Impact of Initial Conditions versus External Forcing in Decadal Climate Predictions: A Sensitivity Experiment*

SUSANNA CORTI,[+,#] TIM PALMER,[+,@] MAGDALENA BALMASEDA,[+] ANTJE WEISHEIMER,[+,@]
SYBREN DRIJFHOUT,[&] NICK DUNSTONE,[**] WILCO HAZELEGER,[&,++,##] JÜRGEN KRÖGER,[@@]
HOLGER POHLMANN,[**,@@] DOUG SMITH,[**] JIN-SONG VON STORCH,[@@] AND BERT WOUTERS[&]

[+] *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*
[#] *Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche, Bologna, Italy*
[@] *Department of Physics, National Centre for Atmospheric Science, University of Oxford, Oxford, United Kingdom*
[&] *Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*
[**] *Met Office, Exeter, United Kingdom*
[++] *Wageningen University, Wageningen, Netherlands*
[##] *Netherlands eScience Center, Amsterdam, Netherlands*
[@@] *Max Planck Institut für Meteorologie, Hamburg, Germany*

ABSTRACT

The impact of initial conditions relative to external forcings in decadal integrations from an ensemble of state-of-the-art prediction models has been assessed using specifically designed sensitivity experiments (SWAP experiments). They consist of two sets of 10-yr-long ensemble hindcasts for two initial dates in 1965 and 1995 using either the external forcings from the "correct" decades or swapping the forcings between the two decades. By comparing the two sets of integrations, the impact of external forcing versus initial conditions on the predictability over multiannual time scales was estimated as the function of lead time of the hindcast. It was found that over time scales longer than about 1 yr, the predictability of sea surface temperatures (SSTs) on a global scale arises mainly from the external forcing. However, the correct initialization has a longer impact on SST predictability over specific regions such as the North Atlantic, the northwestern Pacific, and the Southern Ocean. The impact of initialization is even longer and extends to wider regions when below-surface ocean variables are considered. For the western and eastern tropical Atlantic, the impact of initialization for the 700-m heat content (HTC700) extends to as much as 9 years for some of the models considered. In all models the impact of initial conditions on the predictability of the Atlantic meridional overturning circulation (AMOC) is dominant for the first 5 years.

1. Introduction

Decadal climate predictions exploit the predictability of the climate system arising both from the initial condition information and from external forcings arising

Corresponding author address: Susanna Corti, ISAC-CNR, Via Gobetti 101, 40129 Bologna, Italy.
E-mail: s.corti@isac.cnr.it

from changes in atmospheric composition, solar radiation, and land use (Collins 2002; Hawkins and Sutton 2009). As such, decadal climate forecasts represent a hybrid problem at the edge between predictions of the first and the second kind (Lorenz 1975). A prediction of the first kind is essentially the prediction of the evolution of a system, given some knowledge of its initial state. Predictability of the first kind is therefore primarily an initial-value problem, and skill is limited by how uncertainties in the initial state evolve during the forecast (and by inevitable uncertainties in model formulation). On the other hand, in a prediction of the second kind, we estimate how (the attractor of) a given dynamical system responds to a change in some prescribed parameter or external forcing. Uncertainties in such predictions may arise from the accuracy in prescribing the forcings
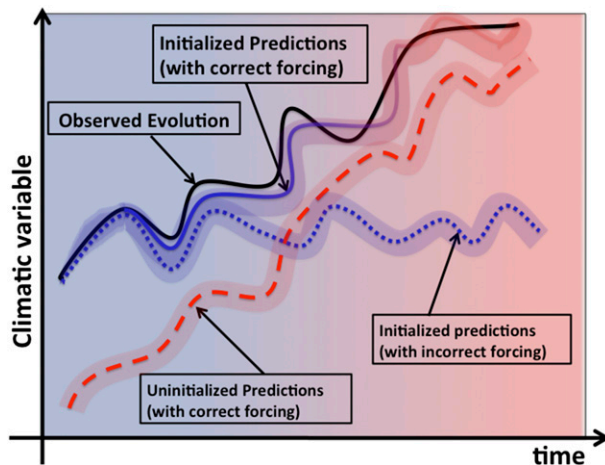
FIG. 1. Schematic illustrating progression from predictions of the first kind (initial-value problems) to predictions of the second kind (boundary condition problems). The black curve represents the observed evolution of a climate variable under the effect of ongoing changes in the eternal forcing (e.g., greenhouse gases). The dotted blue curve and its shading represent the trajectory of the ensemble mean of initialized predictions with an erroneous external forcing. (The shading represents the ensemble spread. It increases with time until it reaches saturation.) The dashed red curve represents the trajectory of an ensemble mean of non-initialized predictions. The solid blue–purple–red curve represents the trajectory of an ensemble mean of initialized predictions with correct forcing.

themselves and the accuracy of the model response to changes in the forcings. Predictability of the second kind is thus a boundary value problem. A numerical weather forecast is clearly a prediction problem of the first kind, as is a forecast of El Niño. By contrast, estimating the effects on climate of prescribed emissions after a volcanic eruption or prescribed anthropogenic changes in atmospheric composition or land use would constitute a climate prediction of the second kind.

Near-term climate predictions from one to several years in advance represent the natural extension of seasonal predictions, so they can be considered as initial-value problems where the correct initialization of the ocean surface and subsurface variables together with the sea ice plays a central role in keeping the trajectory of the system close to the observed one. However, as lead time increases from one season to several years, the component of predictability arising from changes in the external forcing becomes more and more important (Meehl et al. 2009). The progressive change in the main predictability's drivers with time is illustrated in Fig. 1. The schematic shows the time evolution of a given observed climatic variable (black solid line) and three different ensemble predictions with the shading around the lines representing the ensemble spread. The background color changes from blue to red as time

progresses, indicating the increasing dominance of the forcing with respect to the initial conditions for the prediction. The blue dotted line represents the ensemble-mean evolution of an initialized ensemble prediction constrained by a boundary forcing different from the observed one. By contrast, the red dashed line sketches the evolution of the ensemble mean when the prediction is uninitialized (i.e., the model initial conditions are not close to the observations), but the correct observed forcing is applied during the whole length of the integration. For short lead times, the blue line follows the observations, but since the system is chaotic and the forcing is not correct, as time advances the two lines diverge. On the other hand, the red line is far away from the observation at the beginning, but it starts to approach the black line and reproduce its trend at longer lead times, when the system response becomes dominated by the forcing. Between short and long lead times, there is a period when both the blue and red lines fail to reproduce the observed trajectory. The blue–purple–red solid line represents the ensemble-mean evolution when both initialization and forcing are close to the observed ones. In this case, the simulated trajectory lies close to the observed climate all along the forecast length, filling the predictability gap at medium-term lead times.

Since the seminal papers of Griffies and Bryan (1997a,b), whose results suggested that variations in the dominant multidecadal sea surface temperature patterns in the North Atlantic can be predicted if the Atlantic Ocean is adequately monitored, several studies showed the potential added value of initialization for climate predictions (e.g., Collins 2002; Collins and Sinha 2003; Pohlmann et al. 2004; Collins et al. 2006; Dunstone and Smith 2010; Msadek et al. 2010; Hazeleger et al. 2013b; Wouters et al. 2013). However, the question of when exactly the impact on climate forecasts of the initial state becomes smaller than the impact of an imposed external forcing (i.e., when the blue and red lines cross in the schematic) is still a debated issue and an area of active, current research.

The time when the influence of the initial conditions becomes of secondary importance compared to the forced response has been defined as the second limit of initial-value predictability by Branstator and Teng (2010). These authors applied the concept of relative entropy from information theory (Teng and Branstator 2011; Kleeman 2002) to a large ensemble of climate change scenario experiments with the Community Climate System Model, version 3 (CCSM3). They found that, for the upper 300-m ocean temperature on a global scale, the information from initial conditions is of primary importance for about 7 years. When the same methodology was applied to study the predictability of

the Atlantic meridional overturning circulation (AMOC) in CCSM3, it was found that ocean initial conditions are the predominant factor in driving predictability for about a decade (Teng et al. 2011). However, these estimates cannot be easily generalized since model to model variations appear to be substantial (Branstator et al. 2012; Branstator and Teng 2012).

In this study, we investigate this problem using an experimental setup based on decadal ensemble integrations initialized in two different decades, 1965 and 1995, carried out by four different Earth system models. Two reference simulations in which all the models are initialized with states as close as possible to observations and forced with the observed external forcings for the whole 10-yr length of the integration were performed following the protocol from phase 5 of CMIP (CMIP5) for decadal hindcasts (Taylor et al. 2012). (Each set of reference integrations is represented ideally by the blue–purple–red solid curve and its shading in Fig. 1.) Then, following Troccoli and Palmer (2007), two sensitivity ensemble integrations were carried out. Specifically, we integrate the 1965 hindcast starting from correct initial conditions but with the forcing from the 1995 decade and vice versa. By comparing the reference and the sensitivity integrations, we can assess the relative importance of initial conditions compared to the external forcing. In particular, two estimates of the decadal predictability arising from initial conditions only (i.e., the predictability associated with the blue dotted line in Fig. 1) and two estimates of the predictability driven by the external forcing (i.e., the predictability associated with the red dashed curve in Fig. 1) can be obtained.

The paper is structured as follows: Section 2 describes the experimental setup. Section 3 discusses the results of the experiment in terms of sea surface temperature anomalies. The relative effect of external forcing and initial conditions on the AMOC is the subject of section 4. In section 5, we define the "crossover time," that is, the time that separates the region in which predictability is mostly driven by the knowledge of initial conditions (denoted by the blue background in Fig. 1) from the region where predictability is mostly driven by changes in the external forcing (red background in Fig. 1). In section 6, the definition of crossover time is applied to sea surface and subsurface variables. A summary of the main results and some conclusions are presented in section 7.

2. Experimental setup

To investigate the relative roles of initial and boundary conditions in decadal predictability sensitivity experiments consisting of swapping initial and boundary

TABLE 1. Sensitivity experiments (see main text for explanation of IC65F65, IC95F95, IC65F95, and IC95F65). Comparing experiments in the same row gives the impact of forcing. The comparison in the same column gives the impact of initialization.

	Boundary conditions (forcing)	
Initial conditions	1965	1995
1965	IC65F65	IC65F95
1995	IC95F65	IC95F95

conditions for two different decades has been carried out (SWAP experiments). Two initial dates, namely, 1965 (preceding a decade of global cooling) and 1995 (preceding a decade of global warming), were chosen, and the following four 10-yr-long hindcasts were produced:

- IC65F65 for 1965 initial conditions and correct observed forcing from 1965;
- IC95F95 for 1995 initial conditions and correct observed forcing from 1995;
- IC65F95 for 1965 initial conditions and swapped observed forcing from 1995; and
- IC95F65 for 1995 initial conditions and swapped observed forcing from 1965.

By comparing IC65F65 with IC95F65, and IC95F95 with IC65F95, we have two estimates of the decadal predictability arising from having different initial conditions and the same forcing. By comparing IC65F65 with IC65F95, and IC95F95 with IC95F65, we have two estimates of the impact of forcing (since initial conditions are identical) on the predictability of climate variables. A schematic of the four experiments is given in Table 1.

The SWAP experiments were carried out by modeling groups at ECMWF, KNMI, the Met Office (UKMO), and the Max Planck Institute for Meteorology (MPI-M). Different initialization strategies have been employed, and different ocean analyses have been used. ECMWF and KNMI use a full field initialization, while UKMO and MPI-M apply anomaly initialization [see Magnusson et al. (2013), Hazeleger et al. (2013a), and Smith et al. (2013) for a comparison between the two methodologies]. All the models except MPI-M derive atmospheric initial conditions from the reanalyses ERA-40 and ERA-Interim (Uppala et al. 2005; Dee et al. 2011). The ocean initial conditions are derived from the NEMO variational data assimilation (NEMOVAR)–Ocean Reanalysis System, version 4 (ORAS4) (Balmaseda et al. 2012) for ECMWF and KNMI, while MPI-M uses the Ocean Reanalysis System, version 3 (ORAS3) (Balmaseda et al. 2008). The Met Office creates ocean analysis following the strategy described by Smith and Murphy (2007). [See Hazeleger et al. (2013b) for more

TABLE 2. Summary of model setup used for the SWAP experiments.

	Model	Initialization method	Ocean analysis	No. of ensemble members
ECMWF	IFS–NEMO	Full	NEMOVAR–ORAS4	5
KNMI	EC-EARTH	Full	NEMOVAR–ORAS4	3
MPI-M	ECHAM5–MPI-OM	Anomaly	ORAS3	5
UKMO	HadCM3	Anomaly	UKMO analysis	3

details on the models' configurations.] The sea ice conditions have been obtained from NEMO version 2 and LIM, version 2 (LIM2), forced by surface fluxes obtained from the Drakkar forcing set, version 4.3 (Brodeau et al. 2010), for ECMWF and KNMI, while MPI-M does not initialize sea ice. In HadCM3, sea ice concentration was initialized by relaxing (with a 6-h time scale) to monthly mean anomalies obtained from HadISST (Rayner et al. 2003). The external forcing is the same for all models, and it is based on the CMIP5 recommended historical datasets (Taylor et al. 2012), the integrations are carried out in ensemble mode. A summary of the model setups can be found in Table 2. (Expansions of acronyms for the model names are available at http://www.ametsoc.org/PubsAcronymList.)

We focus our analysis on oceanic variables. In particular, we want to investigate the relative roles played by initial and boundary conditions on the predictability of SST, heat content in the upper 700 m (HTC700), and the AMOC.

3. Observed and simulated SST anomalies

Figure 2 shows the difference in sea surface temperature from the ERA-40 and ERA-Interim datasets in the two investigated decades for different averaging windows. In Fig. 2a, the difference between the 1-yr average from November 1995 to October 1996 and from November 1965 to October 1966 is shown. Most of the oceanic basins show a global warming signal, with the important exception of the Pacific Ocean, where the anomaly is consistent with the cold phase of the El Niño–Southern Oscillation (ENSO). In fact, the year 1995/96 is characterized by a strong La Niña event. Figure 2b shows the mean SST difference in the first 5 years of the two decades under consideration. The warming now extends to the equatorial eastern Pacific, but the cold signal is still pronounced in the northern central Pacific. The difference in SST between the two decadal means is shown in Fig. 2c. The global feature of the decadal anomaly is consistent with the global warming trend observed from 1960 to 2010.

The SST maps in Fig. 2 are compared with corresponding maps from the individual model simulations. Figure 3 shows the difference between the mean of the

first year of integration in the ECMWF model. Figure 3a is the difference between the ensemble mean of the two reference integrations (i.e., IC95F95 minus IC65F65). The panels below in the left column are the differences

(a) 1995to1996 minus 1965to1966.
Observed SST anomaly

(b) 1995to2000 minus 1965to1970.
Observed SST anomaly

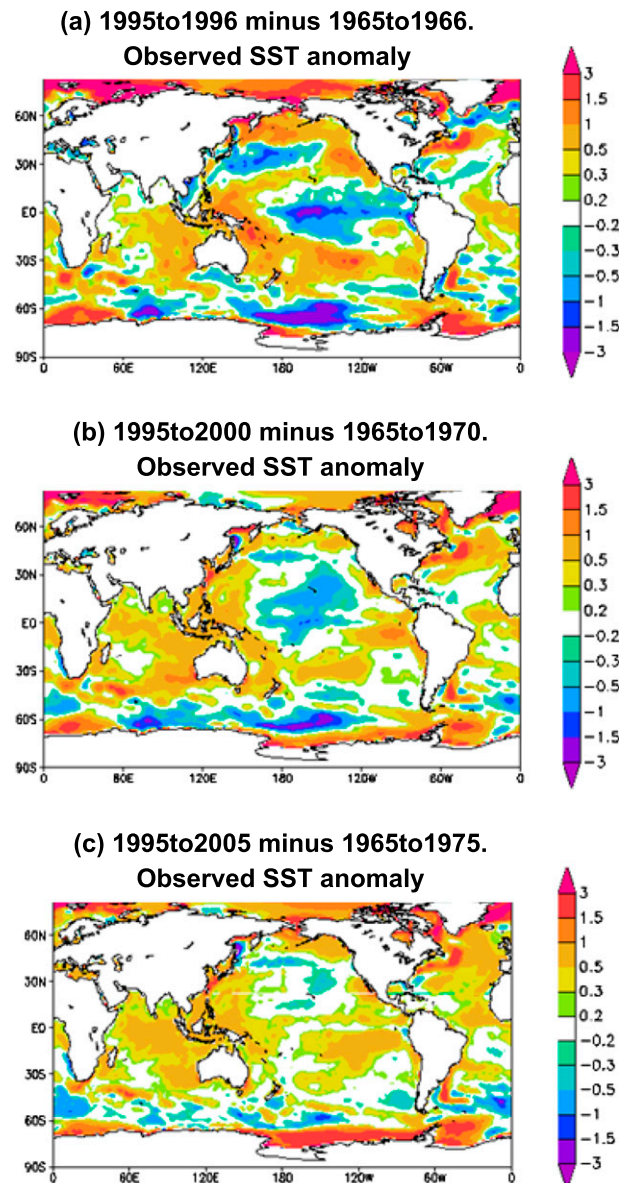(c) 1995to2005 minus 1965to1975.
Observed SST anomaly

FIG. 2. Observed SST differences (K) between decade 1995 minus decade 1965. (a) Differences between the first year mean of the two decades. (b) As in (a), but for the first 5-yr mean. (c) As in (a), but for 10-yr mean.
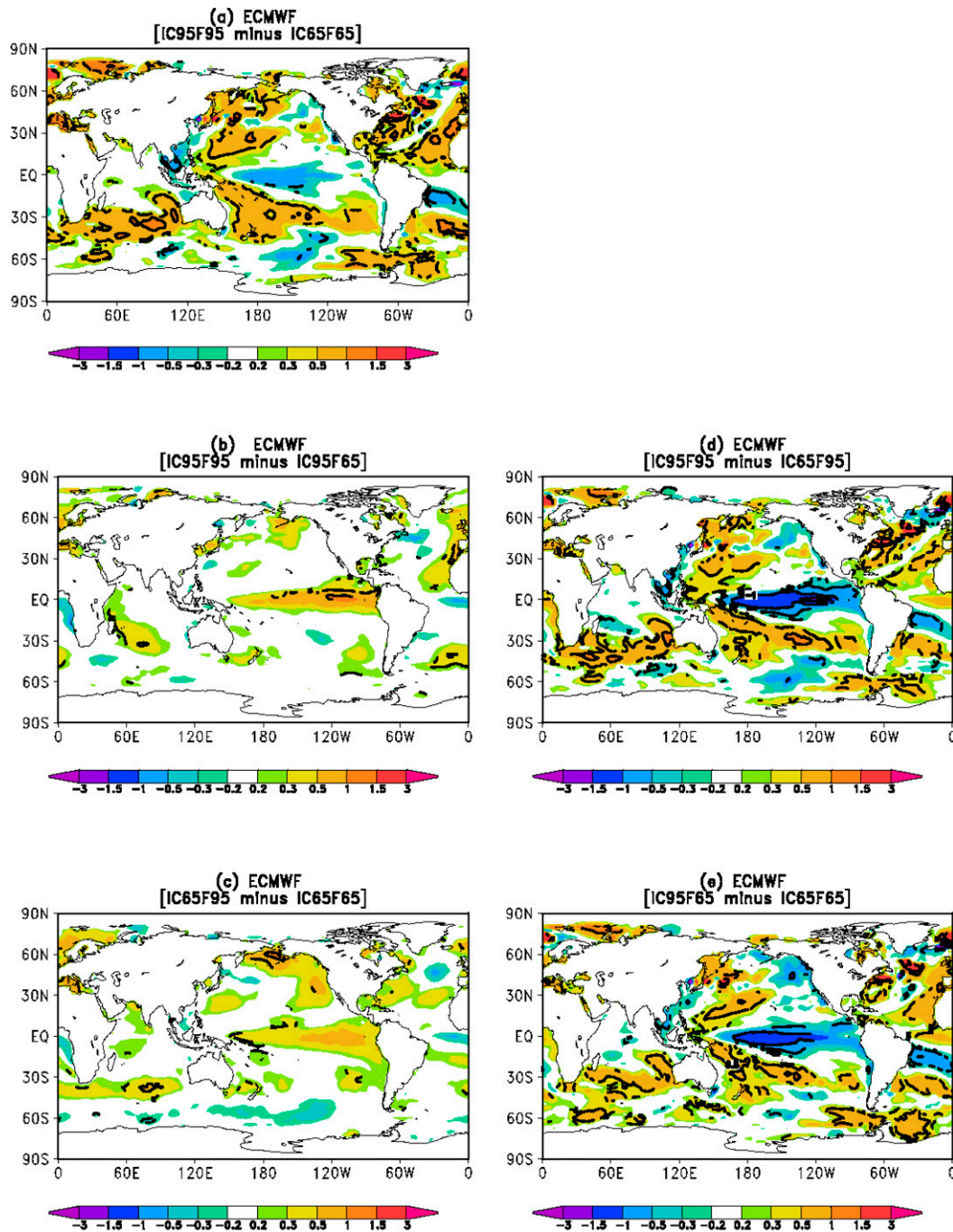
FIG. 3. As in Fig. 2a, but for the difference between the ensemble-mean SST anomalies (K) simulated by the ECMWF model. (a) Reference IC95F95 minus IC65F65. (b),(c) The impact of forcing is shown. (d),(e) The impact of initialization is shown. Shown are (b) IC95F95 minus IC95F65, (c) IC65F95 minus IC65F65, (d) IC95F95 minus IC65F95, and (e) IC95F65 minus IC65F65. Black contours show anomalies statistically significant at 5% level.

IC95F95 minus IC95F65 (Fig. 3b) and IC65F95 minus IC65F65 (Fig. 3c). They show the impact of the forcing. The panels in the right column show the impact of initial conditions [IC95F95 minus IC65F95 (Fig. 3d) and IC95F65 minus IC65F65 Fig. 3e)].

The ECMWF model is able to reproduce the observed anomalies reasonably well in the reference simulation, that is, when both correct initial conditions and forcings are applied (cf. Fig. 2a and Fig. 3a). When only the correct initial conditions are applied (and the forcing

from the other decade is used), the performance of the model compares well with the observations (and the reference simulation). The magnitude of the cold anomaly in the equatorial Pacific is even better simulated in both IC95F95 minus IC65F95 and IC95F65 minus IC65F65 than in the reference. The impact of the forcing seems very modest in the first year and produces a wrong warm signal on the equatorial Pacific (which is also common to the EC-EARTH and HadCM3 models, while the impact of the forcing during the first year of integration is negligible in the MPI-M model). A Student's *t* test was applied to check whether the anomalies in the simulations are significantly different from noise. It was found that the cold signals on the equatorial Pacific shown in Figs. 3d and 3e are statistically significant at 5% level, while the same signal in Fig. 3a is marginally significant. On the other hand, the warm anomalies in the same region (Figs. 3b,c) are statistically significant in Fig. 3b (i.e., IC95F95 minus IC95F65) but cannot be distinguished from noise in Fig. 3c (i.e., IC65F95 minus IC65F65). This seems to indicate that the ECMWF model tends to overestimate the warming effect of climate forcing over the equatorial Pacific. The performance of the other models (not shown) in the first year of integration compares well with the ECMWF model. When the correct initial conditions are applied, all the models are skillful in reproducing the observed anomalies.

Thus, it seems reasonable to conclude that, during the first year, all the models considered are far more sensitive to the initial conditions than to the application of the correct boundary/external forcing. In other words, 1-yr predictions seem to be essentially initial-value problems. This is consistent with previous results (e.g., Matei et al. 2012). Skill in predicting the sea surface temperatures for year one comes mostly from persistence; however, the ENSO high predictability during the first six months of integrations (Stockdale et al. 2011; Magnusson et al. 2013) may also play a role.

The difference between the average SST fields in the first 5 years of integrations for the ECMWF, EC-EARTH, MPI-M, and HadCM3 models can be found in Fig. 4. For each model, we show the difference between the ensemble mean of the two reference integrations (i.e., IC95F95 minus IC65F65), the impact of the forcing (for IC95F95 minus IC95F65), and that of initialization (for IC95F95 minus IC65F95). The corresponding maps for IC65F95 minus IC65F65 and IC95F95 minus IC65F65 are consistent with those shown in Fig. 4.

All the models show a predominant influence of the forcing compared to the impact of the initial conditions over most of the oceanic basins. The tropical Indian

Ocean stands out as the region with the largest impact of the forcing, in agreement with previous studies (see, e.g., Deser et al. 2010; Boer 2011; Guemas et al. 2013). However, in some regions, such as the North Atlantic subpolar gyre, the North Pacific subpolar gyre, and the Southern Ocean, the initialization signal is still pronounced in all models. The role played by the information contained in the initial climate state for the predictions of sea surface temperatures over the North Pacific and North Atlantic Oceans has been found in many studies (see, e.g., Keenlyside et al. 2008; Pohlmann et al. 2009; Smith et al. 2010; Branstator et al. 2012; Chikamoto et al. 2013; Doblas-Reyes et al. 2013). In the reference integrations, EC-EARTH and ECMWF models show statistically significant warm anomalies over the whole North Atlantic region very similar to the observations (cf. Fig. 2b with Figs. 4a,b). The forcing seems to be the main driver of predictability in the tropical Atlantic (see Figs. 4e,f), while the initial conditions influence the subpolar gyre region (see Figs. 4i,l). Here, MPI-M and HadCM3 show cold (but not significant) anomalies. Another region where the effect of initialization is still detectable (especially in the EC-EARTH, ECMWF, and HadCM3 models) is the southern Indian Ocean, where significant warm anomalies persist even when the 10-yr average is considered (see Fig. 5), although there are pronounced differences in regional detail. This simulated warming extends to about 60°S in contrast with the observations, where the warming ends at around 45°S. South of 45°S, cooling is observed, which extends between 40° and 120°W. This negative anomaly is reproduced by the MPI-M model, both in the reference (Figs. 4c and 5c) and in the sensitivity to initial condition simulation (Figs. 4k and 5k). This long memory of initial conditions in the southern Indian Ocean (with some extension to the southwestern Pacific and southern Atlantic) is consistent with the diagnostic potential predictability at decadal time scales found in several studies (Boer 2004; Pohlmann et al. 2004; among others).

Figure 5 shows the differences between the two decades over the whole period for the ECMWF, EC-EARTH, MPI-M, and HadCM3 models, respectively. It can be noticed that in all the models the fingerprint of initialization is still detectable over the North Atlantic and in different regions of the Southern Ocean. However, the four models produce anomalies of opposite sign and different extension. In the Atlantic subpolar gyre region, the EC-EARTH model shows a warm anomaly consistent with the reference simulations and observations. A reduced cold anomaly shifted southward, slightly amplified with respect to the reference integration, is shown in the MPI-M model in the same
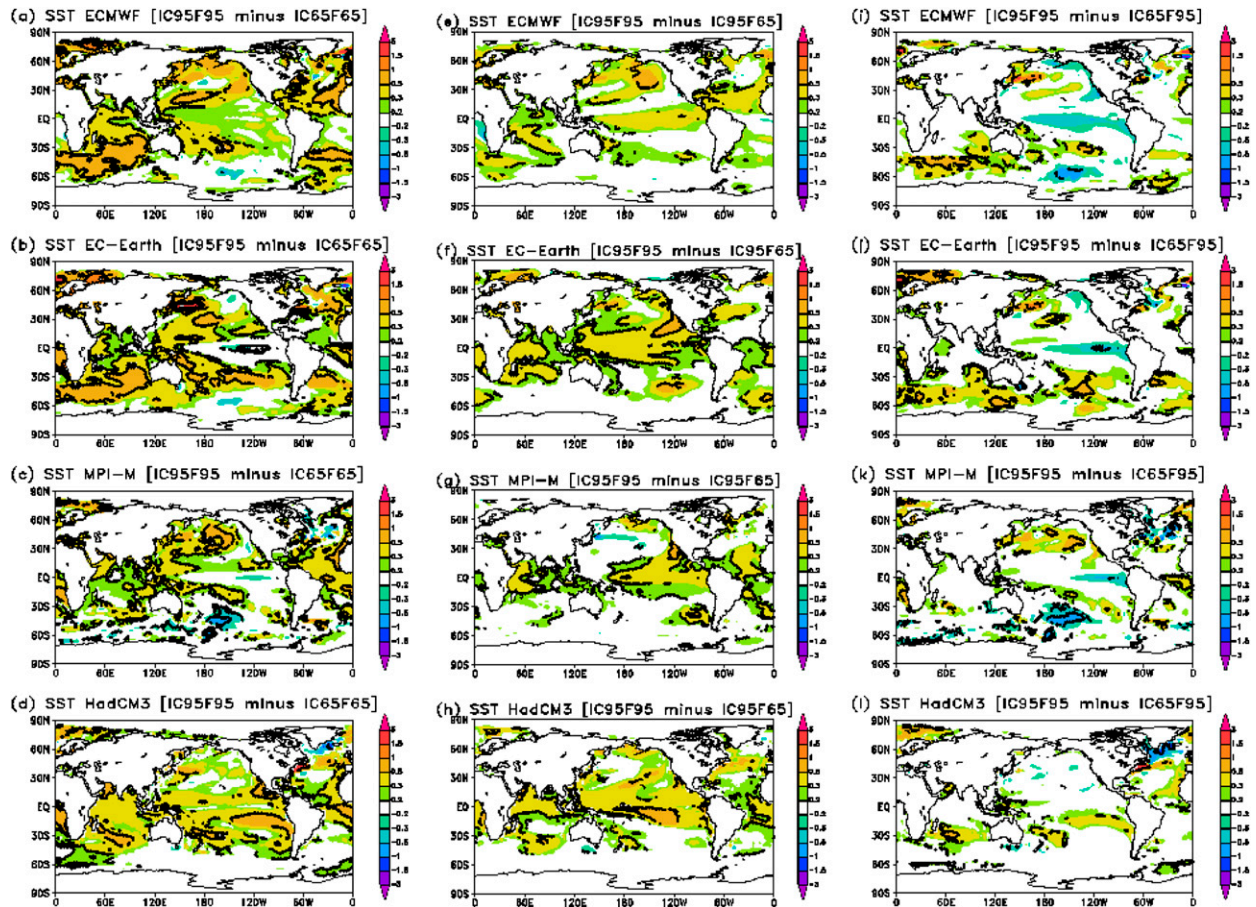
FIG. 4. As in Fig. 2b, but for the difference between the ensemble-mean SST anomalies (K) simulated by (a),(e),(i) ECMWF; (b),(f),(j) EC-EARTH; (c),(g),(k) MPI-M; and (d),(h),(l) HadCM3 models. (left) Reference IC95F95 minus IC65F65. (center) The impact of forcing IC95F95 minus IC95F65. (right) The impact of initialization IC95F95 minus IC65F95. Black contours show anomalies statistically significant at 5% level.

region. The ECMWF and HadCM3 models do not show significant anomalies over the North Atlantic (north of 50°N); however, they both display significant warm SST anomalies over the North Atlantic Current (Gulf Stream region) in agreement with observations and the reference simulation. The fact that only the EC-EARTH model shows a significant signal in the Atlantic subpolar gyre region might be explained by the different representation of the internal variability of SSTs in the different models (shown in Fig. SM1 of the supplementary material). In EC-EARTH, the internal variability of the North Atlantic, estimated from the spread of the ensemble members, is less pronounced than in the other models; therefore, a more coherent signal can emerge. The fingerprint of initial conditions is detectable in the EC-EARTH model also in the tropical North Atlantic (between 15° and 30°N), where a warm SST anomaly extends from the African coast to about 45°W.

A weaker (but still significant) signal of the memory of initial conditions is detectable in EC-EARTH and MPI-M models also in the western part of the North Pacific (ECMWF exhibits a similar feature). This signal seems not well reproduced in HadCM3. Again, the corresponding difference fields for IC65F95 minus IC65F65 and IC95F95 minus IC65F65 compare well with those shown in Fig. 5.

Overall, we can conclude from the comparison between the reference and sensitivity simulations that for SST, in all the models considered, the correct initialization has a strong impact up to about 1 yr on a global domain, but it continues to affect the predictability over specific regions in the North Atlantic over a decade. In the subpolar gyre region, only EC-EARTH shows a significant anomaly persisting over a decade. The long lasting memory of initial conditions over the North Atlantic is consistent with the results of recent analyses performed on the skill (and reliability) of decadal
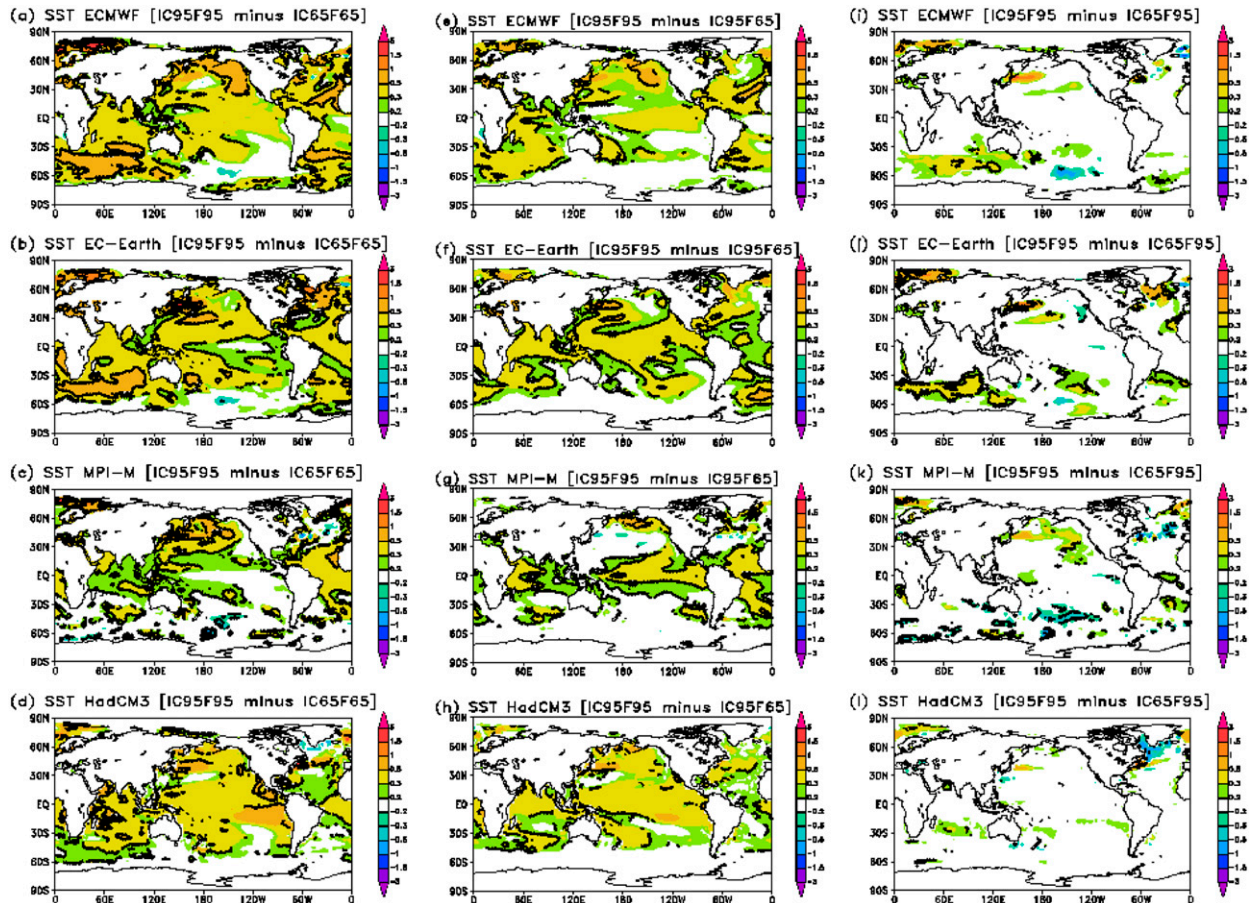
FIG. 5. As in Fig. 4, but for the 10-yr mean.

predictions (see, e.g., van Oldenborgh et al. 2012; Corti et al. 2012; Doblas-Reyes et al. 2013). Other regions also may have a long lasting sensitivity to initial conditions, such as the North Pacific, tropical Atlantic, and southern Indian Oceans. However, the signal over these latter ocean basins varies from model to model.

4. Simulation of the Atlantic meridional overturning circulation

The AMOC is responsible for approximately 90% of the northward ocean heat transport in the subtropical North Atlantic (Johns et al. 2011) and is believed to have a significant influence on the mean state and variability (Latif and Keenlyside 2011) of North Atlantic climate. In addition, modeling studies have shown that the AMOC is an important source of potential predictability (e.g., Collins et al. 2006). Since April 2004, the Rapid Climate Change Programme–U.S. Meridional Overturning Circulation and Heat Flux Array (MOCHA) (hereinafter referred to as RAPID) has made continuous observations of the strength and vertical structure

of the AMOC at 26.5°N (Cunningham et al. 2007), allowing investigation of AMOC variability on subseasonal and interannual time scales. During the RAPID observation period, the AMOC at 26.5°N had a strength of 17.5 ± 2.0 Sv (1 Sv $\equiv 10^6$ m^3 s^{-1}; mean $\pm$ standard deviation of annual means). However 11 yr of observations still represent a record too short to establish robust statistics on the interannual AMOC variability and are definitely insufficient to evaluate AMOC multiannual variations. Moreover, observations are available only at a single latitude. On the other hand, the historical reconstructions of the AMOC by ocean reanalysis products show a nonnegligible inconsistence in the estimation of both trends and year-to-year variability (Munoz et al. 2011).

In this context, considering that the SWAP experiments have been designed to study the relative importance of initial conditions and external forcing in driving decadal predictability, we do not compare the AMOC historical simulations to observations and/or reanalysis and therefore we do not assess the "realism" of such simulations. Our analysis is aimed at a comparison of the
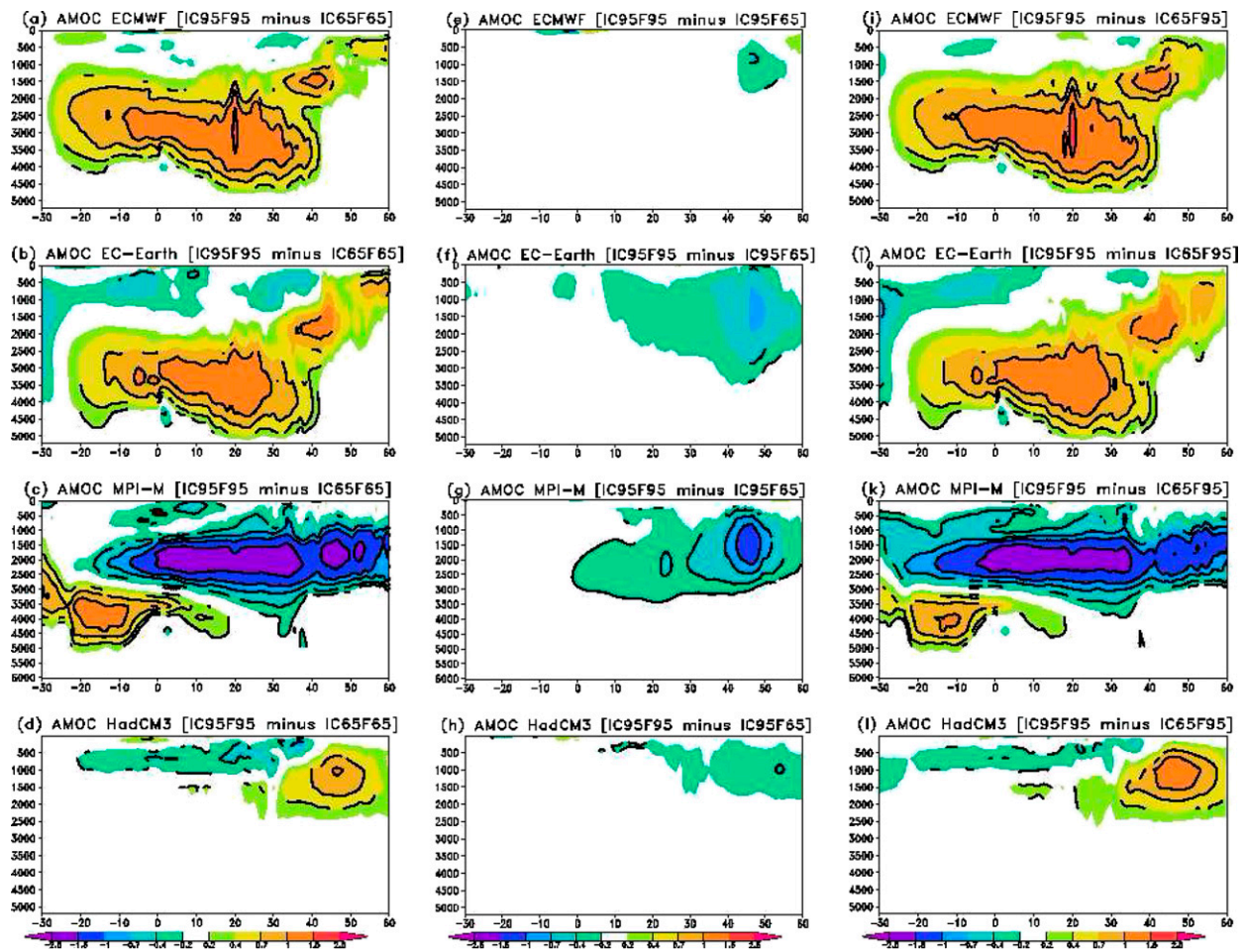
FIG. 6. As in Fig. 5, but for AMOC streamfunction (Sv).

AMOC simulations in the models' reference hindcasts, which is a prerequisite to evaluate the impact of forcing and initial conditions for each model in the sensitivity integrations.

The average Atlantic overturning streamfunctions, the associated year-to-year variability, and the average spread have been computed for all models from their reference simulations (see Fig. SM2 in the supplementary material). (It is worth stressing that the AMOC mean state and the associated variability shown in Figs. SM2a–i refer to the hindcast runs and, as such, do not represent an estimate of the AMOC amplitude and internal variations in the models.) The AMOC in the MPI-M model is considerably stronger than in the other models. EC-EARTH shows two well-pronounced maxima at about 26° and 40°N, while HadCM3 has a shallower maximum at about 45°N. The AMOC in the ECMWF model is very similar to that simulated with EC-EARTH, although weaker, and it does not present evident latitudinal peaks. AMOC year-to-year variability, estimated from the reference simulations, is

shown in Figs. SM2e–h. The different models manifest different depth, intensity, and meridional structure. However, a common feature is the maximum in the high latitudes, at about 45°N, near the AMOC mean state maximum, with an amplitude from 3 Sv (in ECMWF and EC-EARTH) to 1.6 Sv (in MPI-M and HadCM3). AMOC average spread, which gives an estimate of the AMOC internal variability, varies considerably from model to model (Figs. SM2i–l). The HadCM3 model shows the smallest spread with a maximum in the subtropical AMOC and very low internal variability in the North Atlantic. The other systems exhibit the maximum of internal variability between 30° and 50°N, which extends to 80°N in the MPI-M model.

The difference IC95F95 minus IC65F65 for the ensemble-mean average over each one of the 10-yr reference integrations is shown for all models in Fig. 6. The EC-EARTH and ECMWF models exhibit virtually the same anomaly below 2500 m. Closer to the surface, the AMOC anomaly in ECMWF, EC-EARTH, and HadCM3 models is a dipole consistent with a weakening

of the overturning streamfunction south of about 40°N and a strengthening at the northern latitudes in the later decade. In the MPI-M model, the AMOC negative anomaly extends up to 60°N. These results can be compared with the findings of Lozier et al. (2010, see their Fig. 4b). Only the MPI-M model can simulate the rather substantial deep weakening of the AMOC over the subtropical gyre. The other systems, on the other hand, do simulate the reported intensification of the overturning over the subpolar gyre. However, with the exception of the HadCM3 model, they miss the subtropical weakening (the negative anomalies in Figs. 6a,b are not significant).

The different models' representations of the AMOC change between the two decades are not surprising and are somehow expected. Several studies reported on the strong model dependency of AMOC predictability (e.g., Pohlmann et al. 2013; Ba et al. 2014). These variations in predictability may depend on the remarkable differences in the AMOC representation in the state-of-the-art ocean reanalysis products (Munoz et al. 2011; Kröger et al. 2012). Indeed, the plots shown in Fig. SM3 of the supplementary material support this interpretation. Here, the difference between the mean of the first month of integration in 1995 minus the mean of the first month of integration in 1965 for all the models is displayed. These patterns, which are very close to the difference between the correspondent ocean analyses, show no general coherence between systems initialized using different "best observed AMOC estimate." However, the observed model dependency of AMOC predictability may also be related to the different relationship between the AMOC and temperature and salinity in the subpolar gyre region in the four prediction systems considered (Hazeleger et al. 2013b).

Figures 6e–h and 6i–l show the sensitivity of AMOC to the forcing and initial conditions, respectively. The figure is constructed in the same way as Figs. 4 and 5; that is, the left panels show the difference between the ensemble-mean average of two reference integrations over the 10 yr of integration. The impact of the forcing is shown in the center column; the impact of the initialization is shown in the right column. It is evident that in this case, the influence of the initial conditions is far more important than that of the boundary forcing. It is also evident that the forcing contributes to the weakening of the streamfunction north of about 40°N.

This feature is common to all the systems; however, in ECMWF and HadCM3 models the response of the Atlantic streamfunction to the external forcing is noticeably weaker than in the other models. In the MPI-M model, on the contrary, the negative anomaly is particularly pronounced.

Overall, the AMOC predictability seems very sensitive to the initial conditions for the first 10-yr lead time over the North Atlantic subpolar gyre region. The forcing plays a role, even if minor, in weakening the circulation in the northern extratropical Atlantic.

5. Definition of the crossover time

In the previous analyses, temporally averaged fields were analyzed. They provide an estimate of the relative importance of boundary and initial conditions in reproducing the observed difference between the two decades. However, this analysis cannot provide any quantitative estimate of the relative weight of forcing and initial conditions during the forecast evolution.

To give a quantitative estimate of the potential predictability induced by initial conditions with respect to the forced signal as a function of lead time, for each variable x (averaged over a specific area or for each grid point) the crossover time $T_c(x)$ is defined as the time when information from boundary and initial conditions play an equal role in contributing to the forecast error with respect to the reference forecast. The crossover time is computed as follows:

If $X_A(t)$ (t = lead time) is a variable in reference experiment IC65F65 and $X_B(t)$ is the same variable in reference experiment IC95F95, then we can define

$$D_{AB}(t) = |X_A(t) - X_B(t)| \qquad (1)$$

as the evolving absolute distance between these two variables in experiments IC65F65 and IC95F95. In the same way, equivalent distances for sensitivity experiments IC65F95 and IC95F65 can be defined as

$$\begin{aligned} D_{AD}(t) &= |X_A(t) - X_D(t)| \quad \text{and} \\ D_{BC}(t) &= |X_B(t) - X_C(t)|, \end{aligned} \qquad (2)$$

and

$$\begin{aligned} D_{AC}(t) &= |X_A(t) - X_C(t)| \quad \text{and} \\ D_{BD}(t) &= |X_B(t) - X_D(t)|. \end{aligned} \qquad (3)$$

Expressions in (2) give the impact of initialization, while the impact of forcing is given in (3). By construction, both distances in (2) are expected to increase as a function of time as soon as the system gradually loses the memory of initial conditions. On the other hand, the distances in (3) will decrease as time evolves as the result of the constraint imprinted to the system by the radiative forcing. A schematic illustration of this is given in Fig. 7.

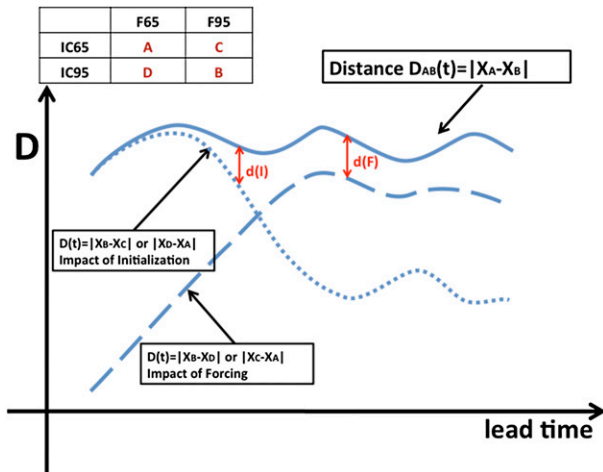Figure 7 also defined two new distances: $d_I(t)$ and $d_F(t)$. The distance $d_I(t)$ measures the departure of the

FIG. 7. Schematic representation of distances D defined in (1), (2), and (3).

FIG. 8. Schematic representation of distances d defined in (4) and (5) and crossover time.

initial conditions–sensitivity curve D_{AD} or D_{BC} from the reference D_{AB} as time evolves. The distance $d_F(t)$, on the other hand, measures the gap between the forcing–sensitivity curves D_{AC} and D_{BD} and the reference:

$$d_I(t) = \begin{cases} |D_{AB}(t) - D_{AD}(t)| \\ |D_{AB}(t) - D_{BC}(t)| \end{cases} \qquad (4)$$

and

$$d_F(t) = \begin{cases} |D_{AB}(t) - D_{AC}(t)| \\ |D_{AB}(t) - D_{BD}(t)| \end{cases} \qquad (5)$$

The distance $d_I(t)$ starts very close to the reference, and it diverges as the lead time increases. In contrast, $d_F(t)$ approaches the reference as time goes by. The distances $d_I(t)$ and $d_F(t)$ will eventually cross at $t = T_c$. The crossover time separates two regions (see the cartoon in Fig. 8 for a schematic illustration). On the left there is a region in which predictability is mostly associated with the knowledge of the initial conditions (initial-value predictability or predictability of the first kind). The region on the right is where predictability is mostly driven by changes in external forcing (boundary value predictability or predictability of the second kind). For times shorter than T_c, information from initial conditions is more important than information about the forcing. For times longer than T_c, the forcing becomes the major driver of predictability.

6. Crossover times for AMOC, HTC700, and SSTs

a. AMOC

In this subsection, the definitions given in section 5 are applied to the AMOC strength (i.e., meridional volume
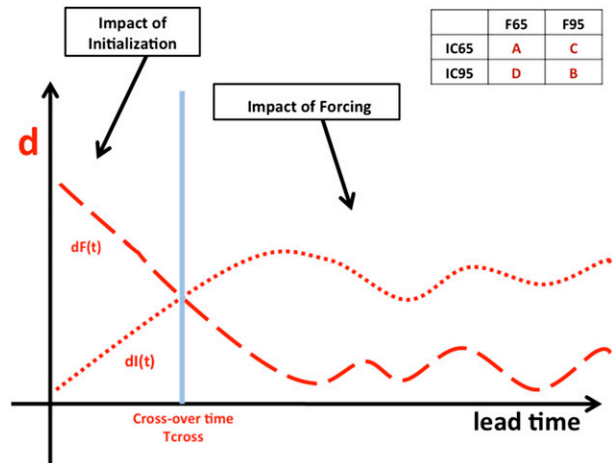
transport at 45°N and 1000-m depth in Sverdrups). Results for each model are shown in Fig. 9 for IC65F95 minus IC65F65 and IC95F65 minus IC65F65. A long crossover time (about 8 yr) is found for the ECMWF model. EC-EARTH and HadCM3 have a crossover time of about 6 yr, while it is of 4 yr for MPI-M. The uncertainty band (estimated from the spread of the ensemble members) is very large in all models except for HadCM3. If one considers this uncertainty, both the ECMWF and the EC-EARTH models have an almost 10-yr crossover time. This is consistent with the anomalies shown in Figs. 6e–h, where only the MPI-M model showed a significant weakening of the AMOC in the subpolar gyre region, with all the other systems relatively insensitive to the external forcing. Crossover times decrease when they are computed at 25°N (see Fig. SM4 in the supplementary material), suggesting that subtropical AMOC variations may have a different predictability horizon. It is interesting to note that the crossover times (and the other general features of the plots in Fig. 9) from the AMOC are similar when the same diagrams are computed for IC95F95 minus IC65F95 and IC95F95 minus IC95F65 (not shown), that is, considering differences with respect to reference periods 1965–75 or 1995–2005. This seems to indicate that the sensitivity to the forcing and/or to the initial conditions does not vary much when the initial conditions and/or the forcing are changed.

b. HTC700

Figure 10 shows the crossover time maps for the average heat content in the 0–700-m layer for IC95F95 minus IC65F95 and IC95F95 minus IC95F65. The corresponding maps for IC65F95 minus IC65F65 and IC95F95 minus IC65F65 are consistent with those shown in Fig. 10. We
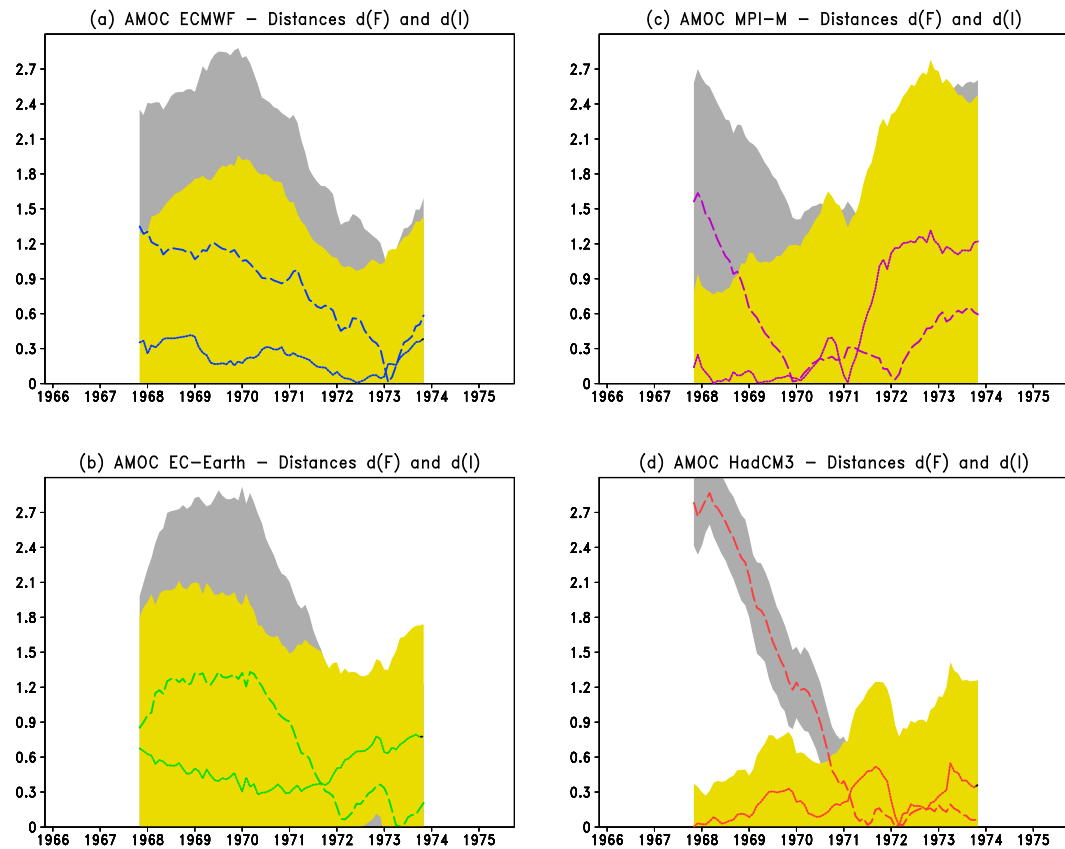
FIG. 9. Distances d_I for |IC95F65 − IC65F65| (solid) and d_F for |IC65F95 − IC65F65| (dashed) for AMOC at 1000-m depth and 45°N for all models (Sv). The gray (yellow) shading represents the uncertainty associated with d_F (d_I) estimated from the ensemble spread. A 4-yr running mean has been applied to d_F and d_I.

computed d_I and d_F at each grid point. Before computing the distances D and d, the time series of the heat content in the 0–700-m layer have been smoothed by applying a 2-yr running-mean time filter. The computations have been done for the ensemble mean for both reference and sensitivity runs. Therefore, the maps do not provide any estimation of the uncertainty.

Extended areas characterized by crossover times exceeding 8 yr are evident in all the models considered. The regions showing the longer crossing times are consistent with those already detected for the SST fields in the analysis described in section 3, although with some exceptions. The Arctic, limited areas in the Atlantic and Pacific Oceans, and large regions in the Southern Ocean stand out as the regions with the longest crossover time (about 9 yr). The long crossing time in the Arctic could be in part related to the long memory of the AMOC initial conditions, which could also affect the Arctic sea ice predictability (Mahajan et al. 2011; Msadek et al. 2014). However, the calculation of the Arctic sea ice extent crossing time (Fig. SM5 in the supplementary material) shows that on average the effect of

initialization is predominant for the first 2–3 yr (4 yr in EC-EARTH). The predominance of the initial condition ranges between 2 and 4 yr both in the North Atlantic basin and, to a lesser extent, in the North Pacific basin. However, the signal over these basins is not spatially homogeneous, and in some limited regions of the North Atlantic the crossover time is longer (almost 9 yr for a few grid points). This result is partially consistent with a recent analysis conducted by Branstator and Teng (2012) on long-term preindustrial control experiments produced within the CMIP5 protocol. These authors show that the impact of initial conditions over the North Atlantic becomes secondary (with respect to the forcing) after 8 yr for the average heat content in the 0–300-m layer. Our analysis shows an average crossover time of about 3 yr over the Atlantic subpolar gyre region, which might appear in contrast with the high skill of these models in predicting the upper-ocean heat content anomaly in this region at 6–9-yr lead time (Hazeleger et al. 2013b, their Fig. 3). However, it is worth pointing out that the results of Hazeleger et al. (2013b) are based on a multimodel ensemble mean over 30 ensemble
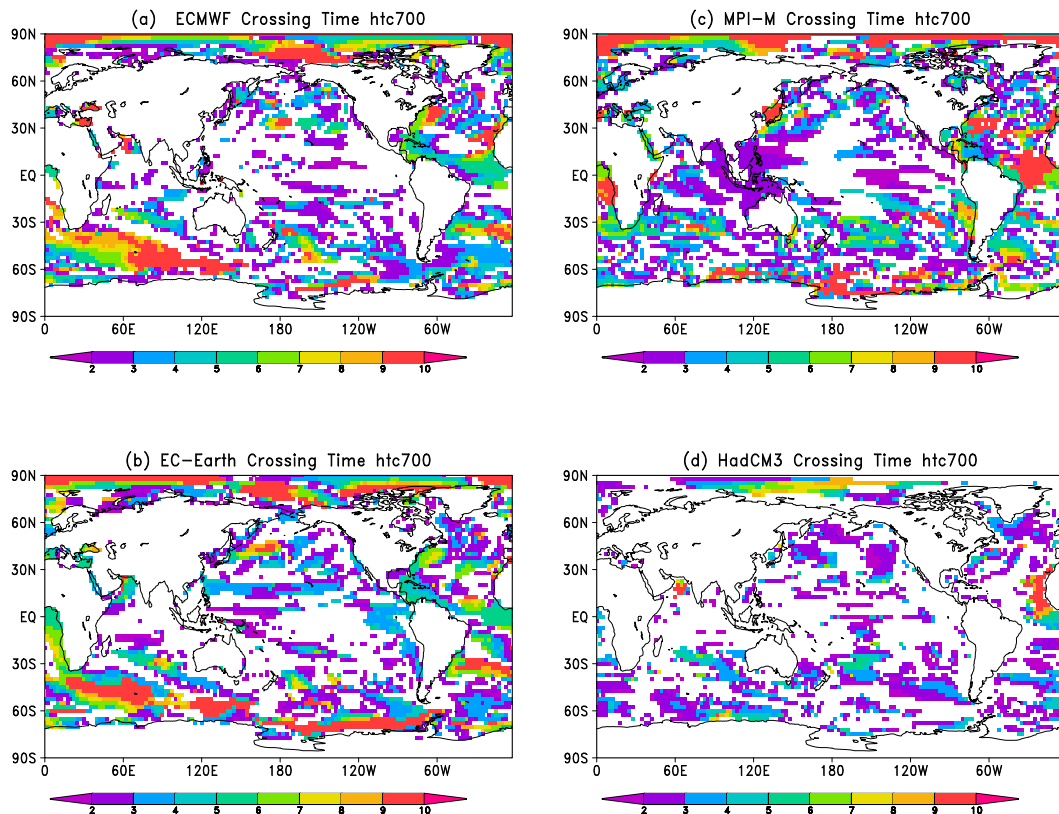
FIG. 10. Crossover times (yr) for d_I for $|IC95F95 - IC65F95|$ and d_F for $|IC95F95 - IC95F65|$ for the mean heat content at layer 0–700-m depth for all models. The crossing times are computed considering for each model the distances d_I and d_F of the ensemble means. The white regions correspond to crossover times shorter than 2 yr.

members, while crossover time estimates are computed here considering single-model ensemble means (over three or five ensemble members). Indeed, combining results from a multimodel ensemble of different prediction systems has been shown to improve the prediction skill (Palmer et al. 2004; Hagedorn et al. 2005).

We found a more prominent influence of the initial conditions in the southwestern and southeastern parts of North Atlantic, between 30° and 45°N. In particular, close to the U.S. East Coast, the crossover time range varies from about 3 (in HadCM3) up to 9 yr (in MPI-M). This last result seems consistent with the high forecast skill (correlation of 0.7 at 9-yr lead time) of the MPI-M model initialized with the ORAS3 reanalysis in predicting the average heat content in the 0–700-m layer over the North Atlantic (equator–60°N) (Kröger et al. 2012).

The map in Fig. 10 shows another region where the influence of initial conditions is more important than the forcing for an extended period. This is the tropical Atlantic, where the impact of initialization is larger than the one from the forcing for about 5 yr (and can reach 9 yr for the MPI-M model). This result is consistent with the findings of Matei et al. (2012). In this study, the skill

score maps of initialized predictions at 2–5-yr lead time display high correlation scores in the tropical Atlantic. The long memory of the heat content initial conditions in this region is apparent in Fig. 11, where the time evolution of d_I and d_F for the average heat content at 0–700 m in the tropical Atlantic (20°S–20°N) is shown. The impact of initial conditions is larger than that of boundary conditions for 10 yr (even considering the uncertainty band) for the MPI-M model and lasts up to about 4–5 yr in ECMWF and EC-EARTH. A similar signal cannot be found in HadCM3.

c. SSTs

The same crossover time maps have been produced for SSTs (shown in Fig. SM6 of the supplementary material). The patterns are broadly consistent with those found for the heat content, but the average crossover times are shorter. This shortening of the duration of information coming from initial conditions when looking at the sea surface temperatures might be expected. In fact, numerical experiments (e.g., Griffies and Bryan 1997b) support the notion that subsurface quantities are more predictable than SSTs. However, a priori one
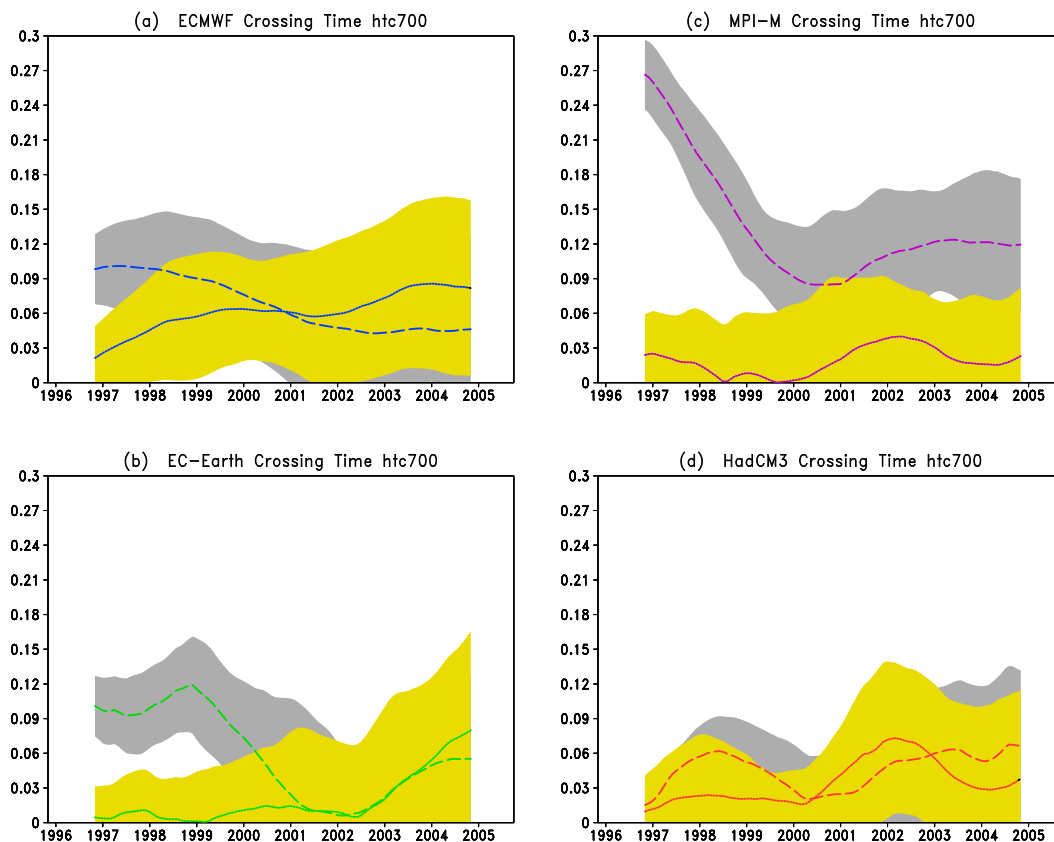
FIG. 11. Distances d_I for |IC95F95 − IC65F95| (dotted) and d_F for |IC95F95 − IC95F65| (dashed) for the heat content at layer 0–700-m depth in the tropical Atlantic (20°S–20°N) for all models (K). The gray (yellow) shading represents the uncertainty associated with d_F (d_I) estimated from the ensemble spread. A 2-yr running mean has been applied to d_F and d_I.

might not expect that the strong signal found over the tropical Atlantic for the heat content at 700 m almost disappears in the SSTs. (The same analysis conducted over the average heat content up to 400 m is consistent with the results found for the deeper layer.) A possible explanation of the lack of signal in sea surface temperature could be related to variations in either the subtropical cells and equatorial undercurrent or AMOC-related cross-equatorial flow (which partly takes place in the North Brazil Undercurrent; e.g., Hazeleger and Drijfhout 2006) that are more subsurface-related processes. In the case of the MPI-M model, this result is consistent with those reported by Matei et al. (2012). These authors found an enhanced predictability of the tropical Atlantic at a lead time of 2–5 yr for the upper-ocean heat content, but they did not find a corresponding signal for SSTs.

The larger area with crossing times longer than 2–4 yr for the sea surface temperatures is the southern Indian Ocean. Over this region (60°–30°S, 30°–120°E), even if the uncertainty is large, it is possible, at least for EC-EARTH, to estimate a lower limit for the crossover

time, which is about 2.5 yr (see Fig. 12). This signal is consistent with the results of van Oldenborgh et al. (2012) for the detrended anomaly correlation maps for 2–5-yr lead time in multimodel decadal hindcasts (see their Fig. 5c). A large and fairly homogeneous signal is found also over the western and tropical Pacific in the MPI-M model (Fig. SM5c), while in the North Atlantic the crossover time varies considerably over the region. When the d_I and d_F for sea surface temperature over the North Atlantic is computed (not shown), it is apparent that the high natural variability over the region induces large uncertainty, which makes estimates of the crossover time difficult.

7. Summary and discussion

Sensitivity experiments have been performed to isolate the impact of the initial conditions from the impact of external forcing in decadal predictions. They consist of two sets of 10-yr-long ensemble hindcasts carried out with four different Earth system models for two initial dates in 1965 and 1995 using either the external forcings
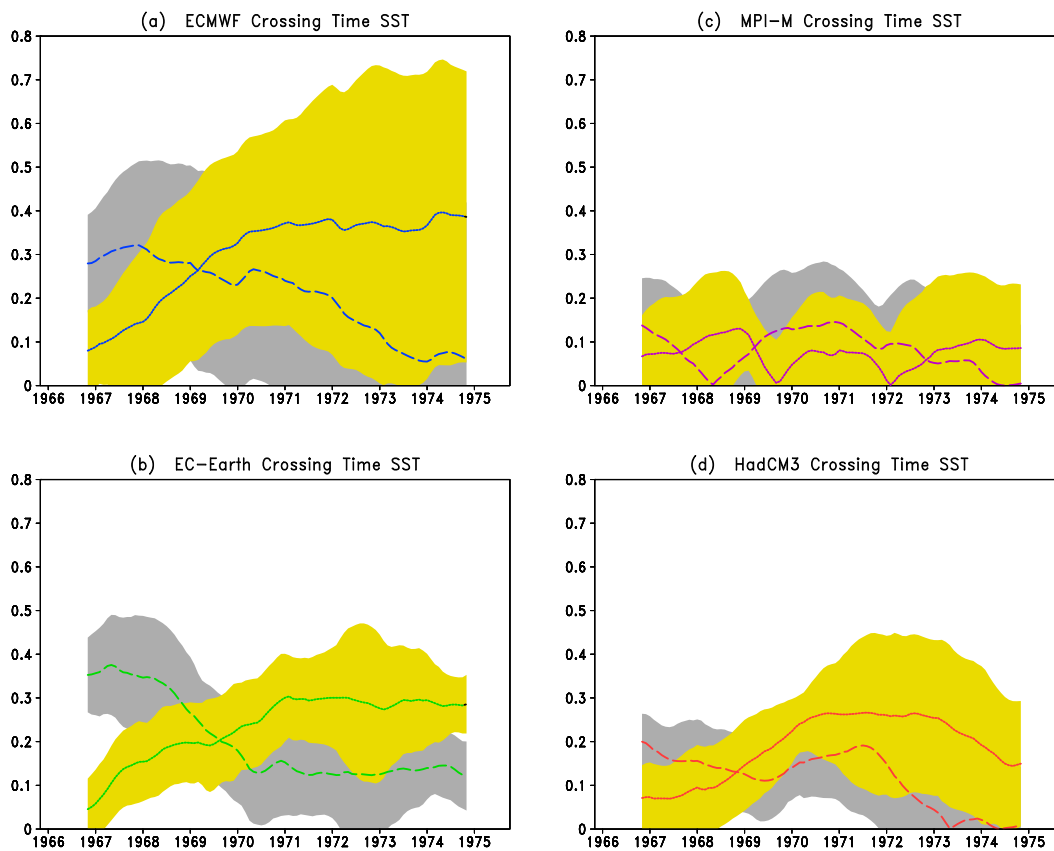
FIG. 12. Distances d_I for |IC95F65 − IC65F65| (solid) and d_F for |IC65F95 − IC65F65| (dashed) for SST in the southern Indian Ocean (60°–30°S, 30°–120°E) for all models (K). The gray (yellow) shading represents the uncertainty associated with d_F (d_I) estimated from the ensemble spread. A 2-yr running mean has been applied to d_F and d_I.

from the "correct" decades or swapping the forcings between the two decades (SWAP experiments). By comparing reference and sensitivity integrations, we estimate the impact of external forcing versus initial conditions on the predictability over multiannual time scales. To give a quantitative measure of the relative contributions, a crossover time is defined as the forecast time range when information from initial conditions and forcings are equal. We focus on the impact of initial conditions versus external forcing in terms of predictability of sea surface (and subsurface) temperatures and the AMOC.

Results from the sensitivity experiment indicate that, over time scales longer than about 1 yr, predictability of SSTs on a global domain arises mainly from the forcing. However, the correct initialization seems to have a longer impact on SST predictability over selected regions around the extensions of the western boundary currents and the Southern Ocean that are likely to be related to changes in the gyre circulation (Figs. 4, 5). There is also a consistent warming signal in the Arctic that persists for the first 5 years in all the models. These results are consistent with the findings of several other

studies (e.g., Boer 2004; Pohlmann et al. 2009; Smith et al. 2010; Branstator et al. 2012).

The impact of initialization is longer and extends to wider regions when subsurface ocean variables are considered. For example, over the southwestern part of the North Atlantic close to the North American coast line, the impact of initialization for the 700-m heat content extends to as much as 9 years for the MPI-M model (this model shows a long memory of initial conditions also in the tropical Atlantic). In all models considered, the impact of initial conditions on the predictability of the AMOC is dominant for the first 5 years. In agreement with previous studies (e.g., Teng et al. 2011), in some models the influence of initial conditions is apparent even up to the 9-yr lead time.

All the models show a similar response to forcing: weakening of the AMOC (Fig. 6) and warming of the tropical oceans (Fig. 5). However, in spite of the consistent response to forcing and to initial conditions, the degree of predictability is model dependent, possibly because the underlying dynamics varies from model to model. Thus, the model with a strong AMOC (MPI-M),

which also shows a strong weakening due to the initial conditions, has a pronounced first kind predictability in the tropical Atlantic 700-m heat content (Fig. 10), consistent with a decrease in the meridional heat transport.

The results shown here have to be tempered by the fact that the band of uncertainty associated with the natural variability is wide. Moreover, the question remains how much these results depend on the use of the selected two starting dates for the SWAP experiments. In particular, one might need a larger sample to filter out the short-term internal variability in either the initial conditions and/or the forcing. Indeed, in order to explore systematically the importance of initial conditions with respect to external forcing in multiannual predictions, the experimental approach described in this paper could be very helpful, when applied systematically to a larger sample of starting dates. The experiments and results described in this paper can be seen as a prototype on which further future testing with more models, larger sample sizes, and different initial dates could be based on.

REFERENCES

Ba, J., and Coauthors, 2014: A multi-model comparison of Atlantic multidecadal variability. *Climate Dyn.,* **43,** 2333–2348, doi:10.1007/s00382-014-2056-1.

Balmaseda, M. A., A. Vidard, and D. L. T. Anderson, 2008: The ECMWF ORA-S3 ocean analysis system. *Mon. Wea. Rev.,* **136,** 3018–3034, doi:10.1175/2008MWR2433.1.

——, K. Mogensen, and A. Weaver, 2012, Evaluation of the ECMWF ocean reanalysis ORAS4. *Quart. J. Roy. Meteor. Soc.,* **139,** 1132–1161, doi:10.1002/qj.2063.

Boer, G. J., 2004: Long time-scale potential predictability in an ensemble of coupled climate models. *Climate Dyn.,* **23,** 29–44, doi:10.1007/s00382-004-0419-8.

——, 2011: Decadal potential predictability of twenty-first century climate. *Climate Dyn.,* **36,** 1119–1133, doi:10.1007/s00382-010-0747-9.

Branstator, G., and H. Teng, 2010: Two limits of initial-value decadal predictability in a CGCM. *J. Climate,* **23,** 6292–6311, doi:10.1175/2010JCLI3678.1.

——, and ——, 2012: Potential impact of initialization on decadal predictions as assessed for CMIP5. *Geophys. Res. Lett.,* **39,** L12703, doi:10.1029/2012GL051974.

——, ——, G. A. Meehl, 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, doi:10.1175/JCLI-D-11-00227.1.

Brodeau, L., B. Barnier, A.-M. Treguier, T. Penduff, and S. Gulev, 2010: An ERA40-based atmospheric forcing for global ocean circulation models. *Ocean Modell.,* **31,** 88–104, doi:10.1016/j.ocemod.2009.10.005.

Chikamoto, Y., and Coauthors, 2013: An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC. *Climate Dyn.,* **40,** 1201–1222, doi:10.1007/s00382-012-1351-y.

Collins, M., 2002: Climate predictability on interannual to decadal time scales: The initial value problem. *Climate Dyn.,* **19,** 671–692, doi:10.1007/s00382-002-0254-8.

——, and B. Sinha, 2003: Predictability of decadal variations in the thermohaline circulation and climate. *Geophys. Res. Lett.,* **30,** 1306, doi:10.1029/2002GL016504.

——, and Coauthors, 2006: Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study. *J. Climate,* **19,** 1195–1203, doi:10.1175/JCLI3654.1.

Corti, S., A. Weisheimer, T. N. Palmer, F. J. Doblas-Reyes, and L. Magnusson, 2012: Reliability of decadal predictions. *Geophys. Res. Lett.,* **39,** L21712, doi:10.1029/2012GL053354.

Cunningham, S., and Coauthors, 2007: Temporal variability of the Atlantic meridional overturning circulation at 26.5°N. *Science,* **317,** 935–938, doi:10.1126/science.1141304.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.,* **137,** 553–597, doi:10.1002/qj.828.

Deser, C., A. Phillips, and M. A. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.,* **37,** L10701, doi:10.1029/2010GL043321.

Doblas-Reyes, F. J., and Coauthors, 2013: Initialized near-term regional climate change prediction. *Nat. Commun.,* **4,** 1715, doi:10.1038/ncomms2704.

Dunstone, N. J., and D. M. Smith, 2010: Impact of atmosphere and sub-surface ocean data on decadal climate prediction. *Geophys. Res. Lett.,* **37,** L02709, doi:10.1029/2009GL041609.

Griffies, S. M., and K. Bryan, 1997a: Predictability of North Atlantic multidecadal climate variability. *Science,* **275,** 181–184, doi:10.1126/science.275.5297.181.

——, and ——, 1997b: A predictability study of simulated North Atlantic multidecadal variability. *Climate Dyn.,* **13,** 459–488, doi:10.1007/s003820050177.

Guemas, V., S. Corti, J. Garcìa-Serrano, F. Doblas-Reyes, M. Balmaseda, and L. Magnusson, 2013: The Indian Ocean: The region of highest skill worldwide in decadal climate prediction. *J. Climate,* **26,** 726–739, doi:10.1175/JCLI-D-12-00049.1.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus,* **57A,** 219–233, doi:10.1111/j.1600-0870.2005.00103.x.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.,* **90,** 1095–1107, doi:10.1175/2009BAMS2607.1.

Hazeleger, W., and S. Drijfhout, 2006: Subtropical cells and meridional overturning circulation pathways in the tropical Atlantic. *J. Geophys. Res.,* **111,** C03013, doi:10.1029/2005JC002942.

——, V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F. J. Doblas-Reyes, K. Wyser, and M. Caian, 2013a: Multiyear

climate predictions using two initialization strategies. *Geophys. Res. Lett.,* **40,** 1794–1798, doi:10.1002/grl.50355.

——, and Coauthors, 2013b: Predicting multiyear North Atlantic Ocean variability. *J. Geophys. Res. Oceans,* **118,** 1087–1098, doi:10.1002/jgrc.20117.

Johns, W., and Coauthors, 2011: Continuous array-based estimates of Atlantic Ocean heat transport at 26.5°N. *J. Climate,* **24,** 2429–2448, doi:10.1175/2010JCLI3997.1.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature,* **453,** 84–88, doi:10.1038/nature06921.

Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.,* **59,** 2057–2072, doi:10.1175/1520-0469(2002)059<2057:MDPUUR>2.0.CO;2.

Kröger, J., W. A. Müller, and J. S. von Storch, 2012: Impact of different ocean reanalyses on decadal climate prediction. *Climate Dyn.,* **39,** 795–810, doi:10.1007/s00382-012-1310-7.

Latif, M., and N. S. Keenlyside, 2011: A perspective on decadal climate variability and predictability. *Deep-Sea Res. II,* **58,** 1880–1894, doi:10.1016/j.dsr2.2010.10.066.

Lorenz, E. N., 1975: Climatic predictability. *The Physical Basis of Climate and Climate Modeling,* GARP Publication Series, Vol. 16, WMO, 132–136.

Lozier, M. S., V. Roussenov, M. S. C. Reed, and R. G. Williams, 2010: Opposing decadal changes for the North Atlantic meridional overturning circulation. *Nat. Geosci.,* **3,** 728–734, doi:10.1038/ngeo947.

Magnusson, L., M. Balmaseda, S. Corti, F. Molteni, and T. Stockdale, 2013: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dyn.,* **41,** 2393–2409, doi:10.1007/s00382-012-1599-2.

Mahajan, S., R. Zhang, and T. L. Delworth, 2011: Impact of the Atlantic meridional overturning circulation (AMOC) on Arctic surface air temperature and sea ice variability. *J. Climate,* **24,** 6573–6581, doi:10.1175/2011JCLI4002.1.

Matei, D., H. Pohlmann, J. Jungclaus, W. Müller, H. Haak, and J. Marotzke, 2012: Two tales of initializing decadal climate predictions experiments with the ECHAM5/MPI-OM model. *J. Climate,* **25,** 8502–8523, doi:10.1175/JCLI-D-11-00633.1.

Meehl, G. A., and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.,* **90,** 1467–1485, doi:10.1175/2009BAMS2778.1.

Msadek, R., K. W. Dixon, T. L. Delworth, and W. Hurlin, 2010: Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys. Res. Lett.,* **37,** L19608, doi:10.1029/2010GL044517.

——, and Coauthors, 2014: Predicting a decadal shift in North Atlantic climate variability using the GFDL forecast system. *J. Climate,* **27,** 6472–6496, doi:10.1175/JCLI-D-13-00476.1.

Munoz, E., B. Kirtman, and W. Weijer, 2011: Varied representation of the Atlantic meridional overturning across multidecadal ocean reanalyses. *Deep-Sea Res. II,* **58,** 1848–1857, doi:10.1016/j.dsr2.2010.10.064.

Palmer, T. N., and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal-To-Interannual Prediction (DEMETER). *Bull. Amer. Meteor. Soc.,* **85,** 853–872, doi:10.1175/BAMS-85-6-853.

Pohlmann, H., M. Botzet, M. Latif, A. Roesch, M. Wild, and P. Tschuck, 2004: Estimating the decadal predictability of a coupled AOGCM. *J. Climate,* **17,** 4463–4472, doi:10.1175/3209.1.

——, J. H. Jungclaus, A. Kohl, 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, doi:10.1175/2009JCLI2535.1.

——, D. M. Smith, M. A. Balmaseda, N. S. Keenlyside, S. Masina, D. Matei, W. A. Muller, and P. Rogel, 2013: Predictability of the midlatitude Atlantic meridional overturning circulation in a multimodel system. *Climate Dyn.,* **41,** 775–785, doi:10.1007/s00382-013-1663-6.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108,** 4407, doi:10.1029/2002JD002670.

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, doi:10.1029/2005JC003172.

——, 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.,* **3,** 846–849, doi:10.1038/ngeo1004.

——, ——, and H. Pohlmann, 2013: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dyn.,* **41,** 3325–3338, doi:10.1007/s00382-013-1683-2.

Stockdale, T. N., and Coauthors, 2011: ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Climate Dyn.,* **37,** 455–471, doi:10.1007/s00382-010-0947-3.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498, doi:10.1175/BAMS-D-11-00094.1.

Teng, H., and G. Branstator, 2011: Initial-value predictability of prominent modes of North Pacific subsurface temperature in a CGCM. *Climate Dyn.,* **36,** 1813–1834, doi:10.1007/s00382-010-0749-7.

——, ——, and G. A. Meehl, 2011: Predictability of the Atlantic overturning circulation and associated surface patterns in two CCSM3 climate change ensemble experiments. *J. Climate,* **24,** 6054–6076, doi:10.1175/2011JCLI4207.1.

Troccoli, A., and T. N. Palmer, 2007: Ensemble decadal predictions from analysed initial conditions. *Philos. Trans. Roy. Soc.,* **A365,** 2179–2191, doi:10.1098/rsta.2007.2079.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012, doi:10.1256/qj.04.176.

van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger, 2012: Decadal prediction skill in a multimodel ensemble. *Climate Dyn.,* **38,** 1263–1280, doi:10.1007/s00382-012-1313-4.

Wouters, B., W. Hazeleger, S. Drijfhout, G. J. van Oldenborgh, and V. Guemas, 2013: Multiyear predictability of the North Atlantic subpolar gyre. *Geophys. Res. Lett.,* **40,** 3080–3084, doi:10.1002/grl.50585.