ORIGINAL ARTICLE

Intercomparison of initialization methods for seasonal-to-decadal climate predictions with the NorCPM

Lilian Garcia-Oliva[1] · François Counillon[1,2] · Ingo Bethke[1] · Noel Keenlyside[1]

Abstract

Initialization is essential for accurate seasonal-to-decadal (S2D) climate predictions. The initialization schemes used differ on the component initialized, the Data Assimilation method, or the technique. We compare five popular schemes within NorCPM following the same experimental protocol: reanalysis from 1980 to 2010 and seasonal and decadal predictions initialized from the reanalysis. We compare atmospheric initialization—Newtonian relaxation (nudging)—against ocean initialization—Ensemble Kalman Filter—(ODA). On the atmosphere, we explore the benefit of full-field (NudF-UVT) or anomaly (NudA-UVT) nudging of horizontal winds and temperature (U, V, and T) observations. The scheme NudA-UV nudges horizontal winds to disentangle the role of wind-driven variability. The ODA+NudA-UV scheme is a first attempt at joint initialization of ocean and atmospheric components in NorCPM. During the reanalysis, atmospheric nudging improves the synchronization of the atmosphere and land components with the observed data. Conversely, ODA is more effective at synchronizing the ocean component with observations. The atmospheric nudging schemes are better at reproducing specific events, such as the rapid North Atlantic subpolar gyre shift. An abrupt climatological change using the NudA-UV scheme demonstrates that energy conservation is crucial when only assimilating winds. ODA outperforms atmospheric-initialized versions for S2D global predictions, while atmospheric nudging is preferable for accurately initializing phenomena in specific regions, with the technique's benefit depending on the prediction's temporal scale. For instance, atmospheric full-field initialization benefits the tropical Atlantic Niño at 1-month lead time, and atmospheric anomaly initialization benefits longer lead times, reducing hindcast drift. Combining atmosphere and ocean initialization yields sub-optimal results, as sustaining the ensemble's reliability—required for ODA's performance—is challenging with atmospheric nudging.

Keywords Climate prediction · Anomaly vs full-field · Data assimilation · Seasonal-to-decadal · Ocean vs atmospheric constrain · Nudging

1 Introduction

Climate prediction is of great socioeconomic importance and is an essential tool for climate services, which help to mitigate the risks caused by climate change (e.g., Mariotti et al. 2020). On S2D time scales, such predictions depend on an accurate initialization of internal variability and the response to external forcing (Smith et al. 2007; Keenlyside et al. 2008; Meehl et al. 2009; Hawkins and Sutton 2009; Pohlmann et al. 2009; Doblas-Reyes et al. 2013). Specifically, the correct initialization of ocean variability and the correct interaction with the atmosphere are essential to achieve skillful predictions at such timescales (Balmaseda and Anderson 2009; Mariotti et al. 2018; Meehl et al. 2021). A dedicated contribution, the Decadal Climate Prediction Project (DCPP, Boer et al. 2016), addressed this topic in the Coupled Model Intercomparison Project (CMIP) organized by the World Climate Research Programme (WCRP) and an operational annual to decadal prediction is organized by the world meteorological Organisation Lead center for annual-to-decadal prediction (Hermanson et al. 2022).

There are various schemes for accurately initializing S2D predictions. One common practice is to initialize each component of the Earth System Models (ESMs) individually,

✉ Lilian Garcia-Oliva
  lilian.garcia@uib.no

1 Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, 5007 Bergen, Norway

2 Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, 5006 Bergen, Norway

Ⓢ Springer

replacing them with an existing reanalysis (Balmaseda et al. 2009), but this can lead to initialization shock. Producing initial conditions with the same ESM used for performing the predictions can overcome this issue (Pohlmann et al. 2009). These techniques can use the data as it is (i.e., full-field; FF), or they can use anomalies about a climatology (i.e., anomaly-field; AF) (Smith et al. 2013; Volpi et al. 2017). Other initialization approaches include atmospheric momentum fluxes initialization, joint atmospheric momentum and heat fluxes initialization (Yeager et al. 2012), ocean data assimilation (ODA) (Wang et al. 2019; Brune and Baehr 2020), and a combination of ODA and atmospheric fluxes initialization (Brune et al. 2018; Polkova et al. 2019; Lu et al. 2020).

There is a debate on whether AF or FF initialization is best (Magnusson et al. 2013; Carrassi et al. 2014). Climate models have biases (climatological error) larger than the signals we aim to predict (Palmer and Stevens 2019), which causes challenges when comparing the two initialization approaches (Dee 2006). FF aims to correct the error in the mean state, which can be important for predictability. However, FF tends to produce a large drift during the prediction as the model reverts to its attractor (Smith et al. 2013; Weber et al. 2015). This technique can be skillful if the drift does not interfere with the signal, as the drift can be subtracted in a post-processing step (Yeager et al. 2012). Conversely, AF assumes that reducing the forecast drift will lead to fewer errors than correcting the mean error in the initial state (Smith et al. 2013; Weber et al. 2015). It thus only constrains the error of the anomaly and reduces prediction drift and initialization shock, i.e., produces a dynamical adjustment that differs from internal variability and external forcing response. Both techniques have strengths and weaknesses, which can be more important depending on the application. For instance, initialization shocks dissipate rapidly in the atmosphere but take much longer in the ocean. Furthermore, FF has other disadvantages when used in data assimilation (DA) methods: (1) When the bias is redundant (reemerging in between the assimilation cycle) and the observation network heterogeneous (e.g., with observations predominantly at the ocean surface), full-field assimilation and multivariate updates propagate the bias to the unobserved regions. (2) DA is designed to correct random, zero-mean errors, i.e., the model and observations are assumed (erroneously) to be unbiased. Consequently, the analysis state with FF still includes part of the bias; finally, (3) with ensemble methods, FF also yields a too strong reduction of ensemble spread (Dee 2006; Anderson 2001). On the other hand, the drawbacks of AF arise when (1) the variability of the model and observations are not comparable (Weber et al. 2015), for example, if the model bias is also characterized by a spatial shift impacting the amplitude

of the variability (Volpi et al. 2017), and (2) the nonlinear relationship between non-observed variables and assimilated variables introduce physical inconsistencies (Robson 2010; Yeager et al. 2012). The choice of initialization technique depends on the prediction's timescale considered. For sub-seasonal-to-seasonal (S2S) predictions, FF is often preferred, while for S2D, about half of the prediction systems are initialized using AF (Meehl et al. 2021). This illustrates such debate.

Most of the predictability in S2D timescales resides in the ocean's slow variability—largely driven by the atmosphere—, and several studies have explored different DA methods, observation networks, and the importance of ocean–atmosphere coupling during initialization. For example, constraining the fluxes at the ocean surfaces of an Ocean General Circulation Model (OGCM, e.g., Yeager et al. 2012) or nudging the atmosphere of the coupled system (Brune and Baehr 2020) can be effective to initialize the ocean component. Another approach having a comparable impact is to nudge the SST, which prescribes the flux at the ocean interface (e.g., Keenlyside et al. 2008; García-Serrano et al. 2015; Smith et al. 2013). It is also possible to focus on the ocean component initialization within the ESM—commonly called coupled initialization—(e.g., Zhang et al. 2009; Pohlmann et al. 2009; Karspeck et al. 2018; Counillon et al. 2016; Brune and Baehr 2020; Bethke et al. 2021). Coupled initialization approaches usually rely on advanced DA methods that can provide multivariate updates of the entire ocean state and take full advantage of the sparse ocean observation network. The joint initialization of the ocean subsurface and atmosphere has been advocated (for example, Smith et al. 2013; Polkova et al. 2019). In idealized studies Zhang et al. (2009, 2010) show that joint assimilation of atmosphere and SST can accurately reproduce the variability of the Atlantic meridional overturning circulation (AMOC) and that complementing the system with subsurface data improved performance in the North Atlantic (NA), proving its potential to initialize decadal predictions. Furthermore, Dunstone and Smith (2010) indicates that the subsurface can skillfully initialize the AMOC and that complementing with atmospheric data improves the initialization during the first lead year.

Isolating the best scheme is challenging since these schemes have been evaluated using different ESMs, reference periods, observational data sets, and experimental designs, which can lead to differences in prediction accuracy. Thus, there is a need to evaluate these schemes under a unified methodology. Here, we evaluate various initialization schemes for S2D predictions using the same prediction system—the Norwegian Climate Prediction Model—and the same experimental design. We will assess the performance of coupled reanalysis, seasonal hindcasts, and decadal hindcasts from 1980 to 2010. We will examine the advantages

of using full-field or anomaly-field initialization and explore the benefits of constraining the atmosphere, the ocean, or both components.

We use the Norwegian Climate Prediction Model (Nor-CPM, Counillon et al. 2014, 2016) that combines the Norwegian Earth System Model (NorESM, Bentsen et al. 2013) and the Ensemble Kalman Filter (EnKF, Evensen 2023) data assimilation method. NorCPM contributes to the CMIP6 DCPP (Bethke et al. 2021) and to the operational Lead Centre for Annual-to-Decadal Climate Prediction (Hermanson et al. 2022). NorESM is a state-of-the-art climate model based on the Community Earth System Model (CESM1, Hurrell et al. 2013), with the difference that it uses an ocean component with isopycnal vertical coordinates, different atmospheric chemistry, and ocean biochemistry. The EnKF is an advanced data assimilation method that corrects unobserved variables through a state-dependent multivariate covariance matrix and the observation error statistics. The model covariances are derived from a Monte-Carlo simulation. NorCPM performs monthly anomaly assimilation of SST, and temperature and salinity profiles. To initialize the atmospheric state, we use the Newtonian relaxation (nudging) towards the ERA-interim reanalysis (Dee et al. 2011).

This paper is organized as follows. Section 2 presents the practical implementation of NorCPM: the description of the ESM, NorESM, the data assimilation method, and the nudging implementation; it also introduces the validation data sets and metrics and describes the experimental setup. Sections 3.1, 3.2.1 and 3.2.2 present and discuss the result of the reanalysis, and the seasonal and decadal hindcasts. Finally, a discussion and summary are presented in Sect. 4.

2 Methods

2.1 Norwegian Earth System Model

The Norwegian Earth System Model (NorESM, Bentsen et al. 2013) is a global, fully coupled climate model based on the Community Earth System Model (CESM1, Hurrell et al. 2013). It uses the same ice and land components as CESM1: Los Alamos Sea Ice Model (CICE4, Bitz et al. 2012) and the Community Land Model (CLM4, Lawrence et al. 2011), respectively. Its atmospheric component is CAM4-OSLO, which is a version of the Community Atmosphere Model (CAM4, Neale et al. 2010) with modifications in the aerosol, chemistry, and cloud-aerosol interaction schemes (Kirkevåg et al. 2012). The ocean component is the Bergen Layered Ocean Model (BLOM, Bentsen et al. 2013; Danabasoglu et al. 2014), a modification of the Miami Isopycnal Coordinate Ocean Model (MICOM, Bleck and Smith 1990; Bleck et al. 1992), using density as its vertical coordinate.

We use the medium-resolution version of NorESM (NorESM1-ME). The atmosphere and land components use a 1.9°×2.5° regular horizontal grid. The atmosphere component uses 26 hybrid sigma-pressure levels. The horizontal resolution for the ocean and ice components is approximately 1°. It is enhanced in the meridional direction at the equator and both zonal and meridional directions at high latitudes. The ocean uses 51 isopycnal vertical levels and includes two additional layers of time-evolving thicknesses and densities representing the bulk mixed layer. External forcings used here comply with CMIP5 historical forcings (Taylor et al. 2012) and the RCP8.5 (van Vuuren et al. 2011) beyond 2005.

While RCP4.5 is more commonly used for climate prediction, we do not expect our results to be sensitive to the scenario choice since inter-scenario differences are still small at the beginning of the 21st Century (Kirtman et al. 2013). In contrast, we expect some impact from the use of CMIP5 versus CMIP6 forcings. Bethke et al. (2021) and Passos et al. (2023) found that upgrading to CMIP6 forcings degraded NorCPM's baseline climate and hindcast performance. Hence, we use CMIP5 forcings in this study.

2.2 Ocean data assimilation with the EnKF

The Ensemble Kalman Filter (EnKF, Evensen 2023) is a sequential data assimilation methodology consisting of a forecast and an update phase (analysis). During the first phase, the ensemble of states (ensemble) is integrated forward in time (forecast) from the previous ensemble of analysis states. During the second phase, observations are used to update (analyze) the ensemble for the next iteration. The method uses the ensemble covariance to provide flow-dependent correction, and it performs a linear analysis update, which preserves the linear properties (such as geostrophy).

We denote the ensemble forecast $\mathbf{X}^f \in \mathbb{R}^{n \times N}$. The superscript f stands for forecast, N is the ensemble size, and n is the dimension of the state. The model error is assumed to follow a Gaussian distribution with zero mean. The ensemble mean is denoted $\mathbf{x}^f$ and the ensemble anomalies are $\mathbf{A}^f = \mathbf{X}^f - \mathbf{x}^f \mathbf{1}^T$, where $\mathbf{1} \in \mathbb{R}^{N \times 1}$ has all its values equal to 1. Under the aforementioned hypothesis, the ensemble covariance $\mathbf{P}$ is an approximation of the forecast error ϵ:

$$\overline{\epsilon \epsilon^T} \approx \mathbf{P} = (N-1)^{-1} \mathbf{A}^f \mathbf{A}^{fT}. \tag{1}$$

We use the Deterministic EnKF (DEnKF, Sakov and Oke 2008), a deterministic formulation of the EnKF. The forecast ensemble mean is updated as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{d} - \mathbf{H}\mathbf{x}^f); \tag{2}$$

and the update of the ensemble anomaly is:

$$\mathbf{A}^a = \mathbf{A}^f - \frac{1}{2}\mathbf{KHA}^f. \tag{3}$$

The superscript a denotes the analysis, and f the forecast. $\mathbf{d} \in \mathbb{R}^{m \times 1}$ is the observation vector with m number of observations, and an associated error covariance $\mathbf{R}$; $\mathbf{H}$ the observation operator which relates the forecast model state variables to the measurements. Finally, $\mathbf{K}$ is the Kalman gain:

$$\mathbf{K} = \mathbf{PH}^T(\mathbf{HPH}^T + \mathbf{R})^{-1}. \tag{4}$$

Then, the full ensemble analysis $\mathbf{X}^a$ can be reconstructed:

$$\mathbf{X}^a = \mathbf{x}^a \mathbf{1}^T + \mathbf{A}^a. \tag{5}$$

We perform a monthly assimilation cycle, which updates the ESM's ocean and sea ice component in the middle of the month as described in Bethke et al. (2021) (the i2 system). The other components (atmosphere and land) adjust dynamically during the assimilation cycle. We assimilate SST from the HadISST2 data set (John Kennedy, personal communication, 2015; Nick Rayner, personal communication, 2015) and hydrographic profiles from EN4.2.1 (Gouretski and Reseghetti 2010). The observation error for the hydrographic profiles and the localization radius varies with latitude, as described by Wang et al. (2017). The localization radius has a bimodal Gaussian function that varies with latitude. At the Equator, the radius has a local minimum of 1500 km, where covariances become anisotropic. In the mid-latitudes, the radius reaches a maximum of 2300 km. There is another minimum in the high latitudes, where the Rossby radius is small (Wang et al. 2017). The observation error of the hydrographic profiles are a function of depth from Stammer et al. (2002), based on Levitus et al. (1994) and Levitus and Boyer (1994). We update the full isopycnal state variable in the vertical. We employ the aggregation method for layer thickness (Wang et al. 2016). The method is a cost-efficient modification of the linear analysis update in data assimilation for physically constrained variables. It ensures that the analysis satisfies physical bounds without changing the expected mean of the update and thus avoids introducing a drift. We use the rfactor inflation method where the observation error is inflated by a factor 2 for the update of the ensemble anomaly (Eq. 3) and the k-factor formulation in which observational error is artificially inflated if the assimilation pushes the update beyond two times the ensemble spread (Sakov et al. 2012). We use an anomaly assimilation technique to remove the climatological monthly difference between the observations and the model. The monthly climatological mean of the model is estimated from the 30-member historical ensemble for the period 1980–2010. The climatological mean for the hydrographic profiles is calculated from the EN4 objective analysis (Good et al. 2013). The EnKF implementation in NorCPM works offline—meaning

that the model is stopped, the state is written on disk, the data assimilation is applied to the files, and the model is restarted.

2.3 Atmospheric nudging

Nudging is a simple method to constrain the evolution of a system towards a prescribed dataset (Hoke and Anthes 1976). It does not consider the uncertainty of the observations and only applies a constraint on the variables nudged (monovariate). However, it is computationally cheap, implemented in most ESMs, and works online. This is beneficial since the time required for initializing the model and writing the input/output is burdensome with large systems. This is the case for the initialization of the atmospheric state that requires 6-hourly updates (see, e.g., Karspeck et al. 2018).

Nudging works by adding a term (nudging tendency) that is applied at the model time step to the prognostic (or tendency) equations:

$$\frac{\partial X_m}{\partial t} = -\frac{X_m - X_p}{\tau}, \tag{6}$$

where X stands for the variable to nudge, and the subscripts m and p identify the model predicted and the prescribed values. The formulation in Eq. (6) corresponds to full-field nudging. The constant τ is the relaxation time scale—how strong the model is attracted to the prescribed dataset. This parameter value is selected to avoid dynamic shocks and to counteract the error growth (Carrassi et al. 2014). The prescribed value can be either from reanalysis data or the model itself (Zhang and Wan 2014).

One can also apply anomaly nudging (Zhang and Wan 2014), where the right-hand side of equation (6) is replaced by the anomaly terms, i.e., $X \rightarrow A$. Thus, $A = X - \overline{X}$ and $\overline{X}$ is the climatological seasonal cycle. The anomaly nudging tendency is:

$$\frac{\partial X_m}{\partial t} = -\frac{A_m - A_p}{\tau}. \tag{7}$$

Considering the model and prescribed data anomalies (A_m and A_p) and re-arranging the terms, the anomaly nudging tendency can be formulated as a function of the model state X_m and a new prescribed term:

$$X_p^* = X_p - \overline{X}_p + \overline{X}_m. \tag{8}$$

Using the new prescribed term, the Eq. (7) can be expressed as:

$$\frac{\partial X_m}{\partial t} = -\frac{X_m - X_p^*}{\tau}. \tag{9}$$

With the formulations of Eqs. (6) and (9), we can perform both full-field and anomaly nudging without having to modify the model code, and by changing only the input data used.

We use the nudging implementation described in Kooperman et al. (2012) and Zhang and Wan (2014). We nudge at every atmospheric model time step (30 min) with relaxation time scale $\tau = 6$ h towards fields from the 6-hourly reanalysis product ERA-Interim (ERA-I, Dee et al. 2011) linearly interpolated in space and time to our model grid. For anomaly nudging, we compute the monthly climatology for the model (from Free, see Table 1) and ERA-I for the period 1980–2010. We interpolate these monthly climatologies linearly to the model time without correcting for biases in the diurnal cycle. Additionally, we nudge surface pressure and apply a correction to the barotropic wind accordingly. In the vertical, nudging is performed on atmospheric model levels below 60 km height with tapering between 50 and 60 km. The surface state is indirectly constrained by the atmospheric state on the model levels, with the lowest level located approximately 60 m above ground.

In CAM, an energy fix is applied to preserve energy in the system during the model integration. When nudging temperature, one modifies the energy in the atmospheric component. A common practice is, thus, to switch off the energy fix and let the energy in the atmosphere converge to that of the target data set. However, when one only nudges winds, energy is no longer sustained. We will therefore consider the impact of nudging the winds without the energy fix activated (default in CAM4) with a version where the energy fix is reactivated.

2.4 Experimental design

We evaluate six different initialization schemes (Table 1), assessing both the accuracy of the reanalyses and the skill of S2D predictions. Two schemes, NudF-UVT and NudA-UVT, use FF and AF atmospheric nudging of horizontal wind and temperature fields (U, V, T). The schemes NudA-UV and NudA-UV (EF) use anomaly atmospheric nudging of the horizontal wind field (U, V), with the difference that the latter imposes energy conservation (EF) in addition (see Sect. 2.3).

A fifth scheme, ODA, constrains ocean variability. We perform anomaly assimilation of SST and vertical temperature and salinity (T, S) profiles with the EnKF (see Sect. 2.2 for details on the practical implementation). Finally, the scheme ODA+NudA-UV combines the ODA and NudA-UV (EF) experiments. We did not combine ODA with full field atmospheric nudging as it would have caused a mismatch of the mean state because our ODA scheme assimilates anomalies (see Counillon et al. 2016, for detailed justification).

All the schemes produce a reanalysis with a 30-member ensemble of NorESM1-ME (Sect. 2.1). The ensemble of initial conditions for all reanalyses is identical and produced by randomly selecting states from a stable pre-industrial simulation and integrating it with historical forcing from 1850 to 1980. The 30-member reanalyses of each initialization method are used as initial conditions for our seasonal-to-decadal hindcasts. The simulation (typical historical ensemble) run without assimilation is called Free and is used to identify the skill associated with external forcing.

The seasonal-to-decadal hindcasts comprise 104 seasonal hindcasts (26 years with four hindcasts per year) and 13 decadal hindcasts for each of the six initialization schemes. The seasonal hindcasts start on the 15th of January, April, July, and October each year during 1985–2010 and run for a year. The decadal hindcasts begin on October 15th every other year between 1985 and 2010 (i.e., 1985, 1987, 1989,...2009), with a duration of 122.5 months, which account for 10 complete calendar years (January–December), plus 2.5 months from the initialization year. Thus, our last decadal hindcast is initialized on October 15th, 2009, covering until 2019. Each hindcast comprises nine realizations (ensemble members). The ensemble size is a compromise between computational cost and skill accuracy. It is close to the 10 members recommended from established climate prediction protocols (e.g., Boer et al. 2016) that has been shown to yield an atmospheric S2S forecast skill that is approximate 90 % of the theoretical maximum skill (Han et al. 2023).

Table 1 Configurations summary

Configuration	Ocean DA	Atmo nud (6 h)	Assimilated variables[a]	E. F.[b]
Free	–	–	–	Yes
NudF-UVT	–	FF	(U, V, T)	–
NudA-UVT	–	AF	(U, V, T)	–
NudA-UV	–	AF	(U, V)	–
NudA-UV (EF)	–	FF	(U, V)	Yes
ODA	AF	–	[SST, T, S]	Yes
ODA+NudA-UV	AF	AF	[SST, T, S] + (U, V)	Yes

[a]Variables in squared brackets (parenthesis) denote ocean (atmosphere) observations

[b]E. F. is for Energy Fix

2.5 Assessment: data and metrics

This section describes the metrics and datasets we used to assess our initialization schemes.

2.5.1 Metrics

We base our analysis on monthly anomalies. We computed all climatologies (seasonal cycles) over the 1980–2010 period. We calculate the anomalies for the reanalyses by subtracting their corresponding climatological seasonal cycle from the monthly average. We obtain the hindcast anomalies after performing a drift correction, which we assume to be lead-time (month or year) dependent. Thus, the hindcast anomalies are computed relative to the average of the N_h hindcasts:

$$X'_{jt} = X_{jt} - N_h^{-1} \sum_{k=1}^{N_h} X_{kt}. \tag{10}$$

X_{jt} and X'_{jt} are the raw and anomalies (drift-corrected) values for hindcast j at the lead time t. The observation anomalies are obtained by removing the corresponding climatology from the dataset.

We assess the system's skill using the following metrics: unbiased root mean squared error RMSE_u, and the anomaly correlation coefficient ACC. The RMSE_u and ACC are defined as:

$$\text{RMSE}_u = \left(N^{-1} \sum_{k=1}^{N} (X'_k - Y'_k)^2 \right)^{1/2}, \tag{11}$$

$$\text{ACC} = \sum_{k=1}^{N} X'_k Y'_k \left(\sum_{k=1}^{N} X'^2_k \sum_{k=1}^{N} Y'^2_k \right)^{-1/2}, \tag{12}$$

where X'_k and Y'_k are the reanalysis (or hindcast) and observation anomalies at month (lead-time) k; and N is the evaluation period's length. Since the assessment is based on the anomalies, the RMSE_u does not penalize if the reanalysis has a bias or if the hindcasts drift with lead time. Similarly, the ACC is insensitive to bias (Wilks 2019). To account for the sampling error in the estimation of the ACCs, we compute the 95 % confidence interval based on Fisher's z-transform (Hv and Zwiers 1999). We used the two-sided Student's t-test to assess the significance of the ACCs estimate at a 95 % confidence level; additionally, we assumed the data to be uncorrelated due to the small sample size (Hughes and Hase 2010).

For the reanalysis, we also computed the climatological change ΔBIAS, defined as the deviation of the reanalysis monthly climatology to that of Free during the reanalysis:

$$\Delta\text{BIAS} = \sum_{t=1}^{N} (\overline{X}_t^R - \overline{X}_t^F). \tag{13}$$

$\overline{X}_t^R$ is the monthly climatology of the reanalyses and $\overline{X}_t^F$ that of Free with $N = 1, \dots, t, \dots, 12$ being the calendar months.

In a reliable system, the total error σ should match RMSE_u (Fortin et al. 2014; Rodwell et al. 2016), thus:

$$\text{RMSE}_u = \sigma = (\sigma_o^2 + \sigma_m^2)^{1/2}, \tag{14}$$

where the total error is the quadratic sum between the ensemble spread σ_m, and the observation error σ_o, and RMSE_u is defined in Eq. (11). For our nudging implementation, since this assimilation method does not consider the uncertainty (error) of the observations, the total error σ is equal to model spread σ_m, thus $\text{RMSE}_u = \sigma_m$.

In the case of the hindcasts (seasonal or decadal), we also used persistence as a benchmark for skill. Thus, the persistence P starting at i-th month (year), at lead-time (month or year) k is:

$$P_{ik} = Y'_i, \tag{15}$$

Y'_i is the observation anomaly of i-th month (year), and $k = 1, \dots, N$; where N is the hindcast length. This is equivalent to constructing the autoregressive order zero model AR(0). Thus, for a seasonal hindcast initialized in April 2000, we constructed the corresponding AR(0) persistence forecast as the April 2000 monthly mean of the deseasoned observation Y'_i.

For the global statistics, we use grid cell area weighting:

$$\text{RMSE}_u = \sum_i a_i \text{RMSE}_{ui} \left(\sum_j a_j \right)^{-1}, \tag{16}$$

and

$$\text{ACC} = \sum_i a_i \text{ACC}_i \left(\sum_j a_j \right)^{-1}. \tag{17}$$

where a_i is the area of the corresponding i-th grid cell. These statistics will measure the effectiveness of our implementations in improving spatial variability. Meanwhile, the regional indices are calculated using area weighting, followed by the estimation of statistics.

2.5.2 Datasets

To validate the reanalysis and hindcasts, we take 2 m temperature (T2M) data from the ERA5 reanalysis (ERA5,

Hersbach et al. 2020), with a horizontal resolution of 0.25° × 0.25°, which we re-grid to the CAM4 model grid. ERA5 is the updated atmospheric reanalysis version of ERA-I and thus cannot be considered independent for the system that uses atmospheric nudging. Observations can be correlated in time, and we can assess the data set's independence in forecast mode by comparing it with persistence,—i.e., the forecasting system must beat persistence (observation auto-correlation) to demonstrate skill.

For the ocean surface temperature, we take SST observations from the Hadley Centre Sea Ice and Sea Surface Temperature dataset HadISST2. We interpolate our ocean outputs towards the HadISST2 horizontal grid. Since Had-ISST2 is assimilated in ODA, it cannot be considered independent for the reanalysis if it does not beat persistence for the hindcast.

We obtain subsurface temperature and salinity data from the EN4.2.1 objective analysis (EN4.2.1, Good et al. 2013). We re-grid and interpolate our ocean subsurface output to EN4 objective analysis resolution for the comparisons. The EN4 objective analysis is not independent of our ODA reanalysis because it uses the same raw observations (i.e., the EN4.2.1 hydrographic profiles Gouretski and Reseghetti 2010). Still, this comparison is of interest because objective analysis and model reanalysis are different in construction. Objective analysis provides a 4D interpolation of the observations without dynamical constraints but reverts to climatology if no observations are available, while model reanalyses provide dynamical reconstructions and can thus propagate improvement to the unobserved regions but are also limited by model error (Storto et al. 2019). The accuracy of the objective analysis is highly dependent on the available observations, which are sparse in the Southern Hemisphere but reasonably good in the North Atlantic. We define the heat and salinity content in the first 500 m, as HC500 and SC500, respectively. We define them as the ocean depth's average temperature (and salinity).

We also use the Atlantic meridional overturning circulation (AMOC) at 26° North from the RAPID dataset (Smeed et al. 2015).

3 Results

In this section, we evaluate the performance of each data assimilation scheme to provide skillful reanalysis (Sect. 3.1), seasonal (Sect. 3.2.1) and decadal (Sect. 3.2.2) predictions.

3.1 Reanalysis

We first compare the quality of the reanalyses using atmospheric nudging with FF (NudF-UVT) and AF (NudA-UVT). Both schemes have similar global ACC and RMSE$_u$ for all evaluated quantities (Fig. 1). Globally, the reanalysis from NudF-UVT is marginally better for SST and T2M (Fig. 1a, b), but yields a degradation for HC500 (Fig. 1c) and SC500 (Fig. 1d). Most of this degradation occurs in the subpolar gyre (SPG), the tropical and South Atlantic, and the Southern Ocean (Fig. 2d). Furthermore, NudF-UVT exhibits a substantial RMSE$_u$ drift of HC500 and SC500 (Fig. 3). Such RMSE$_u$ drift follows a parabolic shape, as the mean climatology (used for computing the metric, Eq. 11) is reached

Fig. 1 Global statistics of the reanalyses computed over 1980–2010, for **a** SST, **b** T2M, **c** HC500, and **d** SC500. The left-hand y-axis (in black) displays units for RMSE$_u$ (magenta), ΔBIAS (cyan), and total error (yellow), while the red right-hand y-axis is for ACC (red). The reanalyses are said to be reliable when the total error (yellow) and RMSE$_u$ (magenta) overlap. The black horizontal line marks zero, and the red dashed line marks the 95 % significance level for ACC
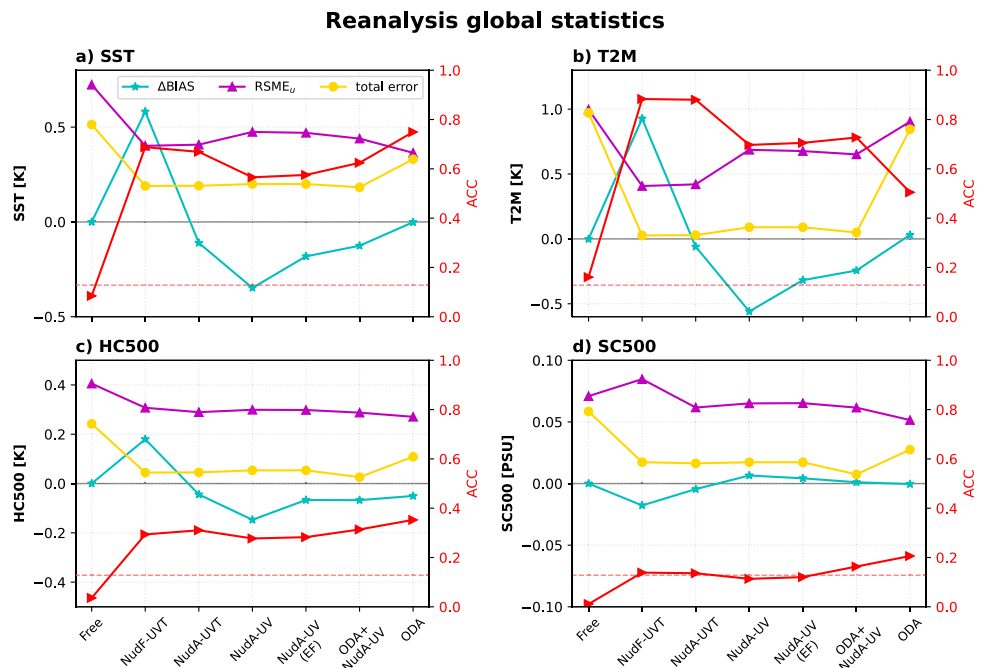
Fig. 2 ACC of monthly HC500 anomalies **a** NudA-UVT, **b** NudA-UV, **c** ODA, **d** NudF-UVT, **e** NudA-UV (EF) and **f** ODA+NudA-UV reanalysis computed against EN4 objective analysis for the period 1980–2010. Green-to-magenta colors indicate positive ACCs, and the cyan color indicates all the negative ACCs
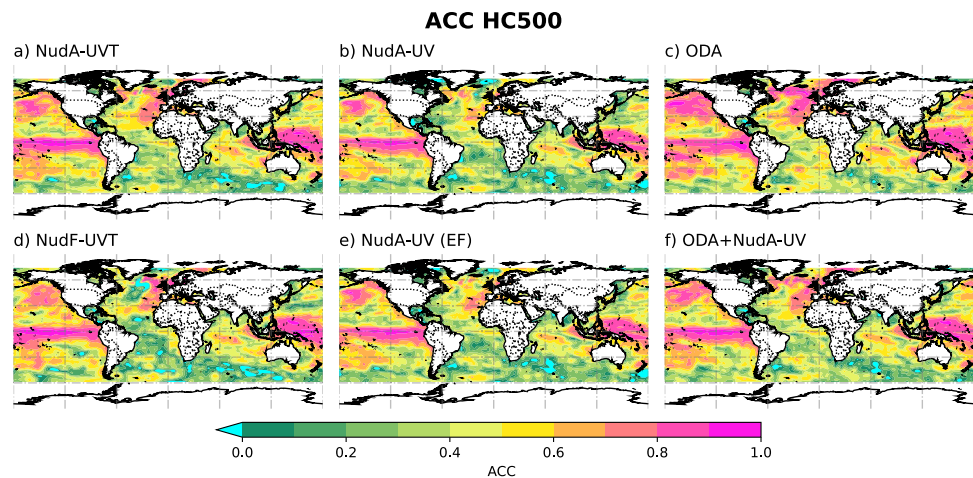

ACC HC500

a) NudA-UVT b) NudA-UV c) ODA

d) NudF-UVT e) NudA-UV (EF) f) ODA+NudA-UV

ACC

Fig. 3 Time series of RMSE$_u$ for **a** HC500 and **b** SC500 in the different reanalyses computed against EN4 objective analysis. Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, magenta to ODA+NudA-UV, and brown is Free

Reanalysis RMSE$_u$ time series

a) HC500 b) SC500
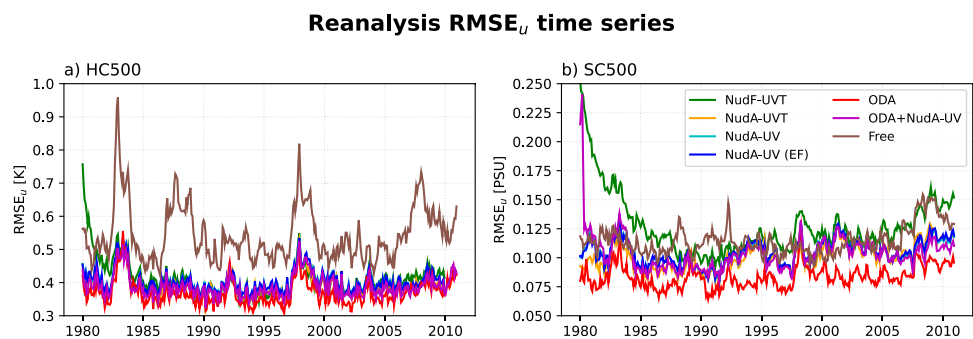
halfway through the reanalysis period. In contrast, the reanalysis provided by NudA-UVT does not have the drift in HC500 RMSE$_u$, while in SC500, the RMSE$_u$ has a much weaker trend than in NudF-UVT. Additionally, the use of FF atmospheric nudging—of U, V, T—introduces a large change in the climatology (ΔBIAS in Fig. 1). Such change in the model climatology is expected as the FF technique shifts the model attractor towards the observed climatological state. For SST and T2M, ΔBIAS is larger than RMSE$_u$. Both schemes yield poor global ensemble reliability near the surface, with the estimated total error (Eq. 14) being much smaller than the RMSE$_u$ (Fig. 1a, b). This implies that the ensemble spread nearly collapses during the reanalyses (Figure 1a and 1b in Supplementary Information, SI). The reliability for HC500 and SC500 is also poor (Fig. 1c, d). It should be acknowledged that the HC500 (and, to a minor extent, SC500) reliability of Free is already too low. Our Free is a pure Monte Carlo simulation where the initial condition samples the possible state in the pre-industrial state and is run forward until the present day. Monte Carlo is well suited to estimating the time evolution of the probability density function of the model state (Evensen 1994). The low reliability in Free suggests that the model has some intrinsic limitations (e.g., signal-to-noise) or that the observation error estimate from EN4 objective analysis is too low.

Still, when applying the nudging, the ensemble uncertainty is reduced more than the error of the ensemble mean, and the reliability is further degraded. In the SPG (Fig. 4a, b), both schemes capture well the timing of the rapid shift in the gyre index in 1995, but only NudA-UVT reproduces the amplitude of the shift correctly. This abrupt shift is linked to the North Atlantic Oscillation (NAO) influence (Häkkinen and Rhines 2004; Yeager and Robson 2017), which induces a preconditioning of the ocean circulation state (Lohmann et al. 2009; Robson et al. 2012). Moreover, both schemes failed to sustain a weak SPG in the 2000s. NudA-UVT achieves overall better performance than NudF-UVT, which exhibits a drift from a too-weak SPG in the 1980s to a too-strong SPG in 2010. This likely relates to the strong decreasing trend in the AMOC in NudF-UVT (Fig. 5b) that affects the poleward heat transport. The verification period with the RAPID data is too short to hold a firm conclusion. Yet, NudA-UVT has a decreasing anomaly from 2005 in good agreement with observations, albeit missing the weakening in 2009, while NudF-UVT has an unrealistic decreasing trend.

We compare the schemes NudA-UV and NudA-UVT to assess the importance of constraining atmospheric temperature in addition to horizontal winds, compared to just constraining horizontal winds. At the surface (SST and

Fig. 4 HC500 anomalies in the SPG box (48°–65°N, 60°–15°S) for **a** NudA-UVT, **b** NudF-UVT, **c** NudA-UV, **d** NudA-UV (EF), **e** ODA and **f** ODA+NudA-UV reanalyses. Solid-colored lines represent the ensemble mean of reanalysis, dash-dotted lines correspond to the ensemble mean of the decadal hindcast, and the solid brown line is Free. Shading denotes ensemble minima and maxima. The solid black line shows the EN4.2.1 objective analysis estimate. The correlation coefficient R between reanalysis and observations is in the top-left-hand corner. Positive values of the index correspond to a weak SPG

North Atlantic SPG index
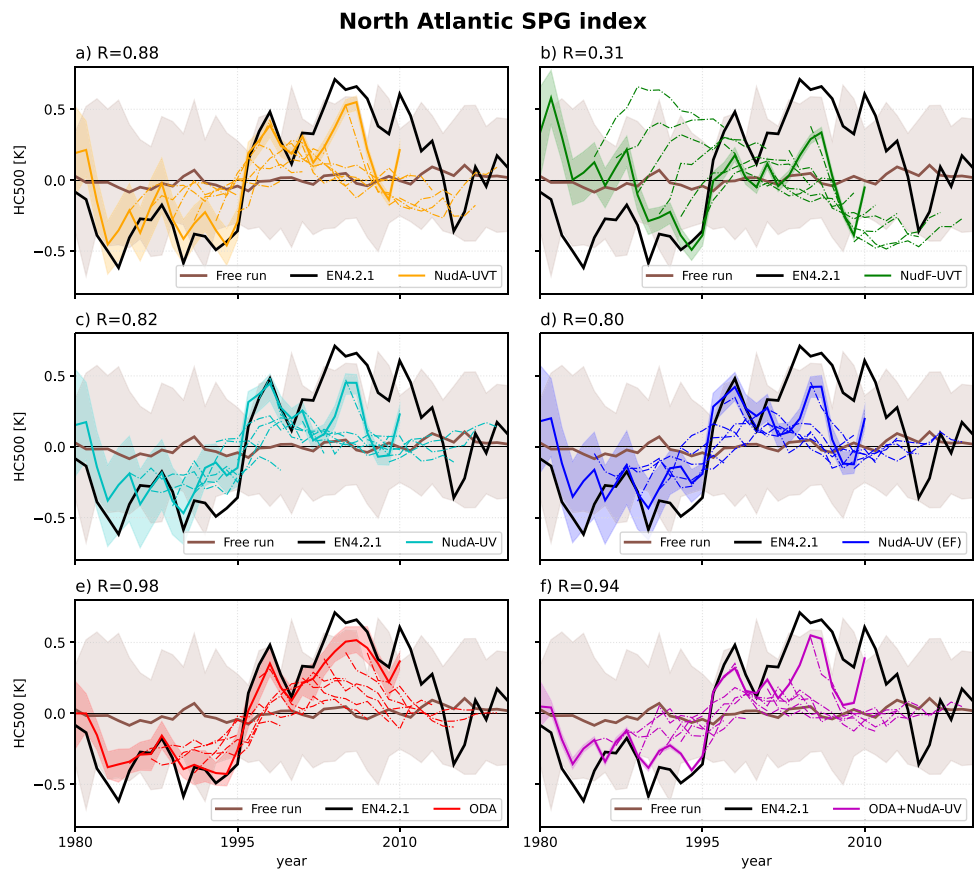
T2M), nudging only horizontal winds degrades performance (Fig. 1a, b). For T2M, for example, NudA-UV reduces error by 0.3 K compared to Free, whereas NudA-UVT reduces it by 0.6 K. Note that this is expected because T2M is nudged in NudA-UVT. The degraded performance of NudA-UV is largest over the tropical band and is less pronounced at mid-to-high latitudes (Fig. 6a, b). The reliability for T2M is slightly improved in NudA-UV compared to NudA-UVT (see also Table 1 in SI). Anomaly nudging imposes that the model remains on the climatological level of the nudged variables. However, deviations can still occur for other variables if they non-linearly respond to the variables that are nudged (e.g., the nudging imposes the observed wind variance, which, in turn, can change the wind-temperature covariance, altering the eddy component of heat transports). In NudA-UV, there is a slight climatological change (ΔBIAS) for SST and T2M, but NudA-UVT sustains ΔBIAS of SST and T2M near 0 K. The ΔBIAS for SST and T2M in NudA-UV is relatively small and much less than with full-field nudging. Below the surface, the global accuracy of NudA-UV and NudA-UVT are similar for HC500 and SC500 (Fig. 1c, d), with NudA-UV being slightly poorer. NudA-UV also impacts ΔBIAS of HC500, giving a larger negative bias than NudA-UVT. Most of the ACC differences for HC500 are in the Atlantic Ocean, specifically in

the Iceland basin (SI, Figures 3c and 3e), North East Atlantic, and South Pacific (Fig. 2a, b). The performance for the SPG (Fig. 4) and AMOC (Fig. 5) variability are comparable, with NudA-UV showing a slightly poorer match in the early 1990s. This suggests that wind-driven variability is not the sole factor determining the amplitude of the SPG, as NudA-UV cannot maintain a strong gyre.

The default implementation of nudging in CAM4 deactivates the energy conservation fix in the atmospheric component (see Sect. 2.3). Anomaly nudging imposes that the model remains on the climatological level of the variables that are nudged, but a deviation can still occur for other variables if there is a non-linear response with the variable that is nudged. Thus, we assess if conserving energy can reduce the climatology change by comparing ΔBIAS in NudA-UV with that of the NudA-UV (EF) experiment for which the global energy fixer is activated (Fig. 1a, d). Overall, the performance ($RMSE_u$, ACCs, and reliability) is unchanged, but the climatological change is reduced by half in NudA-UV (EF). However, we see that the HC500 skill in the Iceland Sea and into the Norwegian Sea differ in these two schemes. An analysis of the HC500 time series for the Iceland Sea further reveals that long-term trend and inter-annual variability contribute to the variability of the region (SI, Figure 3). And comparing NudA-UV and NudA-UV (EF), we find that the

Fig. 5 AMOC transport
anomalies at 26.°N with respect
to the 1980–2010 period for **a**
NudA-UVT, **b** NudF-UVT, **c**
NudA-UV, **d** NudA-UV (EF),
e ODA and **f** ODA+NudA-UV
reanalyses. Solid-colored lines
represent the ensemble mean
of reanalysis, dash-dotted lines
correspond to the ensemble
mean of the decadal hindcast,
and the solid brown line is Free.
Shading denotes ensemble
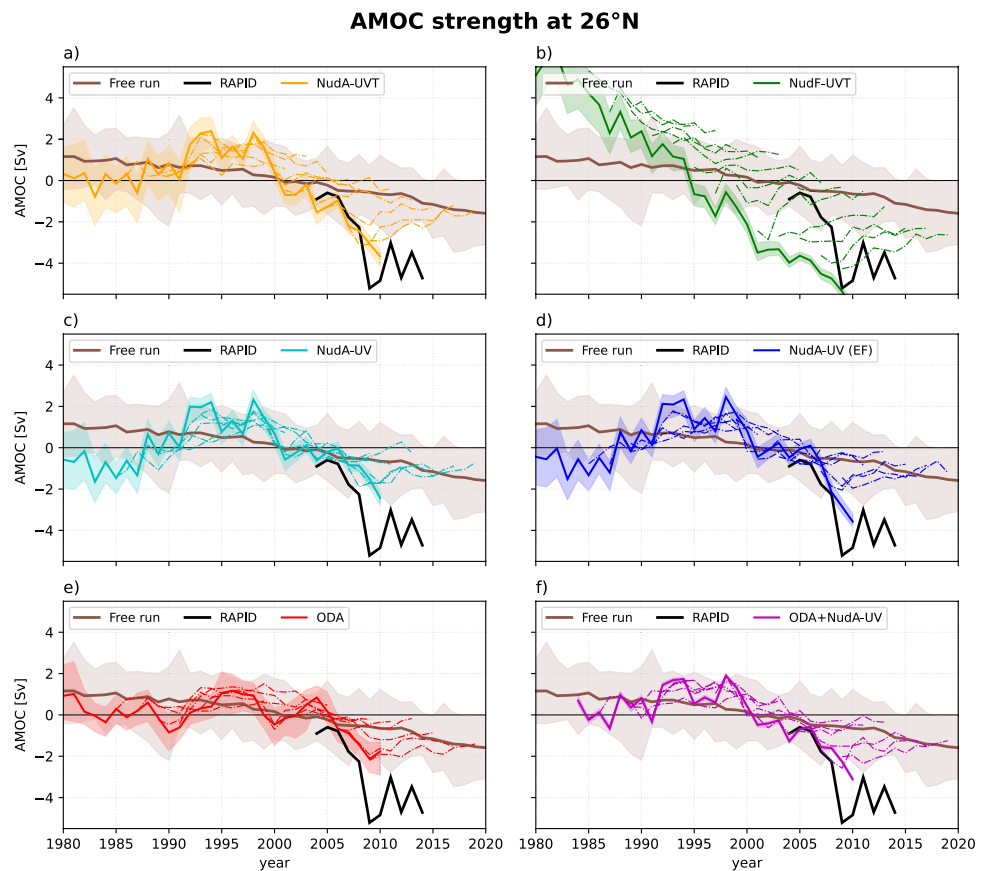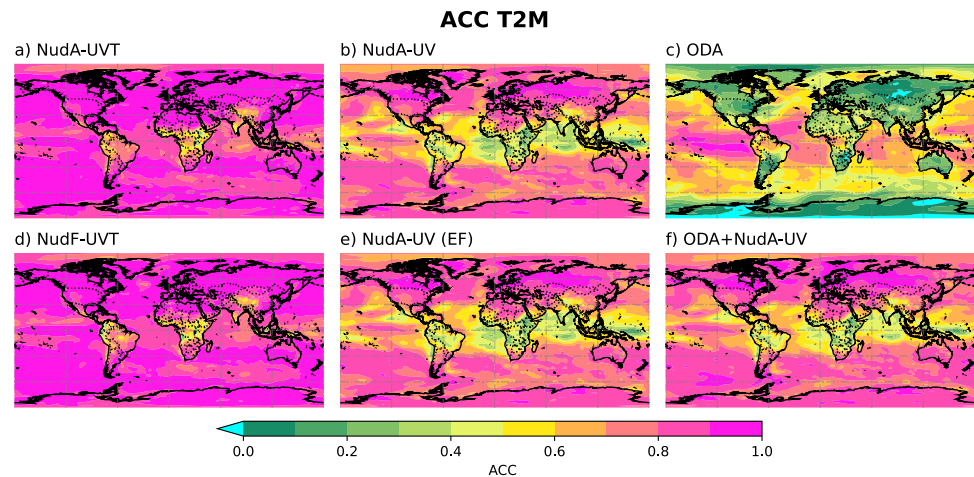minima and maxima. The
solid black line is the RAPID
observations

Fig. 6 ACC of de-seasoned
monthly T2M for **a** NudA-UVT,
b NudA-UV, **c** ODA, **d** NudF-
UVT, **e** NudA-UV (EF) and **f**
ODA+NudA-UV reanalyses
computed against ERA5 for the
period 1980–2010. Green-to-
magenta colors indicate positive
ACC values, and the cyan color
indicates all the negative ACCs

energy fix is very effective in improving the representation
of the trend in the Iceland basin ($R =0.34$ and 0.64, respec-
tively, in Figures 3e and 3 g in SI).

We now compare atmospheric constraints versus ocean
constraints for coupled reanalysis. The skill for T2M
(Fig. 1b) using atmospheric nudging is substantially bet-
ter than using ODA. The ODA system has skill over the
ocean (most pronounced over the tropical band) while skill
over land is poor in the extratropics and polar areas (Fig. 6a,

c). When comparing the T2M skill over the ocean with the
SST skill (not shown), atmospheric nudging works better
than ODA when using T2M. However, for SST, ODA was
found to be more effective. It is important to note that the
correlation between T2M and SST is strong and that the
choice of validation data sets can significantly affect skill
differences. The validation of SST is done against the Had-
ISST2 analysis, which is assimilated in the ODA system.
Meanwhile, the verification of T2M is done against ERA5,

which is similar to the ERA-I product used for atmospheric nudging. This slight contradiction highlights the uncertainties in the observation data sets (Massonnet et al. 2016; Bellprat et al. 2017). In the ocean interior, ODA outperforms all atmospheric nudging schemes (Fig. 1c, d). This is expected because the EN4 Objective analysis is not independent from the EN4.2.1 profiles assimilated. This is also clear from Fig. 3, where ODA has a consistently lower error than the nudging schemes and is the only system with stable $RMSE_u$ for SC500—that does not degrade with time. This stability implies that the strong constraint on the variability of the surface fluxes provided by atmospheric nudging is insufficient to guarantee a stable performance for the ocean interior, such as SC500. The benefit of the ODA over the nudging schemes is largest in the tropical Pacific, the northwestern Pacific, the Indian Ocean, and the SPG (Fig. 2c), where atmospheric nudging introduces a patch of low-skill in the Irminger and Icelandic Seas (see, for example, Fig. 2a, b). The reliability of the system is also better preserved as we see a closer match between $RMSE_u$ and total error σ (Fig. 1c, d; magenta and yellow lines; and SI, Table 1). In the ODA system, the reliability is only marginally degraded from Free and much less than atmospheric nudging. Since our Free is a pure Monte Carlo simulation, we expect it to correctly sample the system's state, and as such, we use it as a benchmark to calibrate our DA system. Therefore, if our DA system degrades reliability, compared to Free, this implies that our DA system is wrongly calibrated or that some of the assumptions made are not satisfied (e.g., observation error underestimated, model bias). In the case of regional indexes, ODA achieves overall the best correlation for the SPG HC500 index ($R = 0.98$, Fig. 4e), and it is the only system that sustains the weak (warm) SPG during the 2000s. However, the shift in 1995 is not as abrupt as in the observations and the atmospheric nudging schemes (see, for example, Fig. 4a). This is because the NAO constraint is very weak in the ODA system, and the system only adjusts a-posteriori through the surface fluxes. Finally, for the AMOC at 26.°N, there is a long-term weakening with a local maximum in 2006 that is underestimated by all systems. ODA is the only system that captured the rebound in 2009; however, it does not capture the local minimum in 2004 as with atmospheric nudging systems (Fig. 5c, e), suggesting that this feature is better constrained with atmospheric variability.

Given the complementary skills of atmospheric nudging and the ODA systems, one would expect their combination to work best as it makes use of two independent observation data sets and that covariance across the two components is not null (Penny et al. 2017). However, comparing the global statistics of ODA and ODA+NudA-UV (Fig. 1), we see that the use of atmospheric nudging in ODA+NudA-UV degrades performance in ocean quantities (SST, HC500,

and SC500). ODA+NudA-UV performs almost identically to NudA-UV. This is more evident at the surface (see T2M in Fig. 6b, c, f). This is because the ODA relies on the reliability of the system—the analysis update depends on the relative importance of the ensemble spread to the observational error— and, in our current implementation, the atmospheric nudging drastically reduces the ocean's ensemble spread (SI, Figure 1). This means that ocean observations have nearly no impact. However, the ODA+NudA-UV performs slightly better than NudA-UV for SST, HC500, SC500, and SPG (Fig. 4f), and AMOC (Fig. 5f) in good agreement with Brune et al. (2018), indicating that ODA yields improvements.

3.2 Predictions

In this section, we evaluate the skill of the seasonal and decadal hindcasts initialized from the reanalysis (see Sect. 2.4).

3.2.1 Seasonal predictions

Our prediction systems have a superior global surface skill compared to persistence starting from the third lead month (Fig. 7a, b). On the other hand, the prediction skill for HC500 is low and only beats persistence after the sixth month, while SC500 never outperforms persistence (Fig. 7c, d). However, it is possible that the skill of persistence is overestimated as it is computed from the same data set used for validation. This is likely the case for HC500 and SC500 since the observation error in the EN4 objective analysis is highly correlated in time due to the sparse in situ measurements. Comparing the different systems, the ODA system performs best for all assessed quantities (Fig. 7). Despite the T2M reanalysis using atmospheric nudging, which reached a better skill than with ODA, this skill is rapidly lost by lead month 1, with the highest skill in the tropical regions (SI Figure 2). This highlights the importance of ocean initialization in the prediction skill achieved, albeit being quite low overall.

While T2M and HC500 global average skill is low, with ACCs below 0.4 (Fig. 7b, c), some regions show enhanced skill (Figs. 8 and 9). The skill is most significant over the ocean and most notably in the tropical band driven by the El Niño–Southern Oscillation (ENSO) (Balmaseda and Anderson 2009; Meehl et al. 2021), the Indian Ocean Dipole (Saji et al. 1999; Webster et al. 1999), and, to a lesser extent, over the Atlantic Niño region (Keenlyside et al. 2020). In agreement with other climate prediction systems (e.g., Kirtman et al. 2014; Wang et al. 2019), a region of significant skill exists in the northern North Atlantic, the SPG, and the Iceland Sea.

Most of our experiments show good skill in predicting T2M and HC500 in the SPG at lead months 2–5. The best

Fig. 7 Global average ACC of the seasonal hindcast with lead month, for: **a** sea surface temperature (SST), **b** 2 m air temperature (T2M), **c** 500 m heat content (HC500), and **d** 500 m salinity content (SC500). Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, and magenta to ODA+NudA-UV. The solid black line is persistence, and the brown line is the Free run. The black dashed line marks the 95 % significance level
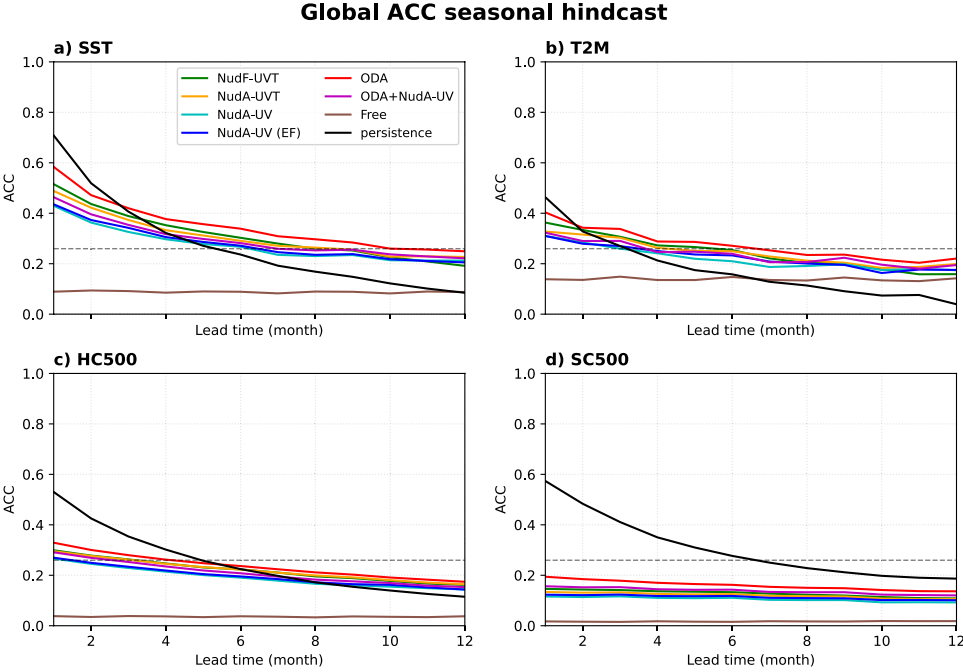

Global ACC seasonal hindcast

Fig. 8 Seasonal hindcast 2–5 lead-month T2M ACC for **a** NudA-UVT, **b** NudA-UV, **c** ODA, **d** NudF-UVT, **e** NudA-UV (EF) and **f** ODA+NudA-UV. Green-to-magenta colors indicate positive ACCs and cyan color indicates all negative ACCs

ACC T2M LM2-5
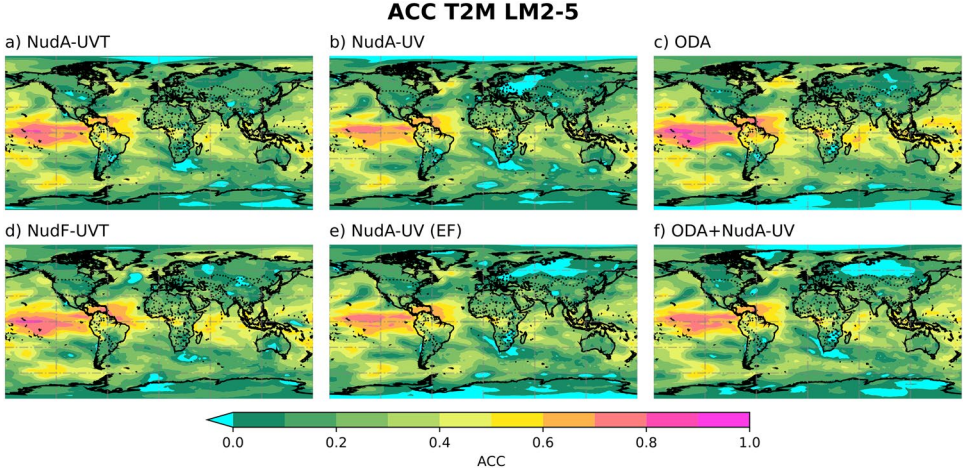
Fig. 9 ACC of the seasonal hindcasts at lead-month 2–5 for HC500 with: **a** NudA-UVT, **b** NudA-UV, **c** ODA, **d** NudF-UVT, **e** NudA-UV (EF) and **f** ODA+NudA-UV computed against EN4 objective analysis. Green-to-magenta colors indicate positive ACCs, and cyan indicates all negative ACCs

ACC HC500 LM2-5
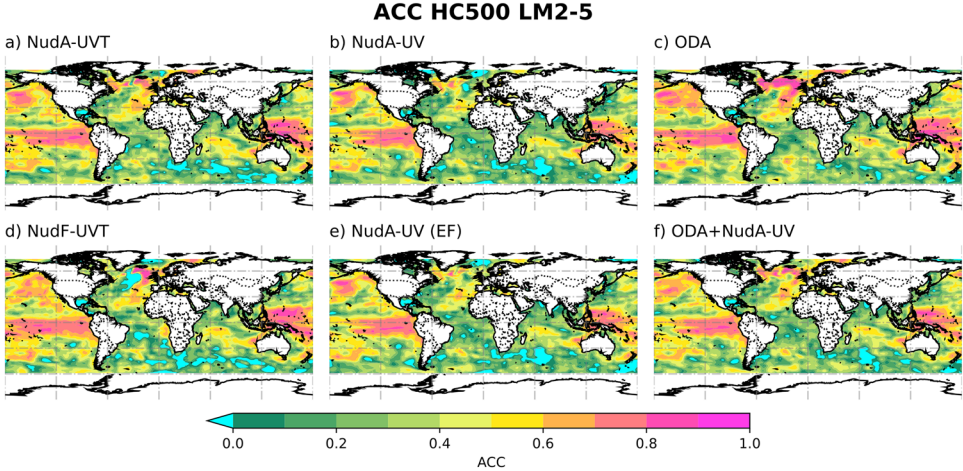
skill is achieved with ODA, and of all the nudging schemes, NudA-UVT performs best (Figs. 8 and 9). NudF-UVT performs poorly and even reaches a negative correlation in the Irminger Sea. This highlights that constraining the mean state error is not critical in this region and that simple lead-dependent drift post-processing is insufficient with our model, unlike in Yeager et al. (2012). On the other hand, in the Iceland Sea and into the Norwegian Sea, ODA again performs best, and it is clear that NudF-UVT and NudA-UVT outperform NudA-UV. This highlights the role of atmospheric heat flux in this region. The comparison between NudA-UV and NudA-UV (EF) highlights that correcting the spurious drift (see Sect. 3.1) in this region is important for predictive skill at seasonal scales.

We assess the prediction skill in the ENSO region by computing $RMSE_u$ and ACC of the Niño 3.4 index (mean SST within the box 5° S–5° N, 120° W–170° W) against HadISST2 observations with lead time (Fig. 10). All prediction systems outperform persistence, with ODA performing best. Note that individual correlation differences between the schemes are not statistically significant (SI Figure 8), but RMSE is substantially and consistently lower in ODA than in the other schemes from lead months 2–9. NudF-UVT and NudA-UVT perform better than NudA-UV, showing the importance of constraining the surface heat flux for predicting ENSO variability. NudF-UVT is initially better

than NudA-UVT, but the skill quickly degrades over time for $RMSE_u$. This nicely highlights the dilemma of full-field versus anomaly-field initialization: the mean state is essential for initialization. However, constraining the bias causes drift and more rapid degradation of predictability performance than anomaly-field initialization. We can also observe ODA's impact in ODA+NudA-UV, which, compared to NudA-UV (EF), has a higher skill, especially after the seventh lead month. These results are valid regardless of the initial season of the hindcasts (Figures 4 and 5 in SI), and no system shows superior performance regarding the May predictability barrier.

For the Atlantic Niño, we analyze the ATL3 index (SST averaged over the region 3° S–3° N, 20° W–0°) $RMSE_u$ and ACC as a function of lead-time (Fig. 11). NudF-UVT outperforms all other systems. NudF-UVT's skill is comparable to persistence during the first six months; the other systems show relatively low ACCs (ACC < 0.4), from which ODA sustains significant ACCs for most of the lead months, except for lead month 4. Note, however, that the correlation differences between the schemes are not statistically significant, but the $RMSE_u$ of the NudF-UVT scheme is consistently lower than for the other schemes during the first 4 months (SI Figure 9). Breaking down the analysis by start season, we see that NudF-UVT performs better among the other schemes for the hindcast starting in May, for which
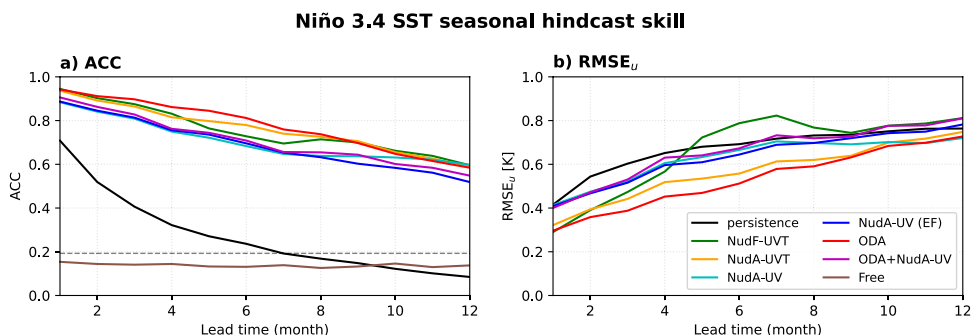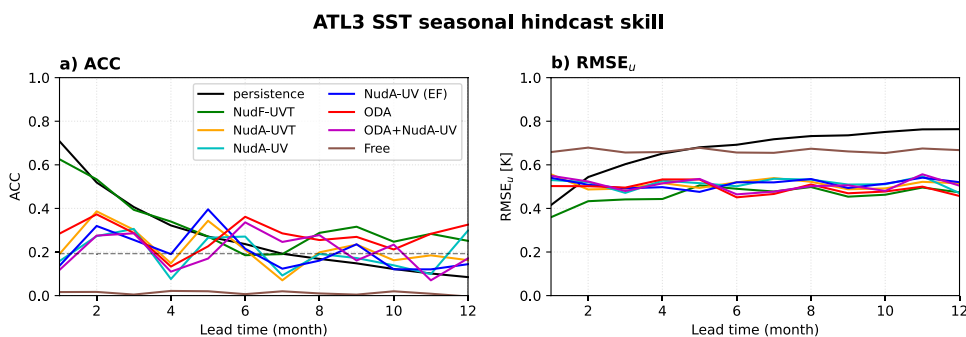

Fig. 10 **a** ACC of Niño 3.4 SST as a function of the lead month and **b** is the same for $RMSE_u$ in K. Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV

(EF), red to ODA, magenta to ODA+NudA-UV, brown to Free, and persistence is the solid black line. The black dashed line marks the 95 % significance level for ACC

Fig. 11 Same as Fig. 10 for ATL3 SST

the skill is significant during the first 3 months and similar to persistence (SI, Figures 6b and 7b). This is at the peak of the Atlantic Niño. Skillfully predicting this event is very challenging, and the NudF-UVT system beats the anomaly-coupled version of NorCPM (Counillon et al. 2021), whose hindcasts starting in May performed poorly. This highlights that constraining the mean seasonal cycle and the wind variability is critical to skillfully predicting the Atlantic Niño (Ding et al. 2015; Dippe et al. 2018; Harlaß et al. 2018). The skill for the other start months is poor (Figures 6 and 7 in SI), in agreement with those shown in Counillon et al. (2021). Overall, the skill remains poor in predicting Atlantic Niño variability.

3.2.2 Decadal predictions

We assess our decadal prediction skill with ACC and $RMSE_u$ as a function of lead year. The whole period analyzed corresponds to 1986–2019 and is lead-year dependent; thus, for instance, at lead year 2, it is 1987–2011, while for lead year 5, the period includes 1990–2014 (see Table 2 in SI).

Figure 12 shows the global average skill with lead years for HC500, and Fig. 13 shows the corresponding pointwise skill for lead-year 2–5. Globally, all systems show higher skill than persistence. ODA performs best and NudF-UVT

worst in comparison of all systems used in this study. NudF-UVT shows comparable skill to NudA-UVT until lead year 2, after which its skill rapidly degrades.

All schemes show a relatively low global skill. Given the small sample of our decadal hindcast (13 start dates), the ACCs pattern is relatively noisy and even negative in some regions (cyan-to-blue colors in Fig. 13). However, compared to the skill of the uninitialized Free experiment (Fig. 13e), all of our schemes show regions of improved skill. The differences between ACC and Free (SI, Figure 10) reveal that all the schemes improve the prediction for HC500 in the Western Pacific Ocean. This improvement is, however, not enough to have significant skill in the region (ACCs < 0) and is further accompanied by the degradation of the skill in the Eastern Pacific Ocean. Another region of improvement is in the North Atlantic, more specifically in the SPG. In this region, the skill is significantly higher (ACCs > 0.5) than in Free for most of the schemes, except NudF-UVT and NudA-UVT. The regions for which skill is improved when compared to Free agree with the NorCPM experiment for CMIP6 DCPP covering the 1950–2020 period (Bethke et al. 2021). The skill is mostly driven by external forcing, and initialization further improves it, in agreement with previous studies (e.g., Choi and Son 2022). The skill is negative in Free at the western coasts of North and South America

Fig. 12 Global **a** ACC and **b** $RMSE_u$ as a function of lead year for HC500. The line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, and magenta to ODA+NudA-UV, brown to Free, and the black line is persistence. The black dashed line marks the 95 % significance level for ACC

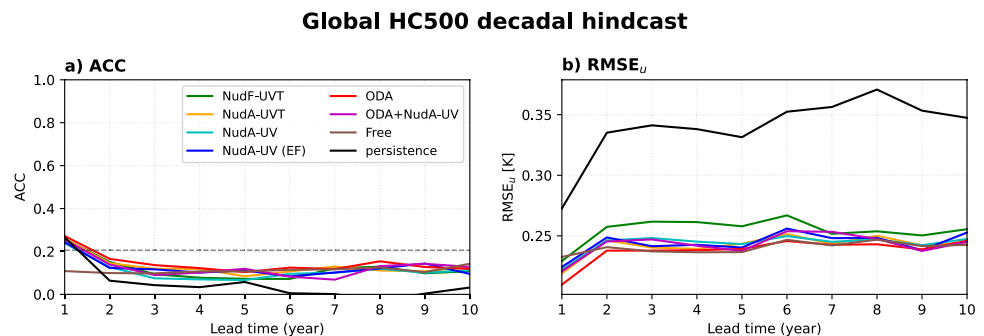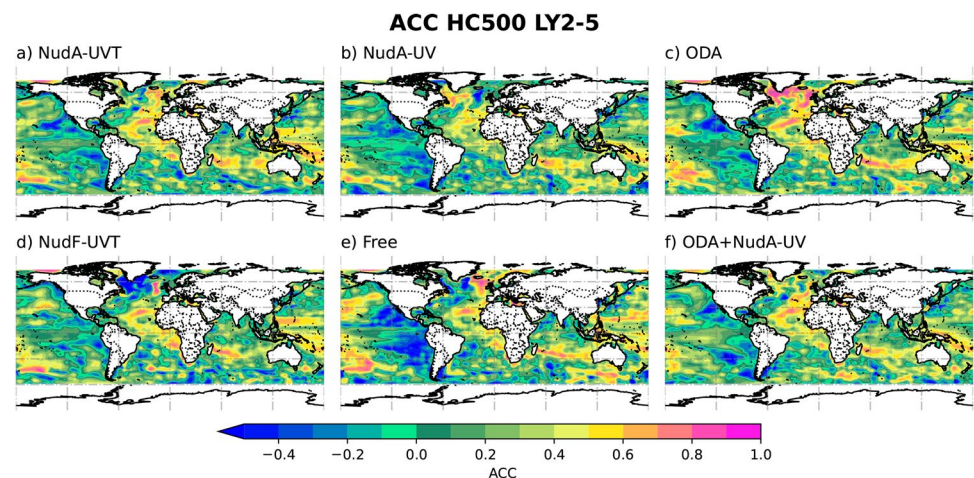

Global HC500 decadal hindcast

Fig. 13 ACC for the decadal hindcast at lead year 2–5 of HC500 **a** NudA-UVT, **b** NudA-UV, **c** ODA, **d** NudF-UVT, **e** Free and **f** ODA+NudA-UV computed against EN4 objective analysis. Green-to-magenta colors indicate positive ACCs, while cyan-to-blue colors indicate negative ACCs

ACC HC500 LY2-5

as the forced response does not agree with the Pacific Decadal Oscillation (PDO) that is predominantly positive during the analysis period 1980–2010 and can be partly related to internal climate variability (Mochizuki et al. 2010). Skill in Free is improved if one considers a longer period, e.g., 1950–2020, see Bethke et al. (2021). The degradation is mitigated by initialization, and overall, the best skill is achieved by NudF-UVT, suggesting that correcting the climate mean state can be important for PDO predictions (e.g., Guemas et al. 2012). Finally, ODA has the largest skill improvement in the SPG region, highlighting the importance of constraining the ocean to initialize decadal variability within the subpolar North Atlantic.

To further analyze the SPG variability, we evaluate the performance of the SPG index based on HC500 with lead-year (Fig. 14). The conclusions are unchanged when using different SPG indices (e.g., based on SSH or SST, not shown). Most systems beat persistence after lead-year 5. ODA provides the best skill and outperforms persistence from the start, while NudF-UVT is the worst. We can also see the benefit that ODA brings in ODA+NudA-UV, which achieves higher skills than NudA-UV only, due to hydrographic profile assimilation. Also, nudging only horizontal winds (NudA-UV) gives better predictions than additionally nudging atmospheric temperature (NudA-UVT) (Fig. 14). In NudA-UV, the dynamical forcing of NAO is well captured, and its effects on predictions are more long-lasting (Lohmann et al. 2009; Häkkinen and Rhines 2004) than additionally applying temperature constrain. The additional constraint of the temperature provides better reanalysis near the surface but introduces a dynamic imbalance with the ocean interior. We can also see that the schemes using NudA-UV give a more steady prediction skill of about 0.6 along the complete forecast. All schemes show a pronounced attraction towards their climatology (dash-dot lines in Fig. 4), showing that the memory of the initial conditions is gradually lost, and the ensemble mean converges with that of Free.

In NudF-UVT, the drift is substantial and overshoots Free. Such a drift is characteristic of dynamic imbalance.

Prediction of AMOC variability at 26.°N is shown in Fig. 5 and compared to the RAPID observation program started in 2004. The validation period is too short to assess robustly which configuration has the most skill. However, most systems tend to agree in their reanalysis, but there is a larger discrepancy for atmospheric nudging, including temperature, and NudF-UVT has, again, a considerable drift.

4 Discussion and summary

In this study, we compared the potential of a large set of initialization schemes to constrain climate variability in an ESM (NorESM1-ME, the model that is used in NorCPM) and to provide skillful initial conditions for climate predictions. This enabled us to assess the strengths and weaknesses of different methodologies and techniques using the same model, setting, and period. We compared anomaly versus full-field atmospheric nudging, together with U, V, and T nudging, compared to only U and V in the atmosphere. We also assessed the importance of conserving energy in atmospheric assimilation, and finally, we tried to combine atmospheric nudging and ocean data assimilation. We assessed the performance for reanalysis and for a set of seasonal and decadal hindcasts for 1980–2010. Our analysis is summarized below:

1. Full-field initialization introduces a large drift in the climate reanalysis and hindcasts, but constraining the mean state error was shown to improve the performance in some regions, such as in the Tropical Atlantic. The initialization with anomalies performs overall best for lead times beyond one month, as the benefit of full field initialization is rapidly lost.
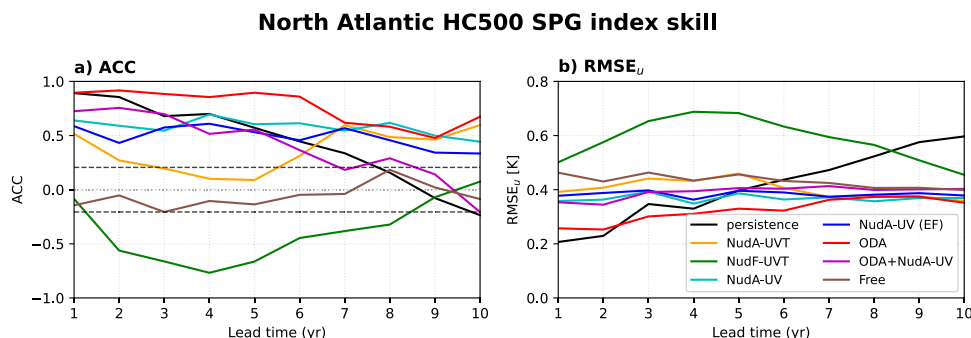
North Atlantic HC500 SPG index skill

Fig. 14 a ACC and **b** RMSE$_u$ of the SPG index (computed from HC500 versus EN4 objective analysis) as a function of lead year. The line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, and magenta to ODA+NudA-UV, the brown line is Free, and the black line is persistence. The black dashed lines mark the 95 % significance level for ACC, and the dotted black line marks the zero ACC

2. Nudging of atmospheric momentum achieves good skill for decadal predictions. It shows little drift in the hindcasts for the North Atlantic Gyre circulation (e.g., SPG or AMOC). Adding a temperature constraint provides more accurate reanalysis and seasonal predictions but degrades decadal predictions.

3. Conserving energy with anomaly atmospheric nudging of horizontal winds limits the climatological change during the reanalysis. This change does not affect the skill for seasonal or decadal predictions.

4. Ocean data assimilation enhances the accuracy of the ocean interior during the reanalysis. It provides a better skill for seasonal and decadal predictions than any atmospheric nudging simulations. For seasonal predictions, it gives better skills for global SST, T2M, HC500, SC500, and the ENSO index, while for decadal hindcasts the better skill for the SPG index. However, atmospheric nudging improves the reanalysis of ocean variability strongly influenced by atmospheric events, such as the 1995 shift in the SPG.

5. While the ocean data assimilation and atmospheric nudging approaches are complementary, and their combination is expected to provide optimal performance, the scheme tested in this study achieved inferior skill. Atmospheric nudging towards a deterministic atmospheric reanalysis causes a near collapse of the ensemble spread at the surface and strongly degrades the influence of the surface ocean data. Still, the assimilation of hydrographic profiles yields slight improvements in decadal predictions.

We have evaluated the accuracy of seasonal and decadal predictions using the CMIP5 forcing scenario. Upgrading to the CMIP6 forcing scenarios should result in a better representation of the external forcing and the system's response to it. This upgrade can have different impacts on global prediction skill. For example, Borchert et al. (2021) found that the use of CMIP6 forcings may lead to improved decadal prediction of the North Atlantic SPG SST, whereas Wu et al. (2023) shows that the inclusion of historic volcanic forcing significantly degrades the forecast skill of multiyear-to-decadal SST in the central-eastern tropical Pacific. Our CMIP6 implementation in NorCPM performed poorer than the one using CMIP5 forcing (Bethke et al. 2021; Passos et al. 2023), which motivated our choice to use CMIP5 forcing here. We do not expect that the use of different forcing would lead to very different conclusions, as the skill of assimilation relies primarily on synchronizing internal variability.

In future work, we will explore ways of preserving the reliability of the ensemble at the ocean–atmosphere interface when combining atmospheric nudging with ocean data assimilation. A substantial limitation of the current approach is that we are nudging toward a deterministic reconstruction of the atmosphere. As such, this approach disregards the atmospheric reanalysis error and causes the ensemble spread to collapse. We will, therefore, nudge toward an atmospheric ensemble reanalysis (e.g., ERA5). Note also that our current study uses ERA-interim, which performs poorer than the latest ERA5. The new simulations will give us a chance to compare the impact of the upgrade of our atmospheric reanalysis data set. Furthermore, models used for producing atmospheric reanalyses have considerably higher resolution than the atmosphere model in our ESM, and representation error (e.g., Janjić et al. 2018) may also induce a collapse of the ensemble spread (Anderson 2001). Therefore, we will complement the system with ad-hoc techniques such as inflation (Anderson 2001; El Gharamti et al. 2021), atmospheric perturbation (Houtekamer and Derome 1995) and consider using a weaker nudging. Once the system is well calibrated for ocean and atmospheric constrain, we will explore the potential of strongly coupled data assimilation (Penny et al. 2017; Sandery et al. 2020; Liu et al. 2013).

We have also seen that full-field and anomaly nudging initialization have advantages. To date, models have biases that are typically larger than the variability being predicted (Palmer and Stevens 2019). However, we foresee that the advantages of the full-field initialization approach will one day out-compete its caveats due to model improvement (for example, using higher resolution (e.g., Hewitt et al. 2017)), and better observational data (more numerous and comprehensive). Furthermore, several methods are being developed to handle climate biases with NorCPM, namely: anomaly coupling (Counillon et al. 2021), multivariate parameter estimation (Singh et al. 2022), super-resolution (Barthélémy et al. 2022) and supermodelling (Counillon et al. 2023; Schevenhoven and Carrassi 2021; Schevenhoven et al. 2023).

Author Contributions All authors contributed to the study's conception and design. Data preparation, model simulations, and analysis were performed by Lilian Garcia-Oliva under the supervision of François Counillon and Ingo Bethke. The first draft of the manuscript was written by Lilian Garcia-Oliva and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

References

Anderson JL (2001) An ensemble adjustment kalman filter for data assimilation. Mon Weather Rev 129(12):2884–2903. 10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

Balmaseda M, Anderson D (2009) Impact of initialization strategies and observations on seasonal forecast skill. Geophys Res Lett 36:1701. https://doi.org/10.1029/2008GL035561

Balmaseda M, Alves O, Arribas A et al (2009) Ocean initialization for seasonal forecasts. Oceanography 22(3):154–159. http://www.jstor.org/stable/24860997

Barthélémy S, Brajard J, Bertino L et al (2022) Super-resolution data assimilation. Ocean Dyn 72(8):661–678. https://doi.org/10.1007/s10236-022-01523-x

Bellprat O, Massonnet F, Siegert S et al (2017) Uncertainty propagation in observational references to climate model scales. Remote Sens Environ 203:101–108. https://doi.org/10.1016/J.RSE.2017.06.034

Bentsen M, Bethke I, Debernard JB et al (2013) The Norwegian Earth System Model, NorESM1-M—Part 1: description and basic evaluation of the physical climate. Geosci Model Dev 6(3):687–720. https://doi.org/10.5194/gmd-6-687-2013

Bethke I, Wang Y, Counillon F et al (2021) NorCPM1 and its contribution to CMIP6 DCPP. Geosci Model Dev 14(11):7073–7116. https://doi.org/10.5194/gmd-14-7073-2021

Bitz CM, Shell KM, Gent PR et al (2012) Climate sensitivity of the community climate system model, version 4. J Clim 25(9):3053–3070. https://doi.org/10.1175/JCLI-D-11-00290.1

Bleck R, Smith LT (1990) A wind-driven isopycnic coordinate model of the north and equatorial Atlantic Ocean: 1. Model development and supporting experiments. J Geophys Res Oceans 95(C3):3273–3285. https://doi.org/10.1029/JC095IC03P03273

Bleck R, Rooth C, Hu D et al (1992) Salinity-driven Thermocline Transients in a Wind- and Thermohaline-forced Isopycnic Coordinate Model of the North Atlantic. J Phys Oceanogr 22(12):1486–1505. 10.1175/1520-0485(1992)022<1486:SDTTIA>2.0.CO;2

Boer GJ, Smith DM, Cassou C et al (2016) The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. Geosci Model Dev 9(10):3751–3777. https://doi.org/10.5194/gmd-9-3751-2016, https://gmd.copernicus.org/articles/9/3751/2016/

Borchert LF, Menary MB, Swingedouw D et al (2021) Improved decadal predictions of north Atlantic subpolar gyre SST in CMIP6. Geophys Res Lett 48(3):e2020GL091307. https://doi.org/10.1029/2020GL091307

Brune S, Baehr J (2020) Preserving the coupled atmosphere-ocean feedback in initializations of decadal climate predictions. Wiley Interdiscip Rev Clim Change. https://doi.org/10.1002/WCC.637

Brune S, Düsterhus A, Pohlmann H et al (2018) Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts. Clim Dyn 51:1947–1970. https://doi.org/10.1007/s00382-017-3991-4

Carrassi A, Weber RJ, Guemas V et al (2014) Full-field and anomaly initialization using a low-order climate model: a comparison and proposals for advanced formulations. Nonlinear Process Geophys 21(2):521–537. https://doi.org/10.5194/npg-21-521-2014

Choi J, Son SW (2022) Seasonal-to-decadal prediction of El Niño–Southern Oscillation and Pacific Decadal Oscillation. npj Clim Atmos Sci 5(1):1–8. https://doi.org/10.1038/s41612-022-00251-9

Counillon F, Bethke I, Keenlyside N et al (2014) Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment. Tellus Ser A Dyn Meteorol Oceanogr. https://doi.org/10.3402/tellusa.v66.21074

Counillon F, Keenlyside N, Bethke I et al (2016) Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. Tellus Ser A Dyn Meteorol Oceanogr 68(1):32,437. https://doi.org/10.3402/tellusa.v68.32437

Counillon F, Keenlyside N, Toniazzo T et al (2021) Relating model bias and prediction skill in the equatorial Atlantic. Clim Dyn 56:2617–2630. https://doi.org/10.1007/s00382-020-05605-8

Counillon F, Keenlyside N, Wang S et al (2023) Framework for an Ocean-Connected Supermodel of the Earth System. J Adv Model Earth Syst 15(3):e2022MS003310. https://doi.org/10.1029/2022MS003310

Danabasoglu G, Yeager SG, Bailey D et al (2014) North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. Ocean Model 73:76–107. https://doi.org/10.1016/J.OCEMOD.2013.10.005

Dee DP (2006) Bias and data assimilation. Q J R Meteorol Soc 131(613):3323–3343. https://doi.org/10.1256/qj.05.137

Dee DP, Uppala SM, Simmons AJ 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

Ding H, Greatbatch RJ, Latif M et al (2015) The impact of sea surface temperature bias on equatorial Atlantic interannual variability in partially coupled model experiments. Geophys Res Lett 42(13):5540–5546. https://doi.org/10.1002/2015GL064799

Dippe T, Greatbatch RJ, Ding H (2018) On the relationship between Atlantic Niño variability and ocean dynamics. Clim Dyn 51(1–2):597–612. https://doi.org/10.1007/S00382-017-3943-Z/FIGURES/12

Doblas-Reyes FJ, Andreu-Burillo I, Chikamoto Y et al (2013) Initialized near-term regional climate change prediction. Nat Commun 4(1):1–9. https://doi.org/10.1038/ncomms2704

Dunstone NJ, Smith DM (2010) Impact of atmosphere and sub-surface ocean data on decadal climate prediction. Geophys Res Lett 37(2):2709. https://doi.org/10.1029/2009GL041609

El Gharamti M, McCreight JL, Noh SJ et al (2021) Ensemble streamflow data assimilation using WRF-Hydro and DART: novel localization and inflation techniques applied to Hurricane Florence flooding. Hydrol Earth Syst Sci 25(9):5315–5336. https://doi.org/10.5194/hess-25-5315-2021

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99(C5):143–153. https://doi.org/10.1029/94JC00572

Evensen G (2003) The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn 53(4):343–367. https://doi.org/10.1007/S10236-003-0036-9

Fortin V, Abaza M, Anctil F et al (2014) Why should ensemble spread match the RMSE of the ensemble mean? J Hydrometeorol 15(4):1708–1713. https://doi.org/10.1175/JHM-D-14-0008.1

García-Serrano J, Guemas V, Doblas-Reyes FJ (2015) Added-value from initialization in predictions of Atlantic multi-decadal variability. Clim Dyn 44(9–10):2539–2555. https://doi.org/10.1007/S00382-014-2370-7/FIGURES/9

Good SA, Martin MJ, Rayner NA (2013) EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J Geophys Res Oceans 118(12):6704–6716. https://doi.org/10.1002/2013JC009067

Gouretski V, Reseghetti F (2010) On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global database. Deep-Sea Res I 57(6):812–833. https://doi.org/10.1016/j.dsr.2010.03.011

Guemas V, Doblas-Reyes FJ, Lienert F et al (2012) Identifying the causes of the poor decadal climate prediction skill over the North Pacific. J Geophys Res Atmos. https://doi.org/10.1029/2012JD018004

Häkkinen S, Rhines PB (2004) Decline of Subpolar North Atlantic Circulation during the 1990s. Science 304(5670):555–559. https://doi.org/10.1126/SCIENCE.1094917/SUPPL/FILE/HAKKINEN.SOM.PDF

Han JY, Kim SW, Park CH et al (2023) Ensemble size versus bias correction effects in subseasonal-to-seasonal (S2S) forecasts. Geosci Lett 10(1):1–12. https://doi.org/10.1186/S40562-023-00292-9/FIGURES/5

Harlaß J, Latif M, Park W (2018) Alleviating tropical Atlantic sector biases in the Kiel climate model by enhancing horizontal and vertical atmosphere model resolution: climatology and interannual variability. Clim Dyn 50(7–8):2605–2635. https://doi.org/10.1007/s00382-017-3760-4

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bull Am Meteor Soc 90(8):1095–1108. https://doi.org/10.1175/2009BAMS2607.1

Hermanson L, Smith D, Seabrook M et al (2022) WMO global annual to decadal climate update: a prediction for 2021–25. Bull Am Meteorol Soc 103(4):E1117–E1129. https://doi.org/10.1175/BAMS-D-20-0311.1, https://journals.ametsoc.org/view/journals/bams/103/4/BAMS-D-20-0311.1.xml

Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049. https://doi.org/10.1002/QJ.3803

Hewitt HT, Bell MJ, Chassignet EP et al (2017) Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales? Ocean Model 120:120–136. https://doi.org/10.1016/j.ocemod.2017.11.002

Hoke JE, Anthes RA (1976) The Initialization of Numerical Models by a Dynamic-Initialization Technique. Mon Weather Rev 104(12):1551–1556. 10.1175/1520-0493(1976)104<1551:TIONMB>2.0.CO;2

Houtekamer PL, Derome J (1995) Methods for ensemble prediction. Mon Weather Rev 123(7):2181–2196. 10.1175/1520-0493(1995)123<2181:mfep>2.0.co;2

Hughes IG, Hase TPA (2010) Measurements and their uncertainties: a practical guide to modern error analysis. Oxford University Press, Oxford

Hurrell JW, Holland MM, Gent PR et al (2013) The community earth system model: a framework for collaborative research. Bull Am Meteor Soc 94(9):1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

Janjić T, Bormann N, Bocquet M et al (2018) On the representation error in data assimilation. Q J R Meteorol Soc 144(713):1257–1278. https://doi.org/10.1002/qj.3130

Karspeck AR, Danabasoglu G, Anderson J et al (2018) A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed. Q J R Meteorol Soc 144(717):2404–2430. https://doi.org/10.1002/qj.3308

Keenlyside NS, Latif M, Jungclaus J et al (2008) Advancing decadal-scale climate prediction in the North Atlantic sector. Nature 453(7191):84–88. https://doi.org/10.1038/nature06921

Keenlyside N, Kosaka Y, Vigaud N et al (2020) Basin interactions and predictability. Cambridge University Press, Cambridge, pp 258–292. https://doi.org/10.1017/9781108610995.009

Kirkevåg A, Iversen T, Seland Ø et al (2012) Aerosol-climate interactions in the Norwegian Earth System Model—NorESM. Geosci Model Dev Discuss 5:2599–2685. https://doi.org/10.5194/gmdd-5-2843-2012

Kirtman B, Power SB, Adedoyin AJ et al (2013) Chapter 11—near-term climate change: Projections and predictability. In: IPCC (ed) Climate change 2013: the physical science basis. IPCC Working Group I Contribution to AR5. Cambridge University Press, Cambridge. http://www.climatechange2013.org/images/report/WG1AR5%5fChapter11%5fFINAL.pdf

Kirtman BP, Min D, Infanti JM et al (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. Bull Am Meteor Soc 95(4):585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Kooperman GJ, Pritchard MS, Ghan SJ et al (2012) Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the Community Atmosphere Model 5. J Geophys Res Atmos 117(D23):23,204. https://doi.org/10.1029/2012JD018588

Lawrence DM, Oleson KW, Flanner MG et al (2011) Parameterization improvements and functional and structural advances in

Version 4 of the Community Land Model. J Adv Model Earth Syst. https://doi.org/10.1029/2011MS00045

Levitus S, Boyer TP (1994) World ocean atlas 1994. Vol. 4, Temperature. https://repository.library.noaa.gov/view/noaa/1381

Levitus S, Burgett R, Boyer TP (1994) World ocean atlas 1994. Vol. 3, Salinity. https://repository.library.noaa.gov/view/noaa/1382

Liu Z, Wu S, Zhang S et al (2013) Ensemble data assimilation in a simple coupled climate model: the role of ocean-atmosphere interaction. Adv Atmos Sci 30(5):1235–1248. https://doi.org/10.1007/s00376-013-2268-z

Lohmann K, Drange H, Bentsen M et al (2009) Response of the North Atlantic subpolar gyre to persistent North Atlantic oscillation like forcing. Clim Dyn 32(2):273–285. https://doi.org/10.1007/s00382-008-0467-6

Lu F, Harrison MJ, Rosati A et al (2020) GFDL's SPEAR seasonal prediction system: initialization and ocean tendency adjustment (OTA) for coupled model predictions. J Adv Model Earth Syst 12(12):e2020MS002149. https://doi.org/10.1029/2020MS002149

Magnusson L, Alonso-Balmaseda M, Corti S et al (2013) Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. Clim Dyn 41(9–10):2393–2409. https://doi.org/10.1007/s00382-012-1599-2

Mariotti A, Ruti PM, Rixen M (2018) Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. npj Clim Atmos Sci 1(1):1–4. https://doi.org/10.1038/s41612-018-0014-z

Mariotti A, Baggett C, Barnes EA et al (2020) Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. Bull Am Meteor Soc 101(5):E608–E625. https://doi.org/10.1175/BAMS-D-18-0326.1

Massonnet F, Bellprat O, Guemas V et al (2016) Using climate models to estimate the quality of global observational data sets. Science 354(6311):452–455. https://doi.org/10.1126/science.aaf6369

Meehl GA, Goddard L, Murphy J et al (2009) Decadal prediction: can it be skillful? Bull Am Meteor Soc 90(10):1467–1485. https://doi.org/10.1175/2009BAMS2778.1

Meehl GA, Richter JH, Teng H et al (2021) Initialized Earth System prediction from subseasonal to decadal timescales. Nat Rev Earth Environ 2(5):340–357. https://doi.org/10.1038/s43017-021-00155-x

Mochizuki T, Ishii M, Kimoto M et al (2010) Pacific decadal oscillation hindcasts relevant to near-term climate prediction. Proc Natl Acad Sci 107(5):1833–1837. https://doi.org/10.1073/pnas.0906531107

Neale R, Richter J, Conley A, et al (2010) Description of the Community Atmosphere Model (CAM 4.0). NCAR Technical Note TN-485+STR

Palmer T, Stevens B (2019) The scientific challenge of understanding and estimating climate change. Proc Natl Acad Sci 116(49):24,390-24,395. https://doi.org/10.1073/pnas.1906691116

Passos L, Langehaug HR, Årthun M et al (2023) Impact of initialization methods on the predictive skill in NorCPM: an Arctic-Atlantic case study. Clim Dyn 60:2061–2080. https://doi.org/10.1007/s00382-022-06437-4

Penny SG, Akella S, Buehner M et al (2017) Coupled data assimilation for integrated earth system analysis and prediction: goals, challenges, and recommendations. Tech. rep., World Meteorological Organization, WWRP_2017_3

Pohlmann H, Jungclaus JH, Köhl A et al (2009) Initializing decadal climate predictions with the GECCO oceanic synthesis: effects on the North Atlantic. J Clim 22(14):3926–3938. https://doi.org/10.1175/2009JCLI2535.1

Polkova I, Brune S, Kadow C et al (2019) Initialization and ensemble generation for decadal climate predictions: a comparison of different methods. J Adv Model Earth Syst 11(1):149–172. https://doi.org/10.1029/2018MS001439

Robson J (2010) Understanding the performance of a decadal prediction system. PhD thesis, https://doi.org/10.13140/RG.2.1.2183.2560

Robson JI, Sutton RT, Smith DM (2012) Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. Geophys Res Lett. https://doi.org/10.1029/2012GL053370

Rodwell MJ, Lang STK, Ingleby NB et al (2016) Reliability in ensemble data assimilation. Q J R Meteorol Soc 142(694):443–454. https://doi.org/10.1002/qj.2663

Saji NH, Goswami BN, Vinayachandran PN et al (1999) A dipole mode in the tropical Indian Ocean. Nature 401(6751):360–363. https://doi.org/10.1038/43854

Sakov P, Oke PR (2008) A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. Tellus Ser A Dyn Meteorol Oceanogr 60 A(2):361–371. https://doi.org/10.1111/j.1600-0870.2007.00299.x

Sakov P, Counillon F, Bertino L et al (2012) TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Sci 8(4):633–656. https://doi.org/10.5194/os-8-633-2012

Sandery PA, O'Kane TJ, Kitsios V et al (2020) Climate model state estimation using variants of EnKF coupled data assimilation. Mon Weather Rev 148(6):2411–2431. https://doi.org/10.1175/MWR-D-18-0443.1, https://journals.ametsoc.org/view/journals/mwre/148/6/mwrD180443.xml

Schevenhoven FJ, Carrassi A (2021) Training a supermodel with noisy and sparse observations: a case study with cpt and the synch rule on speedo-v. 1. Geosci Model Dev Discuss 2021:1–23

Schevenhoven F, Keenlyside N, Counillon F et al (2023) Supermodeling: improving predictions with an ensemble of interacting models. Bull Am Meteor Soc. https://doi.org/10.1175/BAMS-D-22-0070.1

Singh T, Counillon F, Tjiputra J et al (2022) Estimation of ocean biogeochemical parameters in an earth system model using the dual one step ahead smoother: a twin experiment. Front Mar Sci. https://doi.org/10.3389/fmars.2022.775394

Smeed D, McCarthy G, Rayner D et al (2015) Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2014. British Oceanographic Data Centre—Natural Environment Research Council, UK. https://doi.org/10.5285/1a774e53-7383-2e9a-e053-6c86abc0d8c7

Smith DM, Cusack S, Colman AW et al (2007) Improved surface temperature prediction for the coming decade from a global climate model. Science 317(5839):796–799. https://doi.org/10.1126/science.1139540

Smith DM, Eade R, Pohlmann H (2013) A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. Clim Dyn 41(11–12):3325–3338. https://doi.org/10.1007/s00382-013-1683-2

Stammer D, Wunsch C, Giering R et al (2002) Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model estimated from ocean observations and a general circulation model. J Geophys Res 107(C9):3118. https://doi.org/10.1029/2001JC000888

Storto A, Alvera-Azcárate A, Balmaseda MA et al (2019) Ocean reanalyses: recent advances and unsolved challenges. Front Mar Sci 6:418

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteor Soc 93(4):485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

van Vuuren DP, Edmonds J, Kainuma M et al (2011) The representative concentration pathways: an overview. Clim Change 109:5–31. https://doi.org/10.1007/s10584-011-0148-z

Volpi D, Guemas V, Doblas-Reyes FJ (2017) 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 49(4):1181–1195. https://doi.org/10.1007/s00382-016-3373-3

Von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511612336

Wang Y, Counillon F, Bertino L et al (2016) Alleviating the bias induced by the linear analysis update with an isopycnal ocean model. Q J R Meteorol Soc 142:1064–1074. https://doi.org/10.1002/qj.2709

Wang Y, Counillon F, Bethke I et al (2017) Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. Ocean Model 114:33–44. https://doi.org/10.1016/j.ocemod.2017.04.007

Wang Y, Counillon F, Keenlyside N et al (2019) Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF. Clim Dyn 53:5777–5797. https://doi.org/10.1007/s00382-019-04897-9

Weber RJ, Carrassi A, Doblas-Reyes FJ (2015) Linking the anomaly initialization approach to the mapping paradigm: a proof-of-concept study. Mon Weather Rev 143(11):4695–4713. https://doi.org/10.1175/MWR-D-14-00398.1

Webster PJ, Moore AM, Loschnigg JP et al (1999) Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997–98. Nature 401(6751):356–360. https://doi.org/10.1038/43848

Wilks D (2019) Statistical methods in the atmospheric sciences—4th edition. https://www.elsevier.com/books/statistical-methods-in-the-atmospheric-sciences/wilks/978-0-12-815823-4

Wu X, Yeager SG, Deser C et al (2023) Volcanic forcing degrades multiyear-to-decadal prediction skill in the tropical Pacific. Sci Adv. https://doi.org/10.1126/SCIADV.ADD9364/SUPPL_FILE/SCIADV.ADD9364_SM.PDF, https://www.science.org/doi/10.1126/sciadv.add9364

Yeager SG, Robson JI (2017) Recent progress in understanding and predicting Atlantic decadal climate variability. Curr Clim Change Rep 3(2):112–127. https://doi.org/10.1007/s40641-017-0064-z

Yeager SG, Karspeck A, Danabasoglu G et al (2012) A decadal prediction case study: late twentieth-century North Atlantic Ocean heat content. J Clim 25(15):5173–5189. https://doi.org/10.1175/JCLI-D-11-00595.1

Zhang S, Rosati A, Harrison MJ (2009) Detection of multidecadal oceanic variability by ocean data assimilation in the context of a "perfect" coupled model. J Geophys Res Oceans 114(12):12,018. https://doi.org/10.1029/2008JC005261

Zhang S, Rosati A, Delworth T (2010) The adequacy of observing systems in monitoring the Atlantic meridional overturning circulation and North Atlantic climate. J Clim 23(19):5311–5324. https://doi.org/10.1175/2010JCLI3677.1

Zhang K, Wan H et al (2014) Technical note: on the use of nudging for aerosol-climate model intercomparison studies. Atmos Chem Phys 14(16):8631–8645. https://doi.org/10.5194/ACP-14-8631-2014