable trends during the 1958−2007 period; KIEL does not show weakening during the last decade at 26.5°N; and ICTP time series appear quite different than the other models at 45°N. There are also differences among the models in their ranges of anomaly magnitudes with AWI and GISS showing the largest peak-to-peak ranges with about 7 and 9 Sv ($1 \text{ Sv} \equiv 10^6 \text{ m}^3 \text{ s}^{-1}$), respectively, at both latitudes. Nevertheless, the level of general agreement in the characteristics of the AMOC maximum transport time series, e.g., year-to-year variability and long-term trends, among the forward (non-data-assimilating) models participating in this study appears to be substantially greater than among various reanalysis products shown in Karspeck et al. (2015).

We provide a more quantitative assessment of the agreements and disagreements among the models in their representations of AMOC variability in Fig. 3, considering model − model correlations of the AMOC maximum transport time series discussed above. Specifically, the figure shows the high-pass filtered; low-pass filtered with trend; and low-pass filtered but linearly detrended time series correlations between the models. The majority of the models are in agreement in their representations of inter-annual variability at both latitudes (Figs. 3a and 3d). In general, model − model correlations are weaker at 45°N than at 26.5°N. MRI-A is the major outlier at 26.5°N, with ACCESS, ICTP, and INMOM also showing less agreement. ICTP has the lowest correlations at 45°N. Figures 3b and 3e indicate that the model − model correlations are much weaker at decadal and longer time scales than at inter-annual time scales. Again, the disagreement among the models is larger at 45°N than at 26.5°N. At both latitudes, the primary outliers are MRI-A and NOCS with most of their correlation coefficients much less than 0.5. A comparison of the low-pass filtered correlations with trend and with the linear trend removed (Figs. 3b and 3e vs. Figs. 3c and 3f) shows that on decadal time scales the trend is the dominant signal over the 1958−2007 period at both latitudes – but more evident at 26.5°N. We note that although MRI-A emerges as an outlier when compared to the forward models in its representation of several AMOC variability characteristics considered in this paper, it is not an outlier among the reanalysis products analyzed in Karspeck et al. (2015).

The general characteristics of AMOC variability described above with reference to Fig. 3 appear to be consistent with findings of some previous studies (e.g., Biastoch et al., 2008; Yeager and Danabasoglu, 2014). On inter-annual time scales and particularly at lower latitudes, variability is primarily wind-driven as suggested by the strong model − model correlations of Figs. 3a and 3d. Such high model − model correlations from the wind-driven component are expected because all the models are forced by the same wind dataset. On decadal and longer time scales, variability is dominated by buoyancy forcing, and there are larger discrepancies among the models. These latter differences are likely associated with differences in the models' DWF properties.

Figures 4 and 5 show the AMOC EOF1 spatial distributions and the corresponding PC1 time series, respectively, based on the depth−latitude space AMOC. Because we use undetrended time series, the patterns depicted in Fig. 4 are primarily associated with low frequency variability and trends, and PC1 time series are broadly similar to those of Figs. 1 and 2. Thus, most of the time series show strengthening of initially weak AMOC until about the mid- to late-1990s, followed by a weakening trend. The exceptions to this generalization include FSU and, in particular, MRI-A. In general, the EOF1 distributions display a single cell pattern, covering the Atlantic basin south of 60°N. GISS, ICTP, and MRI-A have the largest amplitudes with more than 3.2 Sv per standard deviation. In this EOF measure, AWI does not stand out as one of the models with a large amplitude.

Based on their EOF1 spatial patterns, the models can be separated into three distinct groups. The first group, representing the majority with twelve models, has their maxima in the Northern Hemisphere, mostly between 30° and 50°N. The models in this group are AWI, BERGEN, CERFACS, CNRM, GFDL-GOLD, GFDL-MOM, GISS, GISS2, ICTP, INMOM, KIEL, and NCAR. Particularly for these models, the EOF1 pattern in its positive phase indicates strengthening and deeper penetration of the North Atlantic Deep Water (NADW) cell. The second group of models, i.e., ACCESS, CMCC, FSU, FSU2, MIT, MRI-F, and NOCS, have their maxima in the Southern Hemisphere. With the exception of ACCESS, these are among the models with the weakest mean AMOC transports as shown in Fig. 3 of DY14, Fig. 17, and Table 2. MRI-A is the only member of the third group with its maximum located in the vicinity of the equator. Whether the AMOC EOF1 maximum is located in the Northern or Southern Hemisphere does not appear to be related to the characteristics / properties of the Southern Ocean meridional overturning circulations in these CORE-II simulations in any obvious way (see Farneti et al., 2015).

With its equatorially-enhanced EOF1 spatial structure and associated PC1 time series, MRI-A is one of the models with large differences from the general AMOC behavior described earlier. Similar, large amplitude AMOC variability at or near the equator is also present in other reanalysis products as shown in Karspeck et al. (2015). We think that such a prominent feature in reanalysis products, including MRI-A, may be associated with the mismatches in calendar time between the zonal wind stress used to force the model and the potential temperature and salinity data used in data assimilation. This is because the equatorial circulation represents a balance between the zonal wind stress and zonal pressure gradients and any small discrepancies in this balance can produce anomalous circulation patterns. Thus, we believe that the MRI-A EOF1 likely represents spurious variability.

The EOF1s account for 40% to 70% of the total variances in AMOC. The highest variances occur in BERGEN (70%), ICTP (71%), GFDL-GOLD (74%), GISS (74%), and KIEL (77%). All of these models have their maxima in the Northern Hemisphere. In contrast, the models with the lowest variances, i.e., FSU2(40%), MRI-A (40%), FSU (46%), MIT (46%), and NOCS (47%), have their maxima in the Southern Hemisphere or near the equator.

For comparison purposes, we note that the second EOFs of AMOC (not shown) account for only 7% to 22% of the total variance with fourteen models having variances of < 15%. Not surprisingly, the models with the larger EOF2 variances correspond to the ones with the smallest variances in their EOF1s. With the exception of a few models, the EOF2 spatial patterns can be described as two north − south counter-rotating (dipole) cells, extending from the surface to the ocean bottom (not shown, but see Fig. 2 of Danabasoglu et al. (2012) for an example). The crossover latitude between these two cells varies between 0° and 30°N among the models, but it is near 0° in the models with the largest EOF2 variances. These models are also the ones with their EOF1 maxima in the Southern Hemisphere.

We find qualitatively very similar results when AMOC variability is analyzed in density (σ_2) − latitude space as presented in Fig. 6. For example, relative model differences are largely preserved, with models which have weaker (stronger) AMOC amplitudes in depth space still showing weaker (stronger) amplitudes in density space. In addition to GISS and ICTP, AWI, GFDL-MOM, and KIEL also show variability

> 3.2 Sv per standard deviation. All of the models with their maximum variability in the Northern Hemisphere in depth-space AMOC also retain their maxima in the same hemisphere, but the latitudes of the maxima are shifted northwards in density space. MRI-A has its maximum still near the equator. The models with their maxima in the Southern Hemisphere in depth space display less consistency in density space. For example, while CMCC and FSU have their maxima still in the Southern Hemisphere, the location of maxima is shifted to the Northern Hemisphere in FSU2 and MRI-F. In ACCESS, there is an additional maximum location in the Northern Hemisphere.

The density-space EOF1s account for 35% to 72% of the total variance in AMOC – a very similar spread as in the depth-space analysis. However, the individual model variances are reduced in density space in all models with the exception of AWI, FSU, and GFDL-MOM. The lowest variances occur in MIT (35%), NOCS (36%), FSU2 (37%), and MRI-A (37%) – all among the lowest-variance-models in depth space as well. GFDL-MOM and KIEL have the highest variances with $> 70\%$. The corresponding PC1 time series (not shown) are very similar to those of Fig. 5 for AMOC in depth space, broadly duplicating the low frequency AMOC variability.

As illustrated above, the most prominent features of the AMOC maximum transport and PC1 time series are the strengthening of transports between about the mid-1970s and the mid- to late-1990s, followed by a weakening trend that continues until the end of the integration period. To provide a quantitative assessment of these tendencies, we present the AMOC linear trends in Table 2, calculated using the time series of Figs. 1 and 2 for 26.5° and 45°N, respectively. The trends are calculated for the 1978−1998 and 1998−2007 periods at 26.5°N and for the 1975−1995 and 1995−2007 periods at 45°N to roughly represent the time frames with increases and decreases in AMOC, respectively. The shifts in the time periods between the two latitudes are intended to account for the apparent lag of AMOC changes at 26.5°N in comparison to those at 45°N as alluded to earlier in this section (also see section 9). The trends that meet the 95% confidence level based on a two-sided Student's t-test are shown in bold.

We compute the MMM trends as 0.70 and −1.73 Sv decade^{-1} at 26.5°N and 0.82 and −1.54 Sv decade^{-1} at 45°N. Particularly for the later period, these trends are impacted by the large negative trends in GISS, the exclusion of which reduces the MMM trends at both latitudes to about −1.37 Sv decade^{-1}. A notable feature of the MMM trends is that the weakening rate is nearly double that of strengthening. We note that the models that have their AMOC maximum variability in the Southern Hemisphere or in the vicinity of the equator tend to show weaker and statistically less significant trends. For the 1978−1998 period at 26.5°N, all models show positive trends, ranging from 0.09 (MRI-A) to 1.62 (GISS) Sv decade^{-1}. For the 1998−2007 period at the same latitude, while sixteen models have negative trends – from −0.02 (FSU) to −8.13 (GISS) Sv decade^{-1} – four models, i.e., KIEL, MIT, MRI-A, and NOCS, show positive trends. Except KIEL, these models have their maximum AMOC anomalies in the Southern Hemisphere. At 45°N, the models are again unanimous in their trend signs, all showing AMOC intensification for the 1975−1995 period, ranging from 0.03 (NOCS) to 2.06 (GISS) Sv decade^{-1}. For the 1995−2007 period at 45°N, all but three models show weakening of AMOC with trends ranging from −0.02 (FSU) to −4.81 (GISS) Sv decade^{-1}. The exceptions are MIT, MRI-F, and NOCS, again all with maximum AMOC anomalies in the Southern Hemisphere.

We make the following additional observations based on Table 2: GISS emerges as the model with the largest trends of both signs at both latitudes; the trends in NOCS are positive at both latitudes regardless of the time period; and MRI-A and

NOCS are the only models in which all trends remain below our confidence level.

The spatial patterns of AMOC linear trends are very similar to those of the EOF1s depicted in Fig. 4 and, therefore, not shown. The intensification of AMOC during the earlier period is associated with strengthening and deeper penetration of the NADW cell. We finally note that MRI-A appears to be an outlier in its trend spatial patterns (not shown), revealing strong negative trends in the Southern Hemisphere. Regarding such reanalysis products, Karspeck et al. (2015) show quite diverse representations of AMOC trends over a similar time period among several reanalysis datasets – perhaps even more diverse than those depicted in Table 2 for the present CORE-II simulations.

4. SST variability

An important test for evaluation of the CORE-II hindcast simulations is their ability to reproduce observed spatial patterns and temporal characteristics of SST variability. This is not assured in these simulations as discussed in Doney et al. (2007) where it is shown that ocean processes considerably affect SST and upper-ocean heat content variability. Thus, disagreements with observations can be expected in the North Atlantic where ocean, particularly advective, heat transports are significant.

We show the model SST EOF1 spatial distributions and the associated PC1 time series in Figs. 7 and 8, respectively, including the corresponding distributions from the HadISST observational dataset (Hurrell et al., 2008). The EOFs are obtained for the North Atlantic region bounded by $80°W-10°E$ and $10° - 70°N$. Although we do not use any detrending or low-pass filtering of the SST time series, the EOF1 patterns still produce the familiar AMV pattern (e.g., Sutton and Hodson, 2005) with a basin scale, single-sign SST anomaly (positive in Fig. 7). In HadISST, the maximum variability occurs east of Newfoundland with an amplitude of $> 0.7°C$ per standard deviation. Additional maxima are present in LS and western Irminger Sea. There is an opposite-signed anomaly (negative in Fig. 7) just off the east coast of North America with a small amplitude of about 0.1°C per standard deviation. The CORE-II simulations broadly reproduce observed SST characteristics, but there are many differences from observations in details. Perhaps the most visible of these is the amplitude, location, and spatial extent of the largest SST anomaly. This discrepancy is particularly evident in AWI, CERFACS, CNRM, GFDL-GOLD, GFDL-MOM, GISS, GISS2, and KIEL with maximum anomalies of > 1°C per standard deviation and with substantially broader spatial extent of this maximum in comparison with observations. We think that these discrepancies together with somewhat smaller differences in the details of the negative SST anomalies off the east coast of North America are partly due to the incorrect separation of the models' Gulf Stream and the failure of the subsequent North Atlantic Current (NAC) to reconnect with the topography off the Grand Banks, resulting in a too-zonal path. As discussed in Danabasoglu (2008), these persistent biases can impact model variability in the North Atlantic. The model SST EOF1s account for 29% to 40% of the total SST variance in good agreement with the observational variance of 40%.

The correspondence between the model − model and model − observational SST PC1 time series is remarkably good, both at inter-annual and decadal time scales (Fig. 8). The PC1 time series show large amplitude low frequency variability superimposed onto inter-annual changes. There is an evident warming trend, roughly between the late 1980s and the late 1990s, producing peak SSTs around 1998. Another peak occurs during 2005−2006 after a short-lived cooling in between. Good agreements among all of these PC1 time series, particularly with modeled and observed variability, indicate that the temporal character of the basin-scale SST is primarily dictated by the variability and trends in the atmospheric datasets which already include the impacts of ocean dynamics from nature superimposed onto anthropogenic effects. The role of the *simulated* ocean dynamics in the models, e.g., NAC and AMOC, in influencing smaller-scale SSTs and upper-ocean heat contents is demonstrated by the differences among the models in their SST EOF1 spatial structures (Fig. 7). Indeed, the role of enhanced AMOC transports in the context of the mid-1990s subpolar North Atlantic warming in the upper-ocean has been unequivocally shown in Robson et al. (2012) and Yeager et al. (2012).

5. MLD variability

We assess the variability of the models' DWF regions in the northern North Atlantic, considering the March-mean MLD time series. Following the same procedure as in DY14, we adopt a density-based approach to determine MLDs where they are calculated as the depths at which the potential density (referenced to surface) changes by 0.125 kg m^{-3} from its surface value. MLD is calculated offline using the March-mean potential density obtained from March-mean potential temperature and salinity by each participating group.

Figures 9 and 10 present the March-mean MLD EOF1 spatial distributions and the PC1 time series, respectively. Despite differences in their mean MLDs (see Fig. 13 of DY14), the majority of the models show the area extending from the southeast LS into the Irminger Sea as the region with the largest MLD variability. Such broad regions with deep MLDs appear to be rather extensive in comparison with some observations (e.g., Lavender et al., 2002) which show only relatively small areas of deep mixing, mostly confined to just north of Labrador. In more than half of the models, the maximum amplitude is > 800 m per standard deviation. However, the amplitude and spatial extent of the maximum MLD variability in the LS $-$ Irminger Sea region differ considerably among the models. There are three exceptions to the dominance of this region: in KIEL and MRI-F, the MLD variability is as strong in the Nordic Seas; and NOCS has its largest variability in the Nordic Seas with rather weak variability in the LS region. In some of the models, e.g., BERGEN, CERFACS, CMCC, GFDL-MOM, and, NCAR, the deeper MLDs in the LS region – as depicted in Fig. 9 – are accompanied by shallower MLDs in the northern LS. Small amplitude negative MLD anomalies are also evident in the Nordic Seas in CERFACS, CMCC, CNRM, GFDL-MOM, GISS, GISS2, and MRI-A. The interior white areas in Fig. 9 indicate regions of no variability as the time-mean MLDs reach the ocean bottom in some models. A prominent example is ICTP where the time-mean MLDs are always as deep as the ocean bottom. The MLD EOF1s account for 19% to 49% of the total variance in MLD. While BERGEN (40%) GFDL-GOLD (41%), GFDL-MOM (41%), and NCAR (49%) have the highest variances, INMOM and NOCS have the smallest variances with 19% each. We note that larger (smaller) MLD EOF1 variances do not imply similarly larger (smaller) AMOC EOF1 variances. For example, MLD EOF1 variances are very similar for CMCC, ICTP, AWI, ACCESS, and MIT (35 – 37%), but their AMOC EOF1 variances range from 46% in MIT to 71% in ICTP. Likewise, we do not find any obvious connections between the MLD EOF1 spatial pattern characteristics and where the AMOC EOF1 maxima occur, i.e., Southern vs. Northern Hemisphere.

The PC1 time series (Fig. 10) show general agreement among most of the models, particularly in their representations of low frequency variability. With the sign convention depicted in Fig. 9 and primarily referring to the LS MLDs, MLDs get shallower and stay shallower during the first decade. This is followed by a tendency towards deeper MLDs until the early- to mid-1990s. Finally, we identify a tendency towards shallower MLDs till the end of the integration period. This characterization of the time series is consistent with changes in AMOC and is discussed further in section 9. The exceptions to the generalization include: NOCS with its near-neutral MLDs between the late 1980s and the late 1990s; KIEL, MIT, MRI-F, and NOCS with their mostly positive MLD anomalies after 1998; and GISS with a sharper increase and a sharper decrease of MLDs in the early 1970s and the early 1990s, respectively. We note that KIEL and NOCS deviate significantly from CERFACS, CMCC, and CNRM – the other NEMO-based models – in their PC1 time series, particularly after the mid-1980s.

6. Comparisons with hydrographic data in central LS

Unfortunately, it is rather difficult to verify the fidelity of the simulated MLD variability in the northern North Atlantic discussed above due to very limited observations. Instead, following Yeager and Danabasoglu (2014), we focus on a small central LS region, taking advantage of a compilation of hydrographic observations from Yashayaev (2007) which includes data from research vessels and profiling Argo floats. Specifically, we generate time series of potential temperature (θ), salinity (S), and density (σ_0) by averaging over a region bounded by $49° - 56°$W and $56° - 61°$N. We compute vertical averages in depth coordinates, rather than in density coordinates, for the $150-1000$ m depth range because the observations are available at depths greater than 150 m. We use May-mean θ and S from the models to roughly match the mostly Spring-time observations. Density is calculated using a common equation of state for all models, based on these May-mean θ and S.

We present the resulting model and observational time series for θ, S, and density in Figs. 11, 12, and 13, respectively, as anomalies from the $1958-2007$ period. For this comparison, the data from fourteen of the participating models are available. Also, the observational data are missing for some years roughly between 1975 and 1990. The figures also include the root-mean-square (rms) model $-$ observations time series differences as well as the correlation coefficients between the model and observational time series for each model. These two metrics are evaluated only for years with available observations and, as such, they are less focused on the 1975−1990 period where missing data occurs. We note that low rms differences and high correlation coefficients indicate good agreements with observations.

The observations show decadal-scale variability in θ and S from warm and salty anomalies in the 1960s and the early 1970s to mostly cold and fresh anomalies until about the early 2000s and then back to warm and salty anomalies. There are substantial compensations of θ and S anomalies in their contributions to density, but the density anomalies between about 1985 and 2000 are set primarily by the θ anomalies. The largest positive density anomalies occur in the mid-1990s, roughly coinciding with the deepest MLDs. There is modest agreement between the observational and simulated decadal-scale variability, particularly evident in θ and density time series. We compute the MMM correlation coefficients, i.e., the mean of the correlation coefficients and excluding MRI-A, for θ and density as 0.58 and 0.61, respectively. The corresponding value for S is much lower at 0.26. We note that MRI-A, which assimilates data, usually has the lowest rms and the highest correlation coefficients, producing one of the better agreements with observations by construction. Therefore, in the following discussion, we focus our attention to the performance of the forward models.

In θ (Fig. 11), while the smallest rms differences are in INMOM (0.19°C) and CERFACS (0.20°C), the largest departures from observations are in NCAR (0.31°C), BERGEN (0.32°C), and AWI (0.36°C). ICTP has the lowest correlation coefficient with 0.28. INMOM and CERFACS show the highest correlations with 0.76 and 0.79, respectively. Thus, in these measures, CERFACS and INMOM have the best agreements with observations. We note that, with the exception of NOCS, all models exhibit a prominent cold bias that leads to a positive density bias roughly during the 1983−1985 period. Because such a cold bias also exists in all the reanalysis products analyzed in Karspeck et al. (2015), we speculate that it may indicate a deficiency with the observational data.

In S (Fig. 12), the models with the lowest and highest rms differences are CNRM (0.025 psu), CERFACS (0.026 psu), CMCC (0.028 psu) and GFDL-MOM (0.040 psu), INMOM (0.042 psu), ICTP (0.047 psu), respectively. As indicated above, the correlation coefficients for S are much lower than those of θ and density. Indeed, the correlation is even negative in ICTP (−0.06) and near-zero in three of the models, i.e., GFDL-GOLD (0.00), INMOM (0.04), and GFDL-MOM (0.06). The highest correlation occurs in AWI with only 0.50. Although these metrics do not favor a particular model as better than the others, ICTP, INMOM, and GFDL-MOM produce the largest departures from observations.

While the largest rms density differences (Fig. 13) occur in AWI (0.026 kg m^{-3}), GFDL-MOM (0.026 kg m^{-3}), and INMOM (0.033 kg m^{-3}), the lowest rms differences are in FSU2 (0.014 kg m^{-3}) and CMCC (0.016 kg m^{-3}). The smallest correlations are in MRI-F, NOCS, and AWI with 0.48, 0.53, and 0.54, respectively. CMCC and FSU2 reveal the highest correlation coefficients with 0.69 and 0.73, respectively. Thus, these two models emerge as the models with the best agreements with the observations in density – even better than in MRI-A. In contrast, AWI appears to show the least agreement. As indicated earlier, the density time series include compensating biases in θ and S in their contributions to density. A notable example of this compensation occurs after 1998 where most models show warm and salty biases.

Finally, we compute the linear trends in density for the 1970−1995 period for each model and for the observations as another evaluation metric. The MMM trend of 0.025 kg m^{-3} decade^{-1} compares rather favorably with the observationally-based trend of 0.024 kg m^{-3} decade^{-1}. The range for individual model trends is between 0.009 and 0.049 kg m^{-3} decade^{-1} with NOCS and INMOM at the low and high end of this range, respectively. The simulated trends are within 20% of the observational value in six of the models. These models are (with their trends in kg m^{-3} decade^{-1}) AWI (0.020), NCAR (0.022), BERGEN (0.022), CMCC (0.023), GFDL-MOM (0.027), and FSU2 (0.028). We note that the trend in MRI-A is 0.017 kg m^{-3} decade^{-1}.

7. Gyre and sea surface height variability in the subpolar North Atlantic

Several recent observational and modeling studies highlight the importance and impacts of the North Atlantic SPG circulation variability on the climate of the North Atlantic (e.g., Häkkinen and Rhines, 2004; Böning et al., 2006; Lohmann et al., 2009a; Yeager and Danabasoglu, 2014). Because the SPG transport itself is not easily observed, the satellite-based SSH data (available since 1993) is used instead to determine observed changes in the SPG as well as to evaluate model-based findings (since the strength of the SPG is directly connected to the SSH gradients via geostrophy). As discussed earlier, the previous studies also show that there is a close connection between the SPG / SSH variability and that of AMOC via the NAO-related surface fluxes and associated changes in DWF, i.e., convective events. Indeed, Yeager and Danabasoglu (2014) suggest monitoring of the variations in the LS SSH as a proxy for AMOC changes.

A detailed evaluation of the simulated, global sea level mean and variability for the 1993−2007 period for most of the models participating in CORE-II is presented in Griffies et al. (2014). In the present study, we specifically focus on the SSH – strictly speaking, dynamic sea level – changes in the SPG region defined as the area between 15° − 60°W and 48° − 65°N to provide an assessment of fidelity of model simulations in this important metric in comparison with the data from the AVISO project (Archiving, Validation, and Interpolation of Satellite Oceanographic Data; Le Traon et al., 1998; Ducet et al., 2000). Here, we use a product available from a NASA Jet Propulsion Laboratory web site located at podaac.jpl.nasa.gov/dataset/AVISO_L4_DYN_TOPO_1DEG_1MO. The SSH time series anomalies calculated as the average SSHs for the SPG region with respect to the 1993−2007 mean are given in Fig. 14. The AVISO time series are included in each panel as the black lines. The figure also shows the correlation coefficients between the AVISO and models' time series as well as the linear trends for the 1993−2007 period for each model and from the AVISO data. NOCS clearly emerges as the major outlier in comparison with the AVISO data as the only model with a negative correlation coefficient (−0.19) and as the only model with a negative trend (−0.15 cm yr^{-1}). Half of the models have quite high correlations with the AVISO data with correlation coefficients of 0.96 or higher. The lowest correlations are in MIT and KIEL with 0.69 and 0.75, respectively. The trend in AVISO data is 0.45 cm yr^{-1}. The simulated trends are within 20% of this value in six of the models. These models are (with their trends in cm yr^{-1}) GISS2 (0.38), FSU2 (0.39), GFDL-GOLD (0.39), NCAR (0.39), GFDL-MOM (0.48), and BERGEN (0.51). The largest trend is in ICTP with 0.62 cm yr^{-1} which is within 30% of the AVISO-based value. The smallest positive trends occur in MIT, MRI-F, CNRM, and KIEL with 0.05, 0.07, 0.08, and 0.08 cm yr^{-1}, respectively.

8. AMOC and meridional heat transport variability

AMOC is the principal contributor to the Atlantic Ocean MHT in both observations and model simulations (see, e.g., Böning et al., 2001; Biastoch et al., 2008; Johns et al., 2011; Msadek et al., 2013). Here, we assess the relationships between the AMOC variability and that of the MHT by considering their simultaneous correlations and by performing simultaneous regressions of MHT onto AMOC. For this purpose, we use the AMOC maximum transports and MHT values obtained at 26.5°N for two reasons: i) this latitude is within the range of latitudes for maximum MHTs, and ii) there are observationally-based estimates from the RAPID data (Johns et al., 2011). We note again that the overlap period between the model simulations and the observations is very short: while we analyze the annual-mean data for the 1958−2007 period from the simulations, the observational data are available starting in April 2004 and their analyses usually use 10-day and 30-day means. The implications of such differences are discussed below.

Table 3 summarizes our results. We find that AMOC and MHT variability are very highly correlated with correlation coefficients of ≥ 0.9 in all, but two, of the models. The lowest correlations occur in INMOM and MRI-A with 0.86. These high correlations are consistent with the RAPID-based estimate of 0.97. Such good agreements between the model and RAPID-based AMOC and MHT correlations appear to be independent of the range of time averaging applied in the calculations. For example, we obtain similarly high correlations of 0.93 (for 1958−2007) and 0.96 (for 2004−2007) for NCAR when monthly-mean data are used. The regression coefficients vary between 0.042 and 0.068 PW Sv^{-1} with INMOM at the low end and CMCC, FSU, MRI-F, and NOCS at the high end of this range. We note that the latter four are among the models where the maximum anomalies in AMOC occur in the Southern Hemisphere. The model regression coefficients are all smaller than those of the RAPID-based estimates which are 0.079 PW Sv^{-1} (Johns et al., 2011) and 0.083 PW Sv^{-1} (Msadek et al., 2013) obtained using 10-day and 30-day means, respectively, for the April 2004 − October 2007 period and 0.077 PW Sv^{-1} (W. Johns, personal communication) obtained using 10-day means from April 2004 to mid-2014.

We think that this discrepancy between the model and observationally-based regressions is due to the use of annual-mean vs. 10-day or 30-day mean data in model vs. observational analysis. Specifically, we get 0.074 PW Sv^{-1} (for 1958−2007) and 0.078 PW Sv^{-1} (for 2004−2007) for NCAR when monthly-mean data are employed, both in rather good agreement with the RAPID-based estimates – in contrast with the annual-mean-based regression coefficient of 0.062 PW Sv^{-1}. Similarly, we find that the RAPID-based regression coefficient reduces to 0.067 PW Sv^{-1} when calculated with annual-mean data for the April 2004 − March 2014 period. The models are evenly divided in their intercept values with half above zero and half below zero values. While GISS2 has the highest intercept with +0.177 PW, MRI-F has the lowest value with −0.117 PW. As discussed in Msadek et al. (2013), the differences in regression coefficients and in intercept values among the models can be due to many reasons, and it is beyond the scope of the present study to investigate causes of these differences in each model. However, following Msadek et al. (2013), we offer differences in mean AMOC magnitudes; in correlations between AMOC and temperature fluctuations; and in the gyre component contributions and their variability as possible causes.

9. Variability relationships between AMOC and other fields

In this section, we investigate relationships between the simulated AMOC variability and those of MLD, SPG circulation, SPG SSH, and NAO. We use the AMOC maximum transport at 45°N time series as our primary AMOC index.

We first present in Fig. 15 the low-pass filtered, MMM time series of the AMOC index, March-mean MLD, and SPG barotropic streamfunction (BSF) (top panel), and the AMOC maximum transport time series at 26.5°N and SPG SSH (bottom panel). The top panel also includes a low-pass filtered NAO index, and our primary

AMOC index is repeated in the bottom panel. Here, MLD is calculated as an average for the LS − Irminger Sea region defined as the area between $15° - 60°$W and $48° - 60°$N, thus including the region extending from the southeast LS to the Irminger Sea which contains the largest MLD variability in the majority of the models (see Fig. 9). The SPG BSF and SSH represent average transport and surface height for the SPG domain defined in section 7. For NAO, we adopt the winter (December−March) sea level pressure PC1 time series from the CORE-II data sets as our index. The NAO index shows a stronger-than-normal subtropical high and a deeper-than-normal Icelandic low in its positive phase (NAO+). We note that all models are subject to the same NAO index because it is part of the forcing datasets. All time series are anomalies with respect to the 1958−2007 period, and shadings denote one standard deviation spreads of the models' time series from those of the respective MMM.

The figure shows several noteworthy features. First, changes in MLD tend to lead changes in AMOC. This is particularly evident after 1980: deepening in MLD leads AMOC intensification by a few years with the deepest MLDs and the largest AMOC transports occurring in 1992−1993 and 1995, respectively. Second, the NAO time series similarly lead those of AMOC, with changes in NAO and MLD tending to co-vary. There is a suggestion that NAO slightly leads MLD after about 1990. Third, AMOC and SPG BSF and SSH anomalies appear to be largely in-phase, noting that the negative BSF and SSH anomalies indicate strengthening of the cyclonic SPG circulation. However, the SPG SSH time series suggest that they tend to lead those of AMOC by a few years. In Yeager (2015), these co-variations of AMOC and SPG anomalies are shown to be associated with the bottom pressure torque which emerges as the primary driver in the barotropic vorticity equation responsible for decadal, buoyancy-forced changes in the gyre circulation, thus providing AMOC and SPG coupling. Finally, we note that the two AMOC time series do not show an appreciable lead−lag relationship until about 1985. Thereafter, anomalies at 45°N lead those at 26.5°N by about 5 years. A prominent example is the emergence and strengthening of positive AMOC anomalies at 26.5°N during the 1989−2000 period which follow a similar AMOC intensification at 45°N that occurs during the 1984−1995 period.

To establish the lead−lag relationships between the AMOC index time series and those of the MLD, SPG BSF, SPG SSH, and NAO, we next calculate the correlation functions among these time series. The resulting lead−lag correlations for each model are shown in Fig. 16 where the AMOC index leads for positive lags. The correlations are obtained using the low-pass filtered anomalies with respect to the 1958−2007 period. The figure also includes the MMM correlation function evaluated as the mean of the individual model correlations as well as 95% confidence levels calculated using a parametric bootstrap method (see section 2 for details). As above, MLD and BSF time series are evaluated as spatial averages for their respective regions, and SSH spatial averages use the same domain as in BSF.

We first summarize our analysis considering the MMM correlations shown as the black lines in Fig. 16. The maximum correlations (≈ 0.75) occur when positive MLD anomalies, i.e., MLD deepening, lead AMOC intensification by 2−3 years. As also suggested by Fig. 15, the correlation coefficient between the AMOC index and the SPG BSF time series is a maximum ($\approx |0.7|$) at lag of -1 to -2, again noting that the negative correlations indicate in-phase strengthening and weakening of AMOC and SPG. We see a similar relationship between the AMOC index and the SPG SSH time series with the largest negative correlations of about 0.6 occurring when SSH leads by 2−3 years. These lead−lag relationships between the AMOC index time series and those of SPG BSF and SSH along with the time series plots of Fig. 15 support the idea of monitoring the variations in the LS SSH as a proxy for AMOC changes as suggested by Yeager and Danabasoglu (2014). Lastly, we note that the NAO index leads the AMOC index by 2−4 years with a maximum correlation coefficient of about 0.6.

There are many differences among the individual correlation functions, for example, in their correlation coefficient magnitudes as well as in their lead−lag times for maximum correlations. We discuss only a few of these differences here both to provide some examples of such differences and to identify some models that depart from our MMM characterization. Starting with the AMOC and MLD correlation functions, we note that although INMOM also shows relatively strong correlations when MLD leads AMOC, it is the only model which has its maximum correlation when AMOC leads, indicating that MLDs continue to get deeper while AMOC begins to weaken. The maximum correlations vary between about 0.45 and 0.9 among the models, with ICTP at the low end and AWI, BERGEN, CNRM, INMOM, KIEL, MRI-F, and NCAR at the high end of this range. The low correlations in ICTP that are not statistically significant are likely due to low MLD variability in the LS − Irminger Sea region (Fig. 9) where the time-mean MLDs always remain very deep and the largest variabilities occur in the southern portion. In contrast with the rest of the models, GFDL-GOLD, GISS, MRI-A, and NOCS show earlier transitions to negative correlations starting at lag of 0. Consequently, these models have the largest negative correlation coefficients among the models. Although there does not seem to exist any clear relationships between the AMOC − MLD correlations and where the deepest MLDs occur in the models, we note that in MRI-A and NOCS – two of the models with earlier transitions to negative correlations – AMOC EOF1 anomalies are very weak at 45°N, indeed negative as shown in Fig. 4. Continuing with the AMOC and SPG BSF correlation functions, we find GISS2 and, to some degree, FSU distributions – both below the confidence levels – difficult to interpret due to their pronounced oscillatory behavior with relatively small correlation coefficients. In BERGEN, INMOM, and NCAR, the extrema in SPG transports are attained more than 2 years after the extrema in AMOC. Not surprisingly, there are general similarities in many individual model correlations between the AMOC vs. BSF and AMOC vs. SSH relationships. Only GFDL-GOLD and CNRM appear to have the longest lead times for SSH with 9 to 10+ years. Finally considering AMOC and NAO relationships, we identify MRI-A and NOCS as the major outliers, noting that while MRI-A is below our confidence limit, the minimum in NOCS is very near the 95% limit. They have small or even negative correlations prior to an AMOC maximum, and negative correlations persist through positive lags. As discussed above regarding AMOC − MLD relationships, this behavior in MRI-A and NOCS is likely related to the negative AMOC EOF1 anomalies present at the latitude of our AMOC index (Fig. 4), in contrast with the other models which show positive anomalies. Further, in MRI-A, data assimilation presumably impacts the relationship between AMOC and NOA. To the extent that NAO+ plays an important role in driving AMOC variability through its associated surface fluxes, as discussed previously, the NAO appears to be not a major factor in influencing AMOC variability in these two models. We also note that FSU has its largest positive correlations between AMOC and NAO following an AMOC intensification.

10. Summary and conclusions

We have presented an analysis of the simulated inter-annual to decadal variability and trends in the North Atlantic Ocean for the 1958−2007 period from twenty simulations participating in the CORE-II effort. A major focus has been the representation of AMOC variability. In addition, we have investigated connections between AMOC variability and those of some other fields such as NAO, subpolar MLDs, and

LS hydrographic properties to elucidate some variability mechanisms. This study is Part II of our companion paper, DY14, which documents the mean states in the North Atlantic from the same models, providing a baseline for the present variability analysis.

In general, AMOC variability shows three distinct stages on decadal time scales. During the first phase that lasts from 1958 until the mid- to late-1970s, AMOC remains weaker than its long-term (1958−2007) mean. Thereafter, AMOC intensifies with maximum transports achieved in the mid- to late-1990s. This enhancement is then followed by a weakening trend that continues until the end of our integration period. This sequence of low frequency AMOC variability cannot be directly confirmed by observations. However, it is consistent with the results of many other ocean hindcast simulations (see section 1 for a sampling of references) forced with various historical atmospheric datasets, including NCEP/NCAR and ECMWF ERA-40 reanalysis products.

A prominent and robust feature of the above characterization of the low frequency variability is the strengthening of AMOC between about the mid-1970s and the mid- to late-1990s, distinguished by an intensified and deeper-penetrating NADW cell. Previous studies show that this AMOC intensification is connected to enhanced DWF and associated mixed layer deepening in the subpolar North Atlantic, particularly in the LS region, driven by surface buoyancy fluxes and wind stress resulting from the persistent positive phase of the NAO. Increase in AMOC is then accompanied by more heat transport into the subpolar North Atlantic, contributing to the warming observed in the mid-1990s there. Although an in-depth analysis of AMOC variability mechanisms in the participating models is beyond the scope of the present study, our results support this variability mechanism. In particular, positive density and MLD anomalies precede AMOC intensification, and lead−lag relationships show that both

MLD and NAO indices lead AMOC enhancement by 2−4 years. Such a variability mechanism that suggests an important role for the NAO appears to be very similar to AMOC intrinsic variability mechanisms found in some CGCM control simulations (e.g., Danabasoglu et al., 2012).

The analysis of the mean states presented in DY14 shows that the larger AMOC mean transports are associated with deeper MLDs, resulting from increased salt content in the LS region. In sharp contrast, the increase in AMOC, i.e., the positive AMOC anomaly, discussed above is primarily associated with negative temperature anomalies in the LS region in both model simulations and in observations (see also Yeager and Danabasoglu, 2014). Concerning any links between the Nordic Seas overflow transports and AMOC, DY14 finds no clear links between the mean AMOC and overflow transports. Unfortunately, an investigation of this relationship for variability purposes remains beyond the scope of the present study, requiring a dedicated effort of its own with additional model outputs that are not currently available.

Arguably, the level of general agreement in the representation of AMOC variability, including year-to-year changes and long-term trends, among the forward models participating in CORE-II appears to be substantially greater than among various reanalysis products (Karspeck et al., 2015). Such a general agreement among the models also extends to characterization of MLD and SSH variability in the subpolar North Atlantic. Furthermore, the observed variability of the North Atlantic SSTs is reproduced remarkably well by all the models. These findings suggest that simulated temporal characteristics of the variables considered here are primarily dictated by the variability and trends in the CORE-II atmospheric datasets which include the impacts of ocean dynamics from nature superimposed onto external and anthropogenic effects. The general agreements among the models in their depictions of AMOC, MLD, and SSH variability and trends in the North Atlantic do not necessarily indicate that the models accurately capture variability and trends seen in nature because there are undoubtedly errors in the forcing datasets and the models have errors and common, systematic biases. Indeed, agreements in variability and trends occur in the presence of large mean-state differences among the models – as well as large mean biases from observations – as documented in DY14. In that study, the overarching hypothesis, namely that global ocean – sea-ice models integrated using the same inter-annually varying atmospheric forcing datasets will produce qualitatively similar mean and variability in their simulations, is found to be not satisfied for the mean states in the North Atlantic. In contrast, based on the present results, there appears to be more support for this hypothesis for variability in the North Atlantic. A similar conclusion is also reported in Wang et al. (2015) where the variability in the freshwater content and transports and sea-ice in the Arctic Ocean is found to be represented rather consistently among the models participating in CORE-II in spite of substantial differences in their mean states and mean state biases from observations.

Despite these general agreements, there are many differences – some significant – among the models, particularly in the spatial structures of variability patterns. For example, amplitudes and spatial extents of the largest SST and MLD anomalies differ among the models, reflecting the role of simulated ocean dynamics. Another notable difference occurs in the location of the largest AMOC anomalies (positive as depicted in Fig. 4). While the majority of the models have their maximum variability in the Northern Hemisphere, other models show enhanced variability in the Southern Hemisphere. Whether the maximum anomalies are located in the Northern or Southern Hemispheres does not appear to be related to the properties of the Southern Ocean meridional overturning circulations in these simulations (see Farneti et al., 2015). Similarly, there are no obvious connections between the subpolar

North Atlantic MLDs and where the maximum AMOC variability occurs. We do find, however, that the models that have their maximum variability in the Southern Hemisphere or in the vicinity of the equator tend to show weaker and statistically less significant AMOC trends, and their AMOC EOF1s account for a smaller fraction of their total variance in AMOC in comparison to those models with AMOC maximum variability in the Northern Hemisphere.

As in DY14, the differences among the model solutions do not suggest an obvious grouping of the models based on either their lineage, vertical coordinate representations, or surface salinity restoring strengths. Again, we attribute these differences primarily to use of different subgrid scale parameterizations and their parameter values; differences in horizontal and vertical grid resolutions; and use of different sea-ice models along with diverse snow and sea-ice albedo treatments. Among the forward models, NOCS appears to deviate substantially in some of its low-frequency and trend characteristics from the other models. For example, it is the only model with a negative SSH trend in the subpolar North Atlantic for the 1993−2007 period; it is the only model with positive AMOC trends at both 26.5° and 45°N for the 1975−2007 period; and it shows the lowest trend in its LS upper-ocean density time series for the 1970−1995 period. These NOCS features are certainly in contrast with the solutions from the other NEMO-based models and the reasons for these differences remain unclear. However, several preliminary NOCS simulations that are underway in which the skew-flux form of GM90 is replaced with its advective form and / or associated tapering of both the thickness and isopycnal diffusivities within the surface mixed layer has been modified appear to show low frequency variability and trends that are in much better agreement with the other NEMO-based models.

Based on both our present study and other work (e.g., Yeager et al., 2012; Yeager and Danabasoglu, 2014), we think that the CORE-II experimental protocol and resulting simulations can be confidently used for studies concerning variability and its mechanisms on inter-annual and decadal times scales in the North Atlantic and elsewhere (e.g., Griffies et al., 2014; Farneti et al., 2015). The CORE-II effort has gained unprecedented momentum and exposure over the past few years, attracting participation of many ocean and climate modeling groups worldwide. As such, we think that it has now reached a mature state as the community standard for global ocean – sea-ice simulations. Encouraged by these developments, the CORE-II framework is recently proposed and endorsed as an Ocean Model Inter-comparison Project (OMIP) for inclusion in the Coupled Model Inter-comparison Project phase 6 (CMIP6), again coordinated by the CLIVAR Ocean Model Development Panel (OMDP).

Acknowledgments

NCAR is sponsored by the U. S. National Science Foundation (NSF). The CESM is supported by the NSF and the U. S. Department of Energy. S. G. Yeager was supported by the NOAA Climate Program Office under Climate Variability and Predictability Program grants NA09OAR4310163 and NA13OAR4310138 and by the NSF Collaborative Research EaSM2 grant OCE-1243015 to NCAR. W. M. Kim was supported by the NOAA Climate Program Office under Climate Variability and Predictability Program grant NA13OAR4310136 to Texas A&M University. ACCESS modeling work has been undertaken as part of the Australian Climate Change Science Program, funded jointly by the Department of Climate Change and Energy Efficiency, the Bureau of Meteorology and CSIRO, and was supported by the National Computational Infrastructure facility at the Australian National University. AWI is a member of the Helmholtz Association of German Research Centers. Q. Wang and D. Sidorenko were funded by the Helmholtz Climate Initiative REK-

LIM (Regional Climate Change) project. The BERGEN contribution was supported by the Research Council of Norway through the EarthClim (207711/E10) and NOTUR/NorStore projects, as well as the Centre for Climate Dynamics at the Bjerknes Centre for Climate Research. The CMCC contribution received funding from the Italian Ministry of Education, University, and Research and the Italian Ministry of Environment, Land, and Sea under the GEMINA project. INMOM was sponsored by the Russian Science Foundation (project number 14-27-00126). The KIEL contribution acknowledges support within the Co-Operative Project RACE - Regional Atlantic Circulation and Global Change funded by the German Federal Ministry for Education and Research (BMBF) under grant number 03F0651B and computing resources from the North-German Supercomputing Alliance (HLRN). P. G. Fogli thanks W. G. Large, J. Tribbia, M. Vertenstein, G. Danabasoglu, and D. Bailey for their support and help in bringing NEMO into the CESM framework while vising NCAR. E. Fernandez was supported by the BNP-Paribas foundation via the PRECLIDE project under the CNRS research convention agreement 30023488. We thank M. Harrison and R. Hallberg of GFDL for assistance with defining the GFDL-GOLD configuration, and R. Msadek and Y. M. Ruprich-Robert of GFDL for comments on an earlier version of the manuscript. Finally, we thank both the international CLIVAR and U. S. CLIVAR projects for patiently sponsoring the Working Group on Ocean Model Development (now, Ocean Model Development Panel) over the years as COREs were developed.

Appendix A. Two new HYCOM simulations

The FSU HYCOM used in DY14 was based on an earlier version of HYCOM which advects density and S (instead of θ and S) and therefore does not conserve heat – see Griffies et al. (2014) for a discussion of impacts of this choice on sea level. For the present study, a new HYCOM simulation, denoted as FSU2, has been performed with the formulation that advects θ and S, thus conserving heat. Another new contribution that also uses the heat conserving formulation of HYCOM is GISS2. Here, we give brief summaries of these two new contributions in Appendix A.1 and Appendix A.2 for FSU2 and GISS2, respectively. Appendix A.3 includes a note on the use of σ_1 vs. σ_2 vertical coordinates in HYCOM. A short description of FSU2 and GISS2 time-mean solutions is presented in Appendix A.4, considering only AMOC and MHT distributions.

Appendix A.1. FSU2

FSU2 is a global configuration of HYCOM (Bleck, 2002; Chassignet et al., 2003; Halliwell, 2004). The grid is a tripolar (Mercator grid smoothly connecting to a bipolar grid patch at about 47°N) Arakawa C-grid of 0.72° horizontal resolution with refinement at the equator. There are 500 and 382 grid cells in the zonal and meridional directions, respectively. The bottom topography is derived from the 2-minute NAVO / Naval Research Laboratory DBDB2 global dataset. The vertical discretization combines pressure coordinates at the surface, isopycnic coordinates in the stratified open ocean, and sigma coordinates over shallow coastal regions (Chassignet et al., 2003, 2006). Thirty-two hybrid layers whose σ_2 target densities range from 28.10 to 37.25 kg m^{-3} are used. The initial conditions in θ and S are given by the Polar Science Center Hydrographic Climatology version 2 dataset (PHC2; a blending of the Levitus et al. (1998) dataset with modifications in the Arctic Ocean based on Steele et al. (2001)). The ocean model is coupled to the sea-ice model CICE (Hunke and Lipscomb, 2010) that provides the ocean-ice fluxes. Turbulent air-sea fluxes are computed using the Large and Yeager (2009) bulk formulae. Surface freshwater fluxes are applied as virtual salt fluxes as in FSU. Surface salinity is restored over the entire domain with a piston velocity of 50 m over 4 years everywhere, except for the Antarctic region where the piston velocity is 50 m over 6 months. In addition, a global normalization is applied to the restoring salinity flux at each time step. Vertical mixing is provided by the K-Profile Parameterization (KPP; Large et al., 1994) with a background diffusivity of 10^{-5} m^2 s^{-1} and tracers are advected using a second-order flux corrected transport scheme. Lateral Laplacian diffusion of $0.03\Delta x$ is applied on θ and S and a combination of Laplacian ($0.03\Delta x$) and biharmonic ($0.05\Delta x^3$) dissipation is applied on the velocities. Here, Δx represents grid spacing. Interface pressure smoothing, corresponding to GM90 as discussed in Gent (2011), is applied through a biharmonic operator, with a mixing coefficient determined by the grid spacing (in m) times a velocity scale of 0.02 m s^{-1} everywhere except in the Pacific and Atlantic north of 40°N where a Laplacian operator with a velocity scale of 0.01 m s^{-1} is used. The use of a biharmonic operator differs from GM90, but still ensures conversion from mean available potential energy to eddy potential energy. The interface pressure smoothing tapers off when the generalized vertical coordinate of HYCOM switches from isopycnal to pressure, mostly in the mixed layer and in unstratified regions. In such regions, lateral diffusion is oriented along pressure surfaces rather than rotated to neutral directions. No parameterization has been implemented for abyssal overflows.

We summarize the main differences between FSU2 and the version introduced in DY14 – labeled as FSU – as follows (FSU2 vs. FSU): (i) turbulent air-sea fluxes use Large and Yeager (2009) bulk formulae vs. Kara et al. (2005) bulk formulae; (ii) version 2.2.74 vs. version 2.2.21; (iii) θ and S advection vs. density and S advection; (iv) tripolar grid of finer resolution (0.72° vs. 1°); (v) sea-ice model CICE v4.0 vs. CSIM (Community Sea-Ice Model; Briegleb et al., 2004; Holland et al., 2006); and (vi) surface salinity restoring time scale of 6 months vs. 4 years over 50 m in the

Antarctic region.

Appendix A.2. GISS2

The HYCOM version used at the National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS), denoted as GISS2, represents an updated version of the ocean component of the climate model described in Sun and Bleck (2006). It uses a Mercator grid, which smoothly connects to a bipolar grid patch at about 57°N. The horizontal mesh in the Mercator domain is $1^\circ \times 1^\circ$ cos(latitude), but meridional resolution is enhanced near the equator, resulting in a $1/3^\circ$ meridional mesh size at the equator. There are 360 and 387 grid points in the zonal and meridional directions, respectively (with the Bering Strait being the northernmost grid point in the extended Atlantic). The model is configured with 26 hybrid σ_1 coordinate levels. The adoption of this σ_1 coordinate differs from Sun and Bleck (2006) where a σ_2 coordinate was used. The bottom topography is obtained by spatially integrating ETOPO5 data of 5 minute spatial resolution over each model grid cell, without further smoothing. The initial θ and S are given by the PHC3 climatology. A non-slab KPP mixed layer sub-model (Halliwell, 2004) is employed. GISS2 uses the same prescriptions to specify lateral diffusivity and viscosity as in FSU2 with the exception that the velocity scale used in the biharmonic operator is a global constant set at 0.05 m s^{-1}.

As in the original FSU contribution, GISS2 deviates from the suggested CORE-II protocol in one important aspect. Namely, turbulent air-sea fluxes are computed using the Kara et al. (2005) bulk formulae, instead of the Large and Yeager (2009) bulk formulae. However, the other details of the forcing follow the protocol. Thus, no restoring is applied to SSTs and no additional adjustment of surface heat flux components, e.g., shortwave heat flux, are made. As a consequence, the global-mean

θ in GISS2 increases by 1/3°C over the course of the 300-year simulation. Surface freshwater fluxes are applied as virtual salt fluxes. Surface salinity is restored over the entire domain with a piston velocity of 50 m over 4 years. Precipitation is multiplied by a factor which aims to prevent long-term salinity trends. This factor is updated monthly based on the departure of global-mean salt content from its initial value, using a 1-year time scale. The adjustment factor stabilizes around 0.97, implying a roughly 3% reduction of the imposed precipitation.

The sea-ice model employed in GISS2 is a single-layer thermodynamic model with ice advection by surface currents and a *shaving* device that laterally spreads ice exceeding a prescribed thickness. Thus, it differs from Sun and Bleck (2006), where the coupled ocean-atmosphere climate simulations at GISS use a more realistic sea-ice model. One shortcoming of this highly simplified model is that melting and freezing processes do not involve any exchange of water mass between ice and water; instead, they spawn virtual salt fluxes. Since melting (freezing) reduces (increases) ocean salinity, sea ice in this scheme contributes with a minus sign to the salt budget. When attempting to reconcile surface freshwater fluxes with trends in the overall oceanic salt content, one must be aware of this somewhat counter-intuitive aspect of the sea-ice model.

We identify five major differences between FSU2 and GISS2 configurations. They are (FSU2 vs. GISS2): (i) nominal horizontal resolution of 0.72° vs. 1°; (ii) σ_2 vertical coordinate with 32 layers vs. σ_1 vertical coordinate with 26 layers; (iii) tripolar grid matching at 47°N vs. at 57°N; (iv) CICE4.0 sea-ice model vs. one-layer thermodynamic sea-ice model; and (v) use of Large and Yeager (2009) vs. Kara et al. (2005) bulk formulae.

Appendix A.3. A note on use of σ_1 vs. σ_2 vertical coordinates in HYCOM

A few remarks are in order to explain the choice of σ_1 as vertical coordinate in GISS2 in contrast with the use of σ_2 coordinate in FSU and FSU2. A major problem in models featuring sloping coordinate surfaces is the two-term expression for the horizontal pressure gradient force. In HYCOM, the numerically challenging two-term pressure gradient force is transformed into a more benign, single-term expression by treating sea water as incompressible and, for dynamic consistency with this approximation, by replacing density with a globally referenced potential density (ρ_{pot}) in the equation of state (Spiegel and Veronis, 1960).

One shortcoming of the above approximation is that a water column which is stably stratified in the real ocean may not be stably stratified in ρ_{pot} space. The choice of σ_2 in HYCOM, traditionally regarded as the best compromise, is particularly problematic in the upper Southern Ocean where convection triggered by a reversal of the vertical ρ_{pot} gradient can weaken the seasonal summertime halocline to the point where it becomes hard to form new ice in the fall. Without ice cover, the Southern Ocean acts as a heat source in austral winter, with grave consequences in a coupled climate model.

It is for this reason that in the GISS version of HYCOM, i.e., GISS2, σ_2 has been replaced by σ_1, both in the equation of state and as vertical coordinate. Static stability problems in the abyssal Atlantic due the use of σ_1 as vertical coordinate have been found to be less serious than expected – in the sense that they do not appear to preclude the existence of an abyssal, Southern-Ocean driven overturning cell.

The HYCOM versions in FSU and FSU2 add the thermobaricity treatment of Sun et al. (1999) to the basic Boussinesq-related approximations listed above. Accounting for thermobaric effects has been found to reduce Southern Ocean sea-ice biases in the σ_2-based FSU and FSU2 models. GISS2 does not account for thermobaricity, relying instead on the use of σ_1 to alleviate this problem.

Appendix A.4. Time-mean AMOC and MHT in FSU2 and GISS2

A detailed analysis of the time-mean solutions from FSU2 and GISS2, as was done in DY14 for the other participating models, is beyond the scope of the present study. Instead, we only provide a brief assessment of their time-mean AMOC and MHT distributions, considering the solutions from the fifth cycle of their CORE-II simulations.

Figure 17 shows the time-mean (years 1988−2007 mean) AMOC distributions in depth−latitude space from FSU, FSU2, and GISS2, corresponding to Fig. 3 of DY14. With $<$ 8 Sv, FSU has the weakest NADW maximum transport among all the participating models. This maximum transport is $>$ 14 Sv and $>$ 22 Sv in FSU2 and GISS2, respectively. The NADW penetration depth as measured by the depth of the zero contour line is deeper in FSU2 and GISS2 than in FSU. Indeed, the NADW penetration depth exceeds 5 km in GISS2. In both FSU2 and GISS2, the transports associated with the Antarctic Bottom Water (AABW) are quite weak.

We provide a quantitative comparison of the AMOC profiles from FSU2 and GISS2 to the profile based on the RAPID data (Cunningham et al., 2007) at 26.5°N in Fig. 18a. The figure corresponds to Fig. 5 of DY14 and uses the 4-year mean for years 2004−2007 for the model data while the RAPID data represents the 4-year mean for April 2004 − March 2008. The profile for FSU is also included for reference purposes. We note that the profiles show the total integrated transport between the surface and a given depth, with negative and positive slopes indicating northward and southward flow, respectively. The RAPID estimate for the NADW maximum transport at this latitude is 18.6 Sv, occurring at about 1000-m depth, with about

±1 Sv as its annual-mean range over this short period.

As indicated above, FSU has the lowest NADW maximum transport among all the models with only 5.3 Sv, and its profile deviates quite substantially from the RAPID profile. FSU2 shows major improvements from FSU in both the NADW maximum transport magnitude with 11.5 Sv and the vertical structure of the transport profile. Nevertheless, the NADW maximum transport in FSU2 still remains considerably lower than in RAPID. In GISS2, the NADW maximum transport of about 19.2 Sv is only slightly stronger than in RAPID and its profile captures that of RAPID well, including the NADW penetration depth. As in all the other participating models (see Fig. 5 of DY14), both FSU2 and GISS2 show significant departures from the RAPID profile in their representations of the AABW with near-zero transports at this latitude. We note that the GISS2 profile arguably shows one of the best comparisons with that of RAPID among all the participating models.

We present the time-mean (years 1988−2007 mean) Atlantic Ocean MHT distributions from FSU, FSU2, and GISS2 in Fig. 18b, as in Fig. 6 of DY14. The figure also includes the implied transport estimates from Large and Yeager (2009) calculated using the CORE-II datasets with observed SSTs and sea-ice for the 1984−2006 period, and the direct estimates with their uncertainty ranges from Bryden and Imawaki (2001) and the estimate from the RAPID data (Johns et al., 2011). As a result of its weakest NADW transport, FSU has the lowest MHT among the participating models with about 0.40 PW. In addition, FSU is the only model with southward heat transport in the Atlantic basin. Again, FSU2 represents an improved solution over FSU, with a maximum MHT of 0.86 PW. Still, however, FSU2 MHT distribution remains below the range of the estimates, except south of 10°S. With the exception of north of 60°N, GISS2 distribution is within the bounds of the estimates, with maximum heat transports of about 1.1 PW, occurring at 10°N and 30°N. Including

FSU2 and GISS2, none of the models participating in CORE-II is able to obtain the RAPID based estimate of 1.33 PW at this latitude for this time period – see Msadek et al. (2013) and DY14 for a discussion of lower MHTs in the model simulations.

Appendix B. Departures from the CORE-II protocol

Despite our best efforts, about half of the participating models did not follow the recommendations of the CORE-II protocol exactly. The departures include use of different bulk formulae, modifications of the Large and Yeager (2009) bulk formulae, and changes in the forcing datasets.

For historical reasons, INMOM uses the bulk formulae adopted from the Arctic Ocean Model Inter-comparison Project (AOMIP), while FSU and GISS2 use the Kara et al. (2005) formulae. In MRI-F and MRI-A (data assimilated version of MRI-F), the air-ice neutral bulk transfer coefficients are modified to follow the values in Mellor and Kantha (1989), because the thermodynamic part of their sea-ice model is based on Mellor and Kantha (1989). Specifically, the momentum transfer coefficient is set to 3×10^{-3} and the transfer coefficients for sensible heat and evaporation are set to 1.5×10^{-3}, in contrast with a value of 1.63×10^{-3} used in Large and Yeager (2009).

Regarding the modifications of the forcing datasets, CERFACS, CNRM, and NOCS impose a seasonal cycle to the Antarctic runoff whereby four times the annual-mean value is applied over the summer months, i.e., January, February, and March, and zero runoff is used for the rest of the year. In the CORE-II protocol, the Antarctic runoff is time-invariant.

In addition to using different bulk formulae, INMOM adds 1 m s^{-1} to the CORE-II wind data uniformly, prior to the calculation of the wind stress to improve their sea-ice simulations, particularly in the Arctic basin. As a result, the wind stress for INMOM is larger than in any other model – see Fig. 3 of Farneti et al. (2015).

Finally, KIEL has three differences from the protocol: i) the wind stress near Antarctica is modified to include a parameterization of katabatic winds; ii) a different runoff dataset – though still based on Dai and Trenberth (2002) – is adopted; and iii) model potential temperature and salinity are restored to observed monthly-mean climatology in the Gulf of Cadiz region to improve the representation of the Mediterranean outflow. This restoring is applied within the $627-1297$ m depth range and its strength varies with depth and distance from the coast.

We do not know the impacts of these departures from the CORE-II protocol on model solutions. While some, e.g., transfer coefficient changes, are expected to have minor impacts, the use of different bulk formulae can result in larger changes in model solutions. It is, nevertheless, clear that, despite our best efforts, we are still short of achieving our ultimate goal of having all groups follow the protocol fully. The protocol does not specify a particular recipe for surface salinity restoring; it is left to the modeling groups to choose their optimal salinity restoring procedure. Thus, given the diversity among the models in their use of quite different restoring time scales – see Appendix C of DY14 – it is possible that the differences in model solutions due to their departures from the CORE-II protocol could be substantially masked.

Appendix C. List of Major Acronyms

– ACCESS: Australian Community Climate and Earth System Simulator

– AMOC: Atlantic meridional overturning circulation

– AMV: Atlantic multi-decadal variability

– AVISO: Archiving, Validation, and Interpolation of Satellite Oceanographic

Data

- AWI: Alfred Wegener Institute
- BSF: Barotropic streamfunction
- CERFACS: Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique
- CESM: Community Earth System Model
- CGCM: Coupled general circulation model
- CICE: Sea ice model
- CLIVAR: Climate Variability and Predictability
- CMCC: Centro Euro-Mediterraneo sui Cambiamenti Climatici
- CNRM: Centre National de Recherches Météorologiques
- CORE-II: Coordinated Ocean-ice Reference Experiments phase II
- CSIM: Community Sea Ice Model
- DWF: Deep water formation
- DY14: Danabasoglu et al. (2014)
- ECMWF: European Center for Medium-range Weather Forecasting
- EOF: Empirical orthogonal function
- FESIM: Finite Element Sea-ice Model

- FESOM: Finite Element Sea-ice Ocean Model
- FSU: Florida State University
- FSU2: Version 2 of the FSU contribution
- GFDL: Geophysical Fluid Dynamics Laboratory
- GISS: Goddard Institute for Space Studies
- GISS2: HYCOM contribution from GISS
- GM90: Gent and McWilliams (1990) parameterization
- GOLD: Generalized Ocean Layer Dynamics
- HYCOM: HYbrid Coordinate Ocean Model
- ICTP: International Centre for Theoretical Physics
- INMOM: Institute of Numerical Mathematics Ocean Model
- KIEL: Refers to the contribution from the Helmholtz Center for Ocean Research from Kiel
- KPP: K-Profile Parameterization (Large et al., 1994)
- LIM: Louvain-la-Neuve Sea Ice Model
- LS: Labrador Sea
- MHT: Meridional heat transport
- MICOM: Miami Isopycnal Coordinate Ocean Model

- MIT: Massachusetts Institute of Technology
- MITgcm: Massachusetts Institute of Technology general circulation model
- MLD: mixed layer depth
- MMM: Multi-model mean
- MOM: Modular Ocean Model
- MOVE: Multivariate Ocean Variational Estimation
- MRI: Meteorological Research Institute
- MRI.COM: Meteorological Research Institute Community Ocean Model
- MRI-A: Data assimilated version of MRI-F
- MRI-F: MRI contribution
- NAC: North Atlantic Current
- NADW: North Atlantic Deep Water
- NAO: North Atlantic Oscillation
- NASA: National Aeronautics and Space Administration
- NCAR: National Center for Atmospheric Research
- NCEP: National Centers for Environmental Prediction
- NEMO: Nucleus for European Modelling of the Ocean
- NOAA: National Oceanic and Atmospheric Administration

- NOCS: National Oceanography Centre Southampton
- NorESM-O: Norwegian Earth System Model ocean component
- OMDP: Ocean Model Development Panel
- ORCA: Ocean model configuration of the NEMO model
- PC: Principal component
- PHC: Polar Science Center Hydrographic Climatology
- POP2: Parallel Ocean Program version 2
- RAPID: Rapid Climate Change mooring data
- SIS: GFDL Sea Ice Simulator
- SPG: Subpolar gyre
- SSH: Sea surface height
- SST: Sea surface temperature

References

Beismann, J.-O., Barnier, B., 2004. Variability of the meridional overturning circulation of the North Atlantic: sensitivity to overflows of dense water masses. Ocean Dynamics 54, 92–106.

Bentsen, M., Drange, H., Furevik, T., Zhou, T., 2004. Simulated variability of the Atlantic meridional overturning circulation. Clim. Dyn. 22, 701–720.

Biastoch, A., Böning, C. W., Getzlaff, J., Molines, J.-M., Madec, G., 2008. Causes of interannual - decadal variability in the meridional overturning circulation of the mid-latutude North Atlantic Ocean. J. Climate 21, 6599–6615.

Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modelling 4, 55–88.

Böning, C. W., Dieterich, C., Barnier, B., Jia, Y. L., 2001. Seasonal cycle of the meridional heat transport in the subtropical North Atlantic: a model intercomparison in relation to observations near 25°N. Prog. Oceanogr. 48, 231–253.

Böning, C. W., Scheinert, M., Dengg, J., Biastoch, A., Funk, A., 2006. Decadal variability of subpolar gyre transport and its reverberation in the North Atlantic overturning. Geophys. Res. Lett. 33, L21S01.

Branstator, G., Teng, H., 2010. Two limits of initial-value decadal predictability in a CGCM. J. Climate 23, 6292–6311.

Briegleb, B. P., Bitz, C. M., Hunke, E. C., Lipscomb, W. H., Holland, M. M., Schramm, J. L., Moritz, R. E., 2004. Scientific description of the sea-ice component in the Community Climate System Model, version three. NCAR Tech. Note

NCAR/TN-463+STR, National Center for Atmospheric Research, Boulder, Colorado.

Brodeau, L., Barnier, B., Treguier, A. M., Penduff, T., Gulev, S., 2010. An ERA40-based atmospheric forcing for global ocean circulation models. Ocean Modelling 31, 88–104.

Bryan, F. O., Danabasoglu, G., Nakashiki, N., Yoshida, Y., Kim, D. H., Tsutsui, J., Doney, S. C., 2006. Response of North Atlantic thermohaline circulation and ventilation to increasing carbon dioxide in CCSM3. J. Climate 19, 2382–2397.

Bryden, H., Imawaki, S., 2001. Ocean heat transport. In: Siedler, G., Church, J., Gould, J. (Eds.), Ocean circulation and climate. Vol. 77 of International Geophysical Series. Academic Press, pp. 317–336.

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Wallcraft, A. J., Metzger, E. J., Blanton, B. O., Lozano, C., Rao, D. B., Hogan, P. J., Srinivasan, A., 2006. Generalized vertical coordinates for eddy-resolving global and coastal forecasts. Oceanography 19, 20–31.

Chassignet, E. P., Smith, L. T., Halliwell, G. T., Bleck, R., 2003. North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. J. Phys. Oceanogr. 33, 2504–2526.

Cunningham, S. A., Kanzow, T., Rayner, D., Baringer, M. O., Johns, W. E., Marotzke, J., Longworth, H. R., Grant, E. M., Hirschi, J. J.-M., Beal, L. M., Meinen, C. S., Bryden, H. L., 2007. Temporal variability of the Atlantic meridional overturning circulation at 26.5°N. Science 317, 935–938.

Dai, A., Trenberth, K. E., 2002. Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. J. Hydrometeorology 3, 660–687.

Danabasoglu, G., 2008. On multidecadal variability of the Atlantic meridional overturning circulation in the Community Climate System Model version 3. J. Climate 21, 5524–5544.

Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., Biastoch, A., Böning, C., Bozec, A., Canuto, V. M., Cassou, C., Chassignet, E., Danilov, S., Diansky, N., Drange, H., Farneti, R., Fernandez, E., Fogli, P. G., Forget, G., Fujii, Y., Griffies, S. M., Gusev, A., Heimbach, P., Howard, A., Jung, T., Kelley, M., Large, W. G., Leboissetier, A., Lu, J., Marsland, S. J., Masina, S., Navarra, A., Nurser, A. J. G., Pirani, A., Salas y Mélia, D., Samuels, B. L., Scheinert, M., Sidorenko, D., Treguier, A.-M., Tsujino, H., Uotila, P., Valcke, S., Voldoire, A., Wang, Q., 2014. North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. Ocean Modelling 73, 76–107.

Danabasoglu, G., Yeager, S. G., Kwon, Y.-O., Tribbia, J. J., Phillips, A. S., Hurrell, J. W., 2012. Variability of the Atlantic meridional overturning circulation in CCSM4. J. Climate 25, 5153–5172.

Delworth, T., Manabe, S., Stouffer, R. J., 1993. Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model. J. Climate 6, 1993–2011.

Delworth, T. L., Mann, M. E., 2000. Observed and simulated multidecadal variability in the Northern Hemisphere. Clim. Dyn. 16, 661–676.

Delworth, T. L., Zeng, F., 2012. Multicentennial variability of the Atlantic Meridional

Overturning Circulation and its climate influence in a 4000 year simulation of the GFDL CM2.1 climate model. Geophys. Res. Lett. 39, L13702.

Deshayes, J., Frankignoul, C., 2008. Simulated variability of the circulation in the North Atlantic from 1953 to 2003. J. Clim. 21, 4919–4933.

Doney, S. C., Yeager, S., Danabasoglu, G., Large, W. G., McWilliams, J. C., 2007. Mechanisms governing interannual variability of upper-ocean temperature in a global ocean hindcast simulation. J. Phys. Oceanogr. 37, 1918–1938.

Dong, B., Sutton, R. T., 2005. Mechanism of interdecadal thermohaline circulation variability in a coupled ocean - atmosphere GCM. J. Climate 18, 1117–1135.

Ducet, N., Le Traon, P.-Y., Reverdin, G., 2000. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. J. Geophys. Res. 105, 19477–19498.

Eden, C., Willebrand, J., 2001. Mechanism of interannual to decadal variability of the North Atlantic circulation. J. Climate 14, 2266–2280.

Farneti, R., Downes, S. M., Griffies, S. M., Marsland, S. J., Behrens, E., Bentsen, M., Bi, D., Biastoch, A., Böning, C., Bozec, A., Canuto, V. M., Chassignet, E., Danabasoglu, G., Danilov, S., Diansky, N., Drange, H., Fogli, P. G., Gusev, A., Hallberg, R. W., Howard, A., Ilicak, M., Jung, T., Kelley, M., Large, W. G., Leboissetier, A., Long, M., Lu, J., Masina, S., Mishra, A., Navarra, A., Nurser, A. J. G., Patara, L., Samuels, B. L., Sidorenko, D., Tsujino, H., Uotila, P., Wang, Q., Yeager, S. G., 2015. An assessment of Antarctic Circumpolar Current and Southern Ocean meridional overturning circulation during 1958-2007 in a suite of interannual CORE-II simulations. Ocean Modelling 93, 84–120.

Farneti, R., Vallis, G. K., 2011. Mechanisms of interdecadal climate variability and the role of ocean-atmosphere coupling. Clim. Dyn. 36, 289–308.

Fox-Kemper, B., Danabasoglu, G., Ferrari, R., Griffies, S. M., Hallberg, R. W., Holland, M. M., Maltrud, M. E., Peacock, S., Samuels, B. L., 2011. Parameterization of mixed layer eddies. Part III: Implementation and impact in global ocean climate simulations. Ocean Modelling 39, 61–78.

Gent, P. R., 2011. The Gent-Mcwilliams parameterization: 20/20 hindsight. Ocean Modelling 39, 2–9.

Gent, P. R., McWilliams, J. C., 1990. Isopycnal mixing in ocean circulation models. J. Phys. Oceanogr. 20, 150–155.

Griffies, S. M., Bryan, K., 1997. Predictability of North Atlantic multidecadal climate variability. Science 275, 181–184.

Griffies, S. M., Winton, M., Samuels, B., Danabasoglu, G., Yeager, S., Marsland, S., Drange, H., Bentsen, M., 2012. Datasets and protocol for the CLIVAR WGOMD Coordinated Ocean sea-ice Reference Experiments (COREs). WCRP Report No. 21/2012.

Griffies, S. M., Yin, J., Durack, P. J., Goddard, P., Bates, S. C., Behrens, E., Bentsen, M., Bi, D., Biastoch, A., Böning, C. W., Bozec, A., Chassignet, E., Danabasoglu, G., Danilov, S., Domingues, C. M., Drange, H., Farneti, R., Fernandez, E., Greatbatch, R. J., Holland, D. M., Ilicak, M., Large, W. G., Lorbacher, K., Lu, J., Marsland, S. J., Mishra, A., Nurser, A. J. G., Salas y Mélia, D., Palter, J. B., Samuels, B. L., Schröter, J., Schwarzkopf, F. U., Sidorenko, D., Treguier, A. M., Tseng, Y.-H., Tsujino, H., Uotila, P., Valcke, S., Voldoire, A., Wang, Q., Winton, M., Zhang, X., 2014. An assessment of global and regional sea level for years 1993-2007 in a suite of interannual CORE-II simulations. Ocean Modelling 78, 35–89.

Gusev, A. V., Diansky, N. A., 2014. Numerical simulation of the World ocean circulation and its climatic variability for 19482007 using the INMOM. Izvestiya, Atmospheric and Oceanic Physics 50, 1–12.

Häkkinen, S., 1999. Variability of the simulated meridional transport in the North Atlantic for the period 1951-1993. J. Geophys. Res. 104, 10991–11007.

Häkkinen, S., Rhines, P. B., 2004. Decline of subpolar North Atlantic circulation during the 1990s. Science 304, 555–559.

Halliwell, G. R., 2004. Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM). Ocean Modelling 7, 285–322.

Hibler, W., 1979. A dynamic thermodynamic sea ice model. J. Phys. Oceanogr. 9, 815–846.

Holland, M. M., Bitz, C. M., Hunke, E. C., Lipscomb, W. H., Schramm, J. L., 2006. Influence of the sea ice thickness distribution on polar climate in CCSM3. J. Climate 19, 2398–2414.

Hunke, E. C., Lipscomb, W. H., 2010. CICE: the Los Alamos Sea Ice Model documentation and software users manual version 4.1. Los Alamos National Laboratory Tech. Rep. LA-CC-06012, Los Alamos, NM.

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., Rosinski, J., 2008. A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. J. Climate 21, 5145–5153.

Johns, W. E., Baringer, M. O., Beal, L. M., Cunningham, S. A., Kanzow, T., Bryden, H. L., Hitschi, J. J. M., Marotzke, J., Meinen, C. S., Shaw, B., Curry, R., 2011. Continuous, array-based estimates of Atlantic Ocean heat transport at 26.5°N. J. Climate 24, 2429–2449.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Amer. Meteor. Soc. 77, 437–471.

Kara, A. B., Hurlburt, H. E., Wallcraft, A. J., 2005. Stability-dependent exchange coefficients for air-sea fluxes. J. Atmospheric and Oceanic Technology 22, 1080–1094.

Karspeck, A. R., Stammer, D., Köhl, A., Danabasoglu, G., Balmaseda, M., Smith, D. M., Fujii, Y., Zhang, S., Giese, B., Tsujino, H., Rosati, A., 2015. Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products. Clim. Dyn.(in press).

Kushnir, Y., 1994. Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. J. Climate 7, 141–157.

Kwon, Y.-O., Frankignoul, C., 2012. Stochastically-driven multidecadal variability of the Atlantic meridional overturning circulation in CCSM3. Clim. Dyn. 38, 859–876.

Kwon, Y.-O., Frankignoul, C., 2014. Mechanisms of multidecadal Atlantic meridional overturning circulation variability diagnosed in depth versus density space. J. Climate 27, 9359–9376.

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

Large, W. G., Yeager, S., 2004. Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies. NCAR Tech. Note NCAR/TN-460+STR.

Large, W. G., Yeager, S. G., 2009. The global climatology of an interannually varying air-sea flux data set. Clim. Dyn. 33, 341–364.

Lavender, K. L., Davis, R. E., Owens, W. B., 2002. Observations of open-ocean deep convection in the Labrador Sea from subsurface floats. J. Phys. Oceanogr. 32, 511–526.

Le Traon, P.-Y., Nadal, F., Ducet, N., 1998. An improved mapping method of multisatellite altimeter data. J. Atmos. Oceanic Technol. 15, 522–534.

Levitus, S., Boyer, T., Concright, M., Johnson, D., O'Brien, T., Antonov, J., Stephens, C., Garfield, R., 1998. World Ocean Database 1998, volume I: Introduction.

Liu, Z., 2012. Dynamics of interdecadal climate variability: A historical perspective. J. Climate 25, 1963–1994.

Lohmann, K., Drange, H., Bentsen, M., 2009a. A possible mechanism for the strong weakening of the North Atlantic subpolar gyre in the mid-1990s. Geophys. Res. Lett. 36, L15602.

Lohmann, K., Drange, H., Bentsen, M., 2009b. Response of the North Atlantic subpolar gyre to persistent North Atlantic oscillation like forcing. Clim. Dyn. 32, 273–285.

Mellor, L. G., Kantha, L., 1989. An ice - ocean coupled model. J. Geophys. Res. 94, 10937–10954.

Msadek, R., Dixon, K. W., Delworth, T. L., Hurlin, W., 2010. Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. Geophys. Res. Lett. 37, L19608.

Msadek, R., Johns, W. E., Yeager, S. G., Danabasoglu, G., Delworth, T. L., Rosati, A., 2013. The Atlantic meridional heat transport at 26.5°N and its relationship with the MOC in the RAPID array and the GFDL and NCAR coupled models. J. Climate 26, 4335–4356.

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., Tschuck, P., 2004. Estimating the decadal predictability of a coupled AOGCM. J. Climate 17, 4463–4472.

Pohlmann, H., Smith, D. M., Balmaseda, M. A., Keenlyside, N. S., Masina, S., Matei, D., Müller, W. A., Rogel, P., 2013. Predictability of the mid-latitude Atlantic meridional overturning circulation in a multi-model system. Clim. Dyn. 41, 775–785.

Robson, J., Sutton, R., Lohmann, K., Smith, D., Palmer, M. D., 2012. Causes of the rapid warming of the North Atlantic Ocean in the 1990s. J. Climate 25, 4116–4134.

Schlesinger, M. E., Ramankutty, N., 1994. An oscillation in the global climate system of period 65–70 years. Nature 367, 723–726.

Spiegel, E. A., Veronis, G., 1960. On the Boussinesq approximation for a compressible fluid. Astrophys. J. 131, 442–447.

Srokosz, M., Baringer, M., Bryden, H., Cunningham, S., Delworth, T., Lozier, S., Marotzke, J., Sutton, R., 2012. Past, present and future change in the Atlantic meridional overturning circulation. BAMS 93, 1663–1676.

Steele, M., Morley, R., Ermold, W., 2001. PHC: A global ocean hydrography with a high quality Arctic Ocean. J. Climate 14, 2079–2087.

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

Sun, S., Bleck, R., Rooth, C., Dukowicz, J., Chassignet, E., Killworth, P., 1999. Inclusion of thermobaricity in isopycnic-coordinate ocean models. J. Phys. Oceanog. 29, 2719–2729.

Sutton, R. W., Hodson, D. L. R., 2005. Atlantic Ocean forcing of North American and European summer climate. Science 309, 115–118.

Teng, H., Branstator, G., Meehl, G. A., 2011. Predictability of the Atlantic overturning circulation and associated surface patterns in two CCSM3 climate change ensemble experiments. J. Climate 24, 6054–6076.

Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 reanalysis. Q. J. R. Meteorol. Soc. 131, 2961–3012.

Wang, Q., Ilicak, M., Gerdes, R., Drange, H., Aksenov, Y., Bailey, D. A., Bentsen, M., Biastoch, A., Bozec, A., Böning, C., Cassou, C., Chassignet, E., Coward, A. C., Curry, B., Danabasoglu, G., Danilov, S., Fernandez, E., Fogli, P. G., Fujii, Y., Griffies, S. M., Iovino, D., Jahn, A., Jung, T., Large, W. G., Lee, C., Lique, C., Lu, J., Masina, S., Nurser, A. J. G., Rabe, B., Roth, C., Salas y Mélia, D., Samuels, B. L., Spence, P., Tsujino, H., Valcke, S., Voldoire, A., Wang, X., Yeager, S. G., 2015. An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations: Sea ice and freshwater. Ocean Modelling(submitted).

Yashayaev, I., 2007. Hydrographic changes in the Labrador Sea, 1960-2005. Prog. Oceanogr. 73, 242–276.

Yeager, S., 2015. Topographic coupling of the Atlantic overturning and gyre circulations. J. Phys. Oceanogr. 45, 1258–1284.

Yeager, S., Danabasoglu, G., 2014. The origins of late-twentieth-century variations in the large-scale North Atlantic circulation. J. Climate 27, 3222–3247.

Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., Teng, H., 2012. A decadal prediction case study: Late Twentieth-century North Atlantic Ocean heat content. J. Climate 25, 5173–5189.

Yeager, S. G., Danabasoglu, G., 2012. Sensitivity of Atlantic meridional overturning circulation variability to parameterized Nordic Sea overflows in CCSM4. J. Climate 25, 2077–2103.

Zhang, R., 2010. Latitudinal dependence of Atlantic meridional overturning circulation AMOC variations. Geophys. Res. Lett. 37, L16703.

List of Figures

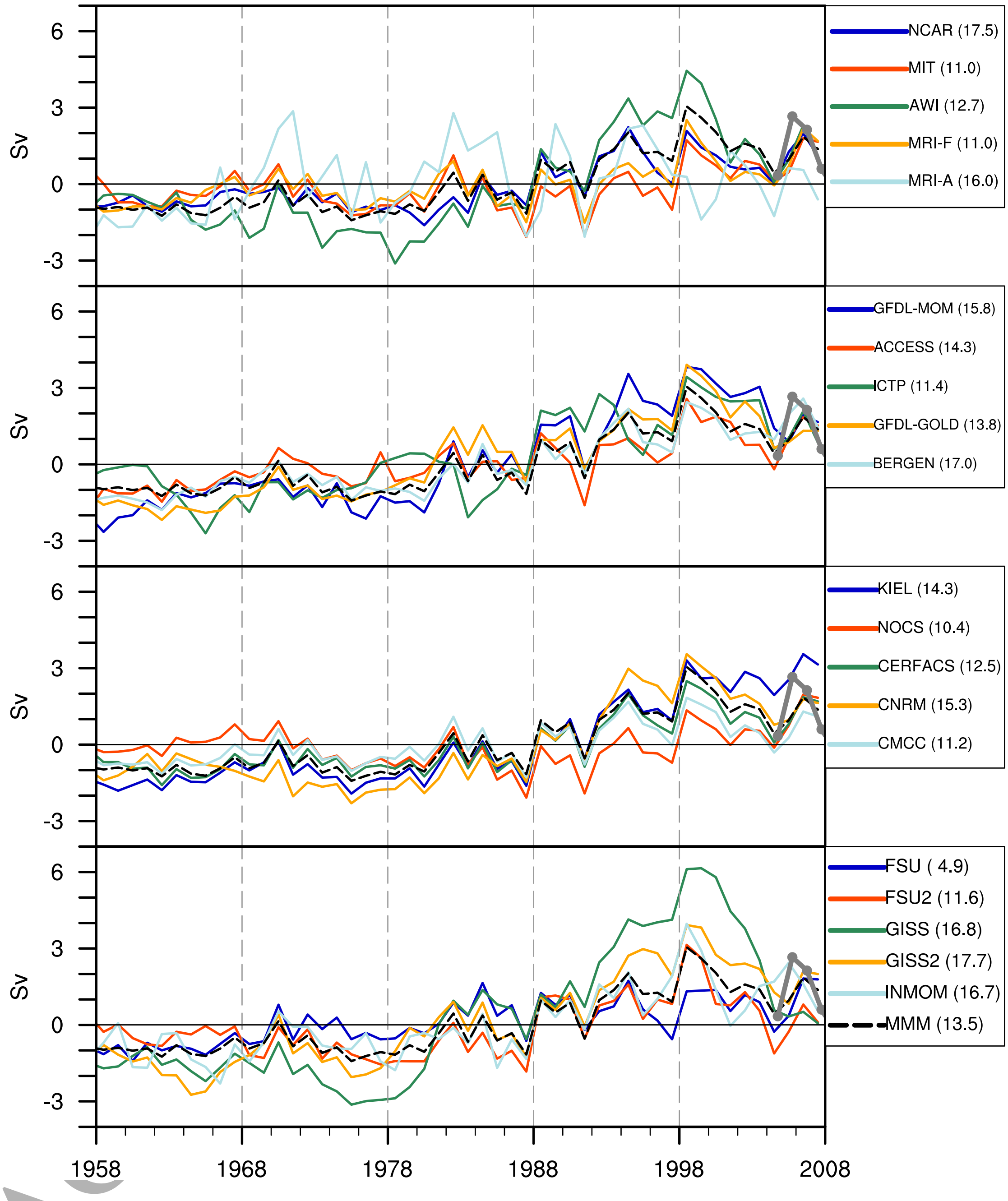

Figure 1: AMOC annual-mean maximum transport time series at 26.5°N for the 1958−2007 period from the last cycle of simulations. The time series are anomalies from the respective 50-year means given for each model in parentheses in the labels. The thick gray lines represent the annual-mean RAPID data from Cunningham et al. (2007). The 4-year mean for the RAPID data is 18.6 Sv. MMM time series are included in all panels as the dashed black lines. MMM does not include MRI-A.

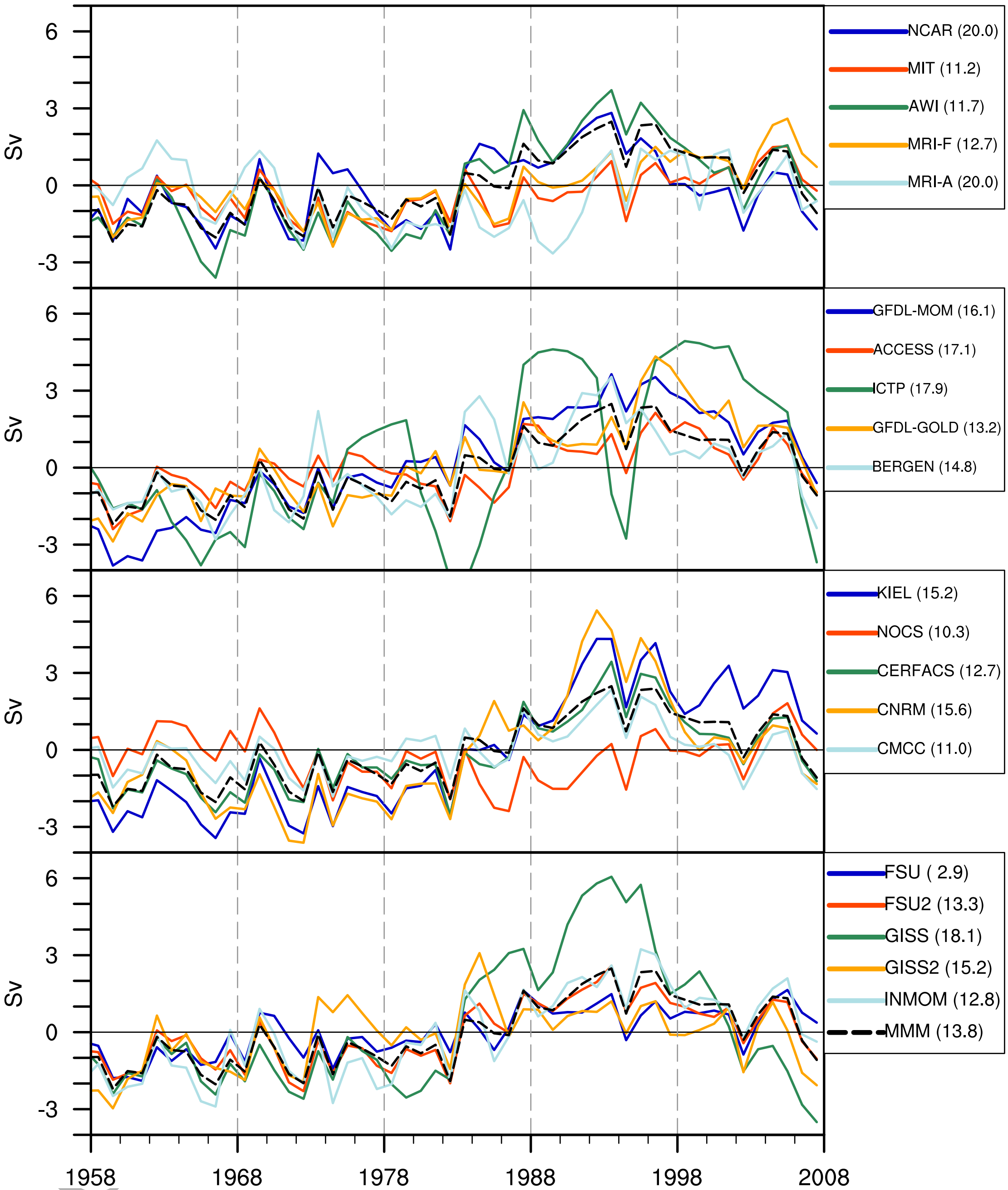

Figure 2: AMOC annual-mean maximum transport time series at 45°N for the 1958−2007 period from the last cycle of simulations. The time series are anomalies from the respective 50-year means given for each model in parentheses in the labels. MMM time series are included in all panels as the dashed black lines. MMM does not include MRI-A.

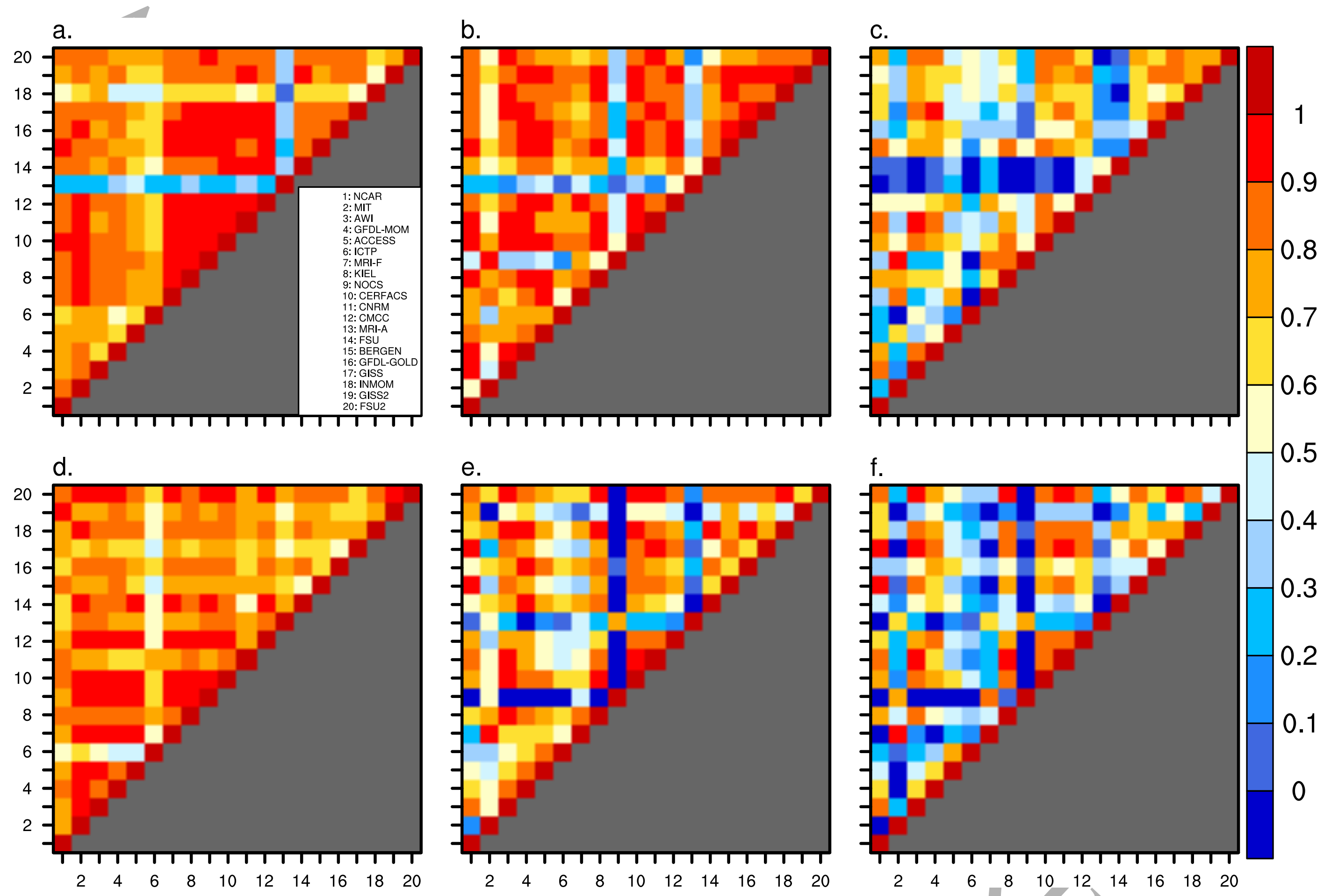

Figure 3: Model − model correlations for the AMOC maximum transport time series at (a-c) 26.5°N and (d-f) 45°N. (left column) High-pass filtered; (middle column) Low-pass filtered with trend; and (right column) Low-pass filtered and detrended. A 7-year cutoff is used for the filters. AMOC in depth and latitude space is used for the 1958−2007 period. All negative correlations are included in the darkest blue color.

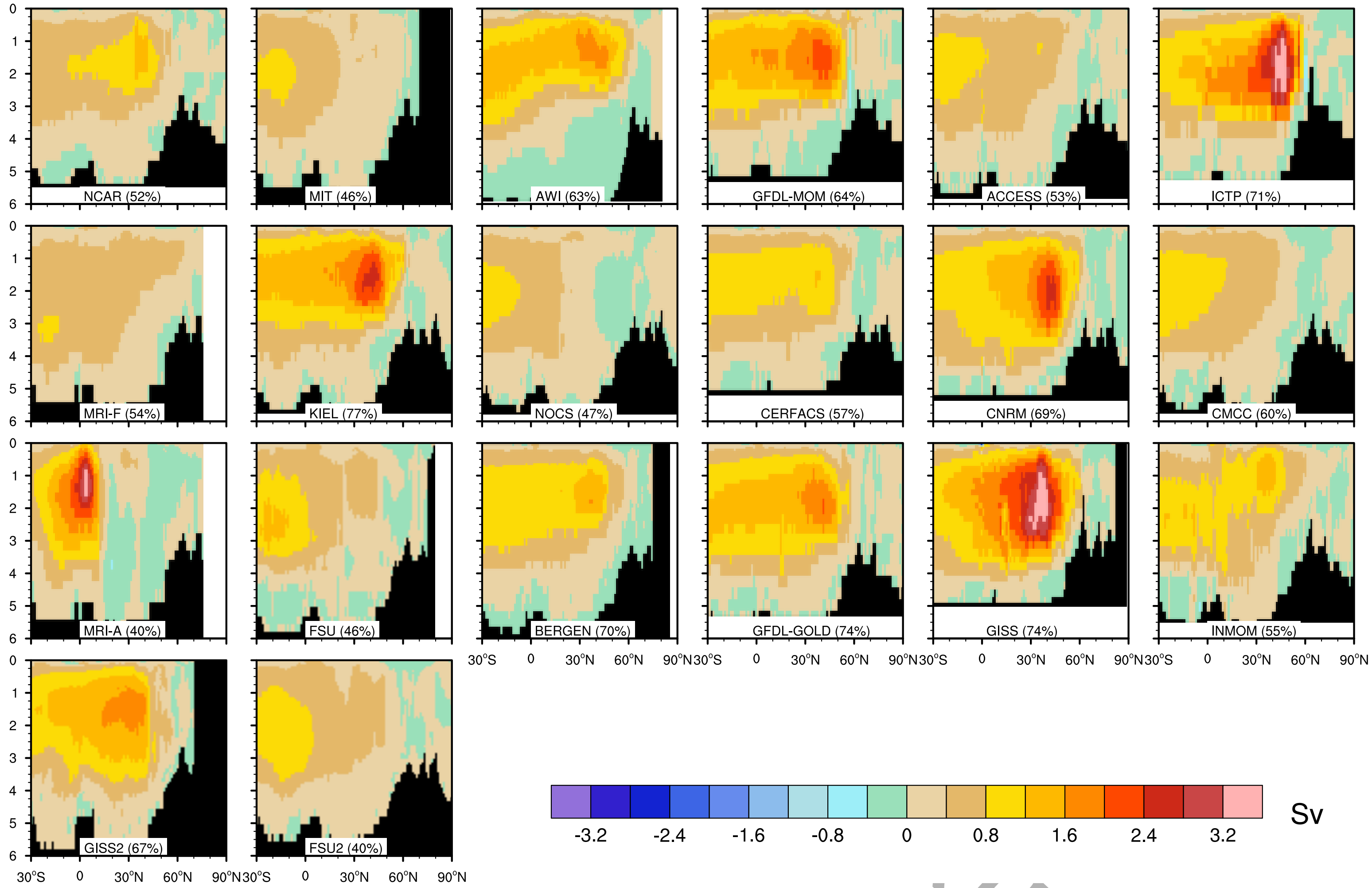

Figure 4: AMOC EOF1 spatial distributions in depth (km) and latitude space for the 1958–2007 period. The associated variances accounted by EOF1 as a percentage of the total AMOC variance are also given. The positive and negative contours indicate clockwise and counter-clockwise circulations, respectively. In MIT, AWI, MRI-F, MRI-A, FSU, BERGEN, GISS, GISS2, and FSU2, the AMOC distributions do not include the high latitude North Atlantic and / or Arctic Oceans, and hence are masked. No detrending is applied.

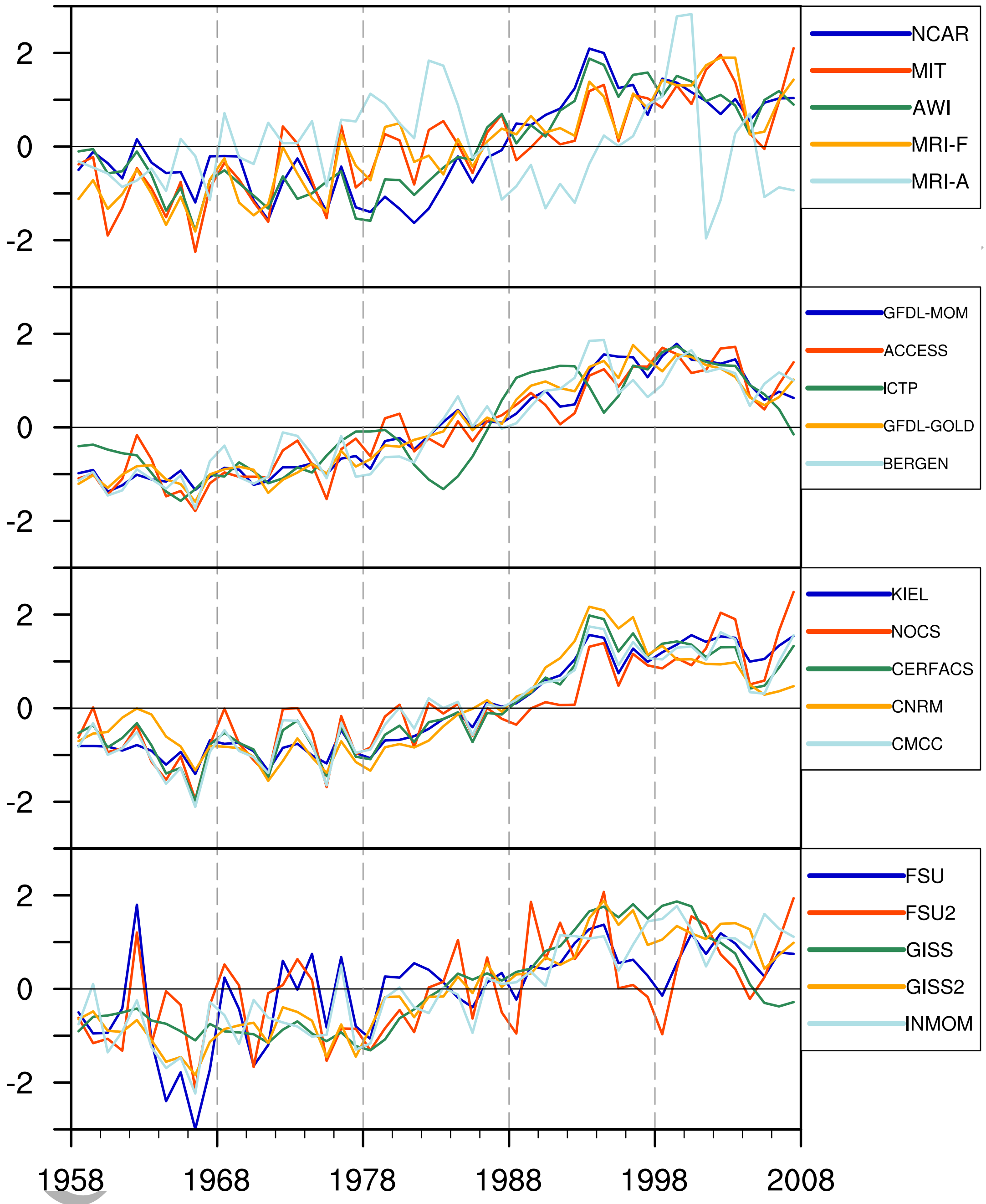

Figure 5: AMOC PC1 time series corresponding to Fig. 4. The time series are normalized to have unit variance, so that the EOF spatial pattern magnitudes correspond to one standard deviation changes in the time series.

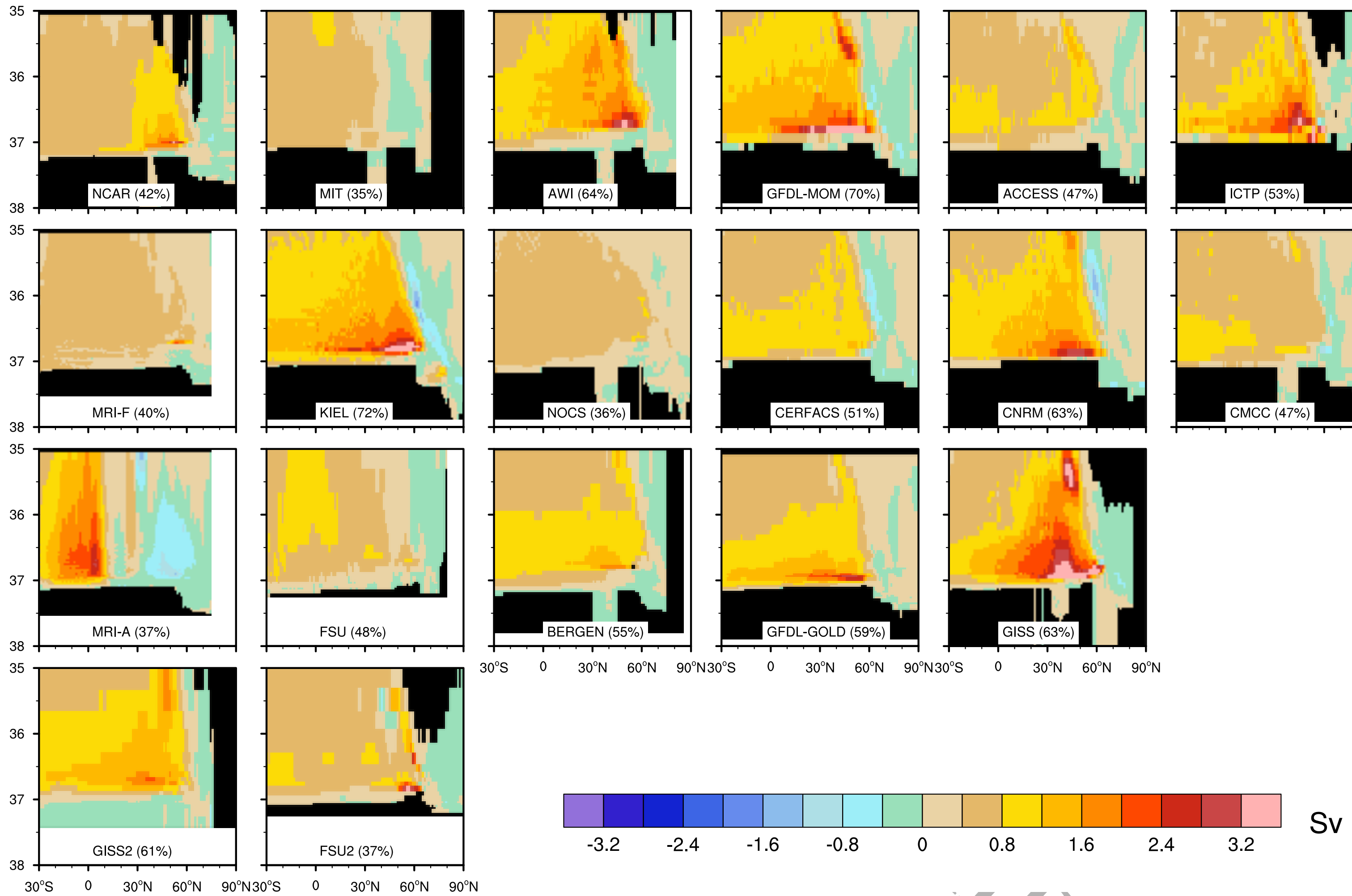

Figure 6: AMOC EOF1 spatial distributions in σ_2 (kg m^{-3}) and latitude space for the 1958–2007 period. The associated variances accounted by EOF1 as a percentage of the total AMOC variance are also given. The positive and negative contours indicate clockwise and counter-clockwise circulations, respectively. INMOM distribution is not available. No detrending is applied.

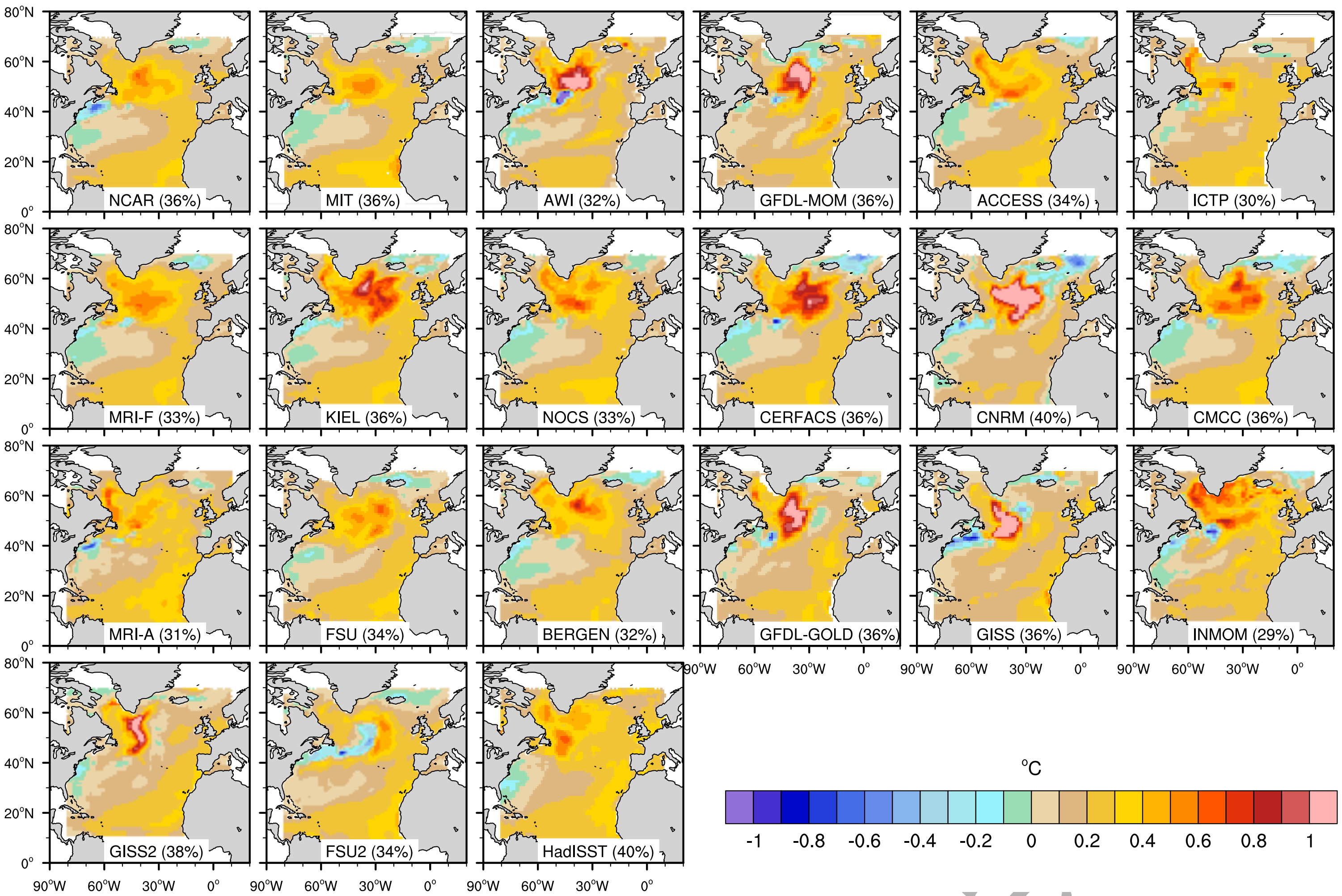

Figure 7: SST EOF1 spatial distributions for the 1958−2007 period for the North Atlantic. The associated variances accounted by EOF1 as a percentage of the total SST variance are also given. The panel to the left of the color bar shows SST EOF1 calculated from the HadISST dataset. No detrending is applied.

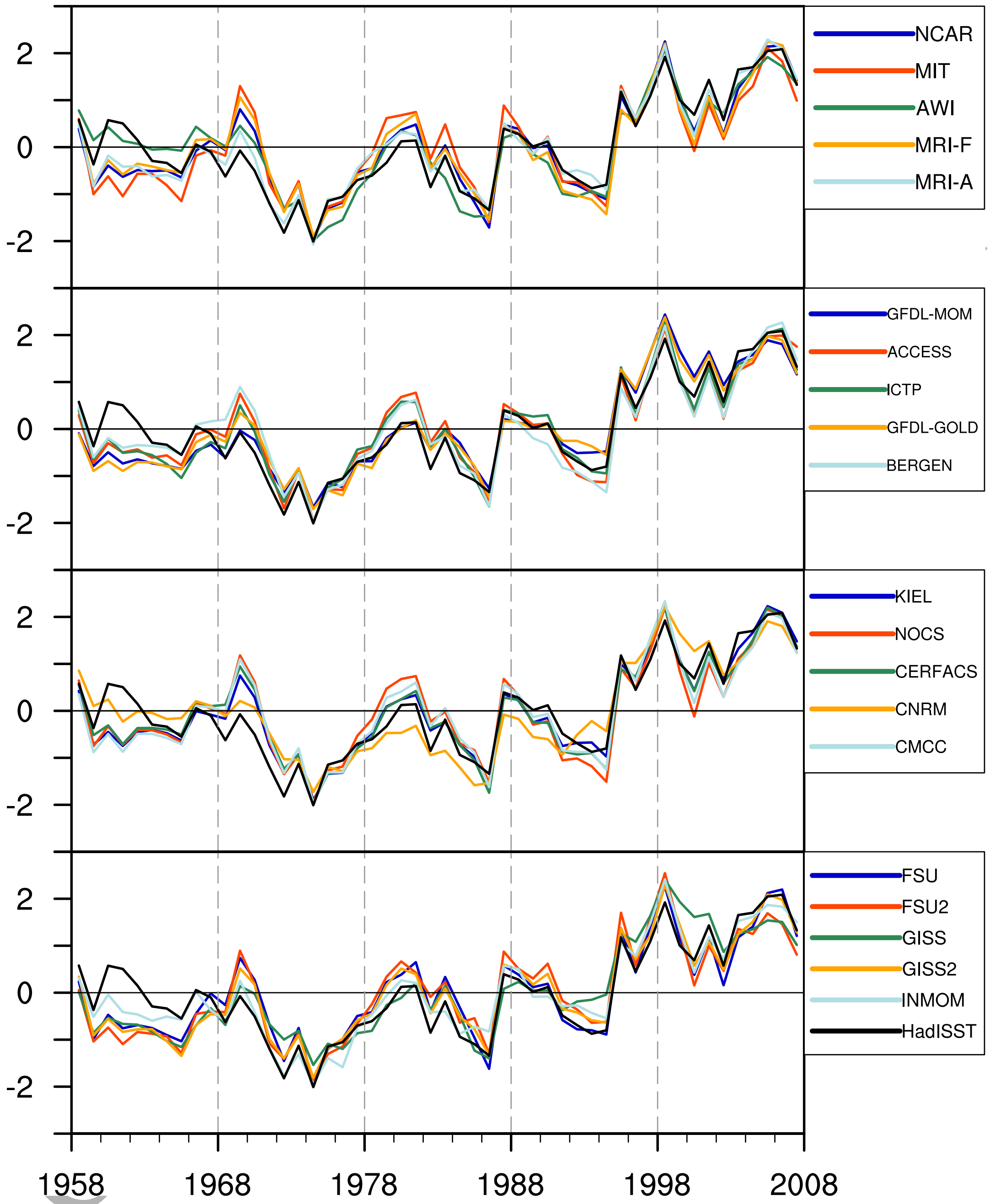

Figure 8: SST PC1 time series corresponding to Fig. 7. The time series are normalized to have unit variance, so that the EOF spatial pattern magnitudes correspond to one standard deviation changes in the time series. The time series from the HadISST dataset are included in all panels as the black lines.

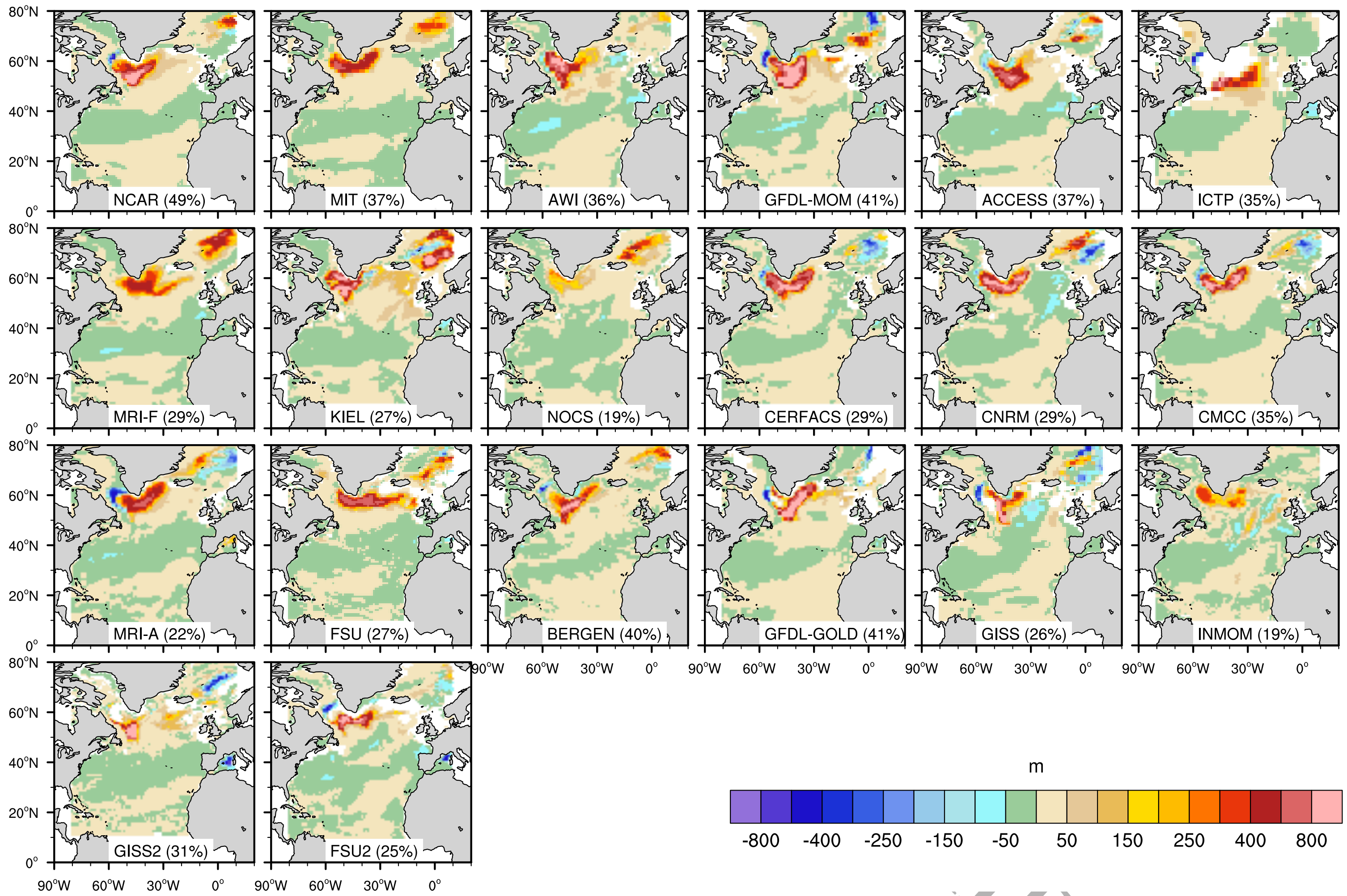

Figure 9: March-mean MLD EOF1 spatial distributions for the 1958−2007 period for the North Atlantic. The associated variances accounted by EOF1 as a percentage of the total MLD variance are also given. MLD is based on a $\Delta\rho = 0.125$ kg m^{-3} criterion. No detrending is applied. The interior white areas (i.e., excluding west of 80°W and east of 10°E) indicate regions of no variability as the time-mean MLDs reach the ocean bottom in some models.

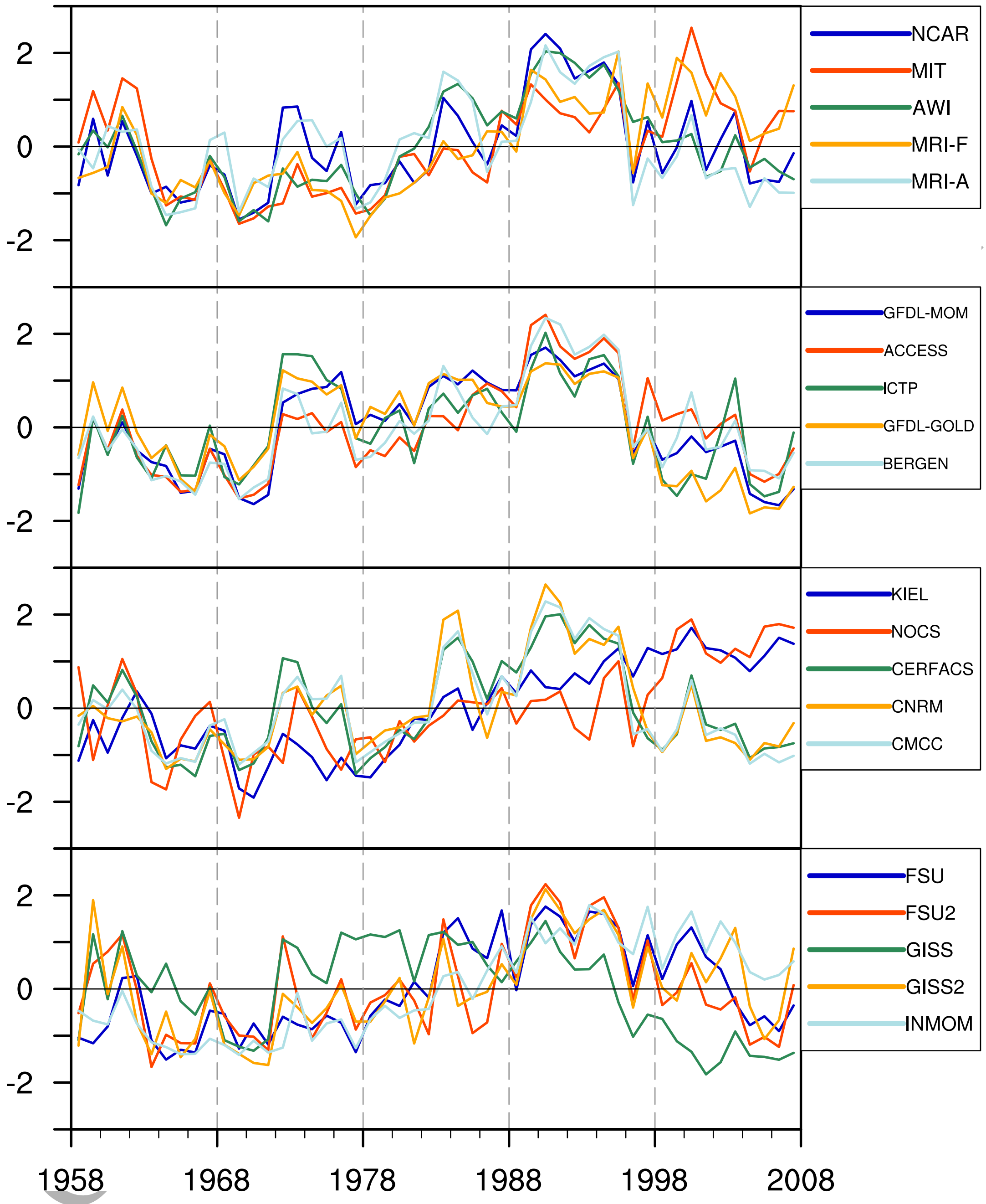

Figure 10: March-mean MLD PC1 time series corresponding to Fig. 9. The time series are normalized to have unit variance, so that the EOF spatial pattern magnitudes correspond to one standard deviation changes in the time series.

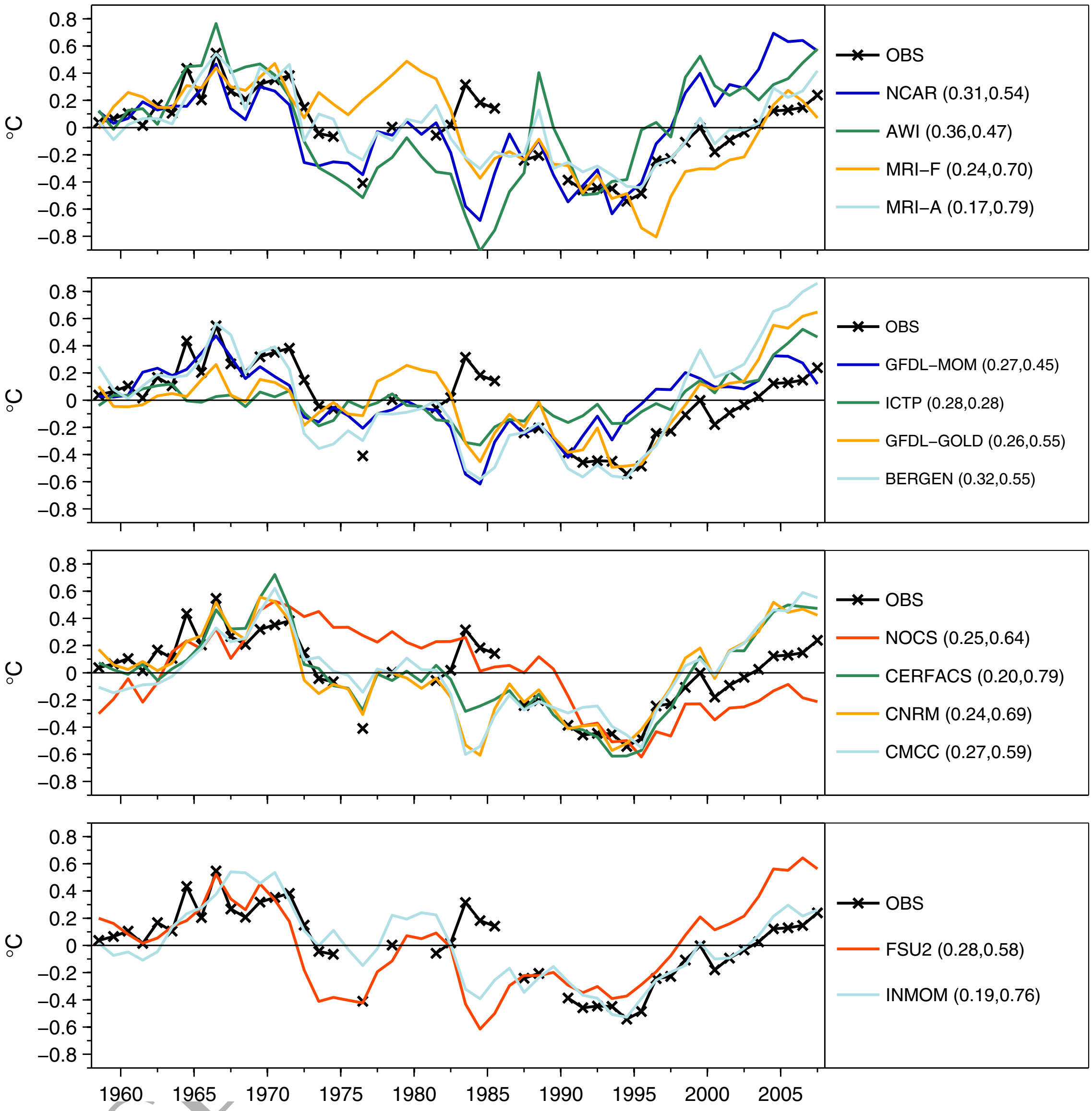

Figure 11: Time series of potential temperature anomalies averaged over the 150−1000 m depth range and within a central Labrador Sea region bounded by 49° − 56°W and 56° − 61°N. The anomalies are with respect to the 1958−2007 period. The black lines show the observational data from Yashayaev (2007) with data missing for some years. May-mean output from the models is used to roughly match the mostly Spring-time observations. For each model, the first number in parentheses gives the root-mean-square model − observations difference of their time series while the second number is the correlation coefficient between the model and observational time series. Data from ACCESS, FSU, GISS, GISS2, KIEL, and MIT are not available.

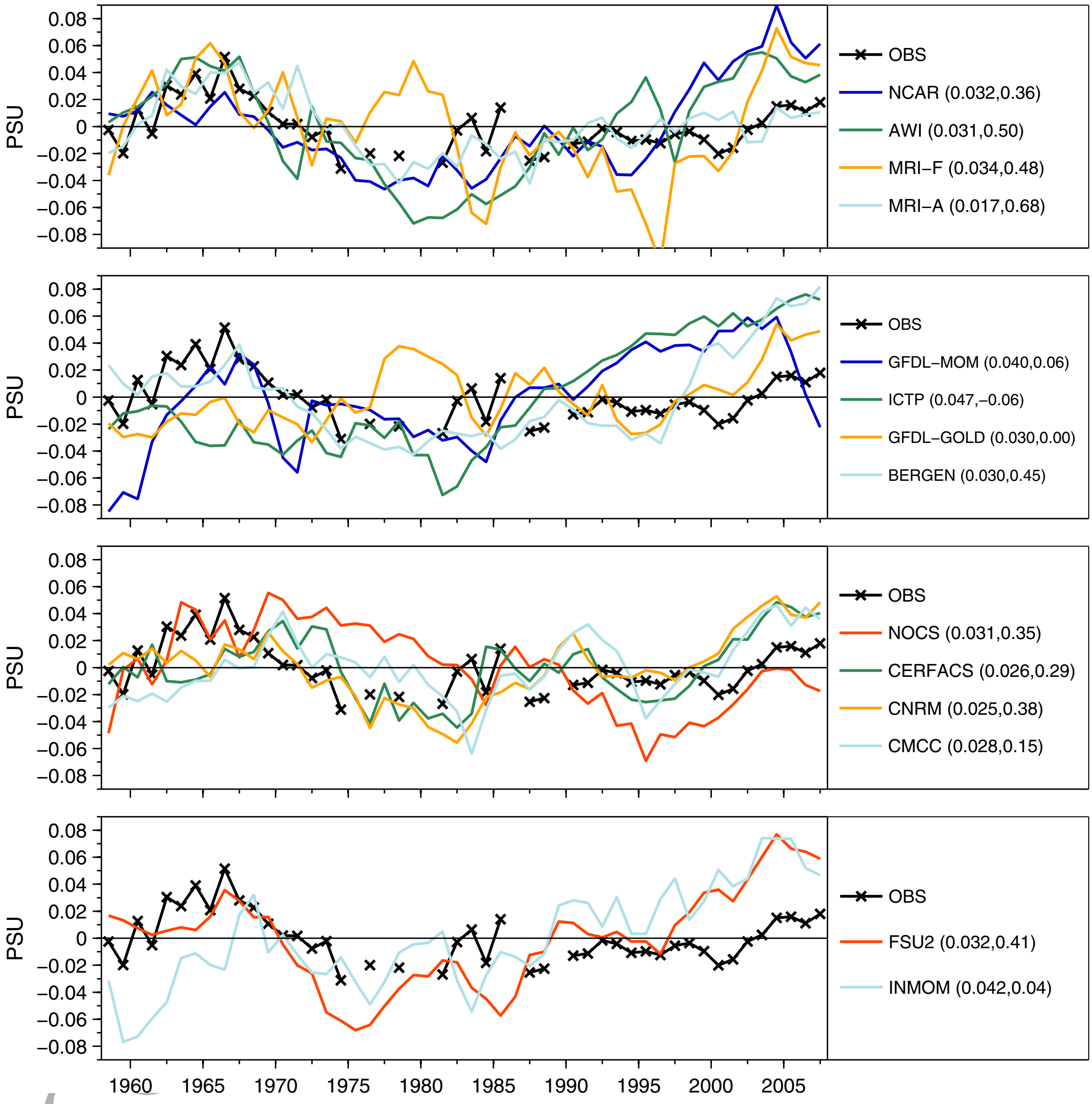

Figure 12: Same as in Fig. 11, but for salinity anomalies.

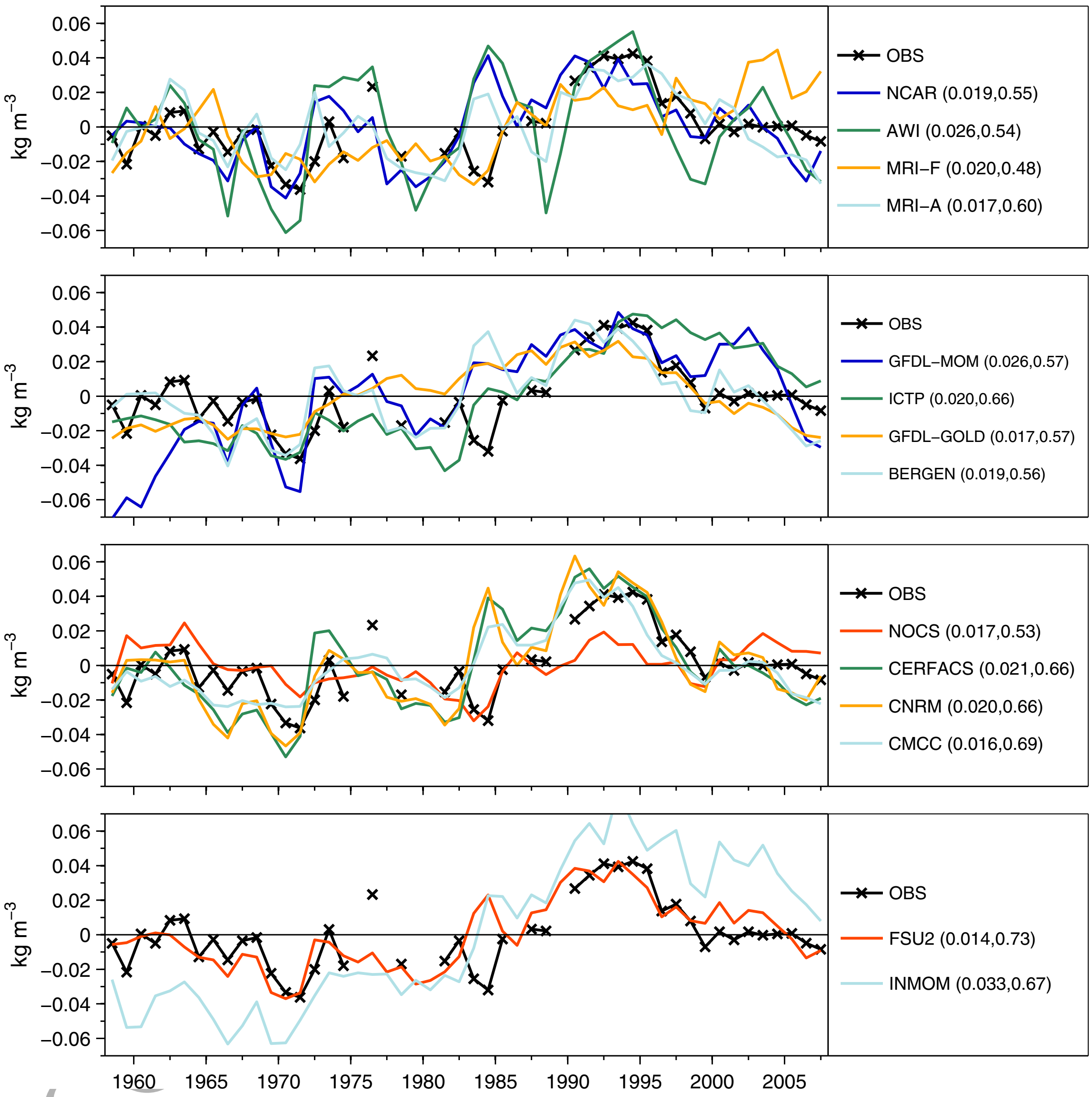

Figure 13: Same as in Fig. 11, but for density anomalies based on σ_0.

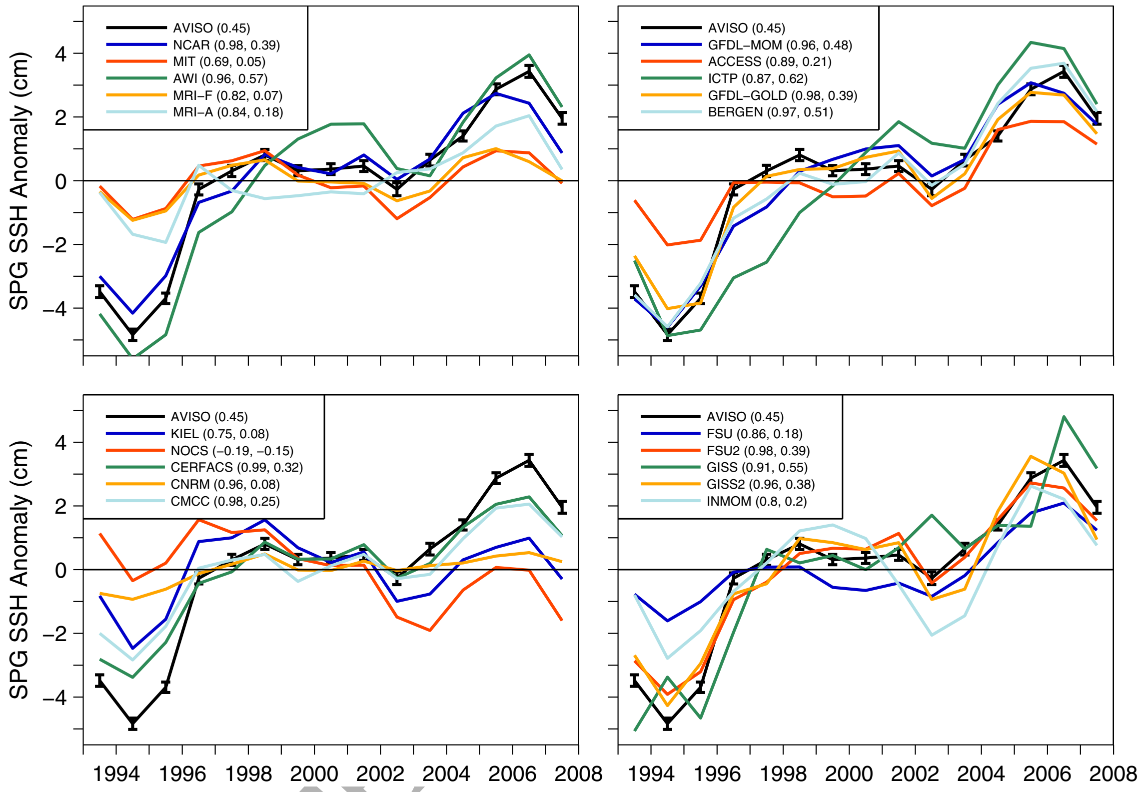

Figure 14: Time series of SPG SSH anomalies with respect to the 1993−2007 mean. SSH time series represent averages for the SPG region defined as the area between $15° - 60°$W and $48° - 65°$N. The SSH anomaly time series from AVISO dataset are also shown in each panel. The AVISO time series include the ranges of the spatially- and annually-averaged standard errors based on the monthly-mean data. The first number in parentheses for each model gives the correlation coefficient between the AVISO and that model's SSH time series. The second number in parentheses and the number for AVISO show the linear trend for the 1993−2007 period in cm yr^{-1}.

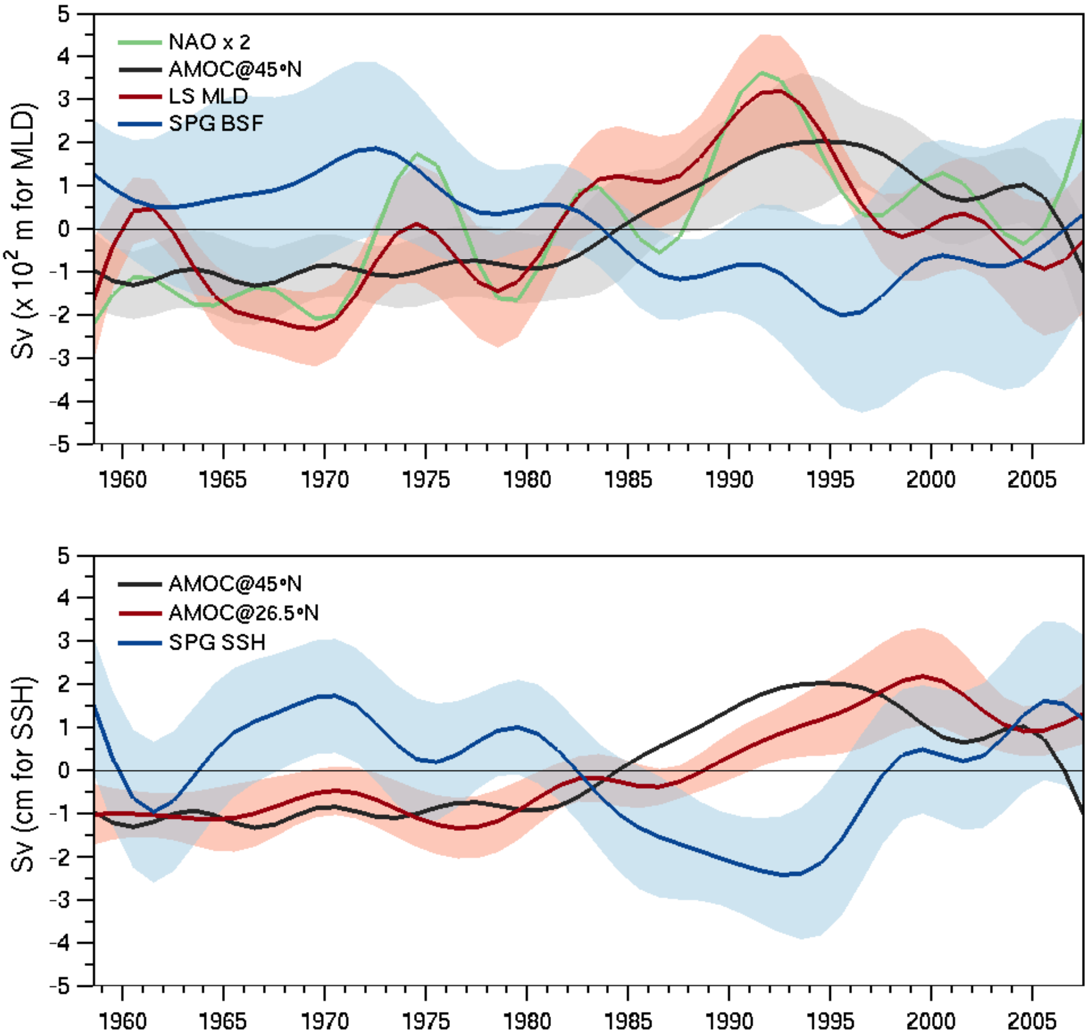

Figure 15: Low-pass filtered, MMM time series of (top) AMOC maximum transport at 45°N, March-mean MLD, and SPG BSF; and (bottom) AMOC maximum transport at 45°N (same as in the top panel), AMOC maximum transport at 26.5°N, and SPG SSH. The top panel also includes low-pass filtered NAO time series whose amplitude is multiplied by a factor of two for clarity. MLD is calculated as an average for the LS − Irminger Sea region defined as the area between $15° - 60°$W and $48° - 60°$N. The SPG BSF and SSH represent averages for the SPG region defined by $15° - 60°$W and $48° - 65°$N. We note that negative SPG BSF and SSH anomalies indicate strengthening of the cyclonic SPG circulation. All time series are anomalies with respect to the 1958−2007 period. A 7-year cutoff is used for the low-pass filter. The respective colored shadings denote one standard deviation spread of the models' time series from those of the respective MMM. The spread for the AMOC transport at 45°N is not repeated in the bottom panel for clarity. MMM does not include MRI-A. Units are Sv for AMOC and BSF; $\times 100$ m for MLD; and cm for SSH.

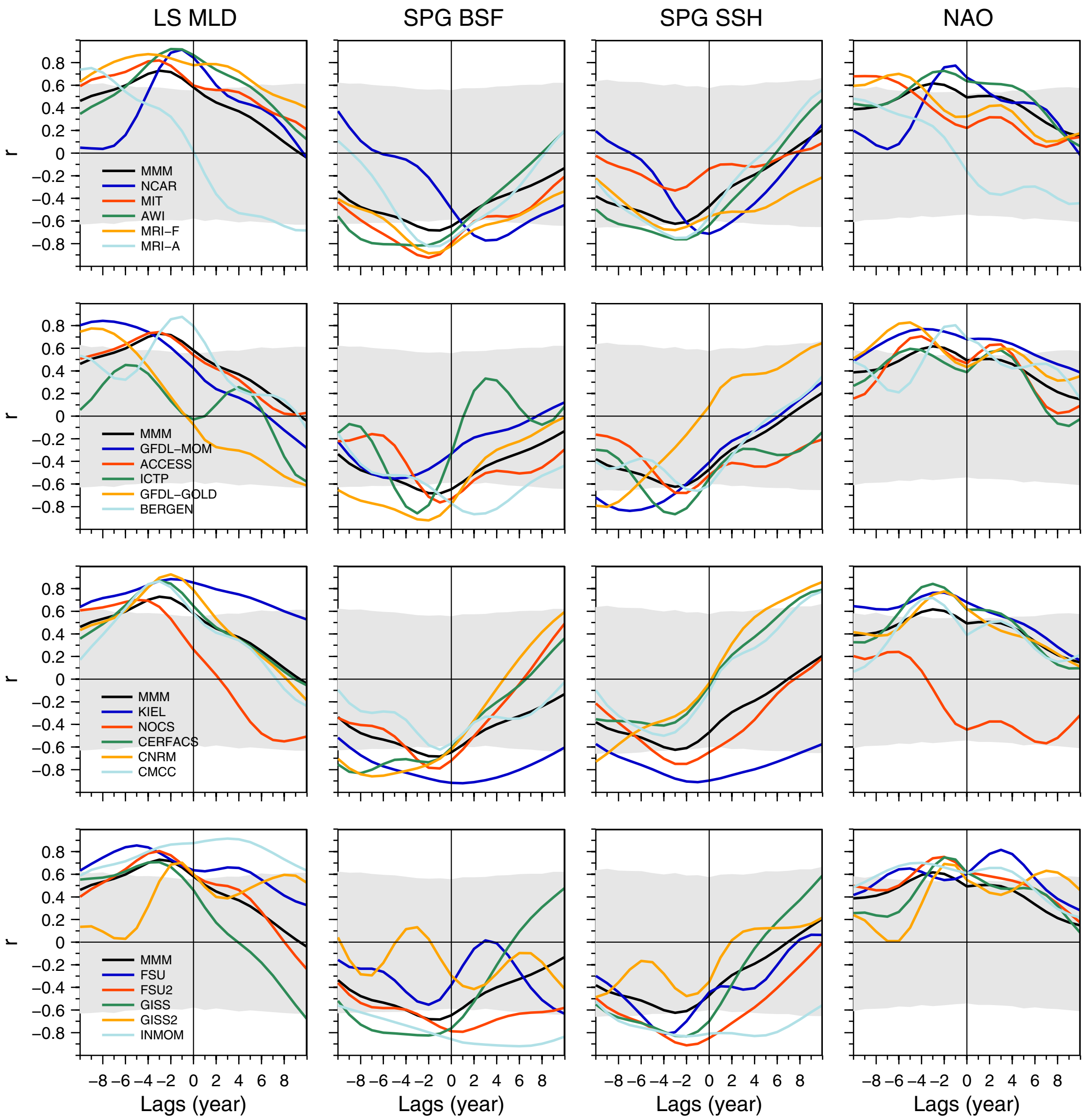

Figure 16: Low-pass filtered AMOC maximum transport at 45°N time series correlations with (first column) March-mean MLD, (second column) SPG BSF, (third column) SPG SSH, and (fourth column) NAO. The black lines in each panel show the MMM correlation functions evaluated as the mean of the individual model correlations. MMM does not include MRI-A. The correlations outside the shaded regions have confidence levels greater than 95% (see section 2 for calculation of confidence levels). Anomalies are with respect to the 1958−2007 period. A 7-year cutoff is used for the low-pass filter. AMOC index leads for positive lags.

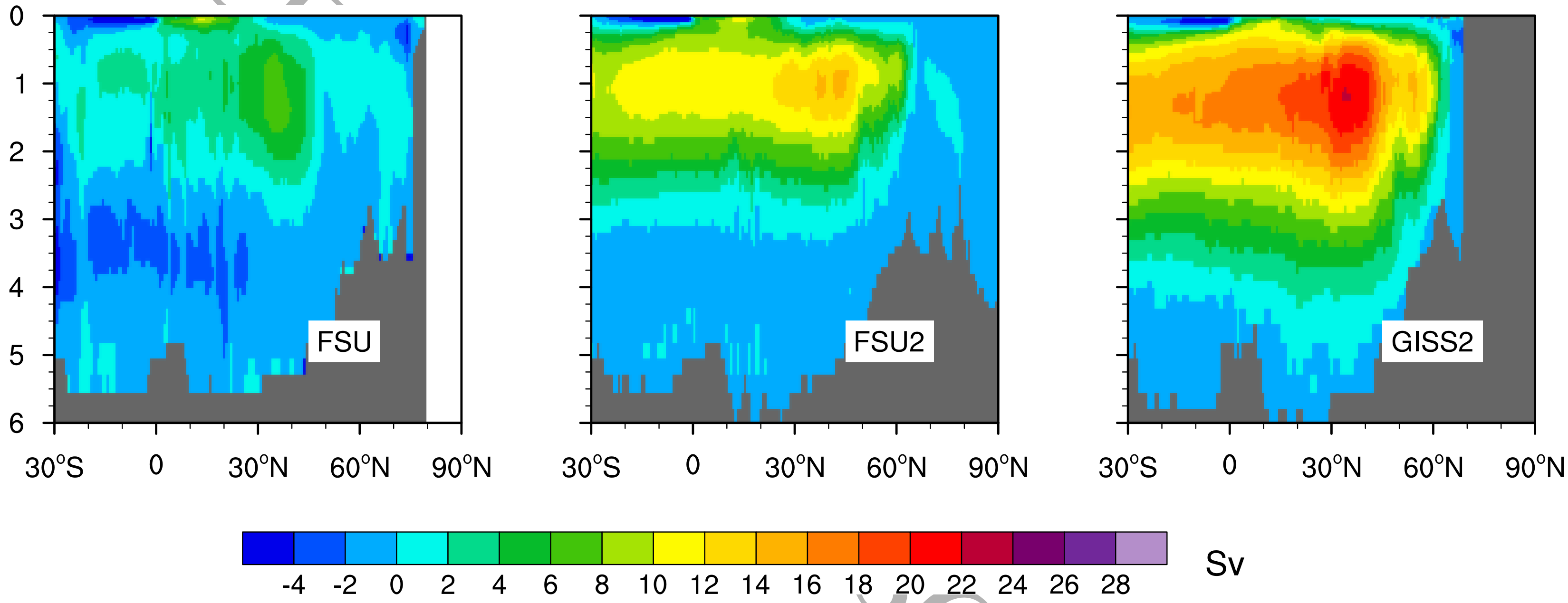

Figure 17: Years 1988−2007 mean AMOC plotted in depth (km) and latitude space from FSU, FSU2, and GISS2. The positive and negative contours indicate clockwise and counter-clockwise circulations, respectively.

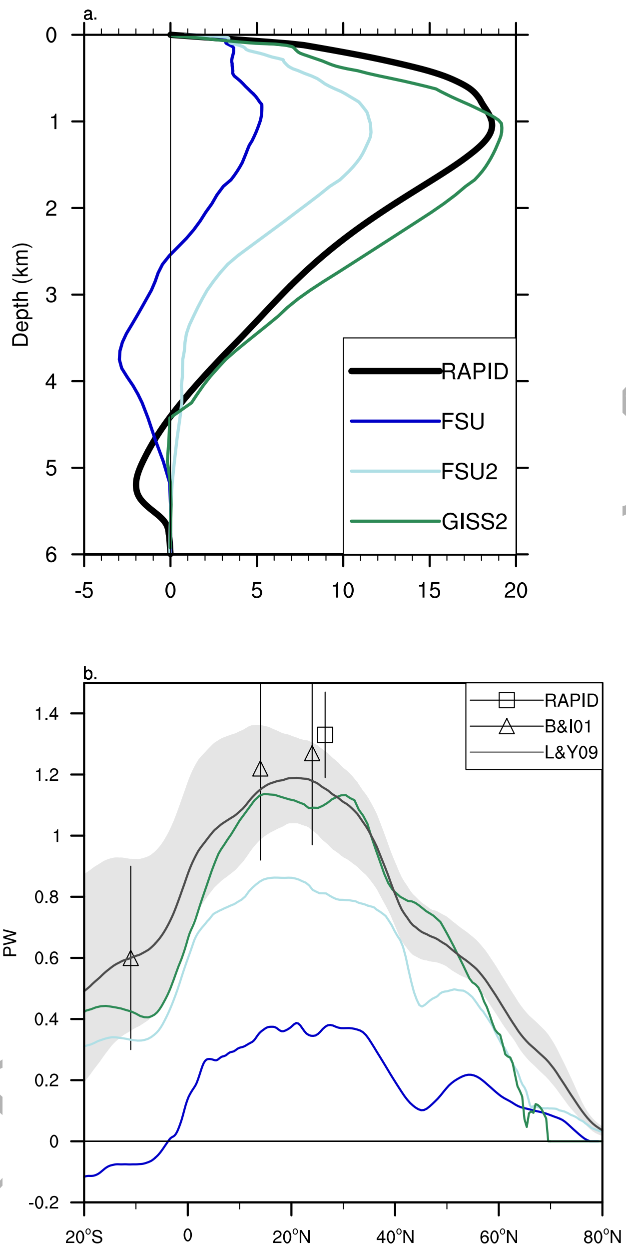

Figure 18: (a) Years 2004−2007 mean AMOC depth profiles at 26.5°N from FSU, FSU2, and GISS2 in comparison with the 4-year mean (April 2004 − March 2008) RAPID data; (b) Years 1988−2007 mean meridional heat transports for the Atlantic Ocean from the three models. In (b), the black line denoted by L&Y09 represents implied time-mean transport calculated by Large and Yeager (2009) with shading showing the implied transport range in individual years for the 1984−2006 period. Direct estimates with their uncertainty ranges from the RAPID data (square; Johns et al., 2011) and from Bryden and Imawaki (2001) (triangle; B&I01) are also shown.

Table 1: Summary of the ocean and sea-ice models in alphabetical order according to the participating group name (first column). The table includes the name of the combined ocean – sea-ice configuration (if any); the ocean model name and its version; the sea-ice model name and its version; vertical coordinate and number of layers / levels in parentheses; orientation of the horizontal grid with respect to the North Pole / Arctic; the number of horizontal grid cells (longitude × latitude); and the horizontal resolution (longitude × latitude). In MRI-A and MRI-F, the vertical levels shallower than 32 m follow the surface topography as in sigma-coordinate models. In FESOM, the total number of surface nodes is given under horizontal grid, because it has an unstructured grid. FESIM is imbedded in FESOM. H79 is Hibler (1979) and MK89 is Mellor and Kantha (1989).

Group	Configuration	Ocean model	Sea-ice model	Vertical	Orientation	Horiz. grid	Horiz. res.
ACCESS	ACCESS-OM	MOM 4p1	CICE 4	z^* (50)	tripolar	360 × 300	nominal 1°
AWI		FESOM	FESIM	z (46)	displaced	126000	nominal 1°
BERGEN	NorESM-O	MICOM	CICE 4	σ_2 (51+2)	tripolar	360 × 384	nominal 1°
CERFACS	ORCA1	NEMO 3.2	LIM 2	z (42)	tripolar	360 × 290	nominal 1°
CMCC	ORCA1	NEMO 3.3	CICE 4	z (46)	tripolar	360 × 290	nominal 1°
CNRM	ORCA1	NEMO 3.2	Gelato 5	z (42)	tripolar	360 × 290	nominal 1°
FSU		HYCOM 2.2.21	CSIM 5	hybrid (32)	displaced	320 × 384	nominal 1°
FSU2		HYCOM 2.2.74	CICE 4	hybrid (32)	tripolar	500 × 382	nominal 0.72°
GFDL-GOLD	ESM2G-ocean-ice	GOLD	SIS	σ_2 (59+4)	tripolar	360 × 210	nominal 1°
GFDL-MOM	ESM2M-ocean-ice	MOM 4p1	SIS	z^* (50)	tripolar	360 × 200	nominal 1°
GISS		GISS Model E2-R		mass (32)	regular	288 × 180	1.25° × 1°
GISS2		HYCOM 0.9		hybrid (26)	regular	360 × 387	nominal 1°
ICTP		MOM 4p1	SIS	z^* (30)	tripolar	180 × 96	nominal 2°
INMOM		INMOM		sigma (40)	displaced	360 × 340	1° × 0.5°
KIEL	ORCA05	NEMO 3.1.1	LIM 2	z (46)	tripolar	722 × 511	nominal 0.5°
MIT		MITgcm	H79	z (50)	quadripolar	360 × 292	nominal 1°
MRI-A	(data assimilation)	MOVE/MRI.COM 3	MK89; CICE	z (50)	tripolar	360 × 364	1° × 0.5°
MRI-F		MRI.COM 3	MK89; CICE	z (50)	tripolar	360 × 364	1° × 0.5°
NCAR		POP 2	CICE 4	z (60)	displaced	320 × 384	nominal 1°
NOCS	ORCA1	NEMO 3.4	LIM 2	z (75)	tripolar	360 × 290	nominal 1°

Table 2: Summary of AMOC maximum transports and linear trends at 26.5°N (columns 3-5) and 45°N (columns 6-8). Models are listed in alphabetical order according to the participating group name (first column). The second column shows whether the AMOC EOF1 maximum occurs in the Northern Hemisphere (N), in the Southern Hemisphere (S), or near the equator (E). The mean transports represent 50-year means for the 1958−2007 period. The linear trends are calculated for the 1978−1998 and 1998−2007 periods for 26.5°N; and for the 1975−1995 and 1995−2007 periods for 45°N based on the annual-mean data. The trends that meet the 95% confidence level based on a two-sided Student's t-test are shown in bold. The mean transports and trends are in Sv and Sv decade^{-1}, respectively. MMM does not include MRI-A.

		26.5°N			45°N		
Group	N/S/E	Mean	1978-1998 trend	1998-2007 trend	Mean	1975-1995 trend	1995-2007 trend
ACCESS	S	14.3	**0.33**	−1.08	17.1	0.32	**−1.35**
AWI	N	12.7	**1.52**	**−3.27**	11.7	**1.37**	**−1.68**
BERGEN	N	17.0	**0.64**	−0.34	14.8	**1.01**	**−1.59**
CERFACS	N	12.5	**0.62**	−1.02	12.7	**0.91**	**−1.86**
CMCC	S	11.2	**0.33**	−0.95	11.0	**0.48**	**−1.51**
CNRM	N	15.3	**1.15**	**−2.30**	15.6	**1.75**	**−2.53**
FSU	S	4.9	0.16	−0.02	2.9	**0.37**	−0.02
FSU2	S	11.6	**0.77**	**−3.15**	13.3	**0.82**	**−1.21**
GFDL-GOLD	N	13.8	**0.62**	**−3.19**	13.2	**0.75**	**−2.67**
GFDL-MOM	N	15.8	**1.08**	**−2.85**	16.1	**0.93**	**−2.06**
GISS	N	16.8	**1.62**	**−8.13**	18.1	**2.06**	**−4.81**
GISS2	N	17.7	**0.88**	**−2.57**	15.2	0.11	**−1.25**
ICTP	N	11.4	**0.66**	**−2.63**	17.9	0.54	**−3.52**
INMOM	N	16.7	**0.82**	**−1.73**	12.8	**1.01**	**−1.52**
KIEL	N	14.3	**0.85**	0.25	15.2	**1.50**	−1.03
MIT	S	11.0	0.13	0.15	11.2	**0.33**	0.15
MRI-A	E	16.0	0.09	0.20	20.0	0.32	−1.10
MRI-F	S	11.0	0.28	−0.30	12.7	**0.48**	0.27
NCAR	N	17.5	**0.66**	−0.38	20.0	**0.88**	**−1.34**
NOCS	S	10.4	0.17	0.72	10.3	0.03	0.29
MMM		13.5	**0.70**	**−1.73**	13.8	**0.82**	**−1.54**

Table 3: Simultaneous correlation and regression relationships between the AMOC maximum transports and meridional heat transports (MHT) at 26.5°N based on the annual-mean transports for 1958−2007. Models are listed in alphabetical order according to the participating group name (first column). The second column gives the correlation coefficients. The regression coefficients and the intercept values obtained when MHT is regressed onto AMOC are listed in the third and fourth columns, respectively.

Group	Correlation	Regression (PW Sv^{-1})	Intercept (PW)
ACCESS	0.93	0.063	−0.095
AWI	0.98	0.065	0.011
BERGEN	0.94	0.055	0.032
CERFACS	0.95	0.061	0.022
CMCC	0.94	0.067	−0.094
CNRM	0.96	0.059	0.000
FSU	0.96	0.067	0.007
FSU2	0.91	0.058	0.082
GFDL-GOLD	0.96	0.064	−0.099
GFDL-MOM	0.96	0.058	−0.070
GISS	0.96	0.051	0.103
GISS2	0.95	0.047	0.177
ICTP	0.97	0.061	0.047
INMOM	0.86	0.042	−0.008
KIEL	0.97	0.056	0.053
MIT	0.92	0.063	−0.026
MRI-A	0.86	0.066	−0.068
MRI-F	0.94	0.067	−0.117
NCAR	0.95	0.062	−0.072
NOCS	0.93	0.068	−0.070