

Predicting multiyear North Atlantic Ocean variability

W. Hazeleger,^{1,5} B. Wouters,¹ G. J. van Oldenborgh,¹ S. Corti,² T. Palmer,⁶ D. Smith,³ N. Dunstone,³ J. Kröger,⁴ H. Pohlmann,^{3,4} and J.-S. von Storch⁴

Received 30 July 2012; revised 5 February 2013; accepted 6 February 2013; published 7 March 2013.

[1] We assess the skill of retrospective multiyear forecasts of North Atlantic ocean characteristics obtained with ocean-atmosphere-sea ice models that are initialized with estimates from the observed ocean state. We show that these multimodel forecasts can skilfully predict surface and subsurface ocean variability with lead times of 2 to 9 years. We focus on assessment of forecasts of major well-observed oceanic phenomena that are thought to be related to the Atlantic meridional overturning circulation (AMOC). Variability in the North Atlantic subpolar gyre, in particular that associated with the Atlantic Multidecadal Oscillation, is skilfully predicted 2–9 years ahead. The fresh water content and heat content in major convection areas such as the Labrador Sea are predictable as well, although individual events are not captured. The skill of these predictions is higher than that of uninitialized coupled model simulations and damped persistence. However, except for heat content in the subpolar gyre, differences between damped persistence and the initialized predictions are not significant. Since atmospheric variability is not predictable on multiyear time scales, initialization of the ocean and oceanic processes likely provide skill. Assessment of relationships of patterns of variability and ocean heat content and fresh water content shows differences among models indicating that model improvement can lead to further improvements of the predictions. The results imply there is scope for skilful predictions of the AMOC.

Citation: Hazeleger, W., B. Wouters, G. J. van Oldenborgh, S. Corti, T. Palmer, D. Smith, N. Dunstone, J. Kröger, H. Pohlmann, and J.-S. von Storch (2013), Predicting multiyear North Atlantic Ocean variability, *J. Geophys. Res. Oceans*, 118, 1087–1098, doi:10.1002/jgrc.20117.

1. Introduction

[2] Dynamical seasonal prediction systems that have been developed by operational meteorological institutes can produce skilful predictions months ahead, in particular in the tropics [Goddard *et al.*, 2001, van Oldenborgh *et al.*, 2005]. The notion that the climate system contains inherent memory on even longer time scales has led to explorations of the potential for decadal predictions. In particular, the oceanic heat and fresh water content and the ocean circulation associated with the Atlantic Meridional Overturning Circulation (AMOC) and the subtropical and subpolar gyres are thought to provide memory that may impact the climate system on decadal time scales. Mechanisms of AMOC variability have been identified that provide a theoretical framework for AMOC predictability [e.g., Marotzke, 1990, Weaver and Sarachik, 1991, Griffies and Tziperman, 1995, te Raa and Dijkstra, 2002]. Idealized

model studies have shown that aspects of the climate may be predicted decades ahead, with the subpolar gyre in the Atlantic as a hotspot [Griffies and Bryan, 1997, Collins and Sinha, 2003]. Also, statistical estimates of potential predictability highlight the midlatitudes and high latitudes as areas where skill in decadal predictions may be obtained [Boer, 2004].

[3] Realistic forecast experiments are feasible now because of improved coupled atmosphere-ocean-land-sea ice models and the development of ocean analyses products in which ocean observations are assimilated into ocean or coupled models. Following strategies from seasonal forecasts, first attempts for decadal forecasts have been made and evaluated [Smith *et al.*, 2007, Keenlyside *et al.*, 2008, Pohlmann *et al.*, 2009, Mochizuki *et al.*, 2010]. The skill of these predictions over the continents is limited, but several studies show skilful predictions in the North Atlantic [van Oldenborgh *et al.*, 2012, Yeager *et al.*, 2012, Chikamoto *et al.*, 2012, Smith *et al.*, 2010, Kim *et al.*, 2012], in accordance with earlier statistical estimates of potential predictability and coherent predictions of the AMOC [Pohlmann *et al.*, 2012].

[4] In this study, we use a new ensemble of model hindcasts to assess the skill of multiyear forecasts of the Atlantic Ocean in more detail. Rather than focusing on the AMOC, which is not directly observed over decadal time scales, we study the skill of forecasts of oceanographic phenomena that are observed directly in the subpolar gyre. Many of these phenomena have been associated with

¹Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands.

²European Center for Medium-Range Weather Forecasts, Reading, UK.

³United Kingdom Meteorological Office (UKMO), Exeter, UK.

⁴Max Planck Institut für Meteorologie, Hamburg, Germany.

⁵Wageningen University, Wageningen, Netherlands.

⁶Oxford University, Oxford, UK.

Corresponding author: W. Hazeleger, KNMI, PO BOX 201, 3730 AE De Bilt, Netherlands. (wilco.hazeleger@knmi.nl)

variability in the AMOC and poleward heat transport and are thus relevant to climate in the North Atlantic region. For instance, there is well-observed variability in the surface and subsurface salinity in the North Atlantic. *Yashayaev* [2007] has shown variability in Labrador Sea Water properties, with strong convection resulting in a fresher water column in the mid-1990s, a reduction afterward and short periods of resumed convection in the early-2000s. *Dickson et al.* [1988] and *Belkin et al.* [1998] report propagation of salinity anomalies along the pathway of the subpolar gyre, with two strong fresh water events: one in the 1970s and one in the 1980s. These events coincide with a shut-off of convection in the Labrador Sea. On longer time scales, a freshening trend has been observed [*Dickson et al.*, 2002, *Curry and Mauritzen*, 2005].

[5] Sea surface temperature also shows pronounced low-frequency variability. Quasi-periodic warming and cooling of the entire North Atlantic on multidecadal time scales is often referred to as the Atlantic Multidecadal Oscillation [AMO, e.g., *Delworth and Mann*, 2000]. These changes are thought to affect the temperature and precipitation of the adjacent continents [*Sutton and Hodson*, 2005, *Zhang and Delwoth*, 2006, *Knight et al.*, 2006] although this depends strongly on the detrending procedure used [*Trenberth and Shea*, 2006, *van Oldenborgh et al.*, 2009]. Other features of variability include abrupt changes in the interhemispheric temperature gradient [*Thompson et al.*, 2010]. In particular, around 1970, an abrupt shift is observed in the difference between Northern and Southern Hemisphere averaged temperatures. This change is particularly evident in the North and South Atlantic.

[6] In this paper, we address the skill of retrospective forecasts of these observed events and quasi-oscillatory patterns of variability that are thought to be associated with low-frequency AMOC variations. Previous studies already showed skill in predictions of the AMO. Here, we extend the analyses to (sub)surface ocean variability, including the AMOC. We emphasize that we address observed events in the ocean. We quantify the extent to which a multimodel decadal prediction system can, in retrospect, predict such major events. We also investigate whether these changes are related to the AMOC variability of the global climate models by showing the covariability between the AMOC and subsurface temperature and salinity. We use a range of models and initialization strategies as a proxy for model uncertainty and observational uncertainty. In the next section, we describe the model systems and their initialization strategies. In section 3, we assess the predictability of oceanographic phenomena in the North Atlantic that are well observed, and, in section 4, we discuss the results, followed by conclusions in section 5.

2. Methods

[7] Ensemble forecasting methods are commonly employed in order to represent uncertainties in estimates of the initial state of the climate system and to capture uncertainties arising from unpredictable components. Indeed, combining results from a multimodel ensemble of different prediction systems has been shown to improve skill in seasonal predictions [*Doblas-Reyes et al.*, 2005]. Here, we use four different state-of-the-art climate model systems with

different initialization methods to assess multiyear hindcasts (see Table 1). These hindcasts have been made as part of the EU Framework 7 THOR project and will be referred to as THOR in the figures.

[8] Most forecast systems used the same experimental setup proposed for the Coupled Model Intercomparison Project 5 [CMIP5, *Taylor et al.*, 2012]. This means that every 5 years, a hindcast was started on the first of November, starting from 1960 up to 2005. Only the MPI-M model system starts on the first of January. Every hindcast consists of 10 members, except for EC-Earth and European Center for Medium-Range Weather Forecast (ECMWF) that have five members. Each hindcast runs for 10 years. The external forcing of the models (greenhouse gasses, ozone, natural and anthropogenic aerosols, solar activity, land use) is based on the CMIP5 recommended historical datasets. After 2005, the concentrations and land use are based on the RCP 4.5 emission scenario [*Moss et al.* 2010].

2.1. Model Description

2.1.1. The EC-Earth Model System

[9] The EC-Earth V2.3 has been used in this study. The EC-Earth V2.2 model and its main characteristics are described by *Hazeleger et al.* [2010, 2012]. In EC-Earth V2.3, a slightly different aerosol forcing has been used, consistent with the CMIP5 protocol. We use a horizontal spectral resolution of T159 (triangular truncation at wavenumber 159) and 62 layers in the vertical up to 5 hPa. The atmosphere model is derived from the Integrated Forecast System cycle 31r1 of the ECMWF. The ocean model is the NEMO version 2 model [*Madec*, 2008], and the sea ice model is the LIM version 2 model [*Goosse and Fichefet*, 1999]. For details and further references, we refer to *Hazeleger et al.* [2012].

[10] The system employs a full-field initialization. The ocean initial conditions have been produced with NEMOVAR at ECMWF, a multivariate 3D-var data assimilation method for the NEMO ocean model [*Weaver et al.*, 2005, *Mogensen et al.*, 2012, *Balmaseda et al.*, 2012]. Observed three-dimensional temperature and salinity and the sea surface height are assimilated. In particular, the NEMOVAR-ORAS4 five-member ensemble has been used, which is the operational analysis for the new Seasonal forecast system (S4) at ECMWF [*Mogensen et al.*, 2012]. The sea ice conditions have been obtained from NEMO V2 and LIM2 forced by surface fluxes obtained from the Drakkar Forcing Set V4.3 [*Brodeau et al.*, 2010]. The atmosphere and land surface are initialized from ERA40 data [*Uppala and coauthors*, 2005] before 1989 and ERA-interim [*Dee and coauthors*, 2011] thereafter. The atmosphere is perturbed using singular vectors to create an ensemble of five members.

Table 1. Summary of Models and Initialization Techniques Used by the THOR Models (See Section 2.1 for Details)

	Model	Init. Method	Init. Data
MPI-M	ECHAM5/MPI-OM	anomaly	ORAS3
UKMO	HadCM3	anomaly	UKMO anal.
ECMWF	IFS/NEMO	full	ORAS4
KNMI	EC-Earth	full	ORAS4

2.1.2. The DePreSys System

[11] The Met Office Decadal Prediction System, DePreSys [Smith *et al.*, 2007, 2010], is based on the third Hadley Center coupled global climate model, HadCM3 [Gordon and coauthors, 2000, Pope *et al.*, 2000]. The atmosphere component has 19 vertical levels with a lid at approximately 40 km and a horizontal resolution of 2.5° latitude by 3.75° longitude. The ocean component has a horizontal resolution of 1.25 by 1.25° , and 20 vertical levels with 5 in the upper 50 m, the upper layer being 10 m thick.

[12] In order to create initial conditions for hindcasts and forecasts, HadCM3 is run in assimilation mode from December 1958 to the present day, including time-varying radiative forcing from changes in well-mixed trace gases, ozone, sulphate and volcanic aerosol, and solar irradiance. During this integration, the atmosphere and ocean are relaxed towards atmospheric and ocean analyses with a restoring time scale of 3 h in the atmosphere and 6 h in the ocean. The values are assimilated as anomalies with respect to the model climate [see Smith *et al.*, 2010]. The climatological period from which anomalies are computed is 1958 to 2001 for the atmosphere and 1951 to 2006 for the ocean. Atmospheric analyses are taken from ERA-40 [Uppala and coauthors, 2005] and ECMWF operational analyses, while analyses of ocean anomalies are created using an updated version of the scheme developed by Smith and Murphy [2007], based on anomaly covariances calculated from HadCM3, with adjustments to improve the fit to observations. The ensemble consists of 10 members.

2.1.3. The MPI-M Model System

[13] The Hamburg Max Planck Institut für Meteorologie (MPI-M) model version used in this study is the ECHAM5/MPI-OM. The horizontal resolution in the atmosphere is T63 (triangular truncation at wavenumber 63) and 31 layers in the vertical. The ocean model has an average horizontal resolution of 1.5° , but finer around Greenland. It has 40 vertical layers. More details about the same model setup with coarser resolution can be found in Kröger *et al.* [2012] and references therein.

[14] A 10-member ensemble was produced for each start date as outlined above. Initial conditions of the individual ensemble members were generated by shifting the initial states of ocean and atmosphere simultaneously against the radiative forcing (“lagged initialization”, daily intervals). Initial states stem from an assimilation run where anomalies of three-dimensional temperature and salinity fields from ORAS3 [Balmaseda *et al.*, 2008] were nudged into the coupled model. The restoring time scale of the anomalies is 10 days.

2.1.4. The ECMWF Model System

[15] A five-member ensemble of decadal predictions over the period 1960–2000 was carried out with the ECMWF coupled system. The atmosphere model is the Integrated Forecast System, cycle 36r4 [Bechtold *et al.*, 2008, Jung *et al.*, 2010]. The ocean and sea ice modules are the NEMO version 2 and LIM2 models, identical to EC-Earth [Madec, 2008, Goosse and Fichefet, 1999]. The horizontal resolution of the model is the same as for EC-Earth: T159 horizontal resolution in the atmosphere, 1° horizontal resolution in the ocean. ECMWF uses 91 layers in the vertical in the atmosphere and 42 oceanic layers. The model is identical as the ECMWF Seasonal Forecast System 4 [Molteni

et al., 2011], except for the sea ice model which is excluded in the Seasonal Forecast System. Compared to EC-Earth, the model has a higher vertical resolution and a newer IFS model cycle is used.

[16] The system employs full-field initialization that is very similar to EC-Earth. The atmosphere and land surface initialization is derived from the ERA-40 reanalysis [Uppala and coauthors, 2005] for the period 1960 to 1985 and from ERA-Interim [Dee and coauthors, 2011] for the remaining starting dates. The ocean and sea ice initial conditions and initialization strategy are identical to that of EC-Earth (see above).

3. Assessment of Skill

[17] We assess the skill of forecasts of observed surface and subsurface phenomena in the North Atlantic ocean, with a focus on the subpolar regions. For verification, we use a range of gridded data products. We prefer to use objectively analyzed fields that are close to the observations, rather than assimilation products that may be substantially influenced by the ocean model and assimilation method that is used, in particular in data sparse regions. The main argument to use the following datasets is that they are independent from the models. For sea surface temperature, we use the ERSST V3b data [Smith *et al.*, 2008] For subsurface temperatures and heat content, we use World Ocean Database 09 data [Levitus *et al.*, 2009; note that these include recent corrections on in-situ XBT temperature measurements]. For surface and subsurface salinity, the EN3 data is used [Ingleby and Huddleston, 2007]. For atmospheric quantities, we use the NCEP/NCAR reanalysis [Kalnay and coauthors, 1996].

[18] Because there is inherent climate variability, uncertainties in the model system formulations and in observational estimates of the real climate, a probabilistic verification is preferred. However, due to the long time scales considered here and the limited data availability, the use of probabilistic verification metrics is limited. Therefore, we use deterministic skill scores, such as anomaly correlations and root mean square differences between the ensemble mean and observations to quantify the skill of the multimodel ensemble, and we present time series such that individual events can be recognized. All members and start dates are used to assess the skill of the hindcasts as a function of the lead time. These scores are computed relative to anomalies over the verification period. This procedure implies that a drift correction has been applied by subtracting the average drift as a function of lead time as determined from all ensemble members and start dates. This is done for each model separately. Here, we assume that the drift is independent of the climate state, which is a common approach in seasonal forecasting. For the root mean square difference, we also assumed that the amplitude of the model forecast has been amplified to match the observations. When possible the score of a simple first-order autoregressive model (AR1) based on observations is included as a benchmark for the observed variations.

3.1. Atlantic Multidecadal Oscillation

[19] The AMO is often diagnosed as a pattern of multidecadal variability in north Atlantic SST and appears to vary with a time scale of about 70–90 years in the observations. Control simulations of models used in this study

indicate that this pattern of variability is related to AMOC variations [Knight *et al.*, 2006, Wouters *et al.*, 2012]. Although the AMO may vary naturally, it also appears to be influenced by external forcing factors including volcanoes [Otterå *et al.*, 2010] and anthropogenic aerosols [Booth *et al.*, 2012]. Furthermore, global warming from increases in greenhouse gases also affects North Atlantic SSTs. In order to minimize the influence of global warming on the AMO, we follow Trenberth and Shea [2006] and compute an AMO index as the SST anomaly in the North Atlantic (we use 0–60°N, 80–0°W), minus the global mean SST (60°S to 60°N) trend from 1960–2000. This index is almost orthogonal to the global warming signal, as it almost is in ensembles of climate models [van Oldenborgh *et al.*, 2009].

[20] Figure 1 shows time series of the AMO from observations and the spread and ensemble mean of the multimodel hindcasts at lead time of 1 year, 2–5 years, and 6–9 years. The temperature anomalies are averaged over the respective lead times. The deterministic scores indicate that the AMO is well predictable on multiyear time scales. Very high correlations between the observed and the multimodel mean time series are found. This confirms the results of van Oldenborgh *et al.* [2012] who showed a similar result in a different multimodel ensemble. The uninitialized CMIP5 multimodel mean data shows a correlation coefficient of the 2–5 year averaged temperature anomalies of less than 0.2 for 2–5 years (not shown), indicating that the AMO does not vary due to variations in external forcing and initialization with estimates of observed climate enhances skill of predictions. However, it is not possible to say whether initialization improves the skill by predicting natural variability or by correcting the model response to previous external forcing factors such as volcanoes [Otterå *et al.*, 2010] or anthropogenic aerosols [Booth *et al.*, 2012]. Also, the AR1 model shows at one year lead time a high correlation. At longer lead times, the dynamical prediction models show larger correlations, but differences are not significant and the AR1 is more skilful than the uninitialized CMIP5 models, which indicates memory in this part of the climate system due to the initial state.

[21] These scores are comparable with those reported by van Oldenborgh *et al.* [2012]. They found anomaly correlations of 0.84 and 0.57 at 2–5 years and 6–9 years,

respectively. It shows that there is scope for multiyear predictions in the Atlantic Ocean and because the AMO is believed to be related to the AMOC, possibly the AMOC itself, although it should be noted that the AMO does not seem to drive AMOC variations in the models (see section 4 for further discussion). It also implies that related climate phenomena such as hurricane formation in the Atlantic and Sahel rainfall may be predictable [e.g., Smith *et al.*, 2010, Vecchi *et al.*, 2012]. However, analyzing predictions of these phenomena is beyond the scope of this study, which focuses on the ocean.

3.2. Interhemispheric Gradient

[22] A particular event of interest is the abrupt change in interhemispheric gradient of SST that occurred in the early 1970s. This is clearly seen as a cooling in the North Atlantic, visible in the AMO time series as well, and a warming in the South Atlantic [Thompson *et al.*, 2010]. This is one of the most abrupt shifts visible in the observed SST record. Such interhemispheric gradients have often been associated with the AMOC, in particular in studies of paleorecords [Steig *et al.*, 1998].

[23] The uninitialized climate models of CMIP5 do not show skill at predicting the interhemispheric gradient (here defined as SST averaged over the entire northern hemisphere minus the SST averaged over the entire southern hemisphere, not shown). However, the initialized multimodel ensemble does show some skill (Figure 2). At one year lead time, the correlation between the observations and the multimodel mean is 0.7. It appears that most of the skill originates from SST in the North Atlantic, which corresponds to the AMO discussed in the previous section. However, the AR1 model shows very similar skill scores. Also, the largest signal, which is the shift in the early 1970s, is not captured in the multimodel mean. Since also the uninitialized CMIP5 model mean does not indicate such shifts, it either points to missing processes in the models or to an observational artifact.

3.3. Upper Ocean Heat Content

[24] Because the SST predictions associated with the AMO are skilful, it is expected that heat content forecasts,

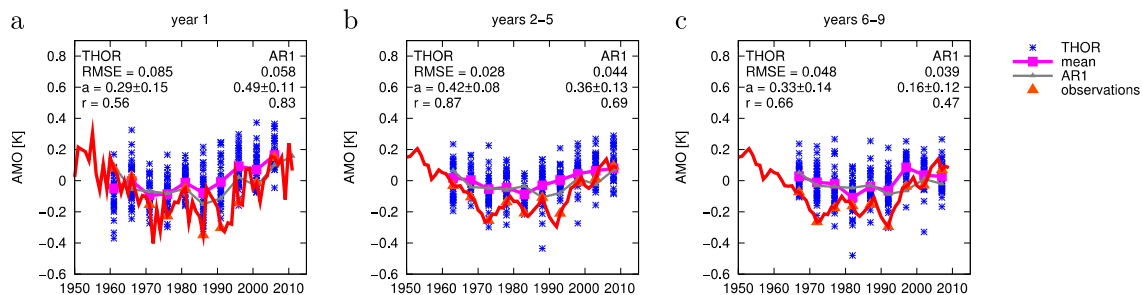


Figure 1. Observed (red curves) and ensemble mean predicted (magenta curves) AMO time series. Gray curves show the predictions from an AR1 model. Left panel 1: year lead time and yearly averaged anomalies, middle panel 2–5 years lead time, and 2–5 year averaged anomalies; Right panel: 6–9 year lead time and 6–9 year averaged anomalies. Individual ensemble members are indicated by asterisks. Top left corner shows the root mean square error of the ensemble mean, the slope A between the ensemble mean and observed time series, and the anomaly correlations. The root mean square error has been computed after correcting for the biases in mean and amplitude (A). On the right-hand side this is shown for the AR1 model.

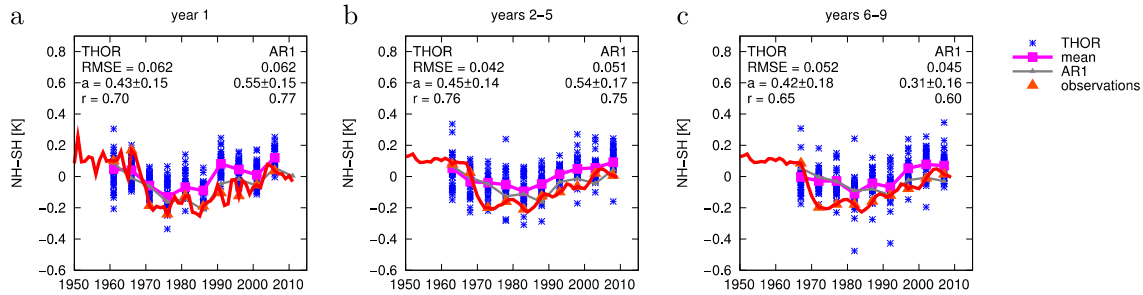


Figure 2. Observed (red curves) and ensemble mean predicted (magenta curves) interhemispheric temperature gradient. Gray curves show the predictions from an AR1 model. Left panel: 1 year lead time and yearly averaged anomalies; Middle panel: 2–5 years lead time and 2–5 year averaged anomalies; Right panel: 6–9 year lead time and 6–9 year averaged anomalies. Individual ensemble members are indicated by asterisks. Numbers in corners as in Figure 1.

which is the integral of temperature to a defined depth, would also be expected to be skilful. Indeed, the heat content of the upper 700 m in the subpolar North Atlantic (defined as 50–65°N, 65–0°W) is extremely well predicted (Figure 3). The persistence has a substantial impact on short lead times, explaining the high score for year-1 predictions. However, the predictions of 6–9 year averaged anomalies are less trivial. The scores are significantly better than those obtained with the AR1 model and indicate that oceanic mechanisms are at play that provide predictability. One of these processes is the formation of deep mixed layers in winter that produce homogeneous subsurface water masses that are advected around the gyres. Because the stratification is present in the initialization, some skill in predictions of the subsurface ocean is expected.

[25] The warming in the mid-1990s has been studied before [Robson *et al.*, 2012; Yeager *et al.*, 2012] and is thought to be generated by a persistently positive North Atlantic Oscillation (NAO) leading to a surge in the AMOC. The current results indicate that even hindcasts that started in 1990 capture part of this warming (consistent with Yeager *et al.* [2012]), although the amplitude of the temperature rise is underestimated. For 2–5 years lead time and 6–9 years lead time, the AR1 model does not capture the warming from 1990 onward. The preconditioning and hence the nonlinear effect of the NAO on the oceanic conditions is captured partly in the initial state. There are indications that the AMOC surge

alone cannot explain the predictability of the event because of the different relationships between AMO and AMOC in the prediction systems (see section 4.2). More start dates and more analysis are needed to study this in more detail.

3.4. Labrador Sea Water

[26] The formation of Labrador Sea Water shows pronounced decadal variability. The production of this water mass is thought to have an impact on the strength of the AMOC. At the end of the 1980s, deep convection occurred up to 2000 m, creating a cold and fresh water layer. Convection ceased in the mid-1990s. This major convection event is thought to be mainly driven by the atmospheric circulation variability. However, ocean processes may also be important. For example, changes in surface freshwater inflow by the East Greenland Current and Davis Strait affect the stability, and the inflow of more salty water of subtropical origin in the Irminger Current affects the stratification. Mesoscale eddies have been shown to be critical for restratification after convection [Gelderloos *et al.*, 2011]. Although these eddies are not resolved in the model systems used here, the eddy parameterizations will cause some influx of fresh water from the boundary currents towards the center of the Labrador Sea. Hence, initialization of ocean conditions might be expected to improve predictions of Labrador Sea water.

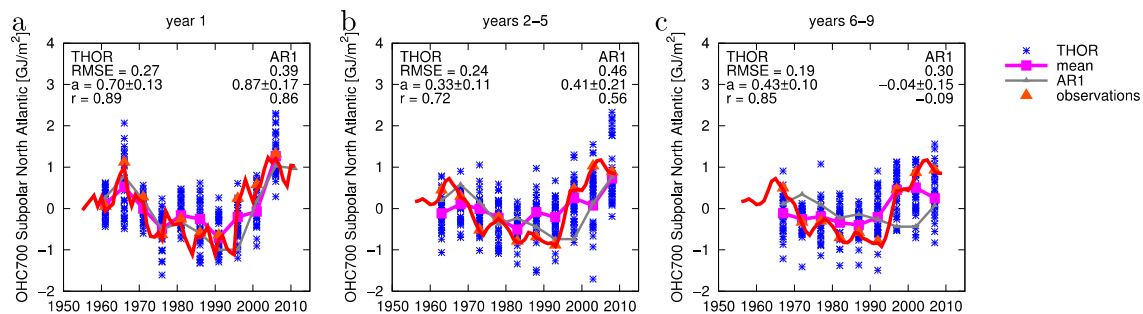


Figure 3. Observed (red curves) and ensemble mean predicted (magenta curves) upper ocean heat content (0–700 m) in the subpolar gyre (50–65°N, 65–0°W). Gray curves show the predictions from an AR1 model. Left panel: 1 year lead time and yearly averaged anomalies; Middle panel: 2–5 years lead time and 2–5 year averaged anomalies; Right panel: 6–9 year lead time and 6–9 year averaged anomalies. Individual ensemble members are indicated by asterisks. Numbers in corners as in Figure 1.

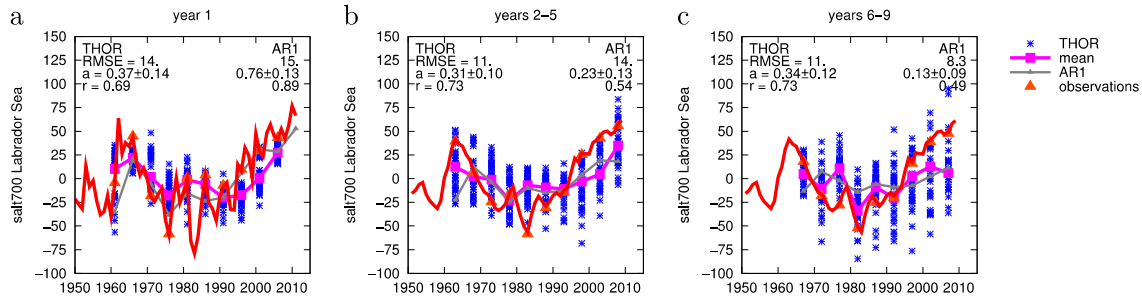


Figure 4. Observed (red curves) and ensemble mean predicted (magenta curves) depth integrated salinity anomalies (0–700 m). Gray curves show the predictions from an AR1 model. Left panel: 1 year lead time and yearly averaged anomalies; Middle panel: 2–5 years lead time and 2–5 year averaged anomalies; Right panel: 6–9 year lead time and 6–9 year averaged anomalies. Individual ensemble members are indicated by asterisks. The Labrador Sea is defined as 45.5–60.5°W, 54–62°N. Numbers in corners as in Figure 1.

[27] Despite the lack of representation of the rich ocean dynamics in the coarse resolution coupled models assessed here, the upper ocean salinity in the Labrador Sea (defined as 60.5–45.5°W, 54.0–62.0°N) is very well predicted (Figure 4). Even with a lead time of 6–9 years, a high anomaly correlation of 6–9 year averaged anomalies of 0.73 is found between the ensemble mean and observed time series. For the AR1 model, we find a correlation of 0.49. Individual convection years, indicated by minima in upper ocean salinity, are not predicted well systematically (as expected), but it is striking that skill in forecasts of the water masses is found at multiyear lead times.

[28] Similar results are obtained for heat content in the Labrador Sea (0.77, 0.73, 0.64 correlations for anomalies averaged at 1 year, 2–5 years, and 6–9 years, respectively, time series not shown). Surface salinity is also highly predictable and shows similar anomaly correlations (0.64, 0.75, 0.74 correlations for anomalies averaged at 1 year, 2–5 years, and 6–9 years, respectively, time series not shown). However, integrated salinity over the upper 2000 m is predicted with less skill. In particular, the increase in the 1990s is not captured at the 6–9 years lead time. Similar results are obtained for the entire subpolar gyre (up to 65°N), but for the Nordic seas (65–80°N), the longer range skill deteriorates (not shown).

[29] Great Salinity Anomalies are one of the most striking examples of oceanic phenomena reported in the literature [Belkin *et al.*, 1998]. In the early 1970s and in the early 1980s minima in upper ocean salinity occurred in the Labrador Sea and there are indications that these propagated along the subpolar gyre on a multiyear time scale. The minimum salinity in the East Greenland current observed in 1982 can perhaps be tracked to the Lofoten Basin in 1988. The salinity anomalies are clearly visible in the EN3 and NEMOVAR data in the Labrador Sea, but coherent propagation is not obvious. Similarly, the predictions assessed here show surface salinity anomalies in the Labrador Sea, but without coherent propagation around the subpolar gyre.

4. Discussion

4.1. Forecasts of Subpolar Gyre Water Mass Variations

[30] The analyses above have shown that there is skill in multiyear predictions of North Atlantic Ocean

characteristics. The difference in skill between the multimodel mean skill scores and those of damped persistence was investigated using a one-sided F -test on the ratio of the root mean square error scores and a T -test on the difference of the Fisher- z transforms of the correlation scores, taking serial autocorrelations of the residuals into account whenever these are significantly different from zero. Only the correlation scores of 3 year averaged anomalies of ocean heat content in the subpolar gyre at 6–9 years lead time are significantly better in the dynamical models using these tests. However, for all quantities investigated here at lead times of 2–5 years and 6–9 years, the correlation scores of the multimodel mean is higher than the score for damped persistence. We conclude that the results indicate that there is predictive skill for 2–5 year and 6–9 year averages beyond simple damped persistence, but the results do not definitely demonstrate that such skill has been achieved.

[31] In the subpolar gyre and Labrador Sea skill on multiyear time scales is found at the surface and up to 700 m depth. The deep mixed layers in winter enhance persistence, but cannot explain the skill of averaged anomalies at 6–9 years, as indicated by the low skill from the damped persistence. An event of particular interest is the warming in the mid-1990s, which is captured by the dynamical models, but not by damped persistence.

[32] Many studies indicate that variability in open ocean convection that produces the water masses is forced by the variations in atmosphere-ocean surface fluxes related to the NAO [Dickson *et al.* [1996], Eden and Willebrand [2001]]. It is therefore important to investigate whether the atmospheric forcing has provided the predictability found here. However, the multimodel mean predictions of mean sea level pressure over the ocean do not show skill for the annual mean wind forcing at multiyear time scales (Figure 5). It is therefore likely that oceanographic processes, such as preconditioning of the stratification and advection, are important for the forecast skill on decadal time scales (see Marshall and Schott, 1999 for a review on mechanisms of ocean convection). Even though the atmospheric forcing shows no skill, it may have contributed to the preconditioning. Doming of isopycnals leads to a weaker stratification so that convective instability can arise easier. Also, during restratification, both the heat fluxes at the surface and the advection of fresh water masses into the subpolar gyre could potentially provide predictability

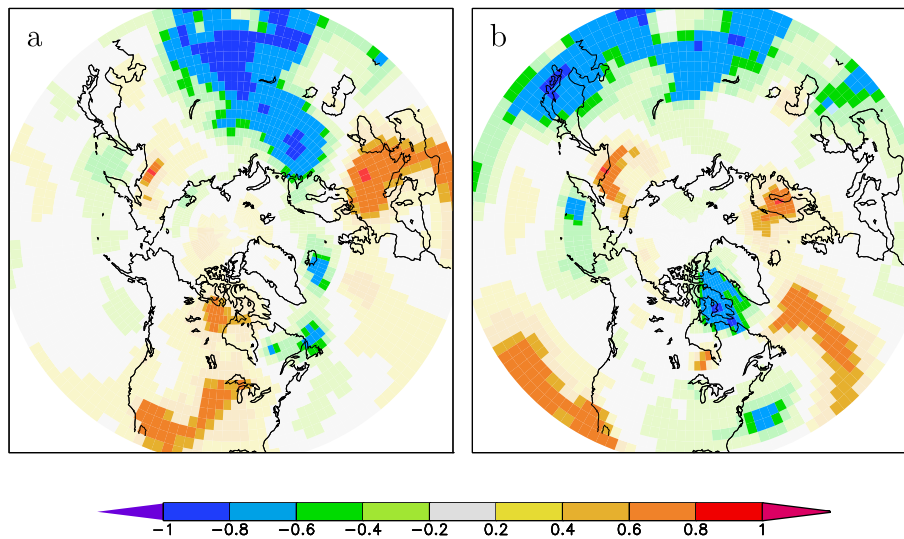


Figure 5. Anomaly correlation between NCEP/NCAR reanalysis mean sea level pressure anomalies and ensemble mean anomalies (detrended) in the hindcasts at year 1 (left) and year 2–5 lead time (right). Lighter shades (lighter than the shading in the color bar) denote areas where the correlations are not significant at $p < 0.1$ (two-sided), taking serial autocorrelations into account.

in the restratification phase. In nonconvective regions, the advection by ocean currents could provide predictability beyond the persistence derived from the initial state. Detailed analysis of these mechanisms would be interesting, but are beyond the scope of this paper.

4.2. AMOC Variations

[33] In the previous sections, we showed that subsurface oceanic thermal and fresh water characteristics that are thought to be related to the AMOC variations can be skilfully predicted at multiyear time scales. Also, the AMO, which is thought to respond to AMOC variations, is predictable at these time scales. This encouraging result may imply that the AMOC can be skilfully predicted.

[34] In Figure 6, we show forecasts of the maximum AMOC strength at 26°N . The ensemble mean of the forecasts at 1 year lead time indicates interannual variability and a longer-term reduction of the AMOC. At longer lead times, a reduction from 1995 onward is present. However, the spread is large, with one model, ECHAM5-OM, standing out with high values. Independent verification is not possible due to

lack of data. We include the AMOC transport data from the RAPID/MOCHA array in the figure for comparison [Cunningham *et al.*, 2007].

[35] A prerequisite of a reliable AMOC prediction is that the relations between the ocean characteristics and responses are robust in the prediction systems. Therefore, we further explore the mechanisms by analyzing lead-lag relations between the AMOC and fresh water and heat content in the subpolar gyre.

[36] Figure 7 shows the -4 year to $+4$ year lead-lag correlations between the maximum annually averaged AMOC strength at 40°N and the subpolar fresh water and heat content up to 700 m in the 10 year forecasts. Heat and fresh water variations contribute to density variations that are thought to drive AMOC variability. For heat content, a robust relationship is found amongst models. The upper subpolar gyre tends to be colder preceding a higher AMOC strength and warmer afterwards. This is consistent with, for example, Wouters *et al.* [2012] who find that a positive contribution of temperature on density variations in the subpolar gyre leads AMOC changes. Furthermore, the

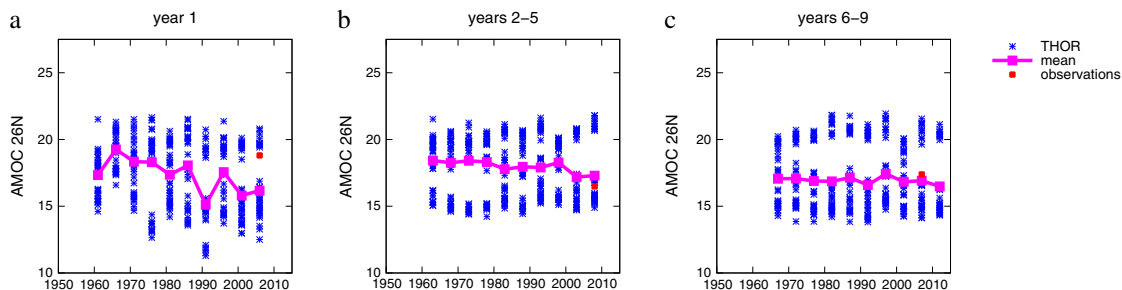


Figure 6. Observed (red dot) and ensemble mean predicted (magenta curves) AMOC strength at 26°N in Sv ($10^6 \text{ m}^3 \text{ s}^{-1}$). Left panel: 1 year lead time and yearly averaged anomalies; Middle panel: 2–5 years lead time and 2–5 year averaged anomalies; Right panel: 6–9 year lead time and 6–9 year averaged anomalies. Individual ensemble members are indicated by asterisks.

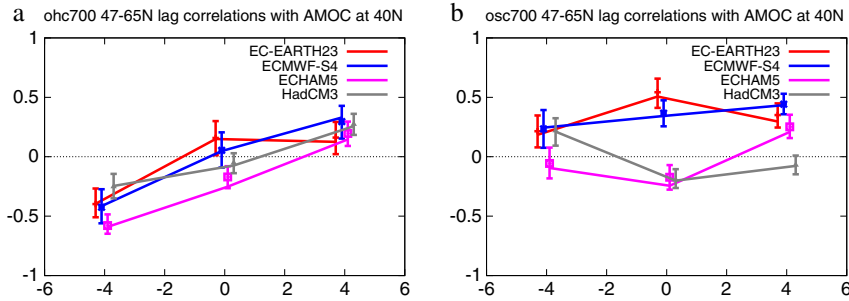


Figure 7. Lead-lag relations between maximum AMOC strength at 40°N and 0–700 m integrated temperature (left) and salinity (right) in the subpolar gyre in the Atlantic (47–65°N). The different prediction systems are indicated in Table 1. The error bars denote the 33%, 50%, and 67% quantiles of a nonparametric bootstrap estimate of the uncertainty.

reverse (but weaker) relationship of a warmer subpolar gyre following increased AMOC is consistent with a stronger heat transport by the AMOC

[37] For vertically integrated salinity, the signal is not consistent among the prediction systems. EC-Earth V2.3 and ECMWF S4 show similar relationships between fresh water and the AMOC strength. This is not surprising because these models have many common modules. In both models, a high salinity at subpolar latitudes leads to a stronger AMOC. At zero lag, there is a positive correlation. In contrast, in ECHAM5-OM, there is a reverse relationship between subpolar salinity and the AMOC when salinity leads. Also, when the AMOC leads different characteristics are found among

the models. For instance, the HadCM3 model shows that the subpolar gyre freshens, while the other systems indicate that a stronger AMOC will lead to a more saline subpolar gyre. It should be noted that correlations are low, and therefore a small amount of variance of AMOC variations can be explained by salinity variations. Still, the differences between the models are well outside the sampling accuracy denoted by 67% error bars (estimated with a nonparametric bootstrap).

[38] The details of the lead-lag relations can be seen in Figure 8. Here, spatial distributions of the correlations between integrated fresh water content and the AMOC are shown (ECMWF S4 is not shown as it is very similar to EC-Earth V2.3). Differences between models are found when

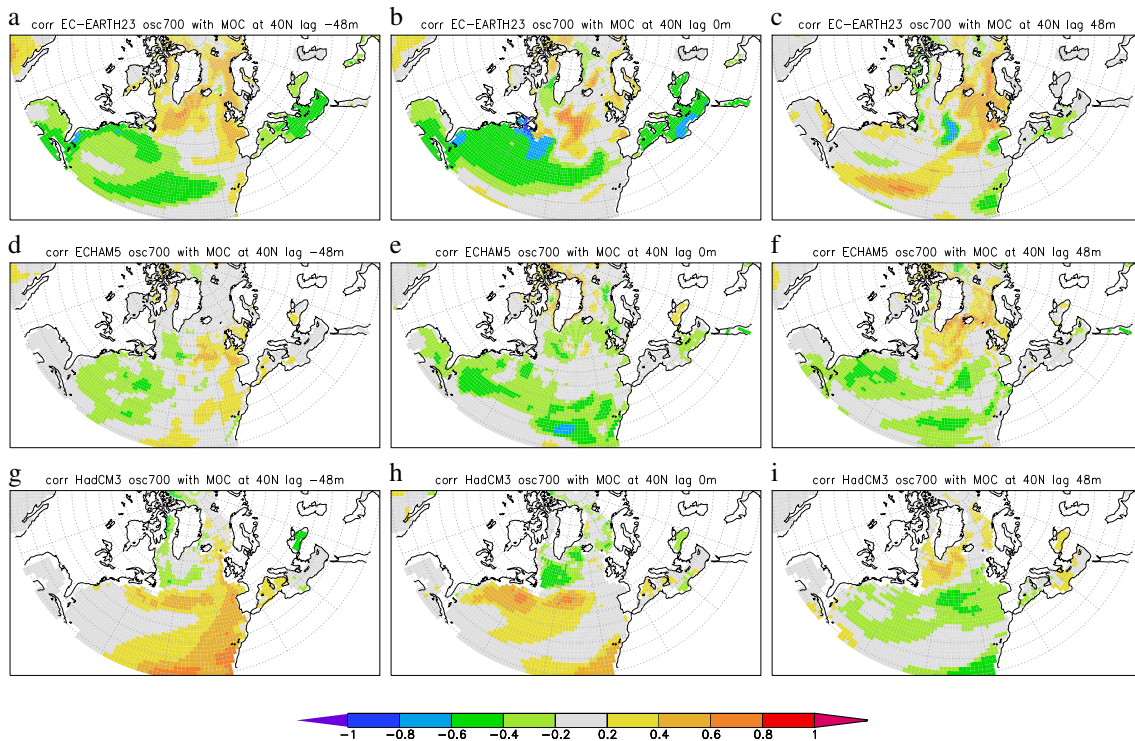


Figure 8. Correlation between maximum AMOC strength at 40°N and integrated salinity (0–700 m) at lag -4 years (left, salinity leads), lag 0 years, and lag 4 years (AMOC leads). Top row: EC-Earth V2.3, middle row: ECHAM5-OM, bottom row: HadCM3.

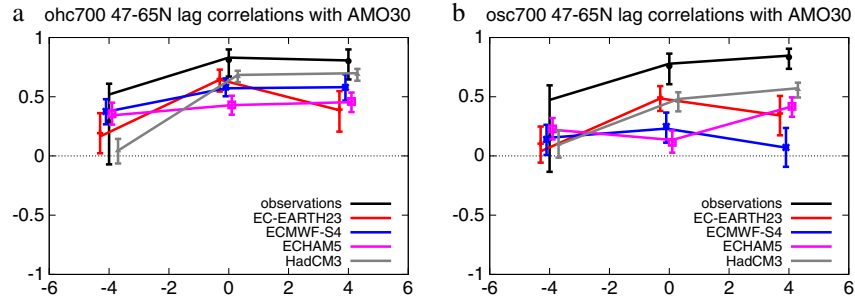


Figure 9. Left: lagged correlations between integrated temperature (0–700 m) and the AMO (northward of 30°N) for the different prediction systems. Right: lagged correlations between integrated salinity (0–700 m) and the AMO (northward of 30°N) for the different prediction systems. The error bars denote the 33%, 50%, and 67% quantiles of a nonparametric bootstrap estimate of the uncertainty.

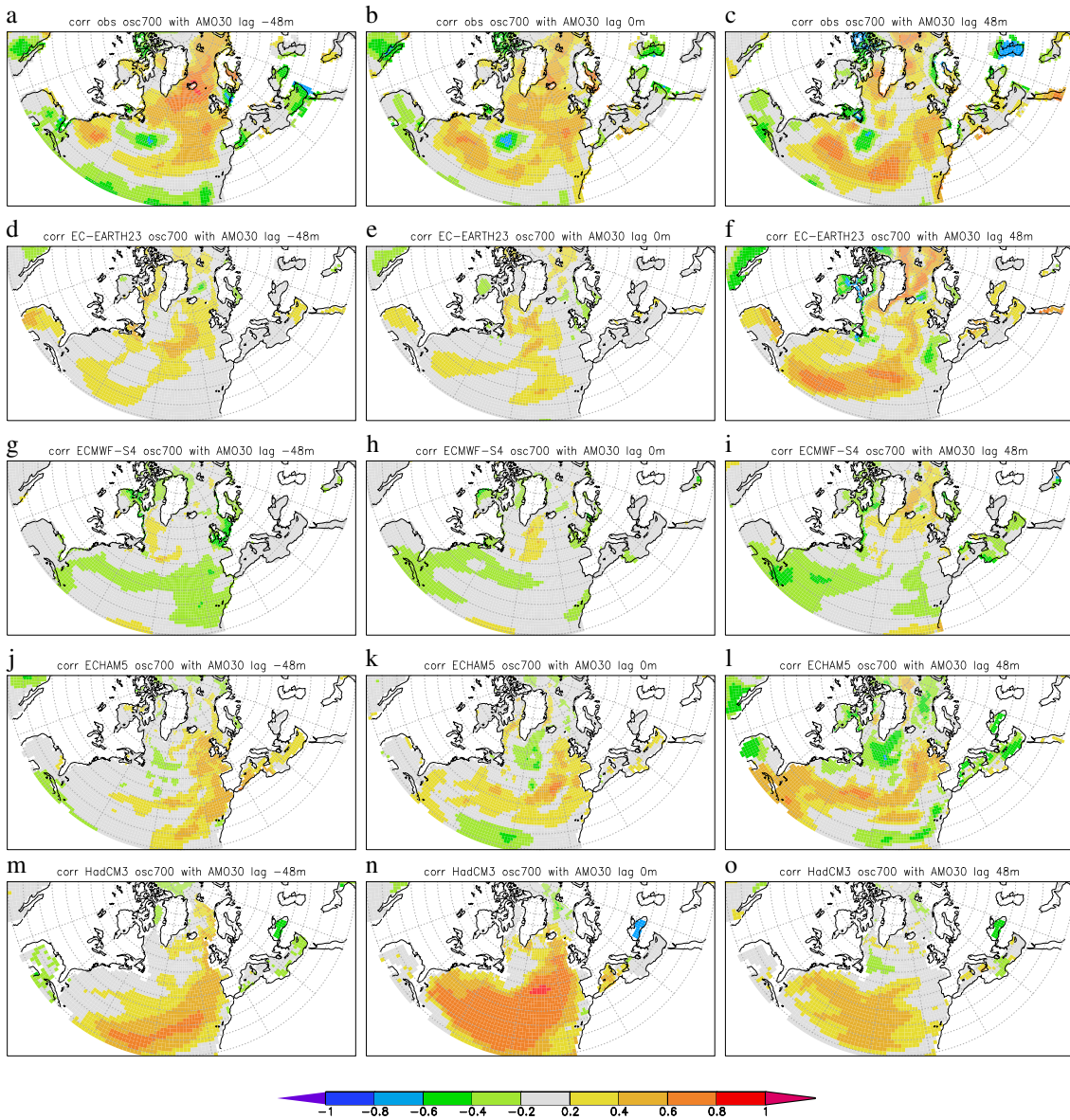


Figure 10. Correlation between the AMO (north of 30°N) and integrated salinity (0–700 m) at lag -4 years (left, salinity leads), lag 0 years, and lag 4 years (AMOC leads). Top to bottom: EN3 observations, EC-Earth V2.3, ECMWF S4, ECHAM5-OM, HadCM3.

the AMOC lags. Both EC-Earth V2.3 and ECHAM5-OM show a negative correlation in the Gulf Stream extension while HadCM3 is positively correlated at lag -4 years. At zero lag, the results of Figure 7 are reflected with positive salinity anomalies in the subpolar gyre in EC-Earth V2.3 and ECMWF S4 (not shown) and negative for HadCM3 and ECHAM5-OM. Apparently, salinity plays a more important role in EC-Earth V2.3 and ECMWF S4 in generating AMOC variability than in HadCM3 and ECHAM5-OM.

[39] It is hard to judge a priori which model shows the correct mechanisms of AMOC variability. The lack of direct observations prior to the RAPID/MOCHA array implies that verification of the AMOC beyond annual time scales is nearly impossible. Therefore, we further investigate the robustness of potentially relevant mechanisms of AMOC variability using observed ocean characteristics. In particular, we relate the integrated heat content and fresh water content in the subpolar gyre to the AMO. Only SST north of 30°N is used to exclude the tropical regions and focus on the North Atlantic only.

[40] Figure 9 shows the lead-lag relations for the different prediction systems and those of the observations. The general characteristics are similar, that is, all correlations are positive, but all prediction systems underestimate the correlation between the subsurface and surface variability at all lags compared to observations. These differences are expressed in the spatial correlation patterns as well (Figure 10). There is a positive correlation between the AMO and upper ocean salinity variations in the subpolar Atlantic at negative and zero lags that is only partially captured by some of the prediction systems. From these results, it is hard to judge the differences in quality of the individual models because all show deficiencies. It also shows that there is room for improvement of the predictions by improving on the mechanisms of variability in the prediction systems. Currently, the multimodel approach averages out some of the compensating biases. Improving the models will likely lead to enhanced skill.

5. Summary and Conclusions

[41] In this paper, we assessed the multiyear predictive skill in the North Atlantic of a multimodel ensemble of coupled atmosphere-ocean-sea ice hindcasts following the CMIP5 protocol. We investigated the hindcast skill of surface and subsurface characteristics, focusing on observables that are thought to be related to the AMOC, such as the AMO, stratification in the subpolar gyre, and Labrador Sea water mass characteristics. The multimodel ensemble shows skill up to 6–9 years ahead in the surface and subsurface temperature and salinity. This is consistent with potential predictability estimates [Boer, 2004] and previous studies [Yeager *et al.*, 2012, van Oldenborgh *et al.*, 2012, Chikamoto *et al.*, 2012, Smith *et al.*, 2010]. There is enhanced skill in the subpolar gyre provided by the initialization of the prediction systems that exceeds skill obtained from damped persistence. However, except for 3 year averaged subpolar heat content at lead times of 6–9 years, differences between skill of damped persistence and initialized dynamical prediction systems are not significant. We conclude that the results indicate that there is predictive skill for 2–5 year and 6–9 year averages beyond simple damped persistence, but the results do not definitely demonstrate that such skill has been achieved. The external

forcing by greenhouse gasses and aerosols provides some predictability as well. However, the skill exceeds that of coupled simulations which have not been initialized with an estimate of the observed state of the climate. The skill is comparable with that obtained in a study of earlier prediction systems [van Oldenborgh *et al.*, 2012]. Although individual events and abrupt shifts are not very well predicted, the skill for multiyear variability is an encouraging result.

[42] We extend previous assessments by including verification of subsurface ocean characteristics. Water masses in the subpolar gyre, including regions of active open ocean convection such as the Labrador Sea are highly predictable on multiyear time scales, with improved skill through initialization compared to uninitialized models in which water masses tend to be too homogenized in the subpolar gyre [de Jong *et al.*, 2009]. However, the skill of dynamic atmospheric variables is low, and there is no indication of long-term predictability of the sea level pressure. This indicates that the skill of predictions in our multimodel ensemble is of oceanic origin. Preconditioning to convection and advection of water masses that are properly initialized provides memory to the climate system that can lead to skilful predictions, even in models that do not resolve oceanic mesoscale eddies.

[43] It remains an open question whether the AMOC can be skilfully predicted as well beyond seasonal time scales. Pohlmann *et al.* [2012] found consistent predictions of the AMOC up to a few years ahead when assessed against the AMOC from a multimodel ocean analysis.

[44] The prediction systems appear to have different relationships between the AMOC and the temperature and salinity in the subpolar gyre. Also, the correlations between the AMO, which is well related to the AMOC in models, and the well-predicted integrated subpolar temperature and salinity are not consistent among models. Observations are used to explore the relationships between the AMO and upper ocean thermal and fresh water variations. It appears that all models show deficiencies and underestimate the correlation between surface and subsurface variations.

[45] The initialized predictions with global coupled models on multiyear time scales are relatively new. The scientific community has recently started to perform and analyze such predictions and optimize prediction systems. It is very encouraging that these prediction systems are already capable of providing skilful multiyear predictions in the North Atlantic and possibly associated climate phenomena. However, there is clearly scope for further improvement. The small number of start dates assessed here (every 5 years) can cause sampling errors and more start dates are needed. Also, the ocean analyses that are used to initialize the prediction systems include limited subsurface ocean data. There are indications that a better coverage of deep ocean data improves the predictions [Dunstone and Smith, 2010]. The methods to perturb the initial states could be further improved [Hawkins and Sutton, 2011]. Also, the model systems themselves contain large biases. The lagged correlations between the AMO and ocean characteristics show that the mechanisms of variability differ among the models [Branstator *et al.*, 2012 and this study]. It is possible that initialization corrects biases in the external forcing. There are indications that variations in aerosol concentration provide predictability [Booth *et al.*, 2012]. Improving on the representation of mechanisms of variability, external forcing

and confronting the models with observations will likely lead to improved predictions of the Atlantic region.

[46] **Acknowledgments.** This work has been sponsored by the EU Framework 7 Program THOR project (GA212643, 2008–2012). We thank the reviewers for constructive remarks on the manuscript.

References

- Balmaseda, M. A., A. Vidard, and D. L. T. Anderson (2008), The ECMWF ORA-S3 ocean analysis system, *Mon. Weather Rev.*, *136*, 3018–3034.
- Balmaseda M. A., K. Mogensen, and A. Weaver (2012), Evaluation of the ECMWF ocean reanalysis ORAS4, *Quart. J. Roy. Met. Soc.*, doi:10.1002/qj.2063.
- Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. Rodwell, F. Vitart, and G. Balsamo (2008), Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, *Quart. J. Roy. Meteor. Soc.*, *134*, 1337–1351.
- Belkin, I. M., S. Levitus, J. I. Antonov, and S.-A. Malmberg (1998), “Great Salinity Anomalies” in the North Atlantic, *Prog. Oceanogr.*, *41*, 1–68.
- Boer, G. J. (2004), Long time-scale potential predictability in an ensemble of coupled climate models, *Climate Dynamics*, *23*, 29–44. doi:10.1007/s00382-004-0419-8.
- Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N. (2012), Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability, *Nature*, *484*, 228–232.
- Branstator, G., M. Kimoto, J. R. Knight, M. Latif, and A. Rosati (2012), Systematic estimates of initial value decadal predictability for six AOGCMs, *J. Climate*, *25*, 1827–1846.
- Brodeau, L., B. Barnier, A.-M. Treguier, T. Penduff, S. Gulev (2010), An ERA40-based atmospheric forcing for global ocean circulation models, *Ocean Modelling*, *31*, 88–104.
- Chikamoto, Y., et al. (2012), An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC, *Climate Dynamics*, doi:10.1007/s00382-012-1351-y.
- Collins, M., and B. Sinha (2003), Predictability of decadal variations in the thermohaline circulation and climate, *Geophys. Res. Lett.*, *30*(6), 1955–8007, doi:10.1029/2002GL016504.
- Cunningham, S. A., et al. (2007), Temporal variability of the Atlantic meridional overturning circulation at 26.5°N, *Science*, *317*, 935–938.
- Curry, R., and C. Mauritzen (2005), Dilution of the northern North Atlantic Ocean in recent decades, *Science*, *308*, 1772–1774.
- Dee, D. P., and coauthors (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597.
- Delworth, T. L., and M. E. Mann (2000), Observed and simulated multidecadal variability in the Northern Hemisphere, *Climate Dynamics*, *16*, 661–676.
- Dickson, R. R., J. Meincke, S.-A. Malmberg, and A. J. Lee (1988), The “Great Salinity Anomaly” in the northern North Atlantic, 1968–1982, *Prog. Oceanogr.*, *20*, 103–151.
- Dickson, R. R., J. Lazier, J. Meincke, P. Rhines, and J. Swift (1996), Long-term-coordinated changes in the convective activity of the North Atlantic, *Prog. Oceanogr.*, *38*, 241–295.
- Dickson, R. R., I. Yashayev, J. Meincke, W. Turrell, S. Dye, and J. Holford (2002), Rapid freshening of the deep North Atlantic Ocean over the past four decades, *Nature*, *416*, 832–837.
- Doblas-Reyes, F. J., R. Hagedorn, and T. N. Palmer (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: Calibration and combination, *Tellus, Ser. A*, *57*, 234–252.
- Dunstone, N. J., and D. Smith (2010), Impact of atmosphere and sub-surface ocean data on decadal climate prediction, *Geophys. Res. Lett.*, *37*, L02709.
- Eden, C., and J. Willebrand (2001), Mechanism of interannual to decadal variability of the North Atlantic Circulation, *J. Climate*, *14*, 2266–2280.
- Gelderloos, R., C. A. Katsman, and S. S. Drijfhout (2011), Assessing the roles of three eddy types in restratifying the Labrador Sea after deep convection, *J. Phys. Oceanogr.*, *41*, 2102–2119.
- Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane (2001), Current approaches to seasonal-to-interannual climate predictions, *Int. J. Climatol.*, *21*, 1111–1152.
- Goosse, H., and T. Fichefet (1999), Importance of ice-ocean interactions for the global ocean circulation: A model study, *J. Geophys. Res.*, *104*, 23, 337–23 355.
- Gordon, C., and coauthors (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Climate Dynamics*, *16*, 147–168.
- Griffies, S. M., and K. Bryan (1997), A predictability study of simulated North Atlantic multidecadal variability, *Climate Dynamics*, *13*, 459–487.
- Griffies, S. M., and E. Tziperman (1995), A linear thermohaline oscillator driven by stochastic atmospheric forcing, *J. Climate*, *8*, 2440–2453.
- Hawkins, E., and R. Sutton (2011), Estimating climatically relevant singular vectors for decadal predictions of the Atlantic Ocean, *J. Climate*, *24*, 109, doi:10.1175/2010JCLI3579.1.
- Hazeleger, W., et al. (2010), EC-Earth: A seamless earth-system prediction approach in action, *Bull Amer Meteor Soc* *91*: 1357–1363. doi:10.1175/2010BAMS2877.1.
- Hazeleger W., et al. (2012), EC-Earth V2.2 Description and validation of a new seamless Earth system prediction model, *Climate Dynamics*, *39*(11), 2611–2629.
- Ingleby, B., and M. Huddleston (2007), Quality control of ocean temperature and salinity profiles - Historical and real-time data., *J. Mar. Syst.*, *65*, 158–175 10.1016/j.jmarsys.2005.11.019.
- de Jong, M. F., S. S. Drijfhout, W. Hazeleger, H. M. Van Aken, and C. A. Severijns (2009), Simulations of hydrographic properties in the northwestern North Atlantic Ocean in coupled climate models, *J. Climate*, *22*, 1767–1786.
- Jung, T., G. Balsamo, P. Bechtold, A. Beljaars, M. Köhler, M. Miller, J.-J. Morcrette, A. Orr, M. J. Rodwell, and A. M. Tompkins (2010), The ECMWF model climate: Recent progress through improved physical parametrizations, *Quart. J. Roy. Meteor. Soc.*, *136*, 1145–1160.
- Kalnay, E., and coauthors (1996), NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.*, *77*, 437–470.
- Keenlyside, N., M. Latif, J. Jungclauss, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, *453*, 84–88.
- Kim H.-M., P. J. Webster, and J. A. Curry (2012), Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts, *Geophys. Res. Lett.* *39*: L10701, doi:10.1029/2012GL051644.
- Knight, J. R., C. K. Folland, and A. A. Scaife (2006), Climate impacts of the Atlantic Multidecadal Oscillation, *Geophys. Res. Lett.* *33*(L17), 706, doi:10.1029/2006GL026242.
- Kröger, J., W. Müller, and J.-S. von Storch (2012), Impact of different ocean reanalyses on decadal climate prediction, *Clim. Dyn.*, doi:10.1007/s00382-012-1310-7.
- Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov (2009), Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems, *Geophys. Res. Lett.*, *36*, L07608, doi:10.1029/2008GL037155.
- Madec, G. (2008), NEMO ocean engine. Note du Pôle de Modélisation, Institut Pierre-Simon Laplace (IPSL), No 27.
- Marotzke, J. (1990), Instabilities and multiple equilibria of the thermohaline circulation. Ph.D. thesis, Berichte aus dem Institut für Meereskunde, 126 pp.
- Marshall, J., and F. Schott (1999), Open-ocean convection: Observations, theory, and models, *Rev. Geophys.*, *37*(1), 1–64, doi:10.1029/98RG02739.
- Mochizuki, T., et al. (2010), Pacific decadal oscillation hindcasts relevant to near-term climate prediction, *Proc. Natl. Acad. Sci.*, *107*, 1833–1837.
- Mogensen, K., M. A. Balmaseda, and A. Weaver (2012), The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for system 4, *ECMWF Technical Memorandum* 668.
- Molteni, F., et al. 2011, The new ECMWF seasonal forecast system (System 4), *ECMWF Technical Memorandum* 656.
- Moss, R. H., Edmonds, J. A., Hibbard, K. A., et al. 2010, The next generation of scenarios for climate change research and assessment, *Nature*, *463*, doi:10.1038/nature08823.
- van Oldenborgh, G. J., M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. T. Anderson (2005), Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *J. Climate*, *18*, 16, 3240–3249.
- van Oldenborgh, G. J., L. A. te Raa H. A. Dijkstra, and S. Y. Philip (2009), Frequency- or amplitude-dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean, *Ocean Sci.* *5*, 293–301.
- van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger (2012), Skill in the trend and internal variability in a multi-model decadal prediction ensemble, *Clim. Dyn.*, *38*, 7, 1263–1280.
- Otterå, O. H., M. Bentsen, H. Drange, and L. Suo (2010), External forcing as a metronome for Atlantic multidecadal variability, *Nat. Geosci.*, doi:10.1038/NNGEO955.
- Pohlmann, H., J. Jungclauss, A. Köhl, D. Stammer, and J. Marotzke (2009), Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic, *J. Climate*, *22*, 3926–3938.
- Pohlmann, H., D. M. Smith, M. A. Balmaseda, N. S. Keenlyside, S. Masina, D. Matei, W. A. Müller, and P. Rogel (2012), Predictability of the mid-latitude Atlantic meridional overturning circulation in a multimodel system, *Clim. Dyn.*, doi:10.1007/s00382-013-1663-6.
- Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A. (2000), The impact of new physical parametrizations in the Hadley Centre climate model - HadAM3, *Clim. Dyn.* *16*, 123–146.

- te Raa, L. A., and Dijkstra H. A. (2002), Instability of the thermohaline circulation on inter decadal time scales, *J. Phys. Oceanogr.*, *32*, 138–160.
- Robson, J., R. Sutton, K. Lohmann, D. Smith, and M. D. Palmer (2012), Causes of the rapid warming of the North Atlantic ocean in the mid-1990s, *J. Climate*, *25*, 4116–4134, doi: 10.1175/JCLI-D-11-00443.1
- Smith, D. M., and J. M. Murphy (2007), An objective ocean temperature and salinity analysis using covariances from a global climate model, *J. Geophys. Res.*, *112*, C02022.
- Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model, *Science*, *317*, 796–799, doi:10.1126/science.1139540.
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA’s historical merged land-ocean surface temperature analysis (1880–2006). *J. Climate*, *21*, 2283–2296.
- Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife (2010), Skilful multi-year predictions of Atlantic hurricane frequency, *Nat. Geosci.*, doi:10.1038/NCEO1004
- Steig E. J., E. J. Brook, J. W. C. White, C. M. Sucher, M. L. Bender, S. J. Lehman, D. L. Morse, E. D. Waddington, and G. D. Clow (1998), Synchronous climate changes in Antarctica and the North Atlantic, *Science*, *282*, 92–95.
- Sutton, R. T., and D. L. R. Hodson (2005), Atlantic Ocean forcing of North American and European summer climate, *Science*, *309*, 115–118.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An Overview of CMIP5 and the experiment design, *Bull. Amer. Meteor. Soc.*, *93*, doi:10.1175/BAMS-D-11-00094.1.
- Thompson, D. W. J., J. M. Wallace, J. J. Kennedy, and P. D. Jones (2010), An abrupt drop in Northern Hemisphere sea surface temperature around 1970, *Nature* *467*, 444–447.
- Trenberth, K. E., and D. J. Shea (2006), Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.*, *33*, L12704.
- Uppala, S. M., and coauthors (2005), The ERA-40 reanalysis, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012.
- Vecchi, G. A., et al. (2012), Multi-year Predictions of North Atlantic Hurricane Frequency: Promise and Limitations, *J. Climate*, doi:10.1175/JCLI-D-12-00464.1.
- Weaver, A. J., and E. S. Sarachik (1991), Evidence for decadal variability in an ocean general circulation model: an advective mechanism, *Atmos. Ocean*, *29*, 197–231.
- Weaver, A. T., C. Deltel, E. Machu, S. Ricci, and N. Daget (2005), A multivariate balance operator for variational ocean data assimilation, *Q. J. R. Meteorol. Soc.*, *131*, 3605–3625.
- Wouters, B., S. S. Drijfhout, and W. Hazeleger (2012), Interdecadal North-Atlantic meridional overturning circulation variability in EC-EARTH, *Clim. Dyn.*, *39*, 2695–2712.
- Yashayaev, I. (2007), Hydrographic changes in the Labrador Sea, 1960–2005, *Prog. Oceanogr.*, *73*, 242–276, doi:10.1016/j.pocean.2007.04.015.
- Yeager, S., A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng (2012), A decadal prediction case study: Late 20th century north Atlantic Ocean heat content, *J. Clim.*, doi: http://dx.doi.org/10.1175/JCLI-D-11-00595.1
- Zhang, R., and T. L. Delwoth (2006), Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes, *Geophys. Res. Lett.*, *33*(L17), 712, doi:10.1029/2006GL026267.