

# Hindcast skill for the Atlantic meridional overturning circulation at 26.5°N within two MPI-ESM decadal climate prediction systems

Vasco Müller<sup>1,2</sup> · Holger Pohlmann<sup>3</sup> · André Düsterhus<sup>1</sup> · Daniela Matei<sup>3</sup> · Jochem Marotzke<sup>3</sup> · Wolfgang A. Müller<sup>3</sup> · Mathias Zeller<sup>4</sup> · Johanna Baehr<sup>1</sup>

Received: 13 October 2014 / Accepted: 29 November 2016 / Published online: 30 December 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** We analyse the hindcast skill for the Atlantic meridional overturning circulation (AMOC) against 10 years of RAPID/MOCHA AMOC observations, which are now long enough to remove the mean seasonal cycle prior to the hindcast skill analysis. We analyse AMOC hindcast skill in two hindcast ensembles generated with two differently initialised decadal prediction systems that are both based on the earth system model MPI-ESM. We evaluate the hindcast skill for the AMOC and its components in both prediction systems against RAPID/MOCHA observations both with and without the mean seasonal cycle removed using anomaly correlation (COR) and root-mean-square error as skill measures. We find significant hindcast skill for most lead years up to 5 for monthly-mean AMOC variations only in the newer of the two prediction systems and only using COR, but with and without the mean seasonal cycle removed. In both systems and for all analysed lead years, the two geostrophic transport components (the upper-mid-ocean transport and Florida Strait combined, that is: AMOC minus Ekman) are the main source of hindcast skill. In the present model setup and with the currently available observational time series, we

cannot relate AMOC hindcast skill to the upper-mid-ocean transport alone. Yet, we can show that the seasonal variability of the upper-mid-ocean transport in the free coupled model originates from eastern boundary density variability. Overall, our results indicate modest yet robust AMOC hindcast skill above the uninitialized simulation, independent of the treatment of the seasonal cycle, although we cannot directly link this hindcast skill to the initialisation of the density field with either initialisation method.

**Keywords** Decadal predictions · Meridional overturning circulation · Hindcast skill

## 1 Introduction

Due to the important influence of the Atlantic meridional overturning circulation (AMOC) on European climate (e.g., Rhines et al. 2008; Johns et al. 2011) and the climate in other regions (e.g., Sutton and Hodson 2005; Pohlmann et al. 2006), possible techniques of predicting the AMOC are of great interest. Matei et al. (2012a) found hindcast skill of the monthly-mean AMOC strength for up to 4 years in advance, using the observational time series of 5 years available at that time, and hence their metrics could not exclude the seasonal cycle. Whether this leads to artificially increased skill was subsequently discussed by Vecchi et al. (2012) and Matei et al. (2012b). With the AMOC observations from the Rapid Climate Change-Meridional Overturning Circulation and Heatflux Array (RAPID/MOCHA, e.g., Cunningham et al. 2007; Kanzow et al. 2007, 2010; Johns et al. 2011; McCarthy et al. 2012; Smeed et al. 2014; McCarthy et al. 2015) now spanning 10 years, we repeat the analysis of hindcast skill for the AMOC with a separate consideration of the mean seasonal cycle. Our analysis is

✉ Vasco Müller  
vasco.mueller@uni-bremen.de

<sup>1</sup> Institute of Oceanography, Bundestr. 53, 20146 Hamburg, Germany

<sup>2</sup> MARUM, Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, 28359 Bremen, Germany

<sup>3</sup> Max Planck Institute for Meteorology, Bundesstr. 53, 20146 Hamburg, Germany

<sup>4</sup> School of Earth, Atmosphere and Environment, Monash University, Melbourne, VIC, Australia

further expanded to a new model version, considering two differently initialised hindcast ensembles.

The AMOC transports approximately 17.5 Sv ( $1\text{Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$ ) of warm saline water northwards towards high latitudes (Smeed et al. 2014). Associated with the transport of water masses is a northward heat transport of 1.3 PW, responsible for up to 25% of the global combined atmosphere-ocean heat flux (e.g., Ganachaud and Wunsch 2000; Bryden and Imawaki 2001; Cunningham and Marsh 2010). Several studies have analysed predictability of the AMOC in a perfect-model framework using uninitialised climate models (e.g., Collins and Sinha 2003; Collins et al. 2006; Hawkins and Sutton 2008; Msadek et al. 2010) and initialised climate models (e.g., Pohlmann et al. 2013).

Hindcast skill can be established if hindcast simulations are compared against observations (e.g., Boer et al. 2013). Continuous estimates of the AMOC based on full-depth measurements exist only from the RAPID/MOCHA array at 26.5 starting in 2004.

Using RAPID/MOCHA data from 2004–2009, Matei et al. (2012a) related the AMOC hindcast skill mostly to hindcast skill in predicting the upper-mid-ocean transport component of the AMOC and its seasonal cycle. The upper-mid-ocean transport is strongly connected to the zonal density difference across the basin and thus to the initialisation of the density field (Matei et al. 2012a).

Here, we investigate hindcast skill for the AMOC and its components in a new model version (Giorgetta et al. 2013; Jungclaus et al. 2013) and consider the impact of two different methods of initialisation: (i) ocean initialisation obtained from forcing an ocean model with the NCEP-NCAR atmospheric reanalysis (baseline-0 system) as used for the CMIP5 simulations and by Matei et al. (2012a), and (ii) combined initialisation from the ORAS4 ocean reanalysis (Dee et al. 2011) and the ERA-Interim atmospheric reanalysis (Balmaseda et al. 2013) (baseline-1 system) of the decadal climate prediction system (MiKlip) of the Max Planck Institute for Meteorology (Pohlmann et al. 2013). We assess hindcast skill using root-mean squared-errors (RMSE) and correlation of the transport anomalies. Similar to Matei et al. (2012b) we evaluate both correlation and RMSE against the skill of the model's climatology, but we also remove the mean seasonal cycle from observations and model.

## 2 Methods

### 2.1 Observations

The hindcast ensembles are verified against the observations from the RAPID/MOCHA array at 26.5°N. Integrated

transport datasets are currently available at nominally twice-daily resolution from April 2004 to March 2014. The AMOC transport is calculated as the sum of three components: (i) the wind-driven Ekman transport  $T_{EK}$  derived from wind stress data from the ERA-interim reanalysis (Dee et al. 2011), (ii) the Florida Strait transport  $T_{FC}$ , calculated from voltage fluctuations in a submarine telephone cable (Meinen et al. 2010), and (iii) the geostrophic upper-mid-ocean transport  $T_{UMO}$  derived from the mooring array deployed across the basin (Cunningham et al. 2007; Kanzow et al. 2007). The transport is mass-balanced to ensure zero net mass flux across 26.5°N on timescales longer than 10 days (Kanzow et al. 2007). A detailed description of the calculation of the individual components is provided by Rayner et al. (2011). For the present study we use monthly averages of the transports.

### 2.2 Model setup and prediction system

We use output from the coupled climate model MPI-ESM (Giorgetta et al. 2013) with the ocean component MPIOM (Marsland et al. 2003; Jungclaus et al. 2013) and the new atmospheric component ECHAM6 (Stevens et al. 2013). Here, we only summarise the most important characteristics of the ocean component MPIOM. MPIOM solves the primitive equations in the hydrostatic and Boussinesq approximations with a free surface on an Arakawa-C grid in horizontal space and a z-grid in the vertical. The low-resolution configuration (LR) used in this study has a nominal horizontal resolution of 1.5° and 40 vertical levels, with a vertical spacing ranging from 12 m at the surface to several hundred meters in the deep ocean. The bipolar grid with the South Pole over Antarctica and the North Pole shifted to Greenland avoids singularities at the geographic North Pole and leads to a high horizontal resolution in the deepwater formation regions around Antarctica and the Nordic Seas.

We use two sets of hindcasts, each initialised from a different assimilation experiment. The first set of hindcasts are the decadal prediction experiments performed for CMIP5 (Taylor et al. 2012). The initial conditions are produced as follows: first the ocean model (MPIOM) alone is forced with daily NCEP-NCAR reanalysis atmosphere data (NCEP-run). In the subsequent assimilation experiment, the oceanic temperature and salinity fields of the coupled model are relaxed towards the 3-D anomaly from the NCEP-run added to the coupled model's ocean climatology (Matei et al. 2012c; Müller et al. 2012). The three-dimensional ocean and atmospheric fields from the assimilation experiment are then used as initial conditions for the decadal hindcasts (Table 1). With this anomaly initialisation the drift of the model back to its own imperfect climatology is reduced.

**Table 1** Overview of the different assimilation experiments, hindcasts and reference climatologies used for the present study

Experiment	Type	Time analysed here	Comment
RAPID/MOCHA	Observations	2004–2012	
Baseline-0 assimilation	Assimilation ocean 1948–2012	2004–2012	CMIP5
Baseline-1 assimilation	Assimilation ocean and atmosphere 1960–2012	2004–2012	MiKlip
Baseline-0 hindcast	Hindcasts started from baseline-0 assimilation	2001–2012, annual January start dates with 10 ensemble members each	CMIP5
Baseline-1 hindcast	Hindcasts started from baseline-1 assimilation	2001–2012, annual January start dates with 10 ensemble members each	MiKlip
Uninitialised historical/ rcp4.5 simulation	Twentieth century simulation (1850–2005) continued with rcp4.5 scenario (2006–2012)	2004–2012	CMIP5/MiKlip
ClimRef CMIP5	Mean annual cycle from uninitialised historical simulation (1850–2005)		CMIP5
ClimRef M12	Mean annual cycle from uninitialised historical simulation in ECHAM5/ MPIOM (1850–2005)		ECHAM5/ MPIOM, only for AMOC

Since the CMIP5 experiments, the Max Planck Institute for Meteorology has carried out several improvements for the MiKlip (Mittelfristige Klimaprognosen, “mid-range climate prognoses”) decadal climate prediction system, resulting in the baseline-1 system (Pohlmann et al. 2013). Unlike baseline-0, the term adopted by MiKlip for the CMIP5 initialised simulation, the assimilation experiment for baseline-1 has a combined oceanic and atmospheric initialisation. Surface pressure and spectral temperature, vorticity and divergence from ERA40 between 1960 and 1989 (Uppala et al. 2005) and ERAinterim between 1990 and 2012 (Dee et al. 2011) are directly nudged into the atmosphere model. For the ocean initialisation, 3-dimensional temperature and salinity anomaly fields from the ORAS4 ocean reanalysis (Balmaseda et al. 2013) are added to the model climatology (Table 1, for details see Pohlmann et al. 2013).

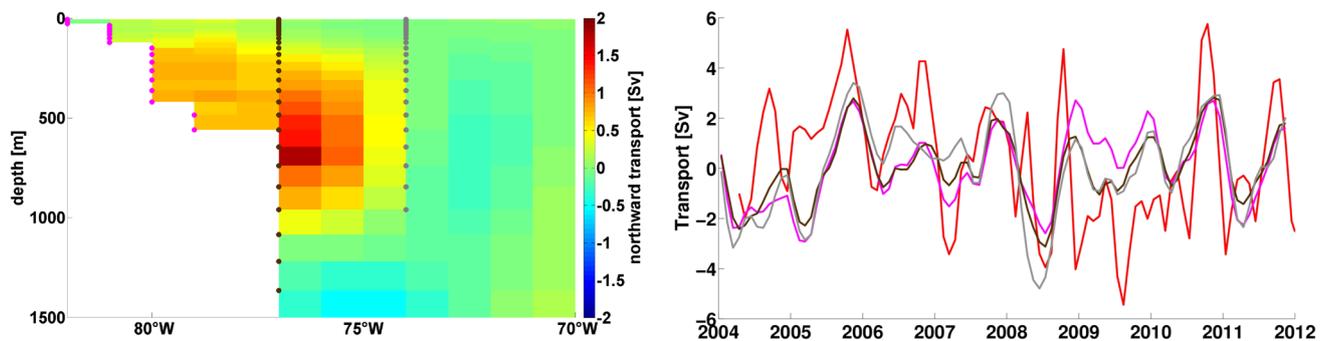
We will use the terms baseline-0 for the CMIP5 initialised simulations and baseline-1 for the MiKlip initialised experiments. For each hindcast of the two systems, 10 ensemble members are generated by one-day-lagged initialisation. For the analysis, monthly means of the ensemble mean are used. Note that the hindcast skill analysis is restricted on the one hand by the RAPID/MOCHA timeseries that only starts in 04/2004, and on the other hand by the start years of the hindcasts ensembles with the latest one being 2012. We consider lead years up to 5 years, and hence we do take into account any hindcast ensemble year that falls into the period covered by the RAPID/MOCHA observations.

### 2.3 Calculation of transport components

In the model we reconstruct the RAPID/MOCHA array to obtain the same independent components as for the observations. The method has been described in several other studies (e.g., Hirschi et al. 2003; Baehr et al. 2004, 2009). Here we describe our adaptation of the method.

Analogous to RAPID/MOCHA, the surface Ekman transport  $T_{EK}$  is calculated from the zonal wind stress, which is taken from the atmospheric component of the model (ECHAM6) and integrated zonally over all grid cells covering the Atlantic at 26.5°N.

The upper-mid-ocean transport  $T_{UMO}$  is derived from the thermal-wind balance using density profiles at the eastern and western boundaries and thus depends on the definition of the density profiles. The eastern density profile is defined at the easternmost grid point at every depth level. At the western boundary, the Bahamas are not resolved in the model, so there is no topographic boundary between the interior and the western boundary current. We test three definitions of the western density profiles. The  $T_{UMO}$  time series obtained with the different western density profiles are similar (Fig. 1b), though not identical. First, we exclude a “Florida Strait box” at the western boundary (similar to Matei et al. 2012a). Second, we define no Florida Strait and use the density profiles directly at the western boundary. Third, we define the western density profile such that only the shelf is excluded from the calculation of the thermal-wind balance (Fig. 1a). With the third definition, the Gulf Stream is partially included in the calculation of the upper-mid-ocean transport (in contrast to Matei et al. 2012a). The resulting upper mid-ocean transport variability includes



**Fig. 1** Three possible locations for the “Florida Strait box” at the western boundary (left). The three possibilities are: (i) excluding the western boundary current (method used by Matei et al. 2012a, grey), (ii) no Florida Strait (magenta), and (iii) excluding the shelf (brown). The corresponding time series of upper-mid-ocean transport anomaly are shown in the right panel. Observations from RAPID/MOCHA

are shown in red, time series for the three different locations of the “Florida Strait box” in the same colors as the respective definitions of the western boundary [excluding the western boundary current, COR: 0.37 (grey), no Florida Strait, COR: 0.27 (magenta), excluding the shelf, COR: 0.43 (brown)]

part of the western-boundary variability, but is much closer to the observed monthly upper-mid-ocean transport variability than the definition by Matei et al. (2012a) or the ‘no Florida Strait’ definition, and hence we used the third definition in our analysis.

The level-of-no-motion for the calculation of the geostrophic velocities is 4670 m, which is the level in the model closest to the 4740 m used for RAPID/MOCHA.  $T_{UMO}$  has to be corrected with a depth-independent compensation transport in order to assure zero net mass flux across the section at 26.5°N (Hirschi et al. 2003).

The crucial difference between RAPID/MOCHA and the calculation in our model is the Florida Strait transport  $T_{FC}$ . For RAPID/MOCHA, the transports  $T_{UMO}$ ,  $T_{EK}$  and  $T_{FC}$  are obtained individually from observations and summed up to calculate  $T_{AMOC}$ . In the model, the three-dimensional velocity field is known and is used to calculate the overturning stream function directly. The AMOC is defined as the maximum of the stream function above 1000 m. The transport through the Florida Strait, which is too narrow to be resolved at the given model resolution, can be then calculated as the residual  $T_{FC} = T_{AMOC} - T_{UMO} - T_{EK}$  (as in Matei et al. 2012a).

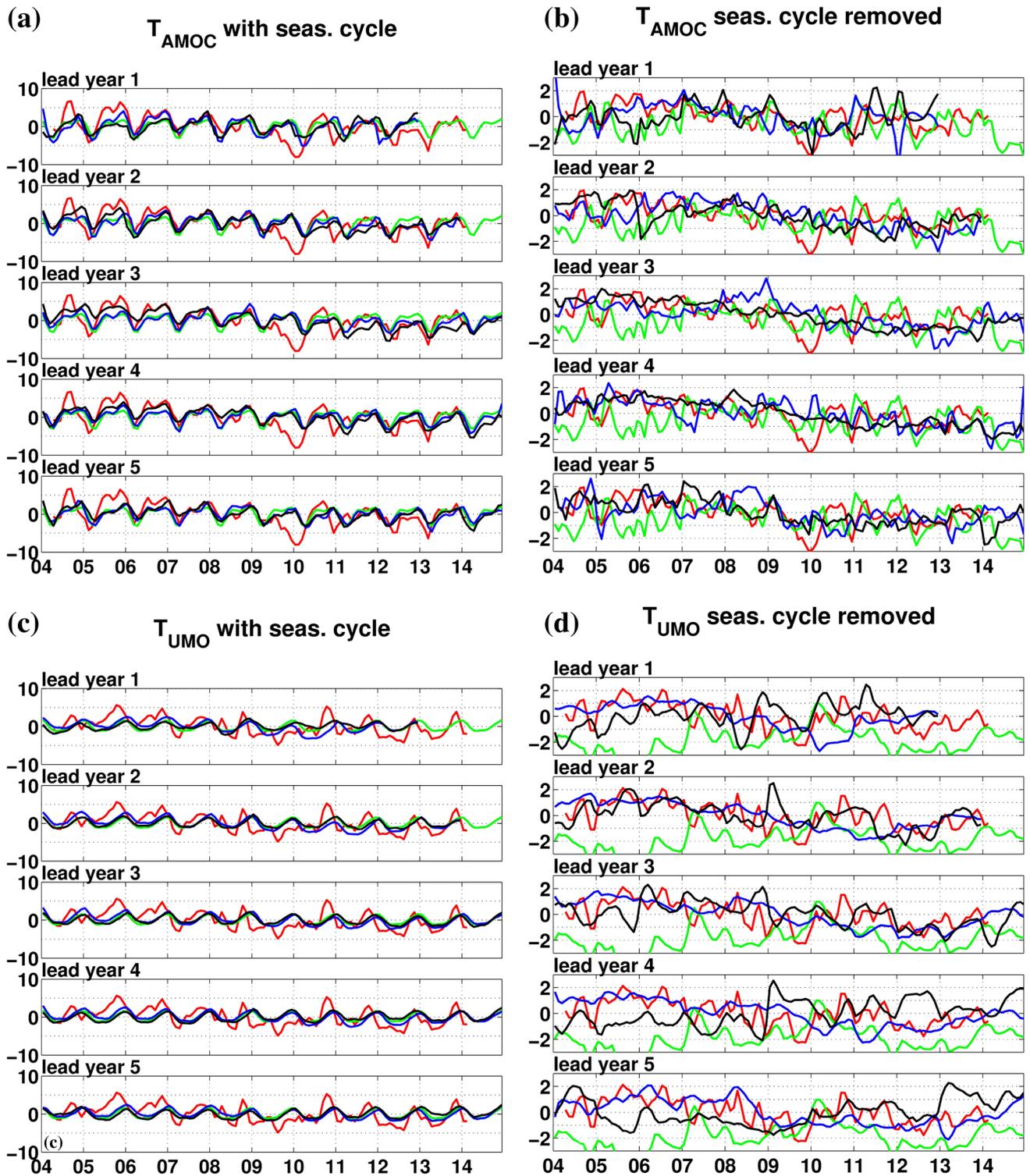
Note that in our adaptation of the method only  $T_{UMO}$  is mass balanced directly. A possible mass imbalance of  $T_{EK}$  is simply compensated by the Florida Strait transport  $T_{FC}$ , due to the residual calculation method of  $T_{FC}$ . Small errors in  $T_{EK}$  will thus only show up in  $T_{FC}$ , which is already the transport component with the highest uncertainty.

## 2.4 Calculation of hindcast skill and quality of assimilation experiments

We calculate hindcast skill for lead-time-dependent time series with monthly resolution. Lead-time-dependent time series are obtained following standard practice for hindcast

ensemble analysis: start-year-dependent time series are separated into individual years and re-merged to obtain a continuous (04/2004 to 03/2014) time series per lead year (Fig. 2a, b). This means that for the lead year 1 time series all first years of the start-year-dependent time series are merged, resulting in a combined time series where each year represents the months 1–12 after initialization. For lead year 2, all second years after the initialisation are merged together, resulting in a time series where each year represents the months 13–24 after initialisation and so on. Note that the time series for lead year 1 only ranges from 04/2004 to 12/2012 and the time series for lead year 2 ranges from 04/2004 to 12/2013, because no hindcast is initialized after 2012. The time series for lead years 3–5 all span the entire RAPID/MOCHA period from 04/2004 to 03/2014. The resulting discontinuities between December and January are smoothed with a two-month running average, which we also apply to the RAPID/MOCHA time series. Compared to Matei et al. (2012a), who calculated hindcast skill for every individual start year, we lose the information about the start year, but the time series to compute correlation and RMSE extends from 12 months to over 100 months.

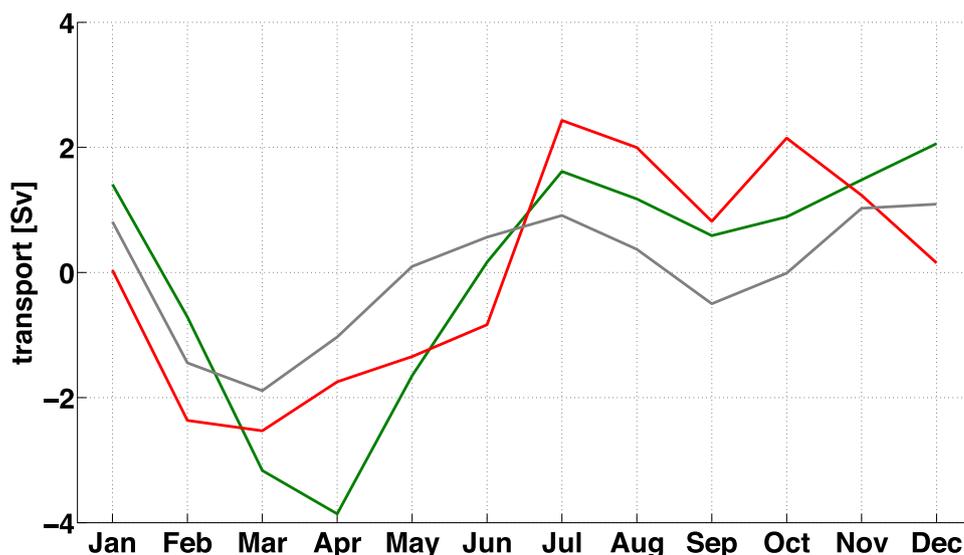
We quantify hindcast skill using anomaly correlation (COR) and RMSE for the hindcast ensemble against the RAPID/MOCHA observations. COR and RMSE are calculated for transport time series that represent anomalies against their respective time mean. Since the transport time series are serially dependent (especially when the seasonal cycle is included), we cannot treat the individual months of the time series as independent values. The statistical significance of the COR is computed using the effective degrees of freedom ( $eDoF$ ). Following the same method as Matei et al. (2012a), we calculate the effective degrees of freedom from the lagged autocorrelations of the respective time series:  $eDoF = n/[1 + 2 \times (r_1 r'_1 + r_2 r'_2 + \dots + r_n r'_n)]$ ,



**Fig. 2** Lead-time-dependent time series of  $T_{AMOC}$  (a, c) and  $T_{UMO}$  (b, d) compared to the respective RAPID/MOCHA observations (red) and the ClimRef CMIP5 (a, b)/the uninitialized run (c, d) in green. Baseline-0 is shown in blue and baseline-1 in black. We compare the

time series with seasonal cycles included (a, b) and the time series with removed seasonal cycle (c, d). The time series with removed seasonal cycle are normalised by their standard deviation

**Fig. 3** Different reference climatologies of  $T_{AMOC}$  from ECHAM5/MPIOM (ClimRef M12, grey) and MPI-ESM (ClimRef CMIP5, green) compared to the annual cycle from RAPID/MOCHA observations (red)



where  $n$  is the length of the time series,  $r_1$  and  $r'_1$  are the lag-1 autocorrelations of the time series,  $r_2$  and  $r'_2$  are the lag-2 autocorrelations of the time series, and so on. This reduces the respective number of degrees of freedom to typical values between 10–20 for time series including the seasonal cycle and somewhat larger values for the time series without seasonal cycle. The critical value for statistical significance for each lead year is then calculated individually from the respective  $eDoF$  using a one-sided t-test (since only positive correlations indicate potential skill). If COR for a lead-year time series is above the 90% significance level, we assume that the respective lead-year time series is significantly related to the RAPID/MOCHA time series.

The confidence intervals for the RMSE are calculated by using a block bootstrapping approach. Random blocks of the length ( $\tau$ ) are picked from the lead-year-dependent time series and the RAPID/MOCHA time series, respectively and resampled to new time series of the same length as the original ones, now containing blocks of random subsets of the original time series.  $\tau$  is determined from the persistence time of the individual time series (i.e. the time until the autocorrelation of both time series reaches zero). The persistence times vary between several months and several years depending on the respective time series. With 10,000 iterations for each time series, the resulting distributions of the RMSE values are Gaussian-shaped and robust enough to estimate the 90% confidence intervals by taking the 10th percentile of the distribution (one-sided).

Because the lead-year time series with monthly resolution contain seasonal variability, we compare the hindcast skill for the complete lead-year time series against the hindcast skill of a repeated seasonal cycle of the uninitialised simulation (ClimRef CMIP5), as discussed in Vecchi et al.

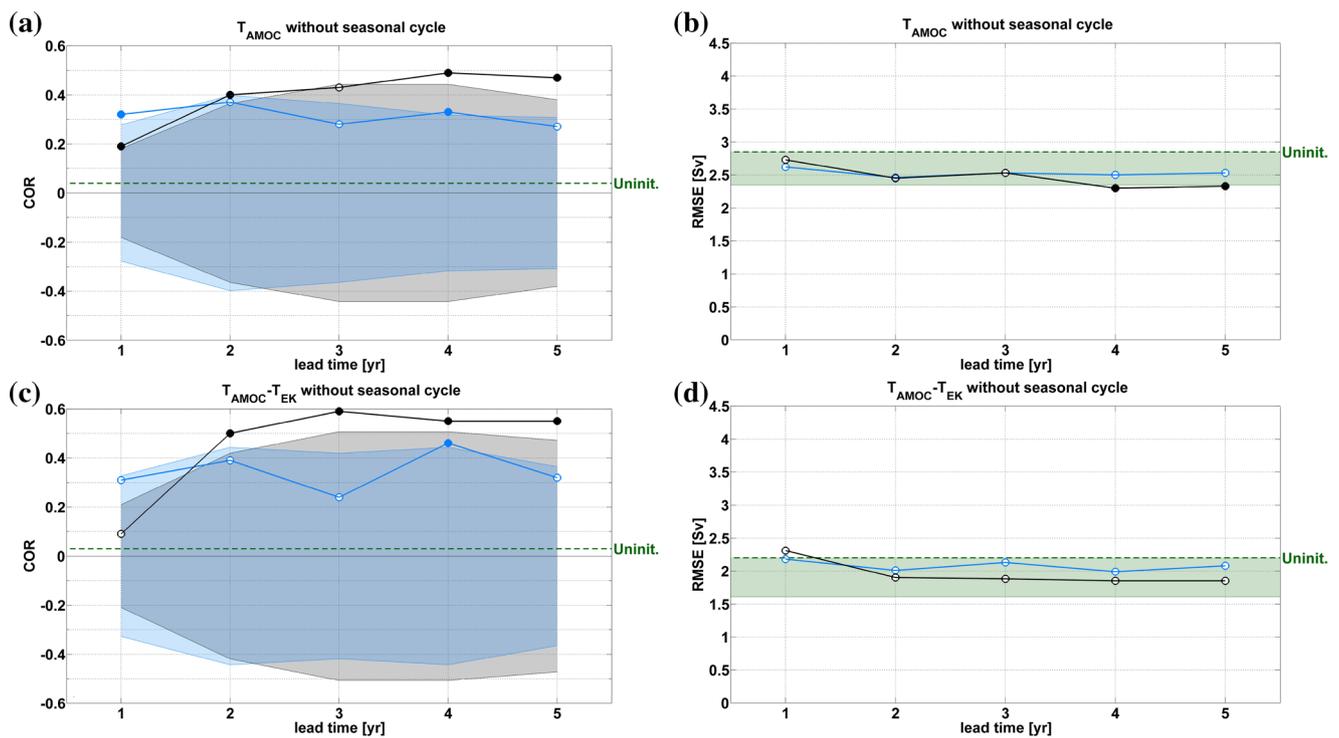
(2012) and Matei et al. (2012b). We use an AMOC reference climatology (ClimRef CMIP5) that is slightly different from the climatology of the uninitialised simulation of the model used by Matei et al. (2012a) (ClimRef M12, Fig. 3). The observed peak-to-peak variability of 5.0 Sv is reproduced better in ClimRef CMIP5 (6.0 Sv) than in ClimRef M12 (3.0 Sv) (Fig. 3). Correlation with the observed climatology is higher (0.78) for ClimRef CMIP5 than for ClimRef M12 (0.68), while the RMSE with the observed seasonal cycle is the same for both reference climatologies (1.20 Sv). For the analysis with the seasonal cycle included, we focus on the ClimRef CMIP5 climatology as a benchmark, because the skill of the ClimRef CMIP5 climatology is higher than or equal to the hindcast skill of the uninitialised simulation.

With the RAPID/MOCHA observations available for 10 years now, we also remove the mean seasonal cycle from every observed and simulated time series, and calculate hindcast skill for each lead-year time series (Fig. 2c, d). We compare the resulting hindcast skill against the hindcast skill of the uninitialised simulation.

### 3 Hindcasts

#### 3.1 AMOC and AMOC minus Ekman hindcasts without seasonal cycle

Hindcast skill for  $T_{AMOC}$  without the mean seasonal cycle is for all lead years better than the uninitialised simulation using both COR and RMSE in both baseline-0 and baseline-1 (Fig. 4). With the exception of lead year 1, baseline-1 shows higher COR and lower or equal RMSE values than baseline-0. COR values for baseline-1 range from 0.2



**Fig. 4** Skill of initialised  $T_{AMOC}$  (a, b) and  $T_{AMOC} - T_{EK}$  (c, d) hindcasts with seasonal cycles removed against uninitialised historical experiment using COR (a, c) and RMSE (b, d). Baseline-0 is shown in blue and baseline-1 in black. The blue (baseline-0) and grey (baseline-1) shaded areas indicate the 90% significance levels for each lead

year. The green dashed line indicates the COR/RMSE of the uninitialised historical experiment (shaded green area 90% confidence interval of the RMSE). Significant COR/RMSE values are marked with solid points

to 0.5 (significant at a 90% level, except lead year 3) with higher values for lead years 2–5. COR values for baseline-0 range from 0.3 to 0.4 and the majority of the values are not significantly different from zero (except lead year 1 and 4, Fig. 4a; Table 2). The RMSE shows a similar pattern with baseline-1 hindcasts lower or equal to baseline-0 hindcasts for lead years 2–5, while baseline-1 RMSEs are lower for lead year 1 (Fig. 4b). Even though all RMSE values are lower than that from the uninitialized run, only lead years 4 and 5 of the baseline-1 hindcast are outside of the 90% confidence interval of the uninitialized run. Overall, hindcast skill in baseline-1 is above or similar to hindcast skill in baseline-0, except for lead year 1, where hindcast skill for baseline-0 is above hindcast skill for baseline-1.

To analyse where hindcast skill stems from, we analyse hindcast skill for  $T_{AMOC}$  minus the Ekman transport component ( $T_{AMOC} - T_{EK}$ ). We find for  $T_{AMOC} - T_{EK}$  an overall similar lead-time-dependence of the hindcast skill as for  $T_{AMOC}$  hindcast skill for both COR and RMSE and for both baseline-0 and baseline-1 (Fig. 4c, d). Hindcast skill for baseline-1 is higher than that of baseline-0 for all lead years except lead year 1 using both COR and RMSE. COR for baseline-1 ranges from 0.5 to 0.6 for the lead years 2–5,

while COR in baseline-0 is mostly not significantly different from zero (except for lead year 4, Fig. 4c). For the lead years 2–5, RMSE values in baseline-1 are also significantly lower than those of the uninitialized run (Fig. 4d).

Because the individual COR and RMSE values indicate higher hindcast skill for  $T_{AMOC} - T_{EK}$  than for  $T_{AMOC}$  and because the lead-year-dependent evolution of the hindcast skill for  $T_{AMOC} - T_{EK}$  is similar to the lead-year-dependent evolution of the hindcast skill for  $T_{AMOC}$ , we conclude that the hindcast skill in  $T_{AMOC}$  mainly stems from  $T_{AMOC} - T_{EK}$ .

### 3.2 AMOC and AMOC minus Ekman hindcasts with seasonal cycle

We now test the robustness of the analysis where the seasonal cycle is removed from all time series, by comparing against an analysis where the seasonal is not removed. Instead of using the uninitialized run as a reference that needs to be outperformed, hindcast skill is tested against a repeated seasonal cycle first for  $T_{AMOC}$  and then  $T_{AMOC} - T_{EK}$ .

The overall lead-year-dependent evolution of the hindcast skill for  $T_{AMOC}$  is remarkably similar to the analysis that does not include the seasonal cycle. COR values in

**Table 2** Ensemble mean RMSE and correlation of hindcasts [with/without seasonal cycle] for AMOC, AMOC minus Ekman, upper-mid-ocean transport, Florida Strait and density differences at 26.5°N

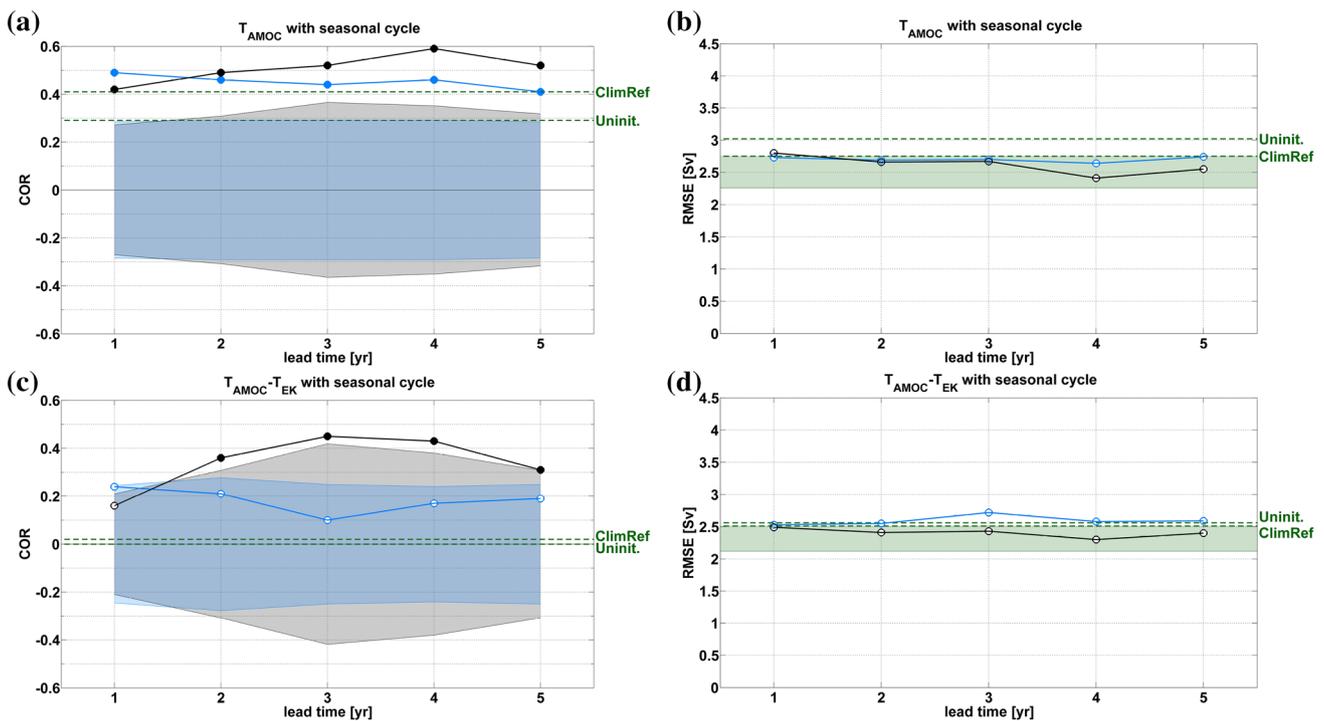
	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	ClimRef/uninit.
<b><math>T_{AMOC}</math></b>						
b-0						
COR	<b>0.5/0.3</b>	<b>0.5/0.4</b>	<b>0.4/0.3</b>	<b>0.5/0.3</b>	0.4/0.3	0.4/0.0
RMSE [Sv]	2.7/2.6	2.7/2.5	2.7/2.5	2.6/2.5	2.7/2.5	2.3/2.4
b-1						
COR	<b>0.4/0.2</b>	<b>0.5/0.4</b>	<b>0.5/0.4</b>	<b>0.6/0.5</b>	<b>0.5/0.5</b>	0.4/0.0
RMSE [Sv]	2.8/2.7	2.7/2.5	2.7/2.5	2.4/2.3	2.5/2.3	2.3/2.4
<b><math>T_{AMOC} - T_{EK}</math></b>						
b-0						
COR	0.2/0.3	0.2/0.4	0.1/0.2	0.2/0.5	0.2/0.3	0.0/0.0
RMSE [Sv]	2.5/2.2	2.5/2.0	2.7/2.1	2.6/2.0	2.6/2.1	2.1/1.6
b-1						
COR	0.2/0.1	<b>0.4/0.5</b>	<b>0.5/0.6</b>	<b>0.4/0.6</b>	<b>0.3/0.6</b>	0.0/0.0
RMSE [Sv]	2.5/2.3	2.4/1.9	2.4/1.9	2.3/1.9	2.4/1.9	2.1/1.6
<b><math>T_{UMO}</math></b>						
b-0						
COR	<b>0.4/0.3</b>	<b>0.5/0.4</b>	<b>0.4/0.5</b>	<b>0.5/0.5</b>	<b>0.4/0.5</b>	0.3/−0.2
RMSE [Sv]	2.3/1.9	2.1/1.8	2.1/1.7	2.1/1.7	2.1/1.7	2.1/1.9
b-1						
COR	0.2/0.0	<b>0.4/0.5</b>	0.3/0.2	0.1/−0.5	0.2/−0.1	0.3/−0.2
RMSE [Sv]	2.3/2.1	2.2/1.8	2.2/1.9	2.4/2.1	2.3/2.0	2.1/1.9
<b><math>T_{FC}</math></b>						
b-0						
COR	0.0/−0.2	0.1/−0.1	0.1/0.0	0.1/−0.2	0.1/0.1	0.1/0.0
RMSE [Sv]	2.0/1.8	2.0/1.8	1.8/1.5	1.8/1.6	1.8/1.5	1.5/1.4
b-1						
COR	0.1/0.1	<b>0.2/0.2</b>	0.1/0.1	0.1/0.1	<b>0.2/0.2</b>	0.1/0.0
RMSE [Sv]	2.8/2.6	2.5/2.3	2.5/2.3	2.3/2.2	1.9/1.6	1.5/1.4
<b><math>\Delta\rho</math></b>						
b-0						
COR	0.3/0.2	0.3/0.1	0.2/0.0	0.3/0.0	0.3/0.2	0.5/0.0
RMSE [ $10^{-2}$ kg m $^{-3}$ ]						
b-1						
COR	−0.1/−0.1	0.0/−0.1	0.2/0.0	0.1/−0.3	0.0/−0.2	0.5/0.0
RMSE [ $10^{-2}$ kg m $^{-3}$ ]	7.2/6.9	7.2/6.8	4.5/3.9	3.9/3.2	6.5/6.4	2.5/2.7

Significant values better than the ClimRef (with seasonal cycle)/uninitialised (without seasonal cycle) simulation are highlighted in bold print

both baseline-0 and baseline-1 range from 0.4 to 0.6. They are all significant and higher than the COR of the ClimRef CMIP5 reference climatology (Table 2; Fig. 5a). For all lead years, except lead year 1, baseline-1 hindcasts outperform baseline-0 hindcasts using both COR and RMSE. The RMSE for baseline-0 are all within  $\pm 0.1$  Sv similar to and hence not outperforming the RMSE of ClimRef CMIP5 (Fig. 5b). In baseline-1, RMSE values for lead years 1, 2 and 3 are insignificant or similar to and hence not outperforming the RMSE of ClimRef CMIP5. For lead years 4 and 5 baseline-1 outperforms the RMSE of ClimRef CMIP5.

Turning now to  $T_{AMOC} - T_{EK}$ , we find again a similar picture as for the hindcasts where the seasonal cycle is removed. Baseline-1 outperforms baseline-0 for all lead years except lead year 1 (Fig. 5c, d). In baseline-0, no COR is significantly different from zero (Fig. 5c), while COR of baseline-1 hindcasts outperform ClimRef CMIP5 for the lead years 2–5 with values around 0.4. Using RMSE, baseline-1 hindcasts outperform ClimRef CMIP5 and baseline-0 for all lead years.

Overall, hindcast skill is (apart from the first lead year) generally higher in baseline-1 than in baseline-0. This advance is more pronounced in COR than in RMSE,



**Fig. 5** Skill of initialised  $T_{AMOC}$  (a, b) and  $T_{AMOC} - T_{EK}$  (c, d) hindcasts with seasonal cycles included against ClimRef CMIP5 using COR (a, d) and RMSE (b, d). Baseline-0 is shown in blue and baseline-1 in black. The blue (baseline-0) and grey (baseline-1) shaded areas indicate the 90% significance levels for each lead year. The

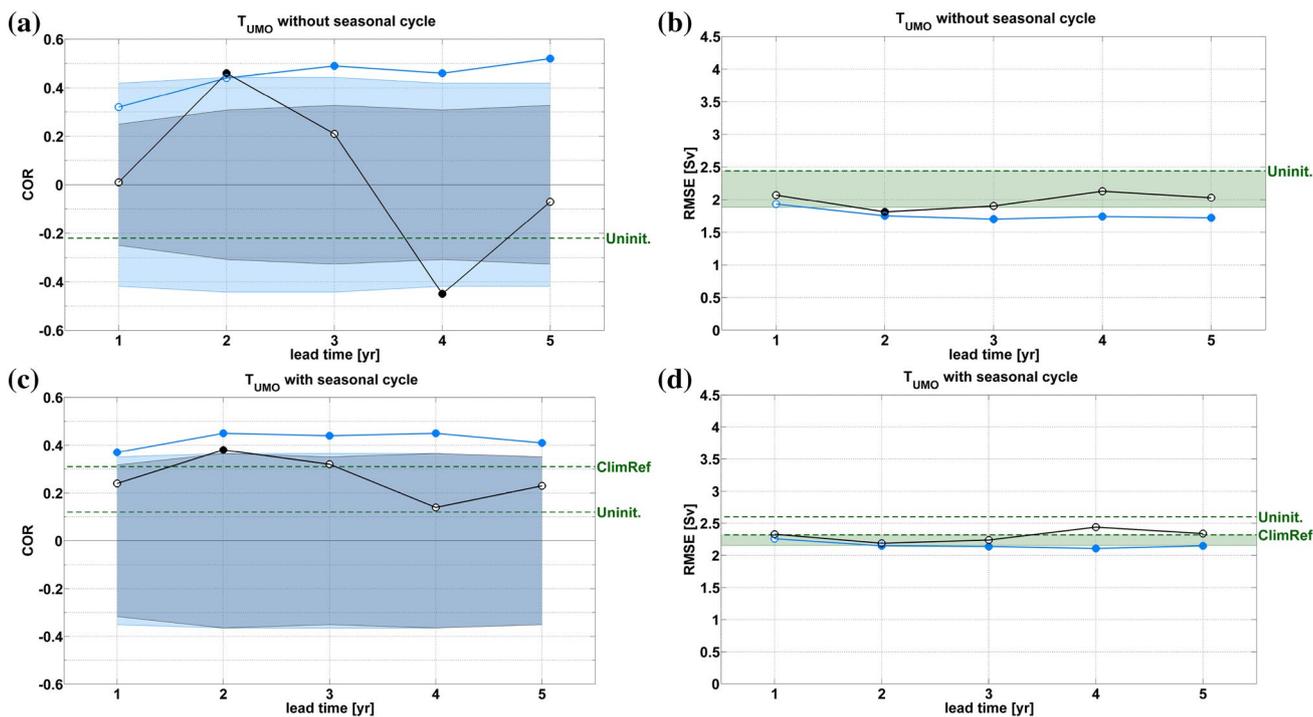
green dashed lines indicate ClimRef CMIP5 and the uninitialised historical experiment, respectively (shaded green area 90% confidence interval of the RMSE). Significant COR/RMSE values are marked with solid points

suggesting an improved phase though not magnitude of variability through initialisation. We find the same lead-time-dependence for the hindcast skill for the analysis with and without the seasonal cycle removed. However, we typically find higher hindcast skill in both COR and RMSE when the seasonal cycle is removed compared to the seasonal cycle included. With either metric for hindcast skill, we find higher hindcast skill for  $T_{AMOC} - T_{EK}$  than for  $T_{AMOC}$ , confirming the analysis of Matei et al. (2012a), and extending their finding for  $T_{AMOC}$  from four years to five years in baseline-1.

### 3.3 Upper-mid-ocean transport hindcasts

Matei et al. (2012a) related the hindcast skill for  $T_{AMOC}$  to hindcast skill for the upper-mid-ocean transport  $T_{UMO}$ . In baseline-0 and with the mean seasonal cycle removed from all time series, we find for  $T_{UMO}$  significant COR around 0.5 in lead years 3–5, with corresponding RMSE around 1.7 Sv (Fig. 6a, b). COR and RMSE indicate for baseline-0 with the seasonal cycle included also higher hindcast skill than ClimRef CMIP5, significantly outperforming ClimRef CMIP5 for lead years 3–5 (Fig. 6c, d).

In baseline-1 and with the mean seasonal cycle removed, we find no significant hindcast skill. COR values range between about  $-0.4$  (lead year 2) and  $0.4$  (lead year 4). The extremely low skill in COR of  $-0.4$  for lead year 4 in the baseline-1 hindcasts is likely to be caused by ambiguities in separating the upper-mid-ocean transport from the Florida Strait transport at the western boundary of the basin at the given  $1.5^\circ$  resolution. The sum of the two transport components—AMOC minus Ekman ( $T_{AMOC} - T_{EK} = T_{UMO} + T_{FC}$ )—shows high hindcast skill for the same lead year (Fig. 4c, d). However, for nearly all lead years, both COR and RMSE hindcasts skill is lower in baseline-1 than in baseline-0 (Fig. 6a, b). Similar  $T_{UMO}$  results are found when the seasonal cycle is not removed (Fig. 6c, d). Both COR and RMSE skill for  $T_{UMO}$  is larger in baseline-0 than in baseline-1. COR in baseline-0 ranges between 0.4 and 0.5 and is significant for all lead years. Baseline-1 again shows a very low COR for lead year 4 (Fig. 6c). However, the very low COR for lead year 4 is less distinct than for the removed seasonal cycle (Fig. 6a, c), which might point to remaining ambiguities when the seasonal cycle is removed resulting from the limited length of the time series.



**Fig. 6** Skill of initialised  $T_{UMO}$  hindcasts with seasonal cycles removed against uninitialised historical experiment (a, b) and  $T_{UMO}$  hindcasts with seasonal cycles included against ClimRef CMIP5 (c, d) using COR (left) and RMSE (right). Baseline-0 is shown in blue and baseline-1 in grey. The blue (baseline-0) and grey (baseline-1)

shaded areas indicate the 90% significance levels for each lead year. The green dashed lines indicate ClimRef CMIP5 and the uninitialised historical experiment, respectively (shaded green area 90% confidence interval of the RMSE). Significant COR/RMSE values are marked with solid points

Overall, we find hindcast skill for  $T_{UMO}$  significantly above the uninitialised simulation for nearly all lead years in baseline-0 using both COR and RMSE. In contrast to  $T_{AMOC} - T_{EK}$ , we find for  $T_{UMO}$  higher hindcast skill in baseline-0 than in baseline-1, and it is only baseline-0 that significantly outperforms ClimRef CMIP5. Although  $T_{UMO}$  is a source for AMOC hindcast skill in both baseline-0 and baseline-1, the significance depends on the chosen metric and analysis.

### 3.4 Density difference hindcasts

Matei et al. (2012a) found the hindcast skill in  $T_{UMO}$  to be related to the vertically averaged zonal density difference. Here we analyse the upper-ocean (between 200 and 1000 m) density differences  $\Delta\rho$ . In both baseline-0 and baseline-1 COR values are either insignificant or outperformed by the uninitialized run for nearly all lead years (Fig. 7a). The RMSE values for both sets of hindcasts are also higher than those of the uninitialized run (Fig. 7b). The same result is found when the seasonal cycle is not removed: neither baseline-0 or baseline-1 outperform ClimRef CMIP5 for COR or RMSE at any lead

time (Fig. 7c, d). Overall, there is no statistically significant hindcast skill for the upper-ocean density differences for any of the two initialisation methods at any lead time, although the ocean density field is directly initialised from the data assimilation of temperature and salinity.

## 4 Assimilation experiments

To better understand the origin of the differences in the hindcast skill between the two differently initialised hindcast ensembles, we examine more closely the representation of the different AMOC components in the assimilation experiments. We conduct this investigation with the seasonal cycle included, to include the representation of the non-trivial dynamics that govern the seasonal cycle.

The time averaged  $T_{AMOC}$  from the RAPID/MOCHA observations is  $17.3 \pm 3.6$  Sv (mean  $\pm$  monthly standard deviation). Over the same period, the baseline-0 assimilation slightly overestimates the strength of  $T_{AMOC}$  with an average transport of  $19.7 \pm 3.3$  Sv, while the baseline-1 assimilation substantially underestimates the strength

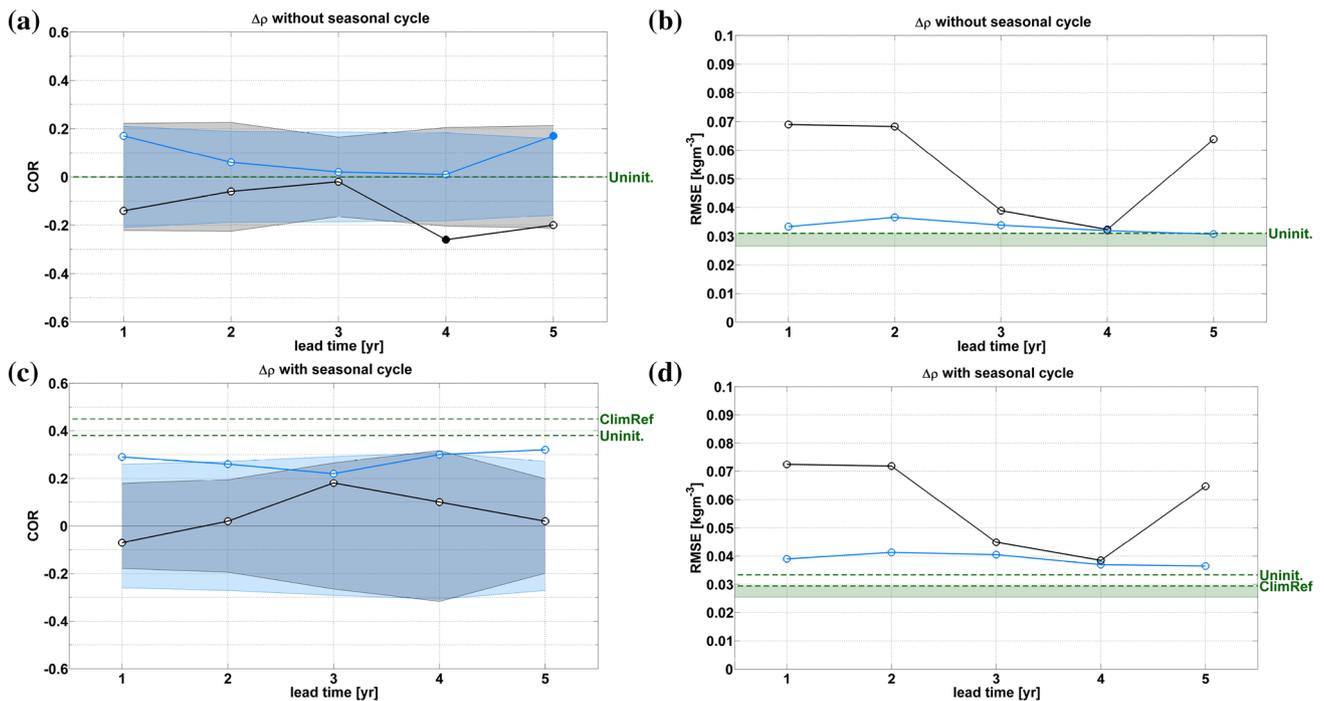
of  $T_{AMOC}$  with  $13.7 \pm 3.1$  Sv. Both systems represent the monthly variability of  $T_{AMOC}$  equally well.

The correlation of  $T_{AMOC}$  with the observations is much lower in the baseline-0 assimilation (0.33) than in the baseline-1 assimilation (0.75, Table 3). In particular, the observed short-term minima of  $T_{AMOC}$ , such as during the minimum of the 2009/2010 winter, are better reproduced in baseline-1 than in baseline-0 (Fig. 8a).

As the atmospheric component is initialised in baseline-1, but not in baseline-0, we analyse  $T_{AMOC} - T_{EK}$  to determine the direct influence of the atmospheric initialisation. For the baseline-0 assimilation, subtracting the Ekman transport from the AMOC reduces the correlation with the observations from 0.33 to 0.23. For the baseline-1 assimilation, the correlation with observations is reduced from 0.75 to 0.45 (Table 3). This suggests that the full-field atmospheric initialisation in the baseline-1 assimilation plays an important role for the initialisation of the short-term variability of  $T_{AMOC}$  and in turn for the hindcast skill of  $T_{AMOC}$  and also  $T_{AMOC} - T_{EK}$ . For both  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$ , the baseline-0 assimilation shows lower agreement with the observations than baseline-1, and the baseline-0 hindcast skill is smaller than the baseline-1 hindcast skill. Hence, in the two systems for both  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$  the level of

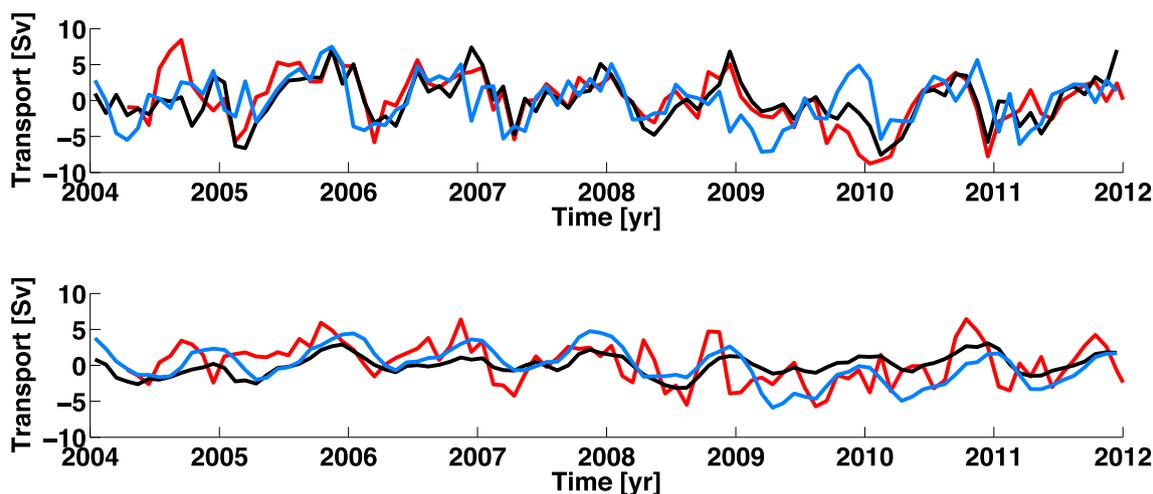
**Table 3** Assimilation experiments (baseline-0, baseline-1) for AMOC, AMOC-Ekman transport, upper-mid-ocean transport, Florida Strait transport and density differences at 26.5°N compared to RAPID/MOCHA observations

Transport component	Mean $\pm$ monthly Std	RMSE	COR
<b><math>T_{AMOC}</math></b>			
RAPID/MOCHA	$17.3 \pm 3.6$	0	1
Baseline-0	$19.7 \pm 3.3$	3.96	0.33
Baseline-1	$13.7 \pm 3.1$	2.41	0.75
<b><math>T_{AMOC} - T_{EK}</math> (Sv)</b>			
RAPID/MOCHA	$13.8 \pm 2.7$	0	1
Baseline-0	$15.2 \pm 2.4$	3.14	0.23
Baseline-1	$10.4 \pm 1.9$	2.51	0.45
<b><math>T_{UMO}</math> (Sv)</b>			
RAPID/MOCHA	$-17.7 \pm 2.9$	0	1
Baseline-0	$-15.2 \pm 2.4$	2.74	0.46
Baseline-1	$-11.7 \pm 1.4$	2.59	0.39
<b><math>T_{FC}</math> (Sv)</b>			
RAPID/MOCHA	$31.5 \pm 2.2$	0	1
Baseline-0	$30.4 \pm 1.7$	2.58	0.18
Baseline-1	$18.3 \pm 1.3$	2.32	0.25
<b><math>\Delta\rho</math> (<math>10^{-2}</math> kg m<math>^{-3}</math>)</b>			
RAPID/MOCHA	$32.9 \pm 4.2$	0	1
Baseline-0	$27.1 \pm 4.5$	5.12	0.29
Baseline-1	$20.3 \pm 2.6$	3.61	0.53



**Fig. 7** Skill of initialised  $\Delta\rho$  hindcasts with seasonal cycles removed against uninitialised historical experiment (a, b) and  $\Delta\rho$  hindcasts with seasonal cycles included against ClimRef CMIP5 (c, d) using COR (left) and RMSE (right). Baseline-0 is shown in blue and baseline-1 in black. The blue (baseline-0) and grey (baseline-1) shaded

areas indicate the 90% significance levels for each lead year. The green dashed lines indicate ClimRef CMIP5 and the uninitialised historical experiment, respectively (shaded green area 90% confidence interval of the RMSE). Significant COR/RMSE values are marked with solid points



**Fig. 8** Anomaly of AMOC (*top*) and upper-mid-ocean transport (*bottom*) at 26.5°N from baseline-1 assimilation (*black*) and baseline-0 assimilation (*blue*), compared to RAPID/MOCHA observations (*red*)

hindcast skill is consistent with the level of representation of the variability in the assimilation experiment.

However, the baseline-0 hindcast skill for  $T_{UMO}$  is larger than the baseline-1 hindcast skill for  $T_{UMO}$ , and we now test whether we find the same consistency between the level of hindcast skill and the quality of the assimilation experiment. The average  $T_{UMO}$  from RAPID/MOCHA observations is  $-17.7 \pm 2.9$  Sv. In both assimilation methods, the southward transport is underestimated, with a larger bias in the baseline-1 assimilation ( $-15.2$  Sv in baseline-0, and  $-11.7$  Sv in baseline-1). The monthly standard deviation of  $T_{UMO}$  is well reproduced in the baseline-0 assimilation ( $-15.2 \pm 2.4$  Sv), while monthly standard deviation of  $T_{UMO}$  is substantially underestimated in the baseline-1 assimilation ( $-11.7 \pm 1.4$  Sv). Correlation with the observations in the baseline-0 assimilation (0.46) is higher than in the baseline-1 assimilation (0.39, Table 3), consistent with the higher hindcast skill for  $T_{UMO}$  in baseline-0 compared to baseline-1.

In the RAPID/MOCHA observations there is a clear connection between  $T_{UMO}$  and  $\Delta\rho$  (COR:  $-0.74$ ). Similarly, in both assimilation experiments the correlation between  $T_{UMO}$  and  $\Delta\rho$  is high:  $-0.93$  for baseline-0 and  $-0.89$  for baseline-1. However,  $\Delta\rho$  in the baseline-0 assimilation shows a correlation with the observations of 0.29, whereas in the baseline-1 assimilation the correlation with the observations is 0.53 (Table 3). This representation of the observed variability of  $\Delta\rho$  is mirrored in hindcast skill for  $\Delta\rho$ , which is smaller in baseline-0 than in baseline-1, but even baseline-1 hindcasts show less hindcast skill for  $\Delta\rho$  than the uninitialised simulation (Fig. 7a, b).

Overall, we find that the level of agreement between the assimilation experiment and the observations for the

AMOC and its individual transport components is often mirrored in the resulting level of hindcast skill. Therefore, a successful data assimilation and thus more realistic initial conditions are a necessary though, as shown for  $\Delta\rho$ , not a sufficient condition for skilful hindcasts.

## 5 Seasonal cycle of upper-mid-ocean transport

While the analysis of the assimilation experiments show consistent results with the analysis of the hindcast ensembles for  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$ , the results for  $T_{UMO}$  and  $\Delta\rho$  are ambiguous. In the observations, the seasonal cycle of  $T_{AMOC}$  arises in the observations predominantly from  $T_{UMO}$  whose seasonal cycle is caused by deep-reaching seasonal density variability at the eastern boundary (Chidichimo et al. 2010). We now test whether the same relation of  $\Delta\rho$ ,  $T_{UMO}$  and  $T_{AMOC}$  governs the seasonal cycle in the model.

We therefore investigate the seasonal cycle of the modelled  $T_{UMO}$ . We use an uninitialised simulation for this analysis because we are interested in the seasonal cycle of the upper-mid-ocean transport in the freely running model, without the influence of data assimilation. We calculate the upper-mid-ocean transport anomaly obtained with either a constant (time averaged) eastern-boundary density profile (ConstEast) or a constant western-boundary density profile (ConstWest) (similar to Kanzow et al. 2010).

In the model, the anomalous seasonal cycle of  $T_{UMO}$  has a maximum of 1.7 Sv in January and a minimum of  $-1.4$  Sv in May. The magnitude of the seasonal cycle in the model is comparable to that of the observations, but the modelled seasonal cycle lags the observed seasonal cycle

by two months; the seasonal cycle from RAPID/MOCHA observations shows a maximum in October and a minimum in March.

Keeping the western-boundary density profile constant (ConstWest) results in an only slightly weaker seasonal cycle, with an anomaly maximum of 1.6 Sv in January and a minimum of  $-1.1$  Sv in May. However, when the eastern-boundary density profile is kept constant, the seasonal cycle of  $T_{UMO}$  almost disappears, and the minimum and maximum are shifted by one month. The maximum of only 0.5 Sv now occurs in December while the spring minimum is reduced to  $-0.3$  Sv in April (Fig. 9). Density variability at the eastern-boundary is responsible for the seasonal cycle of  $T_{UMO}$  in the model as in the observations.

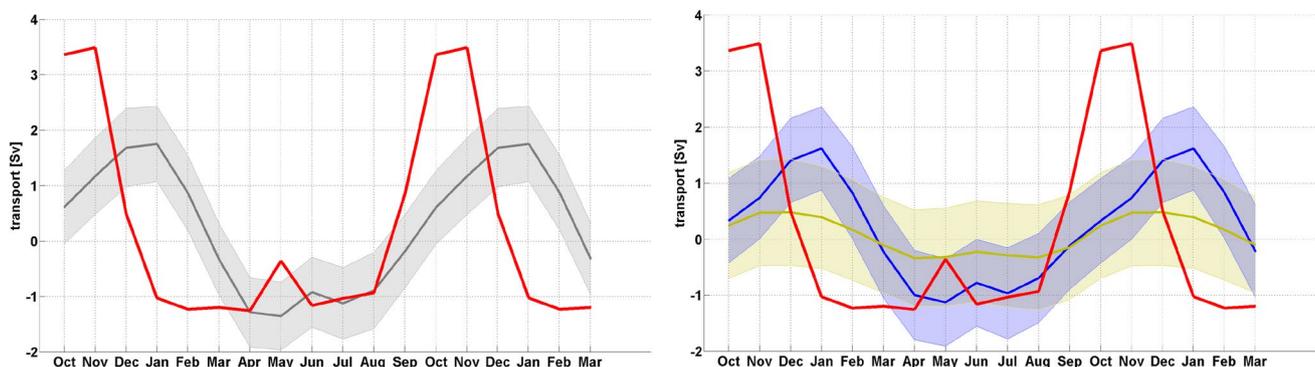
## 6 Discussion

For  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$ , we find robust hindcast skill with and without the seasonal cycle removed. For both  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$  the mean seasonal cycle can be robustly removed, that is the hindcast skill and its temporal evolution does not depend on the removal of the seasonal cycle. While the removal of the seasonal cycle can be robustly done for  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$ , the hindcast skill of  $T_{UMO}$  and  $\Delta\rho$  remains sensitive to the removal of the mean seasonal cycle. Hence, we summarise that the RAPID/MOCHA time series appears long enough to remove the seasonal cycle from  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$ , but the RAPID/MOCHA time series might not be long enough yet to robustly estimate the seasonal cycle of all AMOC components.

Hindcast skill for  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$  is higher in baseline-1 than in baseline-0. One might be tempted to relate this increased hindcast skill to the initialisation of the atmospheric component in baseline-1, in contrast to

baseline-0. The improvement in AMOC hindcast skill for the full-field atmosphere and anomaly-field ocean initialisation with reanalysis data (baseline-1) compared to model-confined ocean initialisation (baseline-0) results in a more realistic representation of the Ekman variability in the assimilation experiment. The additional atmospheric initialisation enables the short-term prediction of wind-driven minima/maxima, for example the wind-driven component of the negative transport anomaly in the 2009/2010 winter, which is also mirrored in higher hindcast skill in lead year 1 for  $T_{EK}$  alone (not shown). However, the atmospheric initialisation, at least in combination with the chosen ocean initialisation, does not generally result in a higher hindcast skill for the first lead year for the full AMOC or its components. In particular at lead year 1, the baseline-1 hindcast skill is very low for both  $T_{AMOC}$  and  $T_{AMOC} - T_{EK}$ .

Another impact of the different initialisation methods might be the difference in  $T_{UMO}$  hindcast skill between the two systems. The initial conditions in baseline-0 come from relaxing the coupled model towards a run of the ocean model component (NCEP-run), while baseline-1 initial conditions come from direct relaxation towards the reanalysis. Initial conditions might therefore be less in balance with the model's own balanced ocean state in baseline-1 than in baseline-0. An indicator that the baseline-1 state in the assimilation experiments is not geostrophically well balanced is the comparison of  $\Delta\rho$  between the baseline-1 assimilation and the respective baseline-1 hindcasts. The correlation between the baseline-1 assimilation and observations is high (0.61, Table 3), but as soon as the model runs without relaxation, initial information on  $\Delta\rho$  is lost and the hindcasts have less skill than the uninitialised simulation (Fig. 7a, b). The hindcast skill for the first lead year in baseline-1 is, particularly for  $\Delta\rho$  and  $T_{UMO}$ , both consistently smaller than the hindcast skill for consecutive years, and consistently smaller than the hindcast skill



**Fig. 9** Mean seasonal cycles of the uninitialised historical (1850–2005) simulation (*left, grey*) and of the sensitivity experiments (*right*) with eastern density profile kept constant (*yellow*) and with western density profile kept constant (*blue*). The *shaded envelopes* represent

the monthly standard deviations of the seasonal cycles. The seasonal cycle of the observed upper-mid-ocean transport from RAPID (*red*) is given as a reference. Note that 18 months are shown to illustrate the two months shift between model and observations

in baseline-0. Hence, an assimilation method that results in initial conditions that are closer to the model's mean state than both baseline-0 or baseline-1 achieve, might improve the hindcast skill of  $\Delta\rho$  and  $T_{UMO}$  particularly in the first year.

Another limitation for  $T_{UMO}$ 's hindcast skill might be the horizontal resolution of the present model that does not permit a clear separation between upper-mid-ocean transport and Florida Strait transport. A higher horizontal resolution that better resolves the bathymetry at the western boundary and the Bahamas would make the separation between upper-mid-ocean transport and Florida Strait transport less ambiguous and more consistent with the RAPID/MOCHA array. However, in a setup with the same model at about  $0.4^\circ$  resolution, Jungclaus et al. (2013) find a Florida Strait transport that is underestimated by about 15 Sv; hence, it remains to be seen whether higher resolution is the key for better hindcast skill of the AMOC and its components.

## 7 Conclusions

Using two ensembles of decadal hindcasts with the MPI-ESM that are initialised from two different assimilation experiments (baseline-1: reanalysis-based atmospheric and oceanic initialisation vs. baseline-0: ocean-only initialisation from a simple model-confined NCEP-forced reanalysis) we conclude:

- Monthly-mean AMOC variations with and without the mean seasonal cycle removed show hindcast skill significantly above the uninitialised simulation for most lead years up to 5 in the baseline-1 prediction system and using COR, but not RMSE. In the baseline-0 prediction system, we only find significant hindcast skill for monthly-mean AMOC variations (above climatology and above the uninitialised simulation) when the seasonal cycle is not removed.
- Monthly-mean AMOC minus Ekman variations with and without the mean seasonal cycle removed show hindcast skill significantly above the uninitialised simulation for lead years 2–5 in the baseline-1 prediction system and using both COR and RMSE. In the baseline-0 prediction system, we only find significant hindcast skill for monthly-mean AMOC minus Ekman variations for individual lead years.
- The upper mid-ocean geostrophic transport is a source for AMOC hindcast skill, though the significance depends on the chosen metric and analysis. Baseline-1 only shows hindcast skill for the upper mid-ocean geostrophic transport for few individual lead years. Baseline-0 shows higher hindcast skill than baseline-1 for the upper mid-ocean geostrophic transport, and signifi-

cant hindcast skill independent of the chosen metric and analysis for lead years 3–5.

- The seasonal cycle of the upper-mid-ocean transport originates from eastern-boundary density variability in the uninitialised simulation as found in the observations. Independent of the chosen metric and analysis, both baseline-0 and baseline-1 show no significant hindcast skill for the zonal density difference at any lead year.
- In the baseline-1 prediction system, hindcast skill for the monthly-mean AMOC variations is predominantly related to hindcast skill in AMOC minus Ekman. In the baseline-0 prediction system, hindcast skill for the monthly-mean AMOC variations is predominantly related to hindcast skill in the upper-mid-ocean geostrophic transport..
- Combined atmospheric and oceanic assimilation results in a more realistic representation of the wind-driven variability, but does not generally result in a higher hindcast skill for the first lead year for the full AMOC or its components.
- We demonstrate for individual AMOC components that a realistic representation of the observed variability in the assimilation experiment is a necessary, but not automatically a sufficient condition for better hindcasts.

**Acknowledgements** We thank two anonymous reviewers for their valuable and helpful comments. This research has been supported by the Cluster of Excellence 'CliSAP' (EXC177), University of Hamburg, funded through the German Science Foundation (DFG) (VM, AD, JB), and by the German Federal Ministry for Education and Research (BMBF) projects MiKlip (<http://www.fona-miklip.de/en/index.php>) (HP, WM) and RACE (<http://race.zmaw.de>) (DM). The climate simulations were performed at the German Climate Computing Centre (DKRZ). Data from the RAPID-WATCH MOC monitoring project are funded by the Natural Environment Research Council and are freely available from <http://www.rapid.ac.uk/rapidmoc>.

## References

- Baehr J, Hirschi J, Beismann JO, Marotzke J (2004) Monitoring the meridional overturning circulation in the North Atlantic: a model-based array design study. *J Mar Res* 62(3):283–312. doi:10.1357/0022240041446191
- Baehr J, Cunningham S, Haak H, Heimbach P, Kanzow T, Marotzke J (2009) Observed and simulated estimates of the meridional overturning circulation at  $26.5^\circ\text{N}$  in the Atlantic. *Ocean Sci* 5(4):575–589. doi:10.5194/os-5-575-2009
- Balmaseda MA, Mogensen K, Weaver AT (2013) Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q J R Meteorol Soc.* doi:10.1002/qj.2063
- Boer GJ, Kharin VV, Merryfield WJ (2013) Decadal predictability and forecast skill. *Clim Dyn* 41(7–8):1817–1833. doi:10.1007/s00382-013-1705-0
- Bryden HL, Imawaki S (2001) Chapter 6.1 Ocean heat transport. In: Gerold Siedler JC, Geophysics JGBTI (eds) Ocean circulation and climate observing and modelling the global ocean, vol 77. Academic Press, pp 455–474. doi:10.1016/S0074-6142(01)80134-0

- Chidichimo MP, Kanzow T, Cunningham SA, Johns WE, Marotzke J (2010) The contribution of eastern-boundary density variations to the Atlantic meridional overturning circulation at 26.5°N. *Ocean Sci* 6(2):475–490. doi:[10.5194/os-6-475-2010](https://doi.org/10.5194/os-6-475-2010)
- Collins M, Sinha B (2003) Predictability of decadal variations in the thermohaline circulation and climate. *Geophys Res Lett* 30(6):1306. doi:[10.1029/2002GL016504](https://doi.org/10.1029/2002GL016504)
- Collins M, Botzet M, Carril AF, Drange H, Jouzeau A, Latif M, Masina S, Otteraa OH, Pohlmann H, Sorteberg A, Sutton R, Terray L (2006) Interannual to decadal climate predictability in the north atlantic: a multimodel-ensemble study. *J Clim* 19(7):1195–1203. doi:[10.1175/JCLI3654.1](https://doi.org/10.1175/JCLI3654.1)
- Cunningham SA, Marsh R (2010) Observing and modeling changes in the Atlantic MOC. *Wiley Interdiscip Rev Clim Change* 1:180–191. doi:[10.1002/wcc.22](https://doi.org/10.1002/wcc.22)
- Cunningham SA, Kanzow T, Rayner D, Baringer MO, Johns WE, Marotzke J, Longworth HR, Grant EM, Hirschi J, Beal LM, Meinen CS, Bryden HL (2007) Temporal variability of the Atlantic meridional overturning circulation at 26.5°N. *Science* (New York, NY) 317(5840):935–8. doi:[10.1126/science.1141304](https://doi.org/10.1126/science.1141304)
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597. doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828)
- Ganachaud A, Wunsch C (2000) Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature* 408(6811):453–457. doi:[10.1038/35044048](https://doi.org/10.1038/35044048)
- Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Böttinger M, Brovkin V, Cruieger 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 Adv Model Earth Syst* 5(3):572–597. doi:[10.1002/jame.20038](https://doi.org/10.1002/jame.20038)
- Hawkins E, Sutton R (2008) Potential predictability of rapid changes in the Atlantic meridional overturning circulation. *Geophys Res Lett* 35(11):L11603. doi:[10.1029/2008GL034059](https://doi.org/10.1029/2008GL034059)
- Hirschi J, Baehr J, Marotzke J, Stark J, Cunningham S, Beismann JO (2003) A monitoring design for the Atlantic meridional overturning circulation. *Geophys Res Lett* 30(7):1413. doi:[10.1029/2002GL016776](https://doi.org/10.1029/2002GL016776)
- Johns WE, Baringer MO, Beal LM, Cunningham SA, Kanzow T, Bryden HL, Hirschi JJM, Marotzke J, Meinen CS, Shaw B, Curry R (2011) Continuous, array-based estimates of atlantic ocean heat transport at 26.5°N. *J Clim* 24(10):2429–2449. doi:[10.1175/2010JCLI3997.1](https://doi.org/10.1175/2010JCLI3997.1)
- Jungclaus J, Fischer N, Haak H, Lohmann K, Marotzke J, Matei D, Mikolajewicz U, Notz D, von Storch J (2013) Characteristics of the ocean simulations in MPIOM, the ocean component of the MPI-Earth system model. *J Adv Model Earth Syst* 5:422–446. doi:[10.1002/jame.20023](https://doi.org/10.1002/jame.20023)
- Kanzow T, Cunningham SA, Rayner D, Hirschi JJM, Johns WE, Baringer MO, Bryden HL, Beal LM, Meinen CS, Marotzke J (2007) Observed flow compensation associated with the MOC at 26.5°N in the Atlantic. *Science* (New York, NY) 317(5840):938–41. doi:[10.1126/science.1141293](https://doi.org/10.1126/science.1141293)
- Kanzow T, Cunningham SA, Johns WE, Hirschi JJM, Marotzke J, Baringer MO, Meinen CS, Chidichimo MP, Atkinson C, Beal LM, Bryden HL, Collins J (2010) Seasonal variability of the Atlantic meridional overturning circulation at 26.5°N. *J Clim* 23(21):5678–5698. doi:[10.1175/2010JCLI3389.1](https://doi.org/10.1175/2010JCLI3389.1)
- Marsland S, Haak H, Jungclaus J, Latif M, Röske F (2003) The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Model* 5(2):91–127. doi:[10.1016/S1463-5003\(02\)00015-X](https://doi.org/10.1016/S1463-5003(02)00015-X)
- Matei D, Baehr J, Jungclaus JH, Haak H, Müller WA, Marotzke J (2012a) Multiyear prediction of monthly mean Atlantic Meridional Overturning Circulation at 26.5°N. *Science* (New York, NY) 335(6064):76–9. doi:[10.1126/science.1210299](https://doi.org/10.1126/science.1210299)
- Matei D, Baehr J, Jungclaus JH, Haak H, Müller WA, Marotzke J (2012b) Response to comment on “Multiyear prediction of monthly mean atlantic meridional overturning circulation at 26.5°N”. *Science* 338(6107):604–604. doi:[10.1126/science.1223200](https://doi.org/10.1126/science.1223200)
- Matei D, Pohlmann H, Jungclaus J, Müller W, Haak H, Marotzke J (2012c) Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. *J Clim* 25(24):8502–8523. doi:[10.1175/JCLI-D-11-00633.1](https://doi.org/10.1175/JCLI-D-11-00633.1)
- McCarthy G, Frajka-Williams E, Johns WE, Baringer MO, Meinen CS, Bryden HL, Rayner D, Duchez A, Roberts C, Cunningham SA (2012) Observed interannual variability of the Atlantic meridional overturning circulation at 26.5°N. *Geophys Res Lett* 39(19):L19609. doi:[10.1029/2012GL052933](https://doi.org/10.1029/2012GL052933)
- McCarthy G, Smeed D, Johns W, Frajka-Williams E, Moat B, Rayner D, Baringer M, Meinen C, Collins J, Bryden H (2015) Measuring the Atlantic meridional overturning circulation at 26°N. *Progn Oceanogr* 130:91–111. doi:[10.1016/j.pocan.2014.10.006](https://doi.org/10.1016/j.pocan.2014.10.006)
- Meinen CS, Baringer MO, Garcia RF (2010) Florida Current transport variability: an analysis of annual and longer-period signals. *Deep Sea Res Part I Oceanogr Res Papers* 57(7):835–846. doi:[10.1016/j.dsr.2010.04.001](https://doi.org/10.1016/j.dsr.2010.04.001)
- Msadek R, Dixon KW, Delworth TL, Hurlin W (2010) Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys Res Lett* 37(19):L19606. doi:[10.1029/2010GL044517](https://doi.org/10.1029/2010GL044517)
- 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(22):L22707. doi:[10.1029/2012GL053326](https://doi.org/10.1029/2012GL053326)
- Pohlmann H, Sienz F, Latif M (2006) Influence of the multidecadal Atlantic meridional overturning circulation variability on European climate. *J Clim* 19(23):6062–6067. doi:[10.1175/JCLI3941.1](https://doi.org/10.1175/JCLI3941.1)
- Pohlmann H, Müller WA, Kulkarni K, Kameswarrao M, Matei D, Vamborg FSE, Kadow C, Illing S, Marotzke J (2013) Improved forecast skill in the tropics in the new MiKlip decadal climate predictions. *Geophys Res Lett* 40(21):5798–5802. doi:[10.1002/2013GL058051](https://doi.org/10.1002/2013GL058051)
- Rayner D, Hirschi JJM, Kanzow T, Johns WE, Wright PG, Frajka-Williams E, Bryden HL, Meinen CS, Baringer MO, Marotzke J, Beal LM, Cunningham SA (2011) Monitoring the Atlantic meridional overturning circulation. *Deep Sea Res Part II Top Studies Oceanogr* 58(17–18):1744–1753. doi:[10.1016/j.dsr2.2010.10.056](https://doi.org/10.1016/j.dsr2.2010.10.056)
- Rhines P, Häkkinen S, Josey S (2008) Is oceanic heat transport significant in the climate system? In: Dickson RR, Meincke J, Rhines P (eds) *ArcticSubarctic ocean fluxes*. Springer Netherlands, pp 87–109. doi:[10.1007/978-1-4020-6774-7\\_5](https://doi.org/10.1007/978-1-4020-6774-7_5)
- Smeed DA, McCarthy GD, Cunningham SA, Frajka-Williams E, Rayner D, Johns WE, Meinen CS, Baringer MO, Moat BI, Duchez A, Bryden HL (2014) Observed decline of the

- Atlantic meridional overturning circulation 2004–2012. *Ocean Sci* 10(1):29–38. doi:[10.5194/os-10-29-2014](https://doi.org/10.5194/os-10-29-2014)
- Stevens B, Giorgetta M, Esch M, Mauritsen T, Crueger T, Rast S, Salzmann M, Schmidt H, Bader J, Block K, Brokopf R, Fast I, Kinne S, Kornblueh L, Lohmann U, Pincus R, Reichler T, Roeckner E (2013) Atmospheric component of the MPI-M Earth System Model: ECHAM6. *J Adv Model Earth Syst* 5:1–27. doi:[10.1002/jame.20015](https://doi.org/10.1002/jame.20015)
- Sutton RT, Hodson DLR (2005) Atlantic Ocean forcing of North American and European summer climate. *Science* (New York, NY) 309(5731):115–8. doi:[10.1126/science.1109496](https://doi.org/10.1126/science.1109496)
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93(4):485–498. doi:[10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1)
- Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Bechtold VDC, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda Ma, Beljaars ACM, Berg LVD, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J (2005) The ERA-40 re-analysis. *Q J R Meteorol Soc* 131(612):2961–3012. doi:[10.1256/qj.04.176](https://doi.org/10.1256/qj.04.176)
- Vecchi GA, Msadek R, Delworth TL, Dixon KW, Guilyardi E, Hawkins E, Karspeck AR, Mignot J, Robson J, Rosati A, Zhang R (2012) Comment on “Multiyear prediction of monthly mean Atlantic Meridional Overturning Circulation at 26.5°N”. *Science* (New York, NY) 338(6107):604. doi:[10.1126/science.1222566](https://doi.org/10.1126/science.1222566)