



# Full-field initialized decadal predictions with the MPI earth system model: an initial shock in the North Atlantic

Jürgen Kröger<sup>1</sup> · Holger Pohlmann<sup>1</sup> · Frank Sienz<sup>1</sup> · Jochem Marotzke<sup>1</sup> · Johanna Baehr<sup>2</sup> · Armin Köhl<sup>2</sup> · Kameswarrao Modali<sup>1</sup> · Iuliia Polkova<sup>2</sup> · Detlef Stammer<sup>2</sup> · Freja S. E. Vamborg<sup>3</sup> · Wolfgang A. Müller<sup>1</sup>

Received: 7 March 2017 / Accepted: 1 December 2017  
© The Author(s) 2017. This article is an open access publication

## Abstract

Our decadal climate prediction system, which is based on the Max-Planck-Institute Earth System Model, is initialized from a coupled assimilation run that utilizes nudging to selected state parameters from reanalyses. We apply full-field nudging in the atmosphere and either full-field or anomaly nudging in the ocean. Full fields from two different ocean reanalyses are considered. This comparison of initialization strategies focuses on the North Atlantic Subpolar Gyre (SPG) region, where the transition from anomaly to full-field nudging reveals large differences in prediction skill for sea surface temperature and ocean heat content (OHC). We show that nudging of temperature and salinity in the ocean modifies OHC and also induces changes in mass and heat transports associated with the ocean flow. In the SPG region, the assimilated OHC signal resembles well OHC from observations, regardless of using full fields or anomalies. The resulting ocean transport, on the other hand, reveals considerable differences between full-field and anomaly nudging. In all assimilation runs, ocean heat transport together with net heat exchange at the surface does not correspond to OHC tendencies, the SPG heat budget is not closed. Discrepancies in the budget in the cases of full-field nudging exceed those in the case of anomaly nudging by a factor of 2–3. The nudging-induced changes in ocean transport continue to be present in the free running hindcasts for up to 5 years, a clear expression of memory in our coupled system. In hindcast mode, on annual to inter-annual scales, ocean heat transport is the dominant driver of SPG OHC. Thus, we ascribe a significant reduction in OHC prediction skill when using full-field instead of anomaly initialization to an initialization shock resulting from the poor initialization of the ocean flow.

## 1 Introduction

In decadal or near-term climate prediction, proper initialization with observations is important for improving forecast quality (e.g., Cox and Stephenson 2007; Hawkins and Sutton 2009). A common approach is to nudge the forecast system to a selection of state parameters from ocean and atmosphere reanalyses. In the ocean, typically temperature and salinity are chosen. In this assimilation step, for a long time, nudging to anomalies instead of full fields of the ocean estimates was favored to suppress drift and initial shocks in the forecasts

(e.g., Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009). However, the issue of drift may be tackled successfully by a posteriori bias correction (e.g., Magnusson et al. 2012) and the issue of initial shocks is not exclusive to initialization from full-fields. A variety of reasons such as application of independent, uncoupled atmosphere and ocean reanalyses or application of a reanalysis product of poor quality in the assimilation procedure can lead to an imbalanced initial state in the coupled prediction system that triggers initial shocks in the forecasts (e.g., Balmaseda and Anderson 2009; Mulholland et al. 2015; Pohlmann et al. 2017). In this paper, we investigate how changing our decadal prediction system from anomaly nudging to full-field nudging in the ocean influences our prediction skill for sea surface temperature (SST) and ocean heat content (OHC) in the northern North Atlantic. In addition, we investigate the impact on prediction skill when applying full-field nudging with a different reanalysis data set since a strong sensitivity of skill to the choice of the ocean reanalysis product was

✉ Jürgen Kröger  
juergen.kroeger@mpimet.mpg.de

<sup>1</sup> Max Planck Institute for Meteorology, Bundesstr. 53, 20146 Hamburg, Germany

<sup>2</sup> Institute of Oceanography, CEN, University Hamburg, Bundesstr. 53, 20146 Hamburg, Germany

<sup>3</sup> European Center for Medium-Range Weather Forecasts, Reading, UK

found in an older version of our prediction system (Kröger et al., 2012).

In the field of decadal climate prediction, the North Atlantic has been identified as a key region that reveals pronounced forecast skill for different parameters such as surface air temperatures or OHC (e.g., Pohlmann et al. 2009; van Oldenborgh et al. 2012; Müller et al. 2012). High forecast skill for both SST and OHC in the North Atlantic subpolar gyre (SPG) region arises in part from the correct initialization of the ocean flow (Matei et al. 2012; Robson et al. 2012a, b; Yeager et al. 2012; Robson et al. 2013, 2014; Msadek et al. 2014). Unfortunately, direct observations of oceanic flow are sparse, therefore estimates of initial flow conditions are not well constrained and vary considerably due to the following reasons: First of all, ocean state estimates or reanalyses (in the following we will use both terms interchangeably) do not provide a coherent picture of the ocean flow (e.g., Karspeck et al. 2015); on top of that, it is unlikely that their flow characteristics are well adopted by the prediction system in the assimilation step (Kröger et al. 2012). Furthermore, the choice of the assimilation technique, namely anomaly or full-field nudging, matters, as was demonstrated for the assimilated long-term mean Atlantic Meridional Overturning Circulation (AMOC) by Smith et al. (2013b). Regarding the choice of the ocean reanalysis product, a severe impact on re-forecasts of the AMOC was already identified in different climate prediction systems (e.g., Kröger et al. 2012; Huang et al. 2015).

In the German national project Mittelfristige Klimaprognosen (“mid-term climate forecasts”; MiKlip), a decadal climate prediction system has been established (Müller et al. 2012; Pohlmann et al. 2013; Marotzke et al. 2016). So far, three different generations have been realized: the baseline0, the baseline1, and the prototype system [for details see Marotzke et al. (2016) and references therein]. For all stages the Max-Planck-Institute Earth System Model (MPI-ESM, Giorgetta et al. 2013) was employed using different initialization strategies (see Sect. 2). Both of our baseline systems use anomaly initialization in the ocean and full-field initialization in the atmosphere. In the prototype system, we change to full-field initialization in the ocean, which has been discussed as an alternative to anomaly initialization in recent studies on decadal prediction, although with no clear general recommendation for one or the other (Magnusson et al. 2012; Smith et al. 2013a; Hazeleger et al. 2013; Polkova et al. 2014; Carrassi et al. 2014; Bellucci et al. 2015; Volpi et al. 2016). In addition to the ocean state estimate from ORAS4 (Balmaseda et al. 2013), which was already used in baseline1, another state estimate, GECCO2 (Köhl 2015), was introduced to calculate initial ocean fields in the prototype system.

So far, a systematic evaluation of the skill of the completed prototype system and a comparison to the baseline

systems is still pending, although there have been studies that have partly tackled this question: With the caveat of using only a subset of the currently available ensembles, Kruschke et al. (2016) found no significant differences between anomaly and full-field initialization with respect to Northern Hemisphere winter storms. In addition, the initialization from GECCO2 has been found to yield slightly better results as compared to ORAS4, especially over the Northeast Pacific (Kruschke et al. 2016). Also Thoma et al. (2015) and Brune et al. (2017) have considered baseline1 and a subset of the prototype system in their comparison studies but did not come up with a clear recommendation for either full-field or anomaly initialization. Our study, on the other hand, is based on the completed MiKlip baseline1 and prototype systems, which include 10-year-long multi-ensemble hindcasts, ten members in case of baseline1 and 30 members in case of prototype, performed every year from 1961 to 2014. The goal of our study is to better understand how details in the initialization impact on processes in the northern North Atlantic that govern the evolution of OHC in that region and its predictability.

In Sect. 2, we provide a brief introduction to the applied MiKlip systems, followed by a summary of the data base and metrics that we use in the analysis. In Sect. 3, we investigate the performance of the prototype system and compare results to its predecessor baseline1. The comparison to baseline1 elucidates the role of anomaly versus full-field initialization in the ocean, the prototype system itself delivers additional insight into the dependency on the ocean reanalysis product applied for the initialization. A heat budget analysis in the North Atlantic SPG region is presented, which points out how prediction of OHC depends on the ocean flow and explains severe differences in the performance of the MiKlip systems. A discussion and a summary follow in Sects. 4 and 5.

## 2 Data and methodology

We analyze the two most recent generations of the MiKlip decadal prediction system, the baseline1 and prototype systems. Since all prototype experiments are solely based on the low-resolution version of MPI-ESM (T63 with 47 levels in the atmosphere and nominally 1.5° horizontal resolution and 40 levels in the ocean; Giorgetta et al. 2013), we consider only the low-resolution contribution from the baseline1 system. In the following we introduce briefly the baseline1 and prototype systems; a detailed introduction into the MiKlip prediction system can be found in Marotzke et al. (2016) and references therein. Table 1 provides an overview of the experiments analyzed in this study.

In the baseline1 system, a new initialization is introduced, in which ocean information no longer stems from

**Table 1** Summary of experiments analyzed in this paper with names of experiments, corresponding MiKlip generations and details of initialization

Experiment	Generation	Nudging in the Ocean
ORAS4-ANOM	Baseline1	Temperature and salinity anomalies from ORAS4
ORAS4-FULL	Prototype	Full temperature and salinity fields from ORAS4
GECCO2-FULL	Prototype	Full temperature and salinity fields from GECCO2

Initialization is based on a coupled assimilation run with nudging in ocean and atmosphere. Nudging in the atmosphere always utilizes full fields from ERA-40 and ERA-Interim. All analyzed experiments are based on MPI-ESM in low resolution and encompass ten ensemble members. More details can be found in Sect. 2

an ocean-only run with MPI-ESM but from a state-of-the-art ocean reanalysis (ORAS4; Balmaseda et al. 2013). In addition, the atmosphere is initialized with state estimates from atmospheric reanalyses ERA-40 (Uppala et al. 2005) and ERA-Interim (Dee et al. 2011). An assimilation run with the coupled model is performed, which delivers the initial conditions for the hindcasts. In the baseline1 assimilation run, the model state is nudged towards 3-dimensional ocean temperature and salinity anomalies added to the model climatology and 3-dimensional full fields of temperature, vorticity and divergence, as well as surface pressure in the atmosphere. The restoring time scales are 10 days in the ocean and 2 days to 15 min in the atmosphere depending on the nudging variable. Initial conditions (in form of complete restart data sets) are then extracted from the assimilation run at the desired frequency. Ensemble hindcasts are realized by using the lagged-initialization method, where initial conditions from adjacent days are shifted to the respective start date. Ten ensemble members of 10-year-long hindcasts are performed starting every year at January 1st over the period 1961–2014.

In the prototype system, we apply full-field initialization in the ocean. Two different initialization fields are used, based on the ocean state estimates ORAS4 and GECCO2 (Köhl 2015). The atmospheric initialization is the same as in baseline1. The ensemble size is extended from 10 to 15 ensemble members for each set of hindcasts since larger ensembles lead to higher prediction reliability (Sienz et al. 2016). As in the baseline experiments, the lagged initialization method is used for the ensemble generation, again 10-year-long hindcasts are performed starting at the beginning of every year over the period 1961–2014. We refer to the investigated suites of assimilation runs and corresponding ensemble hindcasts as ORAS4-ANOM (baseline1), ORAS4-FULL (prototype-oras4) and GECCO2-FULL (prototype-gecco2). We analyze ensemble hindcasts with a common number of ten members (the first ten available members) to assure consistency in our comparisons. For the rest of the paper we refer to the ensemble hindcasts as hindcasts, when particular ensemble members are considered it will be noted. All diagnostics are based on annual or multi-annual means.

In our skill analyses, we employ observational estimates of SST from Hadley Centre Sea Ice and Sea Surface Temperature (HadISST, Rayner et al. 2003) and OHC in the upper 700 m from NODC (National Oceanographic Data Center, Levitus et al. 2012). For the rest of the paper we refer to the observational estimates as observations, nevertheless, it should be kept in mind that all data used in this study are gridded.

For investigating the time evolution of heat content in the northern North Atlantic and related heat flux processes, we define an OHC index as the spatial integral between 50°N and 60°N and between the North American and European continents (75°W–9°E). The longitudinal extent from continent to continent allows for directly linking changes in OHC to meridional transport processes in the ocean. We consider this index to be a good estimate of the OHC evolution in the SPG region. Our definition of the SPG region is very close to the definition used by Müller et al. (2015) (50°N–60°N, 60°W–10°W), and our findings regarding integrated OHC changes are virtually the same for both definitions (not shown).

Hindcast quality of the ensemble mean is assessed by calculating skill based on either anomaly correlation coefficients (ACCs) or mean squared errors (MSEs). ACCs and MSEs are defined with respect to the observations. Given a reference forecast, skill scores for our hindcasts can be calculated from ACCs and MSEs (Murphy 1988; Mudelsee 2010), leading to a correlation skill score (CSS):

$$\text{CSS} = (\text{ACC}_{\text{hindcast}} - \text{ACC}_{\text{reference}}) / (1 - \text{ACC}_{\text{reference}})$$

and a mean squared error skill score (MSESS):

$$\text{MSESS} = 1 - \text{MSE}_{\text{hindcast}} / \text{MSE}_{\text{reference}}$$

In addition, the quality of the ensembles in terms of reliability and sharpness is assessed by the continuous ranked probability score (CRPS). The CRPS is a measure of the integrated squared difference between the cumulative distribution function of the forecasts and the step function located at the observed value (e.g., Goddard et al. 2013). A skill score for CRPSs (CRPSS) is calculated in the same way as for MSEs. The skill assessments imply a lead time dependent bias adjustment (ICPO 2011).

Typically, a simple prediction such as climatology derived from the observations would serve as reference forecast, and then a skill score would represent the improvement in quality of one particular hindcast over the climatological forecast model. However, we are also interested in direct comparisons of two hindcasts; in that case, one of the two hindcasts is used as the reference forecast. In general, positive skill scores denote an improvement in quality, negative values denote a reduction. 95% significance levels for correlations and skill scores are calculated by applying a non-parametric statistical test. We choose a moving block bootstrap test with a block length of 5 years to account for time dependencies such as auto-correlation in our time series (e.g., Goddard et al. 2013).

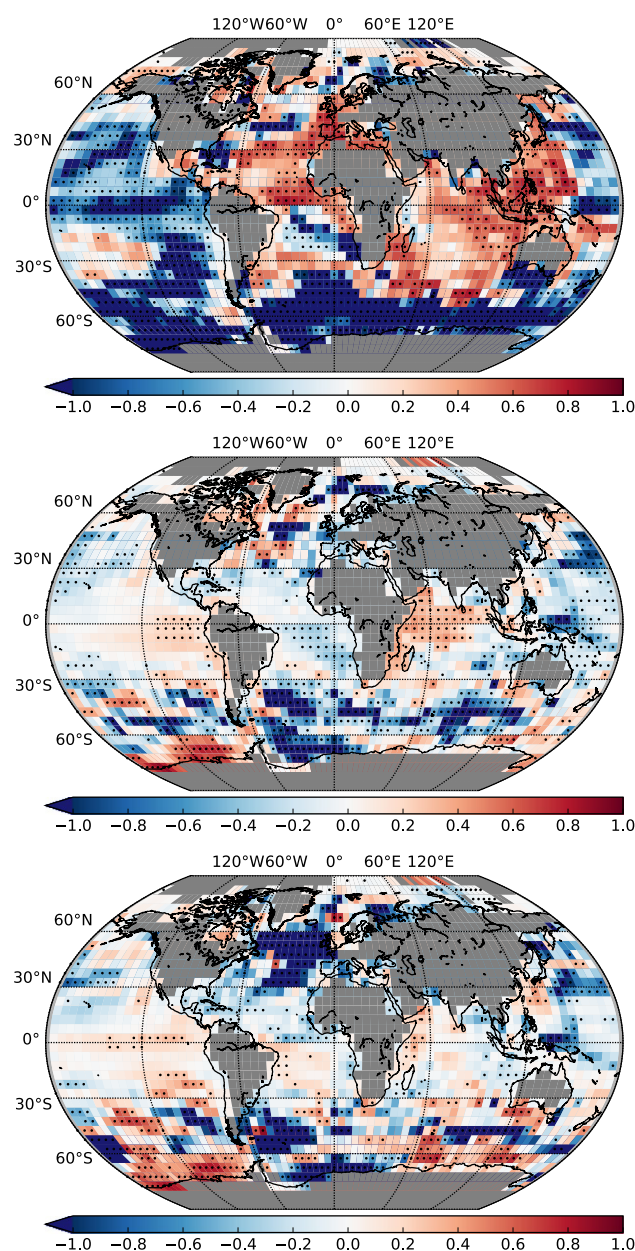
### 3 Results

We provide skill estimates for the two investigated generations of the MiKlip decadal prediction system based on skill scores of hindcasts of SST and North Atlantic OHC. In addition, we analyze heat flux processes that drive changes in OHC, what turns out to be crucial for the understanding of skill differences in the prediction systems.

#### 3.1 Sea surface temperature

We start with a comparison of the prediction quality of global SST (Fig. 1), using MESS of multi-year (lead years 2–5) ensemble-mean SST hindcasts. First, we compare prediction skill between ORAS4-ANOM and a simple statistical model. The forecast of the statistical model is the climatological state of the observations (climatology of SST from HadISST; Rayner et al. 2003). Second, we focus on possible skill improvement from anomaly to full-field initialization by directly comparing ORAS4-ANOM to both ORAS4-FULL and GECCO2-FULL. Here, ORAS4-ANOM hindcasts serve as reference forecast and the MESSs reveal the relative improvement or reduction in skill of each of the full-field systems over ORAS4-ANOM. By choosing MESS of SST, we complement the skill estimate provided by Marotzke et al. (2016), which is based on correlation skill (for lead years 2–5) for global surface air temperatures. Marotzke et al. (2016) reported skill improvement over uninitialized predictions in the North Atlantic in all MiKlip systems, but noted an overall lack of improvement from ORAS4-ANOM to the full-field system.

The ORAS4-ANOM hindcasts show increased skill (positive values) over the climatological forecast model in the Indian Ocean and in large parts of the Atlantic and West Pacific Oceans (Fig. 1, upper). Reduced skill (negative values) prevails in the rest of the Pacific Ocean and in the Southern Ocean. Note that the MESS based on SST agrees



**Fig. 1** Mean squared error skill scores (MSESS) of ensemble hindcasts of SST for start years 1960–2003 and lead years 2–5 in ORAS4-ANOM (upper), ORAS4-FULL (middle) and GECCO2-FULL (lower) computed with observations from HadISST; the reference forecast in the case of ORAS4-ANOM is the climatology of HadISST, in the cases of the full-field hindcasts the reference forecast is ORAS4-ANOM; dots denote skill different from zero exceeding the 95% confidence level; MSESSs are computed with the MiKlip central evaluation system (Illing et al. 2014)

well with the MESS based on surface air temperature (not shown). When compared to ORAS4-ANOM, both full-field hindcasts show distinctive regions of increased skill and of reduced skill. In the Southern Ocean, for both ORAS4-FULL and GECCO2-FULL a significant improvement can be identified in large parts of the Pacific sector, whereas

the opposite is the case in the Atlantic sector. In the Indian Ocean sector, hindcasts of Southern Ocean surface temperatures with ORAS4-FULL (Fig. 1, center) generally perform poorly compared to ORAS4-ANOM. Here, GECCO2-FULL (Fig. 1, lower), in contrast, reveals extended areas with significant improvement. In the Indian Ocean, ORAS4-FULL predictions have improved significantly for large parts, whereas changes from ORAS4-ANOM to GECCO2-FULL appear indifferent and mostly insignificant. In the Pacific Ocean, both full-field systems are characterized by a loss of predictive skill in the northern hemisphere away from the tropics and at the equator in the western part of the basin. In the eastern part of the basin, however, both systems show a moderate but still significant improvement over ORAS4-ANOM in predicting equatorial surface temperatures.

Overall, in contrast to the remarkably strong and consistent gain of skill in the entire tropical band from baseline0 to baseline1 (Pohlmann et al. 2013), low-latitude changes from baseline1 to both full-field systems are less pronounced and point in both directions. In the Atlantic Ocean, in particular north of 30°N, the two full-field systems differ considerably from each other. While changes in prediction skill in the northern North Atlantic appear to be predominantly positive for ORAS4-FULL, the skill significantly drops in case of GECCO2-FULL. In order to better understand the remarkable loss of prediction quality (in GECCO2-FULL) in a region that usually reveals pronounced decadal forecast skill, we now focus on the northern North Atlantic and investigate the sub-surface and deeper ocean. The ocean is the part of our prediction system where the assimilation strategies differ.

### 3.2 Ocean heat content in the North Atlantic

Based on an older version of the MPI climate prediction system, we previously identified a strong sensitivity of prediction skill for upper layer OHC in the North Atlantic to the choice of the ocean reanalysis product applied in the initialization (Kröger et al. 2012). The products used in that study included the direct predecessors of ORAS4 and GECCO2, respectively.

Here, we focus on OHC in the North Atlantic SPG region and look first at the evolution of the integrated heat content in that region depicted by the OHC index as defined in Sect. 2 (Fig. 2). When calculating the integrated heat content, in all experiments, we consider two different depth ranges: The presented OHC indices cover either the upper 700 m (Fig. 2, left column) or the entire water column (Fig. 2, right column). The full depths OHC indices are used in the budget analysis in Sect. 3.4. In all cases, the integrated OHC signal in the assimilation runs is very close to the signal in the respective original reanalyses which is a strong indication of the efficiency of the nudging ( $ACCs \geq 0.99$ ).

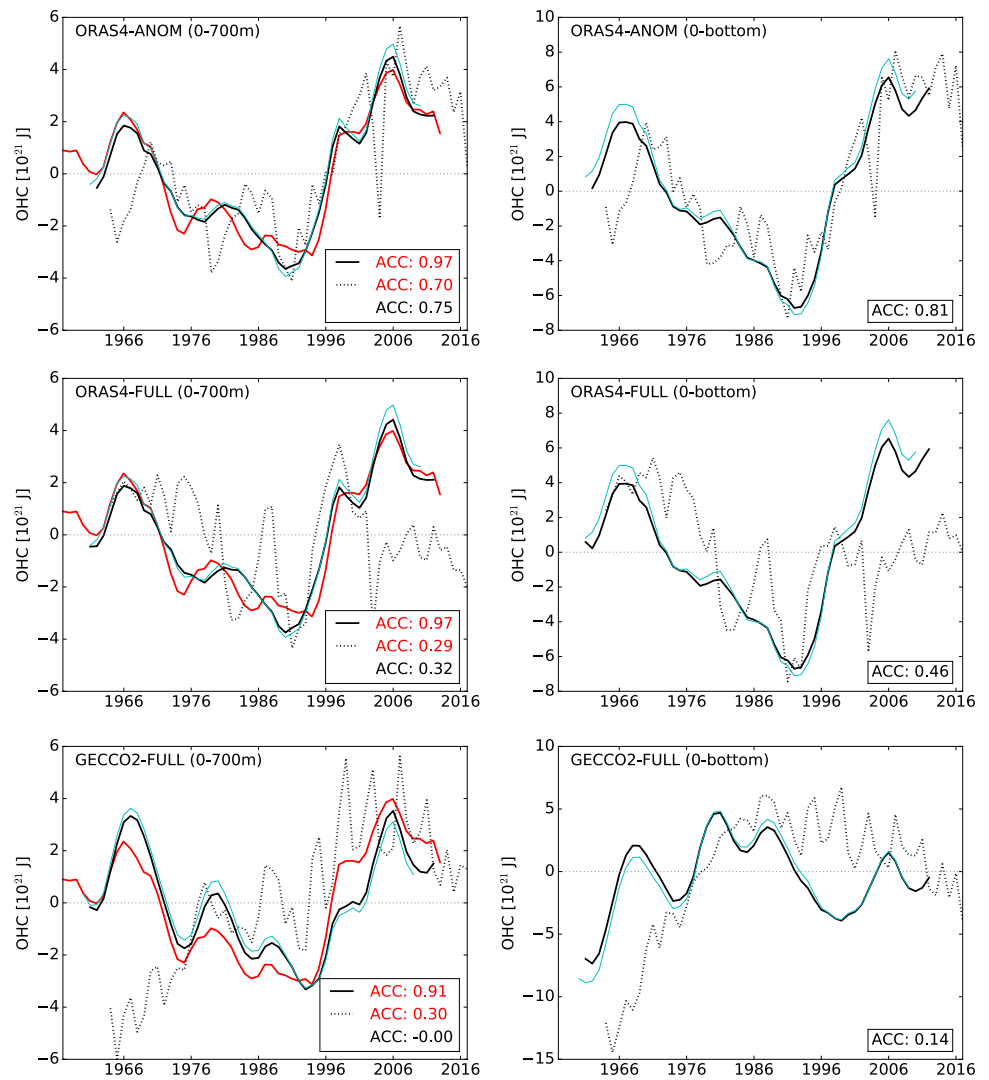
As a reference, all panels in the left column of Fig. 2 contain the observed OHC index in the upper 700 m from the NODC data set (Levitus et al. 2012).

In the upper 700 m, all assimilation runs reveal an evolution of the OHC index that is very close to the one from NODC, resulting in high correlations with the observations for all investigated MiKlip systems: ACCs are 0.97 for both ORAS4-based assimilation runs and 0.91 for GECCO2-FULL. When all depth levels are considered, the OHC indices in both ORAS4-based assimilation runs still reveal high coherence with the observations in the upper 700 m, whereas the interannual to decadal variability in GECCO2-FULL no longer resembles the upper layer NODC signal. However, such a comparison (between different depth levels) does not provide further useful information, in particular, estimates of the authenticity of the full-depth signals can not be inferred. Comparing (same depth) OHC indices in the assimilation runs alone, on the other hand, does provide useful information: First, the full-depth OHC evolution in the SPG region differs substantially between GECCO2-FULL and the ORAS4-based assimilation runs, and, second, the transient behavior of the assimilated OHC signals in ORAS4-FULL and ORAS4-ANOM is virtually the same, regardless of considering the upper 700 m or the entire water column.

A first estimate of prediction skill (for lead years 2–5) is provided by correlating the hindcasted OHC signals in the upper 700 m with the one from NODC (Fig. 2, left column). High correlation can only be found in case of ORAS4-ANOM with  $ACC = 0.7$ , whereas ORAS4-FULL and GECCO2-FULL reveal only moderate correlations with  $ACC = 0.29$  and  $0.3$ , respectively. The considerable rapid warming in the 1990s, which was successfully predicted in the case studies of Robson et al. (2012b), Yeager et al. (2012), and Msadek et al. (2014), is best predicted by the ORAS4-ANOM hindcasts.

Correlating the hindcasts with the corresponding assimilation runs instead of the NODC observations should provide a better estimate of each system's upper limit of predictability, since we compare our hindcasts to an observational estimate that is based on the same coupled model as our hindcasts and that is used for initializing the retrospective forecasts. Correlations with the assimilation runs are higher than those with NODC only in ORAS4-ANOM and ORAS4-FULL ( $ACC = 0.75$  and  $0.32$ , as compared to  $0.7$  and  $0.29$ , respectively). In GECCO2-FULL, in contrast, the correlation disappears when the assimilation run is considered ( $ACC \approx 0$ ). The same calculation for OHC integrated over the entire water column (Fig. 2, right column) leads to slightly higher ACC values of  $0.81$ ,  $0.46$  and  $0.14$  for ORAS4-ANOM, ORAS4-FULL, and GECCO2-FULL, respectively. Possible reasons are the increased size of the volume over which the OHC signal is integrated, which

**Fig. 2** Time series of OHC in the SPG region ( $50^{\circ}\text{N}$ – $60^{\circ}\text{N}$ ,  $75^{\circ}\text{W}$ – $9^{\circ}\text{E}$ ); simulated OHC is integrated either from 0–700 m (left column) or over the entire water column (right column); each panel contains one system's ensemble hindcast (lead years 2–5, black dotted line), the corresponding assimilation run (black line) and reanalysis (cyan line), plus, in the left column, observations from NODC (0–700 m, red line); assimilation runs, reanalyses and observations are 4-year means; ACCs of simulated OHC w.r.t. NODC (red numbers) and of hindcasts w.r.t. the corresponding assimilation runs (black numbers) are included in the panels



would imply less high frequency variability, or the fact that the deeper ocean reveals a higher level of predictability, or both.

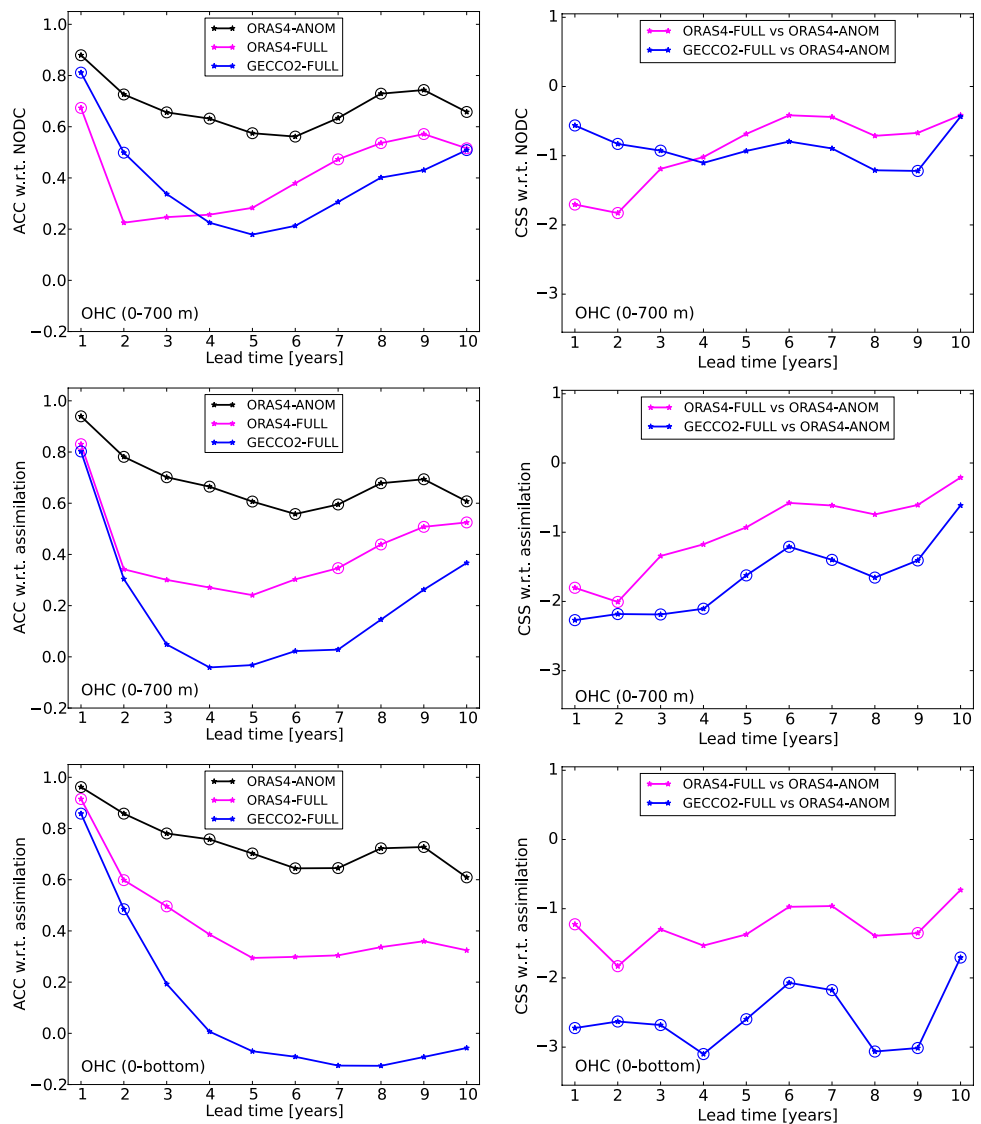
The drop in correlation skill for upper layer OHC in GECCO2-FULL when the assimilation run is considered instead of observations, may not be significant, but the counterintuitive behavior already suggests that we need to take a closer look at the assimilation run that is used for the initialization. As for all assimilation runs, the long-term OHC signal in the GECCO2-FULL assimilation run agrees well with the observations, both show a downward trend until the mid-1990s and an upward trend thereafter; the initialization of OHC seems realistic. The GECCO2-FULL hindcasts, in contrast, reveal an upward trend over the entire hindcast period, which suggests that correct initialization of OHC is not sufficient for its prediction, at least for lead years 2–5. Correct initialization of heat flux into the SPG region, on the other hand, may play a leading role for OHC prediction. The

role of the involved heat flux processes will be investigated in Sect. 3.4.

Next, we look at lead time dependent correlations (ACCs; Fig. 3, left column) and corresponding correlation skill scores (CSSs; Fig. 3, right column) of OHC in the SPG region. Again, we calculate correlations with respect to either observations (NODC; Fig. 3, top row) or the respective assimilation run for the upper 700 m (Fig. 3, center row) and the entire water column (Fig. 3, bottom row). The CSSs provide direct estimates of improvement (or deterioration) of both full-field systems over ORAS4-ANOM since we treat ORAS4-ANOM as reference forecast.

Regardless of whether the NODC data or the respective assimilation run is applied as observational estimate, ORAS4-ANOM hindcasts perform best for all lead times; all correlations are relatively high ( $\text{ACCs} > 0.5$ ) and significant. Both GECCO2-FULL and ORAS4-FULL, in contrast, reveal a strong drop in correlations in the first 2 years, which indicates an initialization shock in these hindcasts.

**Fig. 3** ACCs (left column) and CSSs (right column) of hindcasts of OHC in the SPG region (50°N–60°N, 75°W–9°E) as function of lead time; correlations were calculated with respect to the observations from NODC (0–700 m, top row) and to the respective assimilation run integrated over the upper 700 m (middle row) and the entire water column (bottom row); Negative CSSs indicate an advantage of ORAS4-ANOM over ORAS4-FULL and GECCO2-FULL; circles in all panels indicate significance at 95%



GECCO2-FULL performs worst for all lead times when correlated with its own assimilation run. The performance of ORAS4-FULL generally lies in between the two other systems, with the exception of lead years 1–3 when considering the correlation with NODC. Here, ORAS4-FULL is outperformed by GECCO2-FULL, but only the first 2 of the 3 lead years are significant.

While ORAS4-ANOM reveals significant correlations continuously for 10 lead years, in both full-field systems correlation skill re-emerges after several lead years of no skill: in ORAS4-FULL skill re-emerges for lead years 7–10 and in GECCO2-FULL skill re-emerges only for lead year 10. When correlating the OHC hindcasts in the upper 700 m with the corresponding assimilation runs, both ORAS4-FULL and GECCO2-FULL lack significance in the years after lead year 1, and only in case of ORAS4-FULL does skill re-emerge for lead years 7–10. When the entire water

column is considered, no such re-emergence can be found for either GECCO2-FULL or ORAS4-FULL, and correlations are significant only for the first 2 and 3 lead years, respectively.

The continuously best performance in terms of ACCs in ORAS4-ANOM is also reflected in persistently negative skill scores (CSSs) in GECCO2-FULL and ORAS4-FULL (negative CSSs express an advantage of ORAS4-ANOM over the full-field systems). Significance of negative CSSs for lead years 1 and 2, for example, provides important additional information since all three systems reveal significant ACCs for this prediction horizon. Overall, ORAS4-ANOM beats the hindcasts from GECCO2-FULL for all lead years, whereas ORAS4-FULL is significantly outperformed by ORAS4-ANOM only for particular lead years. Depending on the depth range over that OHC is considered, ORAS4-FULL is significantly outperformed

by ORAS4-ANOM for lead years 1 to 6 when only the upper 700 m are taken into account or for all lead years except year 3 when considering the whole water column.

The drop in performance of GECCO2-FULL when correlating with its own assimilation run instead of the observations from NODC is counterintuitive (Figs. 2 and 3). One may expect the same or better results when the assimilation run is used, because in terms of background state and climate modes the assimilation run is supposed to be a compromise between the observations and the prediction system (in other words between the real world and the model world). However, although based on the same coupled model, the dynamics in the assimilation runs differ from those in the hindcast runs since the assimilation adds nudging terms to the tendency equations of temperature and salinity. Therefore, it is conceivable that estimates from the assimilation runs such as the evolution of OHC or ocean flow, are neither close to real observations or the solution of the unconstrained coupled model nor to be found in between the two.

We have computed additional skill scores of the upper layer OHC in the SPG region, that is, MSESS and CRPSS (Fig. 4). The assessment based on both MSESS and CRPSS is qualitatively in line with the one based on ACCs. As measured by both skill scores and regardless of using climatology or one of the full field systems as reference forecast, ORAS4-ANOM performs best for all lead years. Differences are not always significant though. In particular, when compared to ORAS4-FULL, ORAS4-ANOM is only significantly better for lead year 1, whereas, when compared

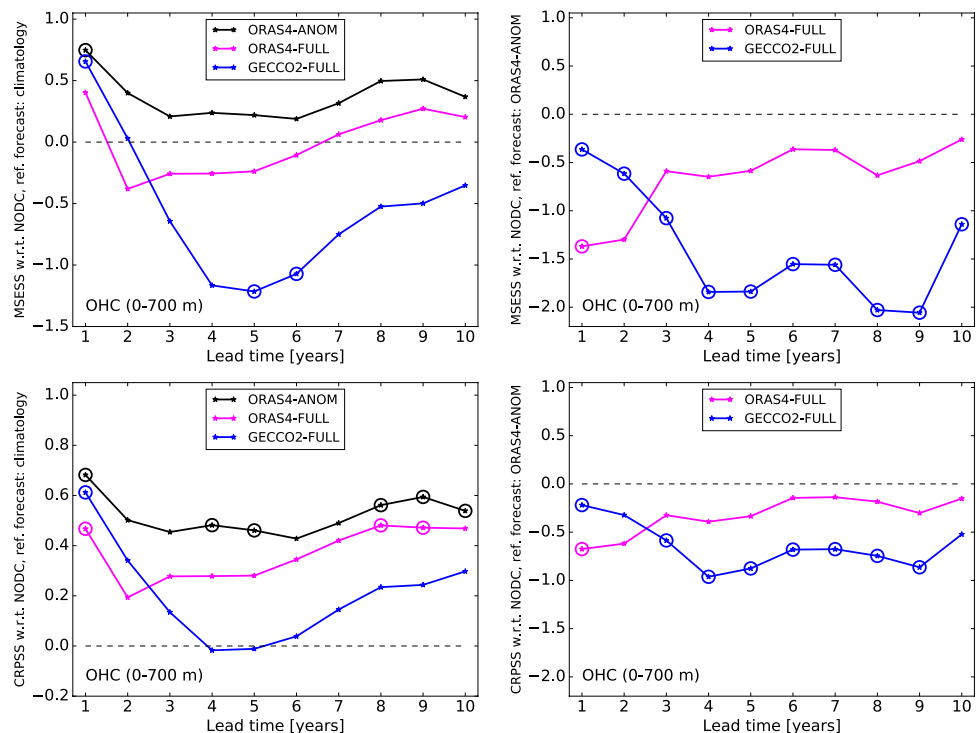
to GECCO2-FULL, differences are significant for almost all lead years. In summary, for the entire prediction horizon, the ensemble mean of the ORAS4-ANOM hindcasts is not only superior in predicting the observed variability but also in resembling the observed magnitude of OHC as indicated by the MSESS. Furthermore, as indicated by the CRPSS, the ensemble properties are always better in the ORAS4-ANOM hindcasts than in the full-field systems.

Overall, the observed changes in upper layer OHC are well reproduced in all assimilation runs (including GECCO2-FULL), allowing for proper initialization of upper layer OHC (Fig. 2, left column). The corresponding hindcasts, on the other hand, differ significantly in their ability to reproduce the observed OHC signal, in particular when comparing ORAS4-ANOM to the full-field systems. In order to understand better these significant differences, we need to understand better the processes in the MiKlip prediction system that drive changes in OHC—a topic we take up next.

### 3.3 Atlantic meridional overturning circulation

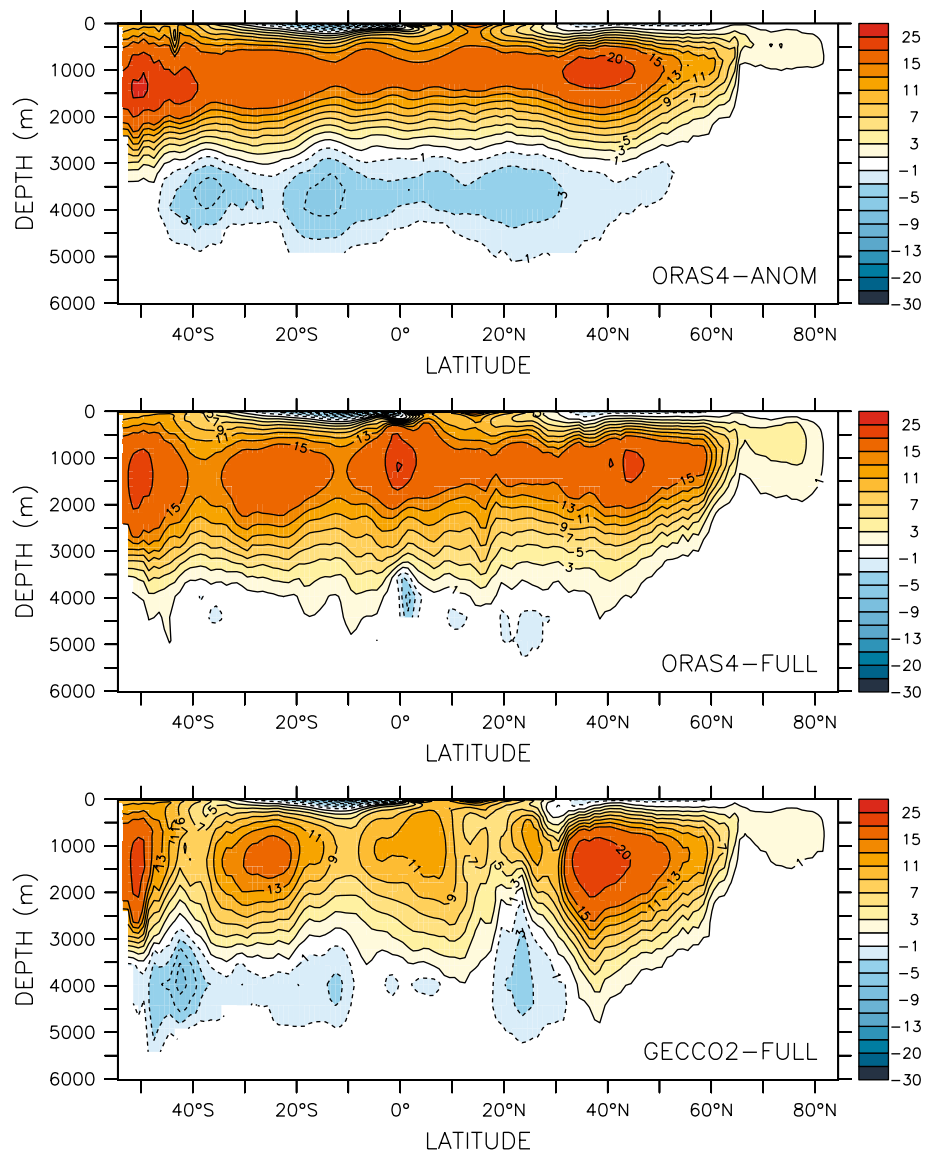
Both SST and OHC in the North Atlantic are thought to be influenced by variations in the Atlantic Meridional Overturning Circulation (AMOC) [see e.g., Knight et al. 2005]. Therefore, we look at the overall structure of the AMOC in the assimilation runs similar to the analysis shown in Smith et al. (2013a). We calculate long-term means of the zonally integrated flow field in the Atlantic for ORAS4-ANOM and both full-field systems (Fig. 5). ORAS4-ANOM reveals

**Fig. 4** MSESSs (upper row) and CRPSSs (lower row) of hindcasts of OHC in the SPG region (50°N–60°N, 75°W–9°E, 0–700 m) as function of lead time; skill scores calculated with respect to the observations from NODC; reference forecast is either the climatology of the observations (left column) or hindcasts from ORAS4-ANOM (right column); circles in all panels indicate significance at 95%





**Fig. 5** Long-term (1960–2011) mean of the zonally integrated flow field in the Atlantic (Sv) from assimilation runs of ORAS4-ANOM (upper), ORAS4-FULL (middle) and GECCO2-FULL (lower)



pronounced North Atlantic deep water (NADW) and Antarctic bottom water (AABW) cells in line with observational estimates (Talley et al. 2003). These basin-wide mono-cells are separated into multiple cells in the full-field assimilation runs. While a strong cross-equatorial NADW cell is still maintained in case of ORAS4-FULL, local cells dominate the upper ocean's overturning in case of GECCO2-FULL. The AABW cell is split into local cells in both full-field assimilation runs.

In order to assess the fidelity of our assimilation runs, that is, the ability of our assimilation procedure to adopt (from the reanalyses) mean state and variability of key climate parameters such as the AMOC (Kröger et al. 2012), we compare the different representations of our assimilated AMOC to the AMOCs in the original ocean reanalyses ORAS4 and GECCO2, which were investigated by Karspeck et al. (2015). For ORAS4, characteristics of the long term

mean AMOC in the original reanalysis such as the depth level of the NADW return flow or the lack of a basin-wide AABW cell are better adopted by ORAS4-FULL than by ORAS4-ANOM. For GECCO2, on the other hand, the full-field assimilation fails to adopt the dominant mean state features from its reanalysis, that is, GECCO2-FULL lacks pronounced, basin wide NADW and AABW cells.

Localized circulation cells similar to those in Fig. 5 (lower) were also found in the comparison of anomaly and full-field assimilations with the British Met Office climate forecasting system and interpreted to be an artifact of the full-field approach (Smith et al. 2013a). In our system, localized overturning cells appear only partially due to the transition from anomaly to full-field initialization; differences between ORAS4-ANOM and ORAS4-FULL are rather moderate, in particular in the upper ocean. Here, the transition from applying full fields from one ocean state

estimate (ORAS4) to applying full fields from another (GECCO2) has a much stronger impact as can be seen when comparing the two full-field assimilation runs alone. A strong sensitivity of the assimilated AMOC to the ocean reanalysis being used has already been found in a predecessor of the current MPI climate forecast system (Kröger et al. 2012). Presumably, the diversity of mean AMOC states in our assimilation runs already translates into a diversity of mean transports of heat in the ocean. However, since we are interested in transport processes that drive changes in OHC in the North Atlantic (cf. Fig. 2), as a next step, we extend our investigation from the time-mean ocean flow to mass and heat transports associated with the transient ocean flow.

First aspects of the impact of ocean flow variability on OHC are provided by a comparison of time series of the AMOC and the advective heat transport at the southern and northern boundaries of the SPG region (Fig. 6). In all assimilation runs, the northward heat transport is always stronger at 50°N than at 60°N; hence, ocean transport delivers a surplus of heat into the SPG region at all times, which is well in line with the climate state in the unconstrained MPI-ESM (Jungclaus et al. 2013) and the observational estimate from (Trenberth and Fasullo 2008).

In general, on annual to multi-annual scales, we find relatively good correlations between mass and heat transport, with ACCs of 0.42, 0.58, 0.36 at 50°N and 0.44, 0.57, 0.75 at 60°N for ORAS4-ANOM, ORAS4-FULL

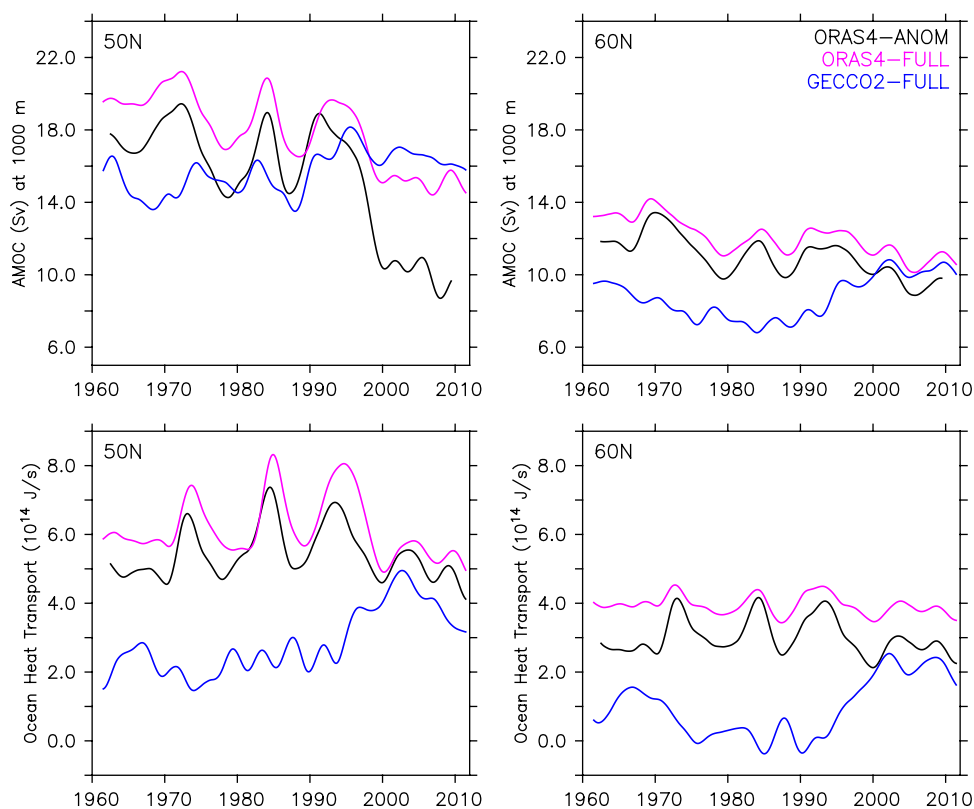
and GECCO2-FULL, respectively. What is remarkable in ORAS4-ANOM is a severe drop in the AMOC of more than 30% at 50°N in the 1990s, which is not accompanied by a comparable drop in heat transport. Such a slowdown of the AMOC in the northern North Atlantic (at 45°N) is also present in the original ORAS4 reanalysis, which, furthermore, coincides with a change in sign of decadal AMOC trends in ORAS4 and the North Atlantic Oscillation (NAO) index (Karspeck et al. 2015). In contrast to ORAS4-ANOM, ORAS4-FULL underestimates the flow reduction by a factor of about 2.

High correlations between ORAS4-ANOM and ORAS4-FULL are, nevertheless, omnipresent for both heat transport and AMOC, including the overturning at the southern boundary (ACCs: 0.91 and 0.83 for AMOC and heat transport at 50°N and 0.94 and 0.88 at 60°N, respectively). Eventually, we are interested in how ocean heat transport impacts the heat budget in the SPG region. Therefore, in the next section, we will deduce from the transports at 50°N and 60°N the heat flux divergence and compare it to the net flux contributions at the surface.

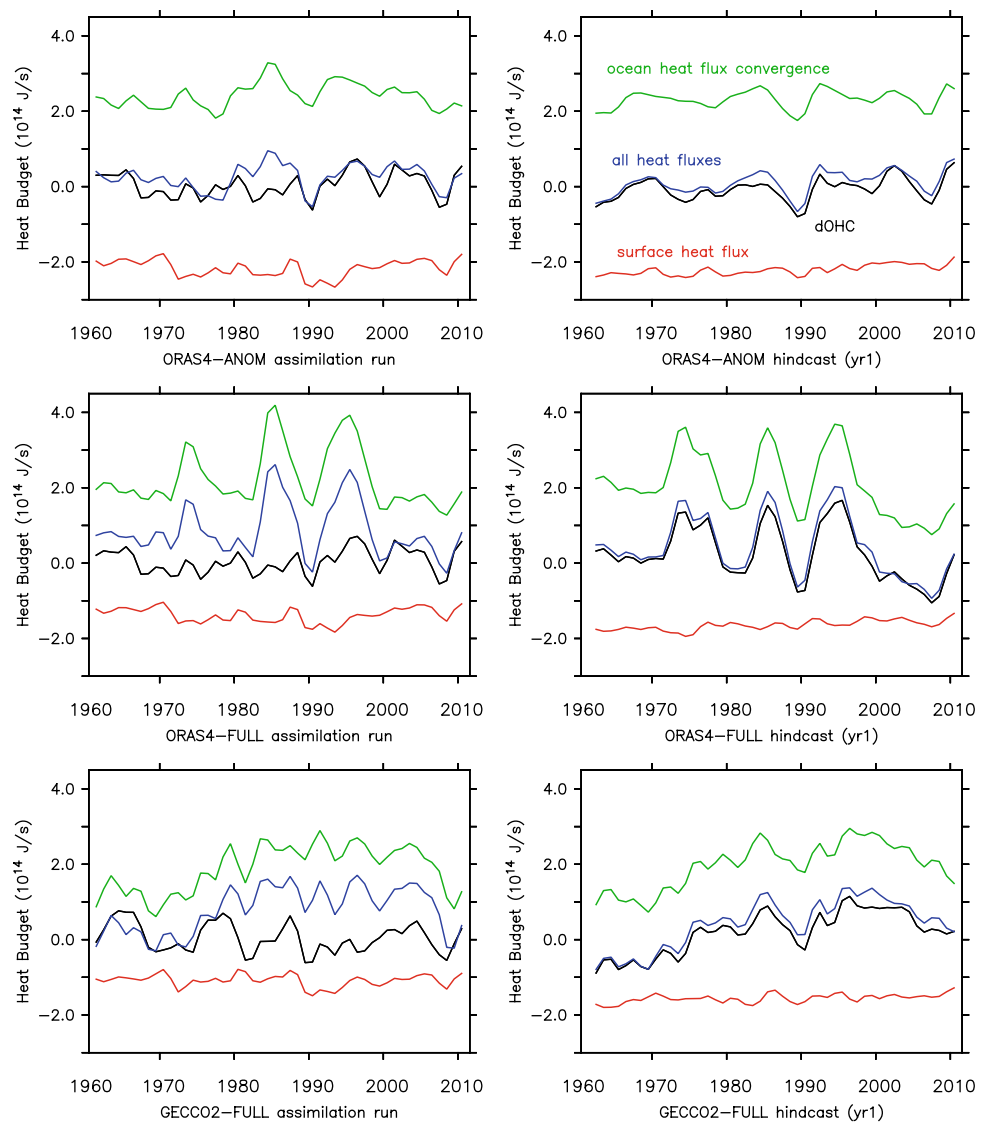
### 3.4 Heat budget in the subpolar gyre

In the assimilation step, the nudging procedure introduces additional (nudging) terms in the tendency equations of temperature and salinity that constitute sources and sinks of

**Fig. 6** Time series of AMOC (upper row) and ocean heat transport (lower row) at 50°N (left column) and 60°N (right column) from assimilation runs of ORAS4-ANOM (black), ORAS4-FULL (magenta) and GECCO2-FULL (blue); a 5 year running mean is applied to all time series



**Fig. 7** Time series of the heat budget in the SPG region (50°N–60°N, 75°W–9°E, entire water column); comparison of assimilation runs (left column) and hindcasts for lead year 1 (right column) from ORAS4-ANOM (upper), ORAS4-FULL (middle) and GECCO2-FULL (lower); each panel contains OHC tendencies (black), net heat flux at the surface (red), heat flux convergence due to ocean advection at northern and southern boundaries (green) and the sum of heat fluxes (blue); a 2 year running mean is applied to all time series



these properties. The effect of these system modifications becomes evident when we calculate the SPG heat budget in the assimilation and hindcast runs (Fig. 7). We calculate the heat budget for the entire water column, similar to the approach from Robson et al. (2012a) and unlike Yeager et al. (2012), who have only accounted for the upper 275 m in their budget analysis. The heat budget is then the sum of net heat flux at the ocean-atmosphere interface (surface heat flux) and heat flux in the interior ocean due to advection and mixing of temperature at lateral boundaries (ocean heat flux). Here, we neglect lateral mixing because its contribution to heat flux changes is at least one order of magnitude smaller than the contribution of the advective transport (not shown).

The assumption of neglecting lateral mixing is supported by the resulting heat budgets in the hindcast runs for lead year 1, where local changes in heat content (also referred to as OHC tendencies) are to first order balanced by the ocean

heat flux convergence due to advection and the net heat flux at the surface (right column of Fig. 7). OHC tendencies are calculated by differentiating in time monthly means of OHC. In all hindcast runs, in particular in the full-field systems, strong variability of the ocean heat flux convergence coincides with strong variability of the OHC tendencies while fluctuations of the surface heat flux are rather moderate. The ocean heat flux convergence appears to be the dominant driver of OHC tendencies. Moreover, the resemblance of ocean heat flux convergence in hindcasts and corresponding assimilation runs points at persisting initial ocean flow conditions in our prediction system (compare left and right columns of Fig. 7).

In contrast to the heat budgets for the hindcast runs, our heat budgets for the assimilation runs are not closed: Net heat flux into the SPG region does not correspond to the region's heat content change (left column of Fig. 7). The reason is that the additional nudging terms in the tendency

equations for temperature and salinity impact on the heat budget; nudging affects the heat budget in several ways: First, OHC is directly altered by the induced changes in temperature in each single grid box (changes due to sources and sinks of heat) and, second, temperature and salinity changes at the lateral boundaries lead to altered temperature gradients across the boundaries and altered density structures along the boundaries, the latter in turn lead to altered geostrophic in- and outflow (dynamical changes). On top of it, the induced changes in temperature at the surface modify the heat exchange between ocean and atmosphere.

Apparently, the sum of the effects from nudging leads to a discrepancy in the budget. In particular in the full-field assimilation runs, the low variability of the OHC tendencies indicates that the strong variability of the net heat flux into the SPG region has to be compensated by strong variability of the nudging-induced heat sources and sinks in the interior. Again, as in the hindcast runs, the variability of the net heat flux is in large parts governed by the variability of the ocean heat flux convergence. To quantify the heat budget discrepancy, that is, the net effect of heat sources and sinks in the interior, we consider the root mean square (rms) difference between the sum of all heat fluxes and the OHC tendencies. The resulting rms differences for the full-field assimilation runs are 2–3 times higher than for the anomaly-based assimilation run ( $1.1$  and  $1.05 \times 10^{14}$  J/s for ORAS4-FULL and GECCO2-FULL, as compared to  $0.44 \times 10^{14}$  J/s for ORAS4-ANOM).

## 4 Discussion

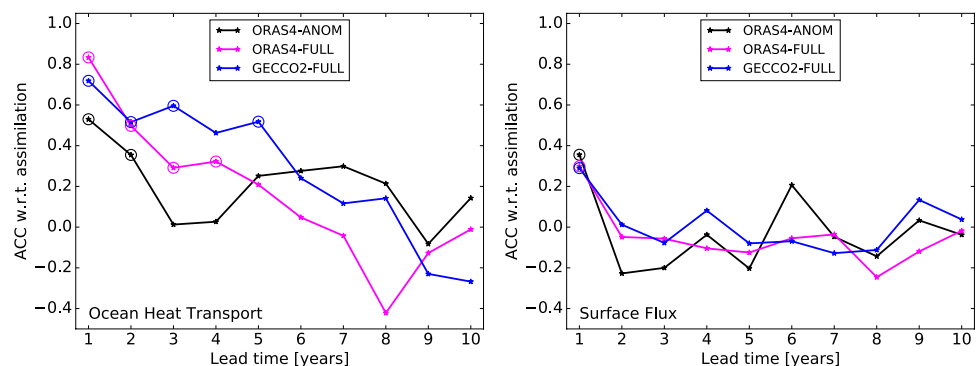
The evolution of North Atlantic OHC is very similar in both ORAS4-based assimilation runs (ORAS4-ANOM and ORAS4-FULL, see Figs. 2 and 7, left column). The ocean transport, in contrast to OHC, does not behave consistently in the ORAS4-based assimilation runs, as indicated by the related heat flux convergence signals in ORAS4-ANOM and ORAS4-FULL, which differ considerably (Fig. 7, left

column). In addition, the variability of the heat flux convergence in both full-field assimilation runs is too strong to be balanced by rather moderate changes in net heat flux at the surface and OHC in the interior. Hence, in the North Atlantic, particular attention has to be paid to the initialization of the ocean flow.

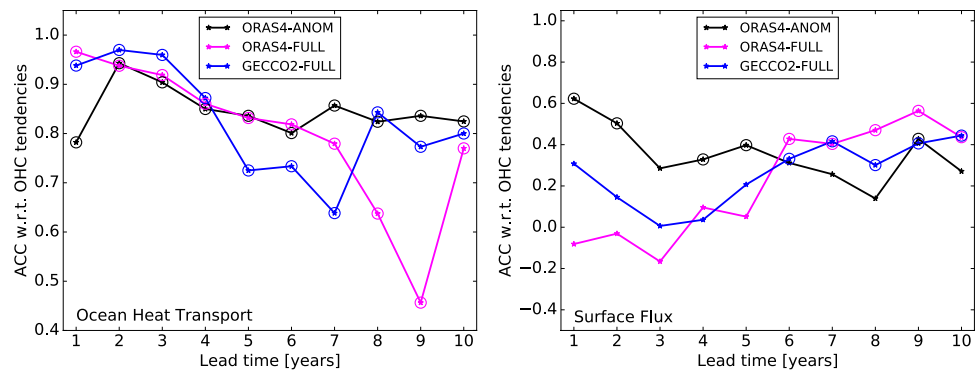
The resemblance of the ocean heat flux convergence in hindcasts and corresponding assimilation runs is striking, most prominently for forecast year 1 (Fig. 7). This is a strong indication of memory in the prediction system, more specifically, of persistent in- and outflow conditions in the hindcast runs, which were beforehand induced by the nudging procedure in the assimilation runs. Correlations of hindcasts with their corresponding assimilation runs provide an estimate of the memory (Fig. 8). For ocean heat flux convergence (Fig. 8, left), significant correlations can be found up to lead year 2 in ORAS4-ANOM, lead year 4 in ORAS4-FULL and lead year 5 in GECCO2-FULL (with the exception of lead year 4). In contrast, we find little indication of longer-term memory in the heat flux signal at the surface; only for lead year 1 correlations are significantly different from zero (Fig. 8, right).

In all hindcasts, OHC tendencies are clearly driven by ocean heat transport as is already indicated for lead year 1 (Fig. 7, right column). Ocean heat transport continues to play an important role over the entire prediction horizon. In all systems and for all lead years, correlations with the respective OHC response are significant and exceed values of 0.6 (with the exception of lead year 9 in ORAS4-FULL) (Fig. 9, left). Heat flux at the surface also co-varies with the OHC tendencies, but correlations are generally lower than for ocean heat transport and significant only in the first pentad in ORAS4-ANOM, as opposed to the full-field systems, where significance prevails in the second pentad (Fig. 9, right). Overall, in all hindcasts, the dominant driver of changes in OHC is the ocean heat transport. The low prediction skill for SPG OHC in the full-field systems (Fig. 3) points to the detrimental effect of the initial flow conditions in these systems. The longer-lasting memory of the initial ocean heat transport in ORAS4-FULL and GECCO2-FULL

**Fig. 8** ACCs of assimilation runs and hindcasts of ocean heat transport convergence (left) and net heat flux at the surface (right) in the SPG region ( $50^{\circ}\text{N}$ – $60^{\circ}\text{N}$ ,  $75^{\circ}\text{W}$ – $9^{\circ}\text{E}$ , 0–bottom) as function of lead time; correlations with the respective assimilation run are calculated for ORAS4-ANOM (black), ORAS4-FULL (magenta) and GECCO2-FULL (blue); Circles indicate significance at 95%



**Fig. 9** ACCs of hindcasts of OHC tendencies and hindcasts of ocean heat transport convergence (left) and net heat flux at the surface (right) in the SPG region (50°N–60°N, 75°W–9°E, 0–bottom) as function of lead time; correlations are calculated for ORAS4-ANOM (black), ORAS4-FULL (magenta) and GECCO2-FULL (blue); Circles indicate significance at 95%



hindcasts (Fig. 8, left) adds to the poor performance of the full-field systems.

In summary, the fast drop in prediction skill for OHC in the SPG region in hindcasts that are based on full-field initialization (Fig. 3) can be attributed to nudging-induced ocean heat flux changes in the assimilation runs that are still present in the hindcasts. These changes are much less pronounced in ORAS4-ANOM leading to a significantly better prediction skill. A similar or better skill for OHC was found by Brune et al. (2017) for hindcasts initialized by an oceanic ensemble Kalman filter using full-field temperature and salinity profiles with identical atmospheric nudging as in ORAS4-ANOM. In our full-field systems, the nudging-induced ocean heat flux signals reveal strong fluctuations that drive OHC tendencies in the hindcasts. In the corresponding assimilation runs, these strong flux fluctuations are compensated by strong fluctuations of heat sources and sinks in the interior that are also induced by the nudging; here, the residual contribution from ocean heat flux together with a moderately fluctuating heat exchange at the surface results in low variability of OHC tendencies. In hindcast mode, on the other hand, the ocean heat flux driven OHC tendencies show strong variability. As a consequence, the predicted OHC tendencies bear little resemblance with those in the assimilation runs.

It is not clear whether the direct modifications of OHC in the assimilation runs play a minor role for the initialized hindcasts or whether their importance is simply obscured by the dominant effect of the particular changes in ocean heat transport on OHC predictions, at least in the first pentad. Re-emergence of significant skill for upper level OHC in the second pentad in ORAS4-FULL, as compared to GECCO2-FULL, may either result from a better initialization of OHC or simply indicate a faster reduction of the detrimental initial flow conditions or be a consequence of both effects.

In this study, we have put the focus on the northern North Atlantic, a key region in the field of decadal climate prediction. Here, our prediction system initialized with full fields is clearly outperformed by predictions based on anomaly initialization when OHC is considered. For SST,

this statement only holds when comparing ORAS4-ANOM to GECCO2-FULL. In other regions such as the central and eastern equatorial Pacific, on the other hand, the quality of SST predictions with both full-field systems is significantly superior to the one based on anomaly initialization, which is in line with the recent study of Bellucci et al. (2015) where multi-model ensemble predictions with both anomaly and full-field initialization were investigated. However, compared to uninitialized hindcasts (historical simulations), prediction skill in the equatorial Pacific is still poor in all MiKlip systems as was already shown by Marotzke et al. (2016) (as measured by surface air temperature correlation skill for lead years 2–5; see their Fig. 1). Here, regardless of using anomalies or full fields, initialization is not able to improve forecast quality.

Initializations from both anomalies and full fields can trigger initial shocks that excite spurious El Niño Southern Oscillation (ENSO) signals (e.g., Pohlmann et al. 2017; Sanchez-Gomez et al. 2016). Such spurious signals have led to an overall lack of prediction skill in the equatorial Pacific in the first generation of MiKlip (baseline0) and could be traced back to a spurious trend in the wind stress used for producing the initialization fields (Pohlmann et al. 2017). The spurious wind stress was discarded after baseline0. Another way to reduce spurious ENSO signals in forecast mode is to omit, in the assimilation step, nudging in the tropics as was demonstrated by Sanchez-Gomez et al. (2016). However, in order to perform better than uninitialized forecasts also in regions such as the equatorial Pacific, we believe that further identification and elimination of spurious effects in the initialization is a more promising strategy than reducing the initialization to limited areas.

In this study, initializations from full fields were identified to trigger initial shocks in ocean transport and related heat advection that degrade upper OHC prediction skill in the North Atlantic SPG region. A detailed analysis of related drift processes in hindcast mode as, for example, provided by Huang et al. (2015) or Sanchez-Gomez et al. (2016) is beyond the scope of the paper. However, the relatively slow adjustment time scales of the AMOC (multiple years

and longer) in the hindcasts from Huang et al. (2015) and Sanchez-Gomez et al. (2016) correspond well to the long memory of ocean heat transport that we find in our system. Here, we demonstrate how a simple budget analysis in the assimilation run alone can be utilized to identify a peculiar ocean heat transport that is the driver of unrealistic OHC tendencies in forecasts. Before investing into an extensive suite of hindcast experiments, it seems advisable to check whether ocean heat transport and changes in OHC are to first order balanced in the initialization.

## 5 Summary

We have investigated the effect of full-field versus anomaly initialization on prediction skill for SST and North Atlantic OHC by analyzing the two most recent generations of the MiKlip decadal prediction system, the baseline1 and prototype systems. Skill analysis of SST and OHC together with a detailed investigation of the OHC budget in the North Atlantic SPG region reveals:

- In the northern North Atlantic, SST hindcasts for lead years 2–5 based on full-field initialization from GECCO2, used in the prototype system, are significantly outperformed by hindcasts based on anomaly initialization from ORAS4, used in the baseline1 system.
- In the assimilation runs, full-field nudging induces pronounced dynamical changes in the North Atlantic, resulting in altered transport of mass and heat.
- In the hindcasts, the dynamical changes are still present in the first 4–5 lead years, result in an initialization shock, and in turn severely degrade SPG OHC predictions.
- As a consequence, SPG OHC hindcasts based on full-field initialization are significantly outperformed by hindcasts based on anomaly initialization for almost all lead years.

**Acknowledgements** Open access funding provided by Max Planck Society. We thank two anonymous reviewers for their constructive and helpful comments for improving the manuscript. This study was funded by the German Federal Ministry for Education and Research (BMBF) project MiKlip under Grant Number FKZ 01LP1519A. The work of Johanna Baehr was supported by the Cluster of Excellence CliSAP (EXC177), Universität Hamburg, funded through the German Research Foundation (DFG). All simulations were carried out at the German Climate Computing Center (DKRZ), which also provided all major data services. Analysis for Fig. 1 is produced with the MURCSS tool (Illing et al. 2014). Primary data and scripts used in the analysis and other supplementary information that may be useful in reproducing the authors' work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Balmaseda M, Anderson D (2009) Impact of initialization strategies and observations on seasonal forecast skill. *Geophys Res Lett* 36:L01701. <https://doi.org/10.1029/2008GL035561>
- Balmaseda MA, Mogenssen K, Weaver AT (2013) Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q J R Meteorol Soc.* <https://doi.org/10.1002/qj.2063>
- Bellucci A, Haarsma R, Gualdi S, Athanasiadis P, Caian M, Cassou C, Germe A, Jungclaus J, Kröger J, Matei D, Müller W, Pohlmann H, y Melia DS, Sanchez E, Smith D, Terray L, Wyser K, Yang S (2015) An assessment of a multi-model ensemble of decadal climate predictions. *Clim Dyn* 46(9):2787–2806. <https://doi.org/10.1007/s00382-014-2164-y>
- Brune S, Düsterhus A, Pohlmann H, Müller WA, Baehr J (2017) Time dependency of the hindcasts skill for the North Atlantic subpolar gyre in initialized decadal hindcasts with the global coupled climate model MPI-ESM. *Clim Dyn.* <https://doi.org/10.1007/s00382-017-3991-4>
- Carrassi A, Weber RJT, Guemas V, Doblas-Reyes FJ, Asif M, Volpi D (2014) Full-field and anomaly initialization using a low-order climate model: a comparison and proposals for advanced formulations. *Nonlinear Process Geophys* 21:521–537. <https://doi.org/10.5194/npg-21-521-2014>
- Cox P, Stephenson D (2007) A changing climate for prediction. *Science* 317:207–208
- Dee DP et al (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597. <https://doi.org/10.1002/qj.828>
- Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Böttinger M, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg HD, Ilyina T, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Mueller W, Notz D, Pithan F, Raddatz T, Rast S, Redler R, Roeckner E, Schmidt H, Schnur R, Segsneider J, Six KD, Stockhause M, Timmreck C, Wegner J, Widmann H, Wieners KH, Claussen M, Marotzke J, Stevens B (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J Appl Meteorol* 5:572–597. <https://doi.org/10.1002/jame.20038>
- Goddard L, Kumar A, Solomon A, Smith D, Boer G, Gonzalez P, Kharin V, Merryfield W, Deser C, Mason SJ, Kirtman BP, Msadek R, Sutton R, Hawkins E, Fricker T, Hegerl G, Ferro CAT, Stephenson DB, Meehl GA, Stockdale T, Burgman R, Greene AM, Kushnir Y, Newman M, Carton J, Fukumori I, Delworth T (2013) A verification framework for interannual-to-decadal predictions experiments. *Clim Dyn* 40:245–272. <https://doi.org/10.1007/s00382-012-1481-2>
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 90:1095–1107. <https://doi.org/10.1175/2009BAMS2607.1>
- Hazeleger W, Guemas V, Wouters B, Corti S, Andreu-Burillo I, Doblas-Reyes F, Wyser K, Caian M (2013) Multiyear climate predictions using two initialisation strategies. *Geophys Res Lett.* <https://doi.org/10.1002/grl.50355>

- Huang B, Zhu J, Marx L, Wu X, Kumar A, Hu Z, Balmaseda MA, Zhang S, Lu J, Schneider EK, Kinter JL (2015) Climate drift of AMOC, North Atlantic salinity and arctic sea ice in CFSv2 decadal predictions. *Clim Dyn* 44:559–583. <https://doi.org/10.1007/s00382-014-2395-y>
- ICPO (2011) Data and bias correction for decadal climate predictions. International CLIVAR Project Office, CLIVAR publication series no, Tech. rep., p 150
- Illing S, Kadow C, Kunst O, Cubasch U (2014) MurCSS: a tool for standardized evaluation of decadal hindcast systems. *J Open Res Softw* 2(1):e24. <https://doi.org/10.5334/jors.bf>
- Jungclaus JH, Fischer N, Haak H, Lohmann K, Marotzke J, Matei D, Mikolajewicz U, Notz D, von Storch JS (2013) Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *J Appl Meteorol* 5:422–446. <https://doi.org/10.1002/jame.20023>
- Karspeck AR, Stammer D, Köhl A, Danabasoglu G, Balmaseda M, Smith DM, Fujii Y, Zhang S, Giese B, Tsujino H, Rosati A (2015) Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products. *Clim Dyn*. <https://doi.org/10.1007/s00382-015-2787-7>
- Keenlyside NS, Latif M, Jungclaus J, Kornbluh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature* 453:84–88. <https://doi.org/10.1038/nature06921>
- Knight JR, Allan RJ, Folland CK, Vellinga M, Mann ME (2005) A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys Res Lett* 32:L20708. <https://doi.org/10.1029/2005GL024233>
- Köhl A (2015) Evaluation of the GECCO2 ocean synthesis: transports of volume, heat and freshwater in the atlantic. *Q J R Meteorol Soc* 141:166–181. <https://doi.org/10.1002/qj.2347>
- Kröger J, Müller W, von Storch JS (2012) Impact of different ocean reanalyses on decadal climate prediction. *Clim Dyn* 39:795–810. <https://doi.org/10.1007/s00382-012-1310-7>
- Kruschke T, Rust H, Kadow C, Müller W, Pohlmann H, Leckebusch G, Ulbrich U (2016) Probabilistic evaluation of decadal prediction skill regarding northern hemisphere winter storms. *Meteorol Z* 25:721–738. <https://doi.org/10.1127/metz/2015/0641>
- Levitus S, Antonov JI, Boyer TP, Baranova OK, Garcia HE, Locarnini RA, Mishonova AV, Reagan JR, Seidov D, Yarosh ES, Zweng MM (2012) World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys Res Lett* 39:L10603. <https://doi.org/10.1029/2012GL051106>
- Magnusson L, Alonso-Balmaseda M, Corti S, Molteni F, Stockdale T (2012) Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Clim Dyn*. <https://doi.org/10.1007/s00382-012-1599-2>
- Marotzke J, Müller WA, Vamborg F, Becker P, Cubasch U, Feldmann H, Kaspar F, Kottmeier C, Marini C, Polkova I, Prömmel K, Rust H, Stammer D, Ulbrich U, Kadow C, Köhl A, Kröger J, Kruschke T, Pinto JG, Pohlmann H, Reyers M, Schröder M, Sienz F, Timmreck C, Ziese M (2016) MiKlip—a national research project on decadal climate prediction. *Bull Am Meteorol Soc*. <https://doi.org/10.1175/BAMS-D-15-00184.1>
- Matei D, Pohlmann H, Jungclaus JH, Müller WA, Haak H, Marotzke J (2012) Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. *J Clim*. <https://doi.org/10.1175/JCLI-D-11-00633.1>
- Msadek R et al (2014) Predicting a decadal shift in North Atlantic climate variability using the GFDL forecast system. *J Clim* 27:6472–6496
- Mudelsee M (2010) Climate time series analysis—classical statistical and bootstrap methods. Springer, Berlin. <https://doi.org/10.1007/978-90-481-9482-7>
- Mulholland D, Laloyaux P, Haines K, Balmaseda MA (2015) Origin and impact of initialization shocks in coupled atmosphere ocean forecasts. *Mon Weather Rev* 143:4631–4644. <https://doi.org/10.1175/MWR-D-15-0076.1>
- Müller WA, Baehr J, Haak H, Jungclaus JH, Kröger J, Matei D, Notz D, Pohlmann H, von Storch JS, Marotzke J (2012) Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology. *Geophys Res Lett* 39:L22707. <https://doi.org/10.1029/2012GL053326>
- Müller WA, Matei D, Bersch M, Jungclaus JH, Haak H, Lohmann K, Compo P, Sardeshmukh D, Marotzke J (2015) A twentieth-century reanalysis forced ocean model to reconstruct the North Atlantic climate variations during the 1920s. *Clim Dyn* 44:1935–1955. <https://doi.org/10.1007/s00382-014-2267-5>
- Murphy AH (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon Weather Rev* 116:2417–2424
- van Oldenborgh GJ, Doblas-Reyes FJ, Wouters B, Hazeleger W (2012) Decadal prediction skill in a multi-model ensemble. *Clim Dyn*. <https://doi.org/10.1007/s00382-012-1313-4>
- Pohlmann H, Jungclaus JH, Köhl A, Stammer D, Marotzke J (2009) Initializing decadal climate prediction with the GECCO oceanic synthesis: effects on the North Atlantic. *J Clim* 22:3926–3938. <https://doi.org/10.1175/2009JCLI2535.1>
- Pohlmann H, Müller WA, Kulkarni K, Kameswarrao K, Matei D, Vamborg F, Kadow C, Illing S, Marotzke J (2013) Improved forecast skill in the tropics in the new MiKlip decadal climate predictions. *Geophys Res Lett*. <https://doi.org/10.1002/2013GL058051>
- Pohlmann H, Kröger J, Greatbatch RJ, Müller WA (2017) Initialization shock in decadal hindcasts due to errors in wind stress over the tropical Pacific. *Clim Dyn* 49:2685–2693. <https://doi.org/10.1007/s00382-016-3486-8>
- Polkova I, Köhl A, Stammer D (2014) Impact of initialization procedures on the predictive skill of a coupled ocean–atmosphere model. *Clim Dyn* 42:3151–3169. <https://doi.org/10.1007/s00382-013-1969-4>
- Rayner NA, Parker DB, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res* 108:4407. <https://doi.org/10.1029/2002JD002670>
- Robson J, Sutton R, Lohmann K, Smith D, Palmer MD (2012a) Causes of rapid warming of the North Atlantic Ocean in the mid-1990s. *J Clim* 25:4116–4134. <https://doi.org/10.1175/JCLI-D-11-00443.1>
- Robson J, Sutton R, Smith D (2013) Predictable climate impacts of the decadal changes in the ocean in the 1990s. *J Clim* 26:6329–6339. <https://doi.org/10.1175/JCLI-D-13-00512.1>
- Robson J, Sutton R, Smith D (2014) Decadal predictions of the cooling and freshening of the North Atlantic in the 1960s and the role of ocean circulation. *Clim Dyn* 42:2353–2365. <https://doi.org/10.1007/s00382-014-2115-7>
- Robson JI, Sutton RT, Smith DM (2012b) Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. *Geophys Res Lett* 39:L19713. <https://doi.org/10.1029/2012GL053370>
- Sanchez-Gomez E, Cassou C, Ruprich-Robert Y, Fernandez E, Terray L (2016) Drift dynamics in a coupled model initialized for decadal forecasts. *Clim Dyn* 46:1819–1840. <https://doi.org/10.1007/s00382-015-2678-y>
- Sienz F, Müller W, Pohlmann H (2016) Ensemble size impact on the decadal predictive skill assessment. *Meteorol Z* 25:645–655. <https://doi.org/10.1127/metz/2016/0670>
- Smith DM, Cusack S, Colman AW, Folland C, Harris GR, Murphy JM (2007) Improved surface temperature prediction for the coming

- decade from a global climate model. *Science* 317:796–799. <https://doi.org/10.1126/science.1139540>
- Smith DM, Eade R, Pohlmann H (2013a) A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Clim Dyn*. <https://doi.org/10.1007/s00382-013-1683-2>
- Smith DM, Scaife AA, Boer GJ, Caian M, Doblas-Reyes FJ, Guemas V, Hawkins E, Hazeleger W, Hermanson L, Ho CK, Ishii M, Kharin V, Kimoto M, Kirtman B, Lean J, Matei D, Merryfield WJ, Müller WA, Pohlmann H, Rosati A, Wouters B, Wyser K (2013b) Real-time multi-model decadal climate predictions. *Clim Dyn* 41:2875–2888. <https://doi.org/10.1007/s00382-012-1600-0>
- Talley LD, Reid LD, Robbins PE (2003) Data-based meridional overturning streamfunctions for the global oceans. *J Clim* 16:3213–3226
- Thoma M, Greatbatch RJ, Kadow C, Gerdes R (2015) Decadal hindcasts initialized using observed surface wind stress: evaluation and prediction out to 2024. *Geophys Res Lett*. <https://doi.org/10.1002/2015GL064833>
- Trenberth KE, Fasullo JT (2008) An observational estimate of ocean energy divergence. *J Phys Oceanogr* 38:984–999. <https://doi.org/10.1175/2007JPO3833.1>
- Uppala SM et al (2005) The ERA-40 re-analysis. *Q J R Meteorol Soc* 131(612):2961–3012. <https://doi.org/10.1256/qj.04.176>
- Volpi D, Guemas V, Doblas-Reyes FJ (2016) Comparison of full field and anomaly initialisation for decadal climate prediction: towards an optimal consistency between the ocean and sea-ice anomaly initialisation state. *Clim Dyn*. <https://doi.org/10.1007/s00382-016-3373-3>
- Yeager S, Karspeck A, Danabasoglu G, Tribbia J, Teng H (2012) A decadal prediction case study: late twentieth-century North Atlantic Ocean heat content. *J Clim* 25:5173–5189. <https://doi.org/10.1175/JCLI-D-11-00595.1>