Intercomparison of initialization methods for Seasonal-to-Decadal Climate Predictions with the NorCPM

Lilian Garcia-Oliva[1], Francois Counillon[2], Ingo Bethke[1], and Noel S Keenlyside[3]

[1]University of Bergen and Bjerknes Centre for Climate Research
[2]Nersc
[3]Geophysical Institute, University of Bergen

September 11, 2023

Abstract

Initialization is essential for accurate seasonal-to-decadal (S2D) climate predictions. The initialization schemes used differ on the component initialized, the Data Assimilation (DA) method, or the technique. We compare five popular schemes within NorCPM following the same experimental protocol: reanalysis from 1980–2010 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 scheme ODA+NudA-UV provides a first attempt at joint initialization of the ocean and atmospheric components. During the reanalysis, atmospheric nudging leads to atmosphere and land components best synchronized with observations. Conversely, ODA best synchronizes the ocean component with observations. The atmospheric nudging schemes are better at reproducing specific events, such as the rapid North Atlantic subpolar gyre (SPG) 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 one-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.

Intercomparison of initialization methods for Seasonal-to-Decadal Climate Predictions with the NorCPM

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

[1]Geophysical Institute, University of Bergen, Bjerknes Centre for Climate Research, 5007 Bergen, Norway
[2]Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, 5006 Bergen, Norway

Key Points:

- Constraining the ocean state to observations produces more skillful predictions than constraining the atmospheric state
- Full-field performs better than anomaly initialization at short-lead times in specific regions, but drift degrades the skill rapidly
- Anomaly nudging of atmospheric momentum can achieve skillful decadal prediction and minimizes hindcast drift

Corresponding author: Lilian Garcia-Oliva, `lilian.garcia@uib.no`

Abstract

Initialization is essential for accurate seasonal-to-decadal (S2D) climate predictions. The initialization schemes used differ on the component initialized, the Data Assimilation (DA) method, or the technique. We compare five popular schemes within NorCPM following the same experimental protocol: reanalysis from 1980–2010 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 scheme ODA+NudA-UV provides a first attempt at joint initialization of the ocean and atmospheric components. During the reanalysis, atmospheric nudging leads to atmosphere and land components best synchronized with observations. Conversely, ODA best synchronizes the ocean component with observations. The atmospheric nudging schemes are better at reproducing specific events, such as the rapid North Atlantic subpolar gyre (SPG) 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 one-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.

Plain Language Summary

This study explores the impact of a wide range of standard initialization schemes on the performance of coupled reanalysis and seasonal-to-decadal predictions produced with the same Earth System Model. We compare atmospherically-driven initialization versus ocean initialization. We also compare full-field initialization —meaning where the observations are used as are—versus anomaly initialization —when the climatological difference between the model and observations is removed. All schemes have strengths and weaknesses. As expected, ocean initialization works best in the ocean, while atmospherically driven initialization works best in the atmosphere and land. Ocean initialization has the best performance overall for seasonal and decadal predictions. Still, the atmospherically driven initialization works better for some specific regions and events—for example, the strong North Atlantic subpolar gyre shift in 1995. Full-field initialization performs better than anomaly initialization at short lead times, and it improves performance in regions where the mean state is important for representing the variability, such as the Tropical Atlantic. Constraining atmospheric temperature is important for reanalysis and seasonal prediction while constraining only the winds works better for decadal prediction.

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; N. S. Keenlyside et al., 2008; Meehl et al., 2009; Hawkins & 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 & 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).

There are various schemes for accurately initializing S2D predictions. One common practice is to initialize each component of the Earth System Models (ESMs) individually, 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 & 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 & 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 initialization shocks and prediction drift. 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 (J. 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) illustrating 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 & 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., N. S. 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., S. Zhang et al., 2009; Pohlmann et al., 2009; Karspeck et al., 2018; Counillon et al., 2016; Brune & 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 S. 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) indicate 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 (NorCPM, Counillon et al., 2014, 2016) that combines the Norwegian Earth System Model (NorESM, Bentsen et al., 2013) and the Ensemble Kalman Filter (EnKF, Evensen, 2003) data assimilation method. 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 summary and conclusions are presented in Section 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 & Smith, 1990; Bleck et al., 1992), using density as its vertical coordinate.

We use the medium-resolution version of NorESM. The atmosphere and land components use a $1.9° \times 2.5°$ 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.

2.2 Ocean data assimilation with the EnKF

The Ensemble Kalman Filter (EnKF, Evensen, 2003) 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 & 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:

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

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:

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

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 & Reseghetti, 2010). The observation error for the hydrographic profiles and the localization radius varies with latitude as described in Wang et al. (2017). 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 (equation 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 & 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 equation (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 et al., 2014).

One can also apply anomaly nudging (Zhang et al., 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 equation (7) can be expressed as:

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

With the formulations of equations (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 et al. (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 below 60 km height with tapering between 50 km to 60 km, while in the land and ocean surfaces, the model is constrained towards the prescribed data.

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 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 Section 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 Section 2.2 for details on the practical implementation). Finally, the scheme ODA+NudA-UV combines the ODA and NudA-UV (EF) experiments. We did 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 (Section 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.

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.

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 15^{th} of January, April, July, and October each year during 1985–2010 and run for a year. The decadal hindcasts start on the 15^{th} of October every other year and run for 11 years each. Each hindcast runs nine realizations (ensemble members). Initial conditions are taken from the first nine members of the 30-member ensemble reanalyses. Note that this choice does not influence the results because all members are equally likely.

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 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. All climatologies are computed using the 1980-2010 period.

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:

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

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

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, Daniel, 2019).

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:

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

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

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

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 equation (11).

For the global (or regional indices) statistics, we use grid cell area weighting:

$$\mathrm{RMSE}_u = \sum_i a_i \mathrm{RMSE}_{ui} \left(\sum_j a_j \right)^{-1}, \tag{15}$$

and

$$\mathrm{ACC} = \sum_i a_i \mathrm{ACC}_i \left(\sum_j a_j \right)^{-1}. \tag{16}$$

where a_i is the area of the corresponding i-th grid cell.

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. For the ocean surface temperature, we take SST observations from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST2, Rayner et al., 2003). We interpolate our ocean outputs towards HadISST2 horizontal grid. We obtain subsurface temperature and salinity data from the EN4.2.1 objective analysis (EN4.2.1, Gouretski & Reseghetti, 2010). We re-grid and interpolate our ocean subsurface output to EN4.2.1 dataset resolution for the comparisons. Furthermore, we consider the heat and salinity content in the first 500 m, named HC500 and SC500 respectively. We define them as the ocean depth's average temperature (and salinity).

For the verification of the decadal hindcasts, we also use the Atlantic meridional overturning circulation (AMOC) at 26° North from the RAPID dataset (RAPID, Johns et al., 2011).

3 Results

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

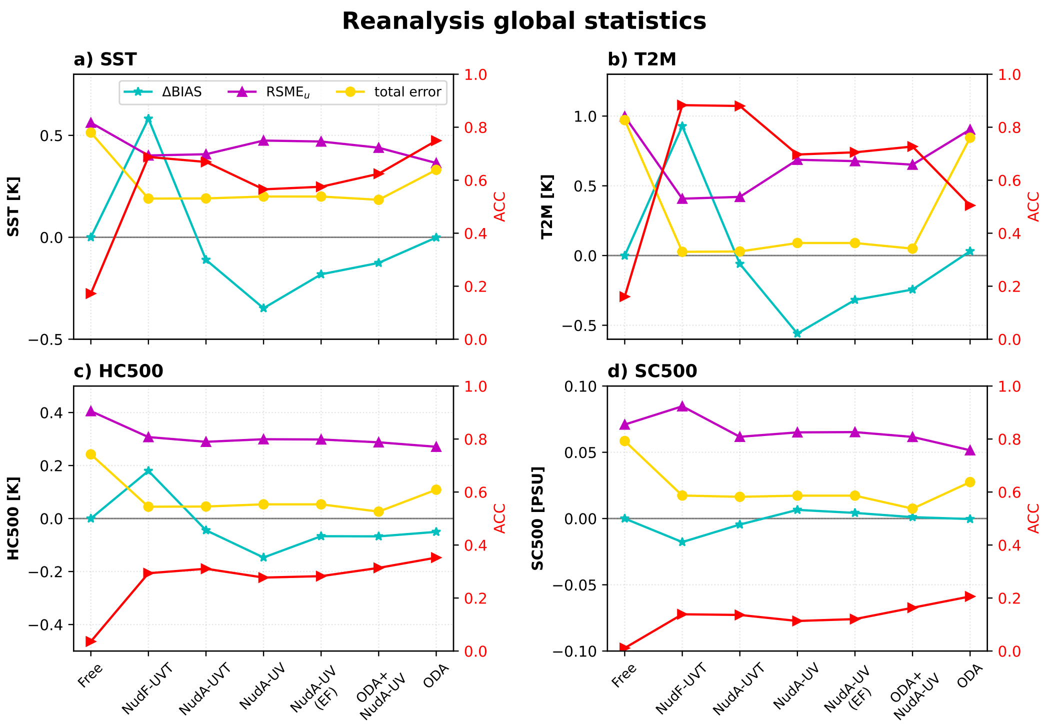

Figure 1. Global statistics of the reanalyses computed over 1980–2010, for a) SST, b) T2M, c) HC500, and d) SC500. The left-hand y-axis (in black) displays units for RMSE$_u$ (magenta), ΔBIAS (cyan), and total error (yellow), while the red right-hand y-axis is for ACC (red). The reanalyses are said to be reliable when the total error (yellow) and RMSE$_u$ (magenta) overlap. The black horizontal line marks zero.

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 (Figure 1). Globally, the reanalysis from NudF-UVT is marginally better for SST and T2M (Figures 1a and 1b), but yields a degradation for HC500 (Figure 1c) and SC500 (Figure 1d). Most of this degradation occurs in the SPG, the tropical and South Atlantic, and the Southern Ocean (Figure 2d). Furthermore, NudF-UVT exhibits a substantial RMSE$_u$ drift of HC500 and SC500 (Figure 3). Such RMSE$_u$ drift follows a parabolic shape, as the mean climatology (used for computing the metric, equation 11) is reached 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 Figure 1). 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 (equation 14) being much smaller than the RMSE$_u$ (Figures 1a and 1b). This implies that the ensemble spread (not shown) collapses during the reanalyses. The reliability for HC500 and SC500 is also poor (Figures 1c and 1d). It should be acknowledged that the HC500 (and to a minor extent SC500) reliability of Free is already too low, although it should, by construction, be satisfied by the experiment. This suggests 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 (Figures 4a and 4b), 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 & Rhines, 2004; Yeager & Robson, 2017), which induces a preconditioning of the ocean circulation state (Lohmann et al., 2009; J. I. Robson et al., 2012). Moreover, both schemes fail 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 (Figure 5b) that affects the poleward heat transport. The verification period with the RAPID (RAPID, Johns et al., 2011) 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.

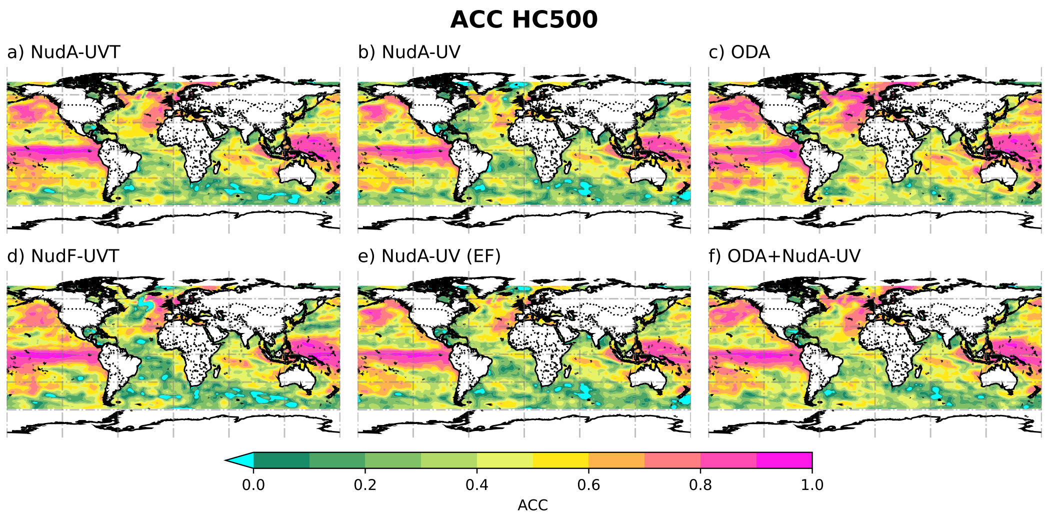

Figure 2. ACC of monthly HC500 anomalies a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV reanalysis computed against EN4 objective analysis for the period 1980–2010. Green-to-magenta colors indicate positive ACCs, and the cyan color indicates all the negative ACCs.

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 T2M), nudging only horizontal winds degrades performance (Figures 1a and 1b). For T2M, for example, NudA-UV reduces error by 0.3 K compared to Free, whereas NudA-UVT reduces it by 0.6 K. The degraded performance of NudA-UV is largest over the tropical band and is less pronounced at mid-to-high latitudes (Figures 6a and 6b). The reliability for T2M is slightly improved in NudA-UV compared to NudA-UVT (see also Table S1). In NudA-UV, there is a significant increase in climatological change ΔBIAS for SST and T2M. On the other hand, NudA-UVT sustains ΔBIAS near 0 K due to temperature nudging. Below the surface, the global skill performance of NudA-UV and NudA-UVT are similar for HC500 and SC500

(Figures 1c and 1d), 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 (Figures S6c and S6e), North East Atlantic, and South Pacific (Figures 2a and 2b). The performance for the SPG (Figure 4) and AMOC (Figure 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.

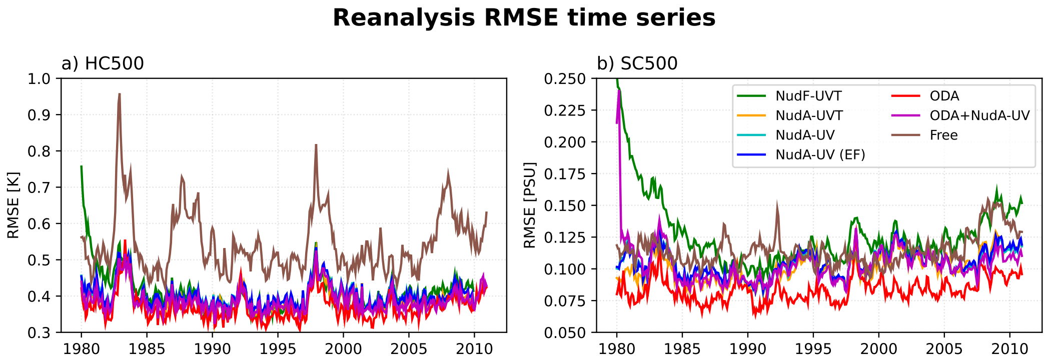

Figure 3. Time series of RMSE$_u$ for a) HC500 and b)SC500 in the different reanalyses computed against EN4 objective analysis. Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, magenta to ODA+NudA-UV, and brown is Free.

The default implementation of nudging in CAM4 deactivates the energy conservation fix in the atmospheric component (see section 2.3). Here, 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 (Figures 1a and 1d). 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 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 (Figure S6). And comparing NudA-UV and NudA-UV (EF), we find that the energy fix is very effective in improving the representation of the trend in the Iceland basin ($R = 0.31$ and 0.61, respectively, in Figures S6e and S6g).

We now compare atmospheric constraints versus ocean constraints for coupled reanalysis. The skill for T2M (Figure 1b) using atmospheric nudging is substantially better 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 (Figures 6a and 6c). 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 HadISST2 analysis, which is assimilated in the ODA system. Meanwhile, the verification of T2M is done against ERA5, 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 (Figures 1c-1d). This is also clear from Figure 3, where ODA has a consis-

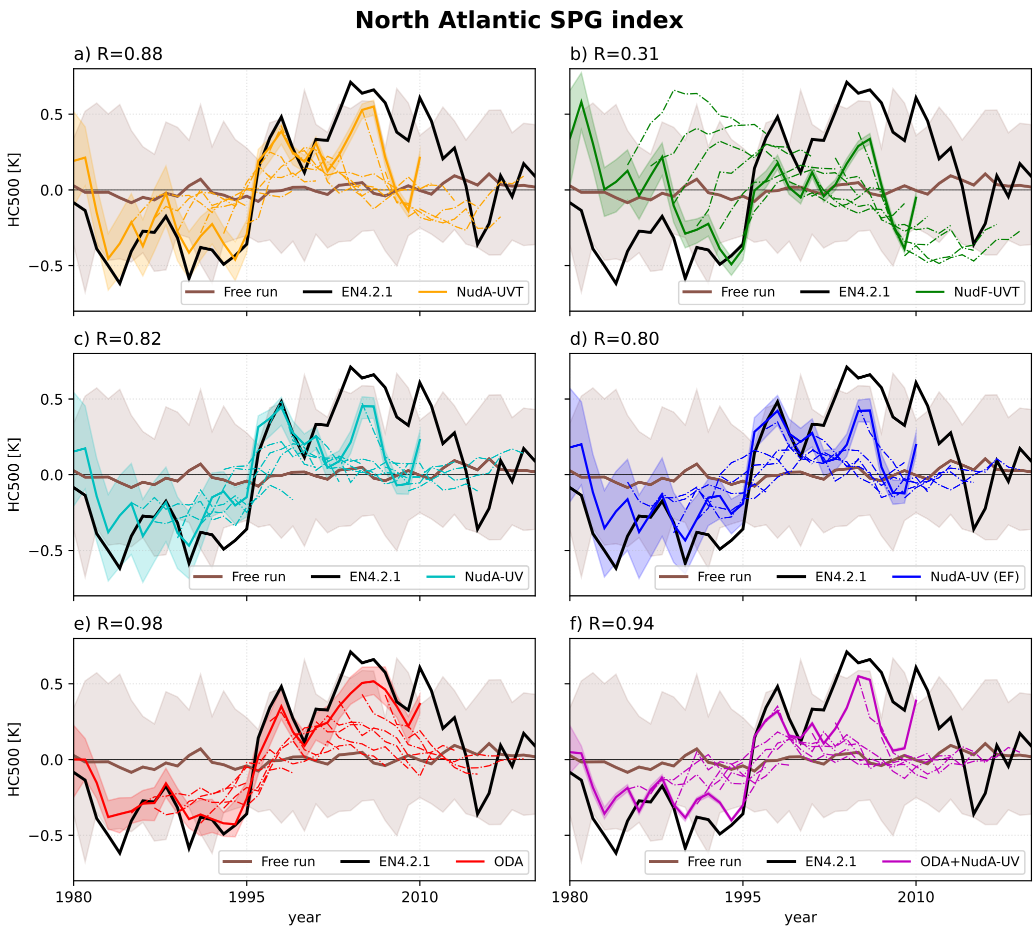

Figure 4. HC500 anomalies in the SPG box (48°-65°N, 60°-15°S) for a) NudA-UVT, b) NudF-UVT, c) NudA-UV, d) NudA-UV (EF), e) ODA and f) ODA+NudA-UV reanalyses. Solid-colored lines represent the ensemble mean of reanalysis, dash-dotted lines correspond to hindcast schemes, and the solid brown line is Free. 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

tently 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 (Figure 2c), where atmospheric nudging introduces a patch of low-skill in the Irminger and Icelandic Seas (see, for example, Figures 2a and 2b). The reliability of the system is also better preserved as we see a closer match between RMSE_u and total error σ (Figures 1c and 1d, magenta and yellow lines). In the ODA system the reliability is only marginally degraded from Free and much less than atmospheric nudging. In the case of regional indexes, ODA achieves overall the best correlation for the SPG index ($R = 0.98$, Figure 4e), and it is the only system that sustains the weak 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, Figure 4a). This is because the NAO constraint is very weak in the ODA system, and the system only adjusts a-posteriori for errors in the atmospheric forcing. Finally, for the AMOC at 26.5°N, there is a long term weakening with a stronger weak anomaly from 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 (Figures 5c and 5e), suggesting that this feature is better constrained with atmospheric variability.

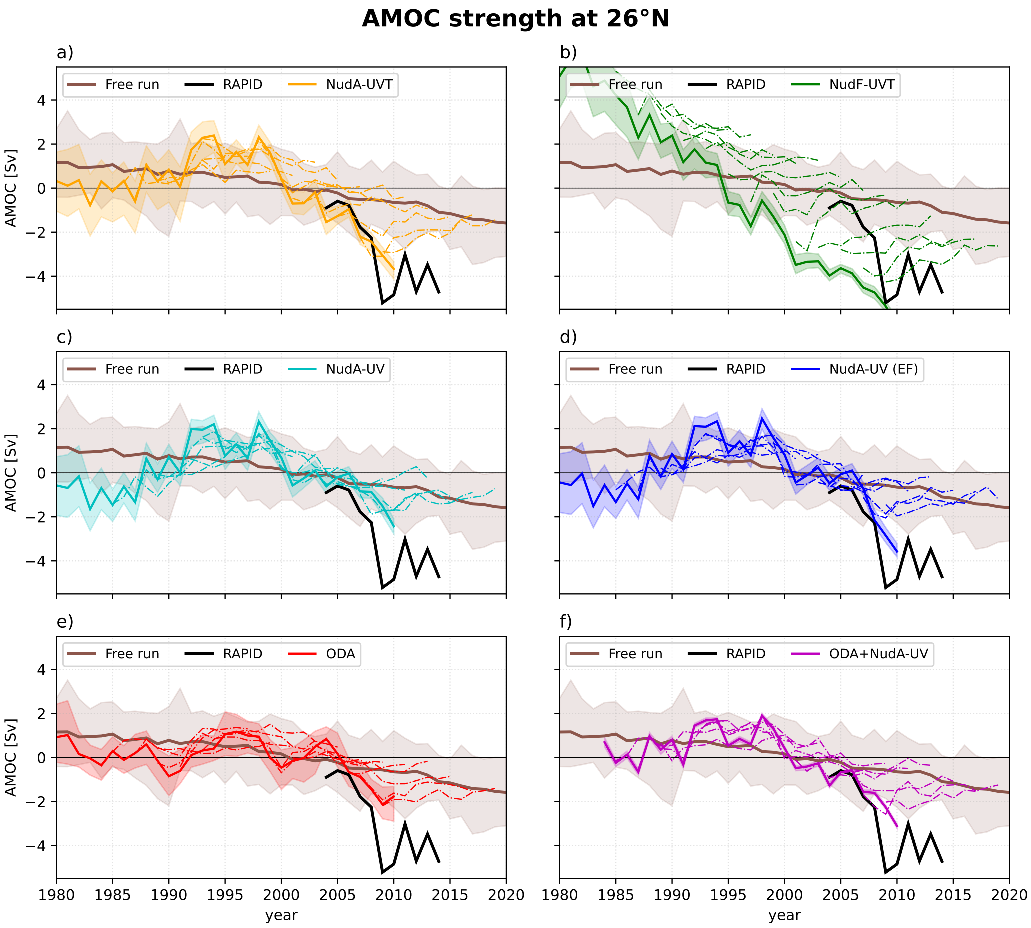

Figure 5. AMOC transport anomalies at 26.5°N with respect to the 2005–2010 period for a) NudA-UVT, b) NudF-UVT, c) NudA-UV, d) NudA-UV (EF), e) ODA and f) ODA+NudA-UV reanalyses. Solid-colored lines represent the ensemble mean of reanalysis, dash-dotted lines correspond to hindcast schemes, and the solid brown line is Free. Shading denotes ensemble minima and maxima. The solid black line is the RAPID observations.

Given the complementary skills of atmospheric nudging and the ODA systems, one would expect their combination to work best. However, comparing the global statistics of ODA and ODA+NudA-UV (Figure 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 Figures 6b, 6c and 6f). 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 collapses the ocean's ensemble spread. 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 (Figure 4f) and AMOC (Figure 5f), in good agreement with Brune et al. (2018), indicating that ODA yields improvements.

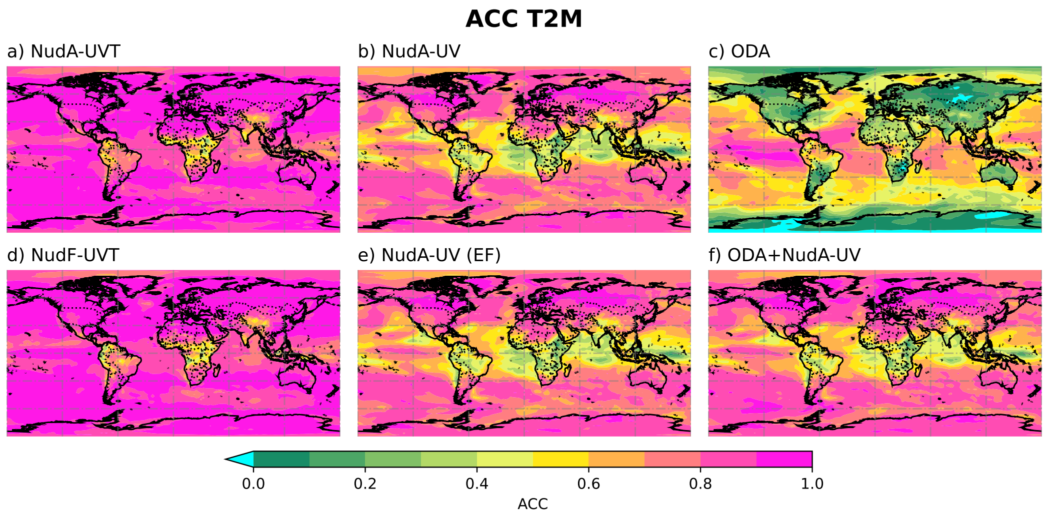

Figure 6. ACC of de-seasoned monthly T2M for a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV reanalyses computed against ERA5 for the period 1980–2010. Green-to-magenta colors indicate positive ACC values, and the cyan color indicates all the negative ACCs

3.2 Predictions

In this section, we evaluate the quality (skill) of the seasonal and decadal hindcasts initialized from the reanalysis (see section 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 (Figures 7a and 7b). On the other hand, the prediction skill for HC500 is low and only beats persistence after the sixth month; while SC500 never outperforms persistence (Figures 7c and 7d). 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 (Figure 7). This highlights the importance of ocean initialization in the prediction skill achieved.

While the globally averaged skill is low (ACCs below 0.4 in T2M and HC500, in Figures 7b and 7c), some regions show enhanced skill (Figures 8 and 9). Skill is most significant over the ocean and most notably in the tropical band driven by the El Niño–Southern Oscillation (ENSO) (Balmaseda & 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 (N. Keenlyside et al., 2020). There is also a region of significant skill in the

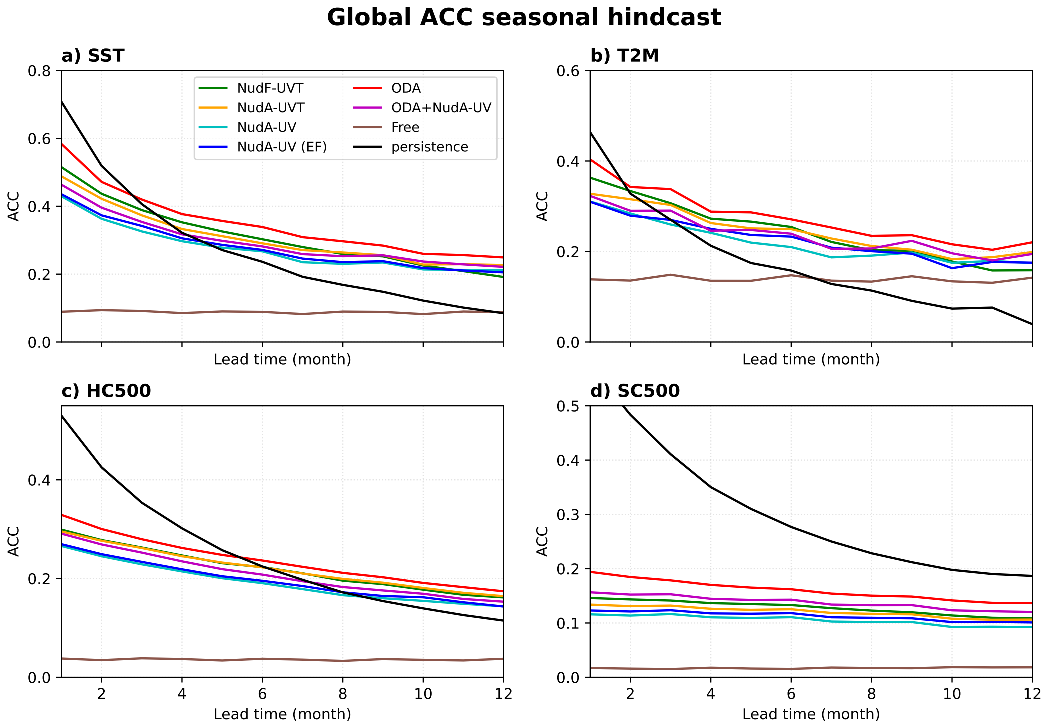

Figure 7. Global average ACC of the seasonal hindcast with lead month, for: a) sea surface temperature (SST), b) 2 m air temperature (T2M), c) 500 m heat content (HC500), and d) 500 m salinity content (SC500). Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV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

northern North Atlantic, the SPG, and the Iceland Sea, in agreement with other climate systems (e.g., Kirtman et al., 2014; Wang et al., 2019).

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 (Figure 10). All prediction systems outperform persistence, with ODA performing best. 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 S2 and S3), 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 (Figure 11). NudF-UVT performs better than all other systems, but it does not beat persistence until month six. Breaking down the analysis by start season (Figures S4b and S5b), we see that NudF-UVT performs best for the hindcast starting in May, slightly beating persistence at lead month 2 (ACC and RMSE$_u$), i.e., at the peak of the Atlantic Niño. Skillfully predict-

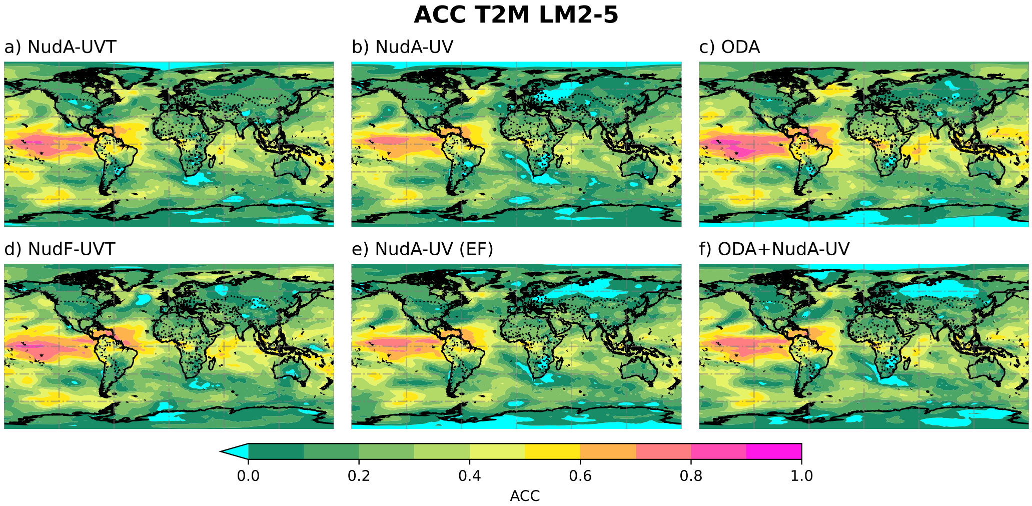

Figure 8. Seasonal hindcast 2-5 lead-month T2M ACC for a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV. Green-to-magenta colors indicate positive ACCs and cyan colour indicates all negative ACCs.

ing 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 S4 and S5), in agreement with those shown in Counillon et al. (2021). Overall, the skill remains poor in predicting Atlantic Niño variability.

Most of our experiments show good skill in predicting T2M and HC500 in the SPG at lead month 2-5. The best skill is achieved with ODA and, of all the nudging schemes, NudA-UVT performs best (Figures 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 Section 3.1) in this region is important for predictive skill at seasonal scales.

3.2.2 Decadal predictions

We assess our decadal predictions skill with ACC and RMSE$_u$ as a function of lead years. Figure 12 shows the global average skill with lead years for HC500, and Figure 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. 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 short period of our decadal hindcast, the ACCs pattern is relatively noisy, and even negative in some regions (cyan-to-blue colors in Figure 13). However, compared to the skill of a non-initialized hind-

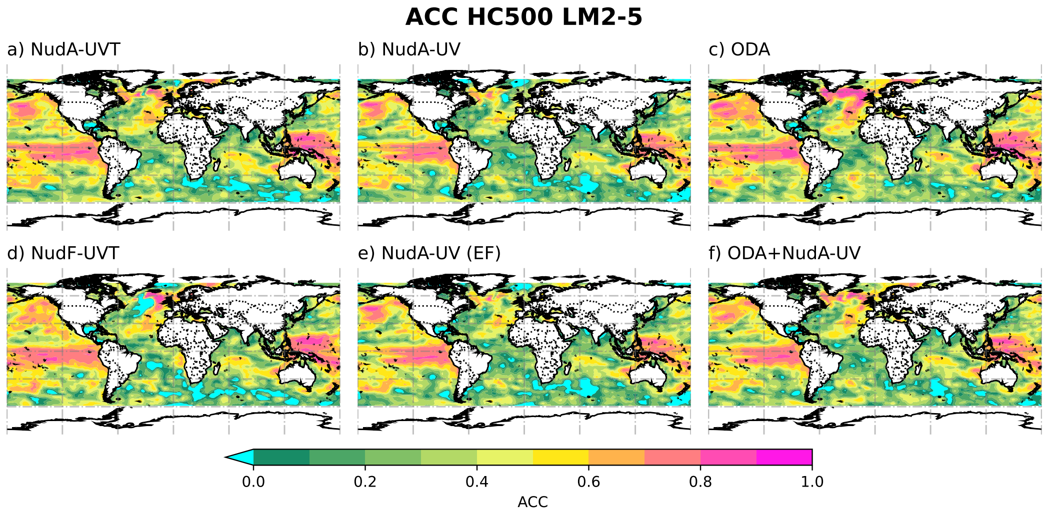

Figure 9. ACC of the seasonal hindcasts at lead-month 2-5 for HC500 with: a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV computed against EN4 objective analysis. Green-to-magenta colors indicate positive ACCs and cyan colour indicates all negative ACCs.

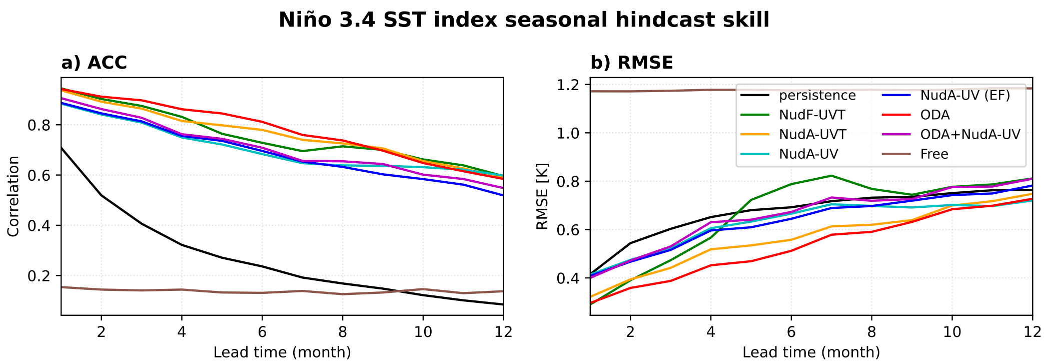

Figure 10. a) ACC of Niño 3.4 SST as a function of the lead month and b) is the same for RMSE$_u$ in K. Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, magenta to ODA+NudA-UV, brown to Free, and persistence is the solid black line.

cast (Figure 13e), all of our schemes show regions of improved skill. These regions are the North Atlantic, the Western Pacific Ocean, and the Indian Ocean. The regions for which skill is improved when compared to Free agree with the NorCPM experiment for CMIP6 DCPP carried for 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 & Son, 2022). The skill is negative in Free at the western coasts of North and South America 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

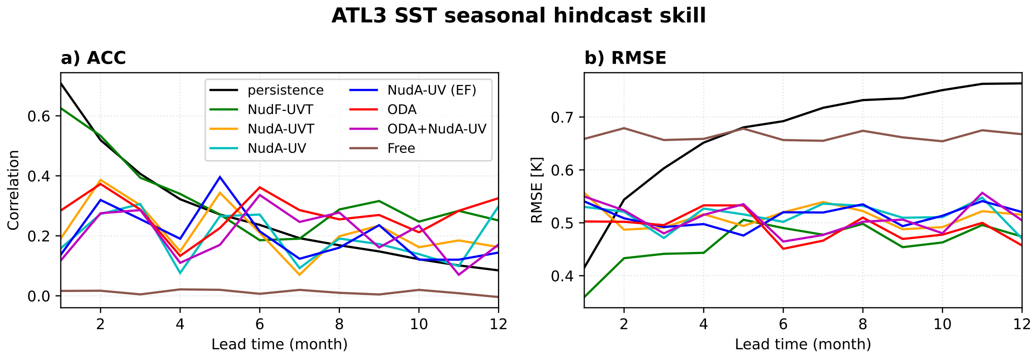

Figure 11. Same as Figure 10 for ATL3 SST.

PDV predictions (e.g., Guemas et al., 2012; Bilbao et al., 2021). Finally, ODA has the largest skill improvement in the SPG region, highlighting the importance of constraining the ocean to initialize decadal variability within the sub-polar North Atlantic.

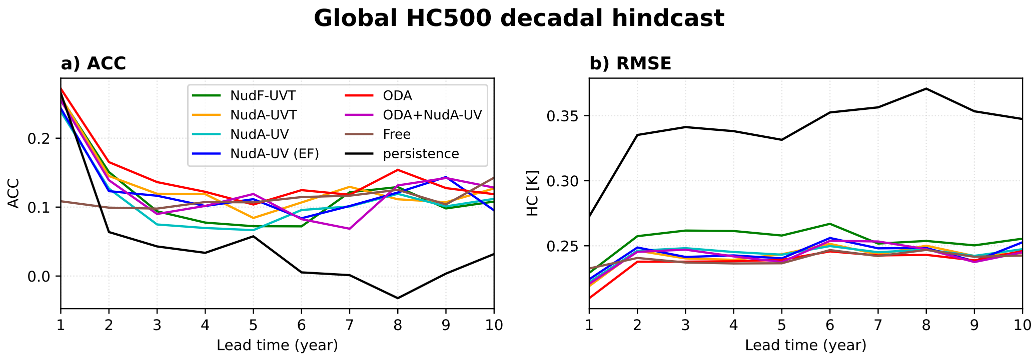

Figure 12. Global a) ACC and b) RMSE_u as a function of lead year for HC500. The line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, and magenta to ODA+NudA-UV, brown to Free, and the black line is persistence.

To further analyze the SPG variability, we evaluate the performance of the SPG index based on HC500 with lead-year (Figure 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) (Figure 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 & 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 Figure 4), showing that the mem-

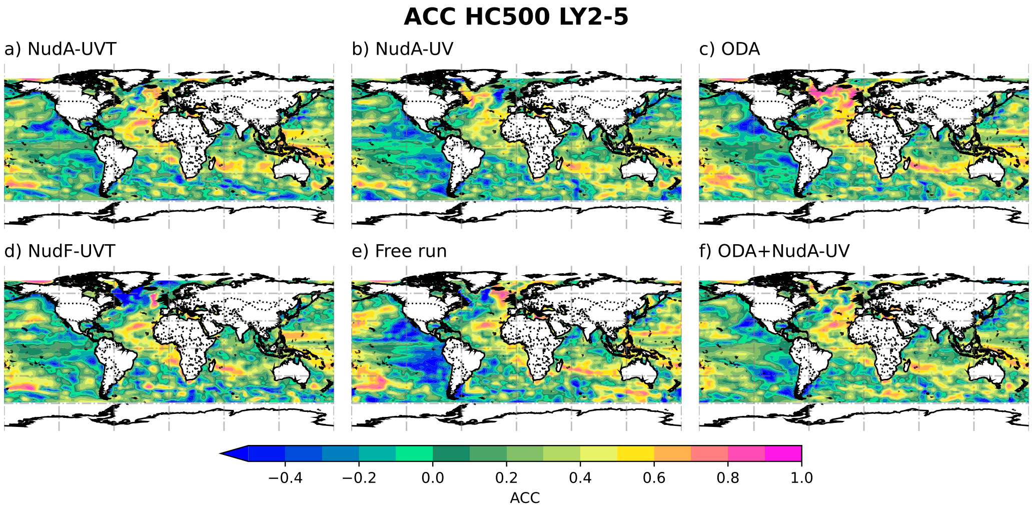

Figure 13. ACC for the decadal hindcast at lead year 2-5 of HC500 a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) Free and f) ODA+NudA-UV computed against EN4 objective analysis. Green-to-magenta colors indicate positive ACCs, while cyan-to-blue colors indicate negative ACCs.

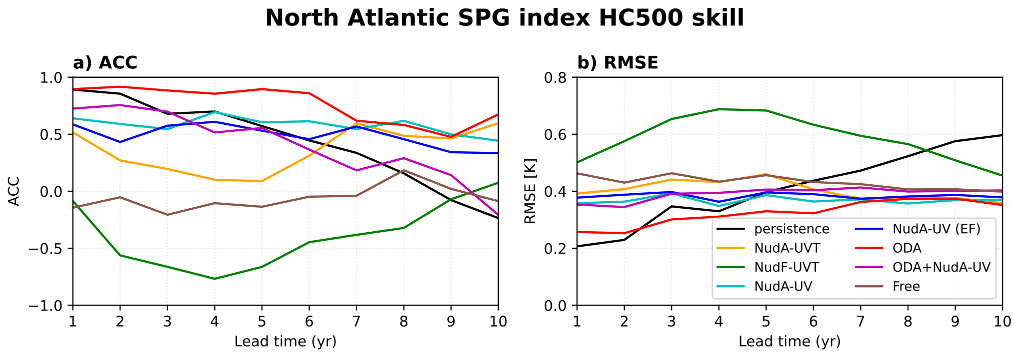

Figure 14. a) ACC and b) RMSE of the SPG index (computed from HC500 versus EN4 objective analysis) as a function of lead year. The line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, and magenta to ODA+NudA-UV, the brown line is Free, and the black line is persistence.

ory 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 26.5°N is shown in Figure 5 and compared to the RAPID observation program (RAPID, Johns et al., 2011) 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 Summary and Conclusions

In this study, we compared the potential of a large set of initialization schemes to constrain climate variability in an ESM 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, and 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. Still, anomaly initialization is performing overall best beyond short lead time.
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 the atmospheric nudging of horizontal winds limits the climatological change during the reanalysis, but very few differences are found during the seasonal hindcasts.
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. 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.

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). 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 & Derome, 1995) and consider using a weaker nudging.

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 & 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; F. J. Schevenhoven & Carrassi, 2021; F. Schevenhoven et al., 2023).

5 Open Research

The reanalysis and seasonal and decadal hindcasts data presented in this article are being organized and archived at `https://ns9039k.web.sigma2.no/lgarcia/initializations/`. The data is organized following the naming convention used in Table 1. Each directory contains the reanalysis and hindcasts monthly ensemble mean for 2 m temperature (T2M), sea surface temperature (SST), and temperature (T) and salinity (S). We also include the AMOC transport at 26.5°N, from annual averages. We provide the data on model grid and using netcdf format. The full simulations will be available on `https://archive.sigma2.no`, with a specific doi upon acceptance of the manuscript.

The code of the Norwegian Earth System Model (NorESM) and the Norwegian Climate Prediction Model (NorCPM version1) are available online on the Norwegian Earth System Modeling hub (`https://github.com/NorESMhub`). Specific details about NorCPM can be found in the website (`https://wiki.app.uib.no/norcpm/index.php/Norwegian_Climate_Prediction_Model`). The temperature and salinity (T, S) vertical profiles from EN4.2.1 objective analysis (Good et al., 2013) can be obtained from the Met Office Hadley Centre observations datasets website (`https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html`). And the sea surface temperature (SST) observations, HADISST2 (Rayner et al., 2003), are available at `https://www.metoffice.gov.uk/hadobs/hadisst2/data/download.html`. The reference data used for 2 m temperature (T2M), from ERA5 (Hersbach et al., 2020), can be obtained the Copernicus web services (`https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form`). The AMOC measurements used are available in the RAPID-AMOC website (`https://rapid.ac.uk`).

Acknowledgments

This study was partly funded by the Trond Mohn Foundation, under project number: BFS2018TMT01, the NFR INES ((INES; 270061), and Climate Futures (309562). This work has also received a grant for computer time from the Norwegian Program for supercomputing (NOTUR2, project number nn9039k) and a storage grant (NORSTORE, NS9039k).

References

Anderson, J. L. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. doi: 10.1175/1520-0493(2001)129⟨2884:AEAKFF⟩2.0.CO;2

Balmaseda, M., Alves, O., Arribas, A., Awaji, T., Behringer, D., Ferry, N., ... Stammer, D. (2009, 9). Ocean Initialization for Seasonal Forecasts. *Oceanography*, *22*(3), 154–159. Retrieved from `http://www.jstor.org/stable/24860997`

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

Barthélémy, S., Brajard, J., Bertino, L., & Counillon, F. (2022). Super-resolution data assimilation. *Ocean Dynamics*, *72*(8), 661–678. Retrieved from `https://doi.org/10.1007/s10236-022-01523-x` doi: 10.1007/s10236-022-01523-x

Bellprat, O., Massonnet, F., Siegert, S., Prodhomme, C., Macias-Gómez, D., Guemas, V., & Doblas-Reyes, F. (2017). Uncertainty propagation in observational references to climate model scales. *Remote Sensing of Environment*, *203*, 101–108. doi: 10.1016/J.RSE.2017.06.034

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., … Kristjánsson, J. E. (2013, 5). The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, *6*(3), 687–720. doi: 10.5194/gmd-6-687-2013

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., … Eldevik, T. (2021, 11). NorCPM1 and its contribution to CMIP6 DCPP. *Geosci. Model Dev*, *14*(11), 7073–7116. doi: 10.5194/gmd-14-7073-2021

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., … Vegas-Regidor, J. (2021, 2). Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth System Dynamics*, *12*(1), 173–196. Retrieved from `https://esd.copernicus.org/articles/12/173/2021/` doi: 10.5194/esd-12-173-2021

Bitz, C. M., Shell, K. M., Gent, P. R., Bailey, D. A., Danabasoglu, G., Armour, K. C., … Kiehl, J. T. (2012, 5). Climate Sensitivity of the Community Climate System Model, Version 4. *Journal of Climate*, *25*(9), 3053–3070. doi: 10.1175/JCLI-D-11-00290.1

Bleck, R., Rooth, C., Hu, D., & Smith, L. T. (1992). Salinity-driven Thermocline Transients in a Wind- and Thermohaline-forced Isopycnic Coordinate Model of the North Atlantic. *Journal of Physical Oceanography*, *22*(12), 1486–1505. doi: 10.1175/1520-0485(1992)022⟨1486:SDTTIA⟩2.0.CO;2

Bleck, R., & Smith, L. T. (1990, 3). A wind-driven isopycnic coordinate model of the north and equatorial Atlantic Ocean: 1. Model development and supporting experiments. *Journal of Geophysical Research: Oceans*, *95*(C3), 3273–3285. doi: 10.1029/JC095IC03P03273

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., … Eade, R. (2016, 10). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3751–3777. doi: 10.5194/gmd-9-3751-2016

Brune, S., & Baehr, J. (2020, 5). Preserving the coupled atmosphere–ocean feedback in initializations of decadal climate predictions. *Wiley Interdisciplinary Reviews: Climate Change*, *11*(3). doi: 10.1002/WCC.637

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J. (2018). Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts. *Climate Dynamics*, *51*, 1947–1970. doi: 10.1007/s00382-017-3991-4

Carrassi, A., Weber, R. J., Guemas, V., Doblas-Reyes, F. J., Asif, M., & Volpi, D. (2014, 4). Full-field and anomaly initialization using a low-order climate model: A comparison and proposals for advanced formulations. *Nonlinear Processes in Geophysics*, *21*(2), 521–537. doi: 10.5194/npg-21-521-2014

Choi, J., & Son, S. W. (2022, 4). Seasonal-to-decadal prediction of El Niño–Southern Oscillation and Pacific Decadal Oscillation. *npj Climate and Atmospheric Science 2022 5:1*, *5*(1), 1–8. doi: 10.1038/s41612-022-00251-9

Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., & Zheng, F. (2014). Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: A twin experiment. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *66*(1). doi: 10.3402/tellusa.v66.21074

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *68*(1), 32437. doi: 10.3402/tellusa.v68.32437

Counillon, F., Keenlyside, N., Toniazzo, T., Koseki, S., Teferi, D., Bethke, I., & Wang, Y. (2021). Relating model bias and prediction skill in the equatorial Atlantic. *Climate Dynamics*, *56*, 2617–2630. Retrieved from https://doi.org/10.1007/s00382-020-05605-8 doi: 10.1007/s00382-020-05605-8

Counillon, F., Keenlyside, N., Wang, S., Devilliers, M., Gupta, A., Koseki, S., & Shen, M.-L. (2023). Framework for an Ocean-Connected Supermodel of the Earth System. *Journal of Advances in Modeling Earth Systems*, *15*(3), e2022MS003310. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003310 doi: https://doi.org/10.1029/2022MS003310

Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., ... Wang, Q. (2014). North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modelling*, *73*, 76–107. doi: 10.1016/J.OCEMOD.2013.10.005

Dee, D. P. (2006, 1). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3323–3343. doi: 10.1256/qj.05.137

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Rosnay, d. P. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society Q. J. R. Meteorol. Soc*, *137*, 553–597. doi: 10.1002/qj.828

Ding, H., Greatbatch, R. J., Latif, M., & Park, W. (2015, 7). The impact of sea surface temperature bias on equatorial Atlantic interannual variability in partially coupled model experiments. *Geophysical Research Letters*, *42*(13), 5540–5546. doi: 10.1002/2015GL064799

Dippe, T., Greatbatch, R. J., & Ding, H. (2018, 7). On the relationship between Atlantic Niño variability and ocean dynamics. *Climate Dynamics*, *51*(1-2), 597–612. doi: 10.1007/S00382-017-3943-Z/FIGURES/12

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., ... Van Oldenborgh, G. J. (2013, 4). Initialized near-term regional climate change prediction. *Nature Communications 2013 4:1*, *4*(1), 1–9. doi: 10.1038/ncomms2704

Dunstone, N. J., & Smith, D. M. (2010, 1). Impact of atmosphere and sub-surface ocean data on decadal climate prediction. *Geophysical Research Letters*, *37*(2), 2709. doi: 10.1029/2009GL041609

El Gharamti, M., McCreight, J. L., Noh, S. J., Hoar, T. J., Rafieeinasab, A., & Johnson, B. K. (2021, 9). Ensemble streamflow data assimilation using WRF-Hydro and DART: Novel localization and inflation techniques applied to Hurricane Florence flooding. *Hydrology and Earth System Sciences*, *25*(9), 5315–5336. doi: 10.5194/hess-25-5315-2021

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

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014, 8). Why Should Ensemble Spread Match the RMSE of the Ensemble Mean? *Journal of Hydrometeorology*, *15*(4), 1708–1713. doi: 10.1175/JHM-D-14-0008.1

García-Serrano, J., Guemas, V., & Doblas-Reyes, F. J. (2015, 5). Added-value from initialization in predictions of Atlantic multi-decadal variability. *Climate Dynamics*, *44*(9-10), 2539–2555. doi: 10.1007/S00382-014-2370-7/FIGURES/9

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), 6704–6716. doi: 10.1002/2013JC009067

Gouretski, V., & Reseghetti, F. (2010, 6). 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. doi: 10.1016/j.dsr.2010.03.011

Guemas, V., Doblas-Reyes, F. J., Lienert, F., Soufflet, Y., & Du, H. (2012, 10).

Identifying the causes of the poor decadal climate prediction skill over the North Pacific. *Journal of Geophysical Research: Atmospheres*, *117*(D20). Retrieved from http://doi.wiley.com/10.1029/2012JD018004 doi: 10.1029/2012JD018004

Häkkinen, S., & Rhines, P. B. (2004, 4). Decline of Subpolar North Atlantic Circulation during the 1990s. *Science*, *304*(5670), 555–559. doi: 10.1126/SCIENCE.1094917/SUPPL/FILE/HAKKINEN.SOM.PDF

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

Hawkins, E., & Sutton, R. (2009, 8). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. doi: 10.1175/2009BAMS2607.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J. N. (2020, 7). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/QJ.3803

Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M., … Roberts, M. J. (2017, 12). Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales? *Ocean Modelling*, *120*, 120–136. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1463500317301774 doi: 10.1016/j.ocemod.2017.11.002

Hoke, J. E., & Anthes, R. A. (1976, 12). The Initialization of Numerical Models by a Dynamic-Initialization Technique. *Monthly Weather Review*, *104*(12), 1551–1556. doi: https://doi.org/10.1175/1520-0493(1976)104⟨1551:TIONMB⟩2.0.CO;2

Houtekamer, P. L., & Derome, J. (1995, 7). Methods for Ensemble Prediction. *Monthly Weather Review*, *123*(7), 2181–2196. doi: 10.1175/1520-0493(1995)123⟨2181:mfep⟩2.0.co;2

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., … Marshall, S. (2013, 9). The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. doi: 10.1175/BAMS-D-12-00121.1

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., … Weston, P. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3130 doi: https://doi.org/10.1002/qj.3130

Johns, W. E., Baringer, M. O., Beal, L. M., Cunningham, S. A., Kanzow, T., Bryden, H. L., … Curry, R. (2011). Continuous, Array-Based Estimates of Atlantic Ocean Heat Transport at 26.5°N. *Journal of Climate*, *24*(10), 2429–2449. doi: 10.1175/2010JCLI3997.1

Karspeck, A. R., Danabasoglu, G., Anderson, J., Karol, S., Collins, N., Vertenstein, M., … Craig, A. (2018). A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed. *Quarterly Journal of the Royal Meteorological Society*, *144*(717), 2404–2430. doi: 10.1002/qj.3308

Keenlyside, N., Kosaka, Y., Vigaud, N., Robertson, A. W., Wang, Y., Dommenget, D., … Matei, D. (2020). Basin interactions and predictability. In C. R. Mechoso (Ed.), *Interacting climates of ocean basins: Observations, mechanisms, predictability, and impacts* (p. 258–292). Cambridge University Press. doi: 10.1017/9781108610995.009

Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L., & Roeckner, E. (2008, 5). Advancing decadal-scale climate prediction in the North Atlantic sector. *Na-*

ture, *453*(7191), 84–88. doi: 10.1038/nature06921

Kirkevåg, A., Iversen, T., Seland, Ø., Hoose, C., Kristjánsson, J. E., Struthers, H., ... Schulz, M. (2012). Aerosol-climate interactions in the Norwegian Earth System Model – NorESM. *Geosci. Model Dev. Discuss.*, *5*, 2599–2685. doi: 10.5194/gmdd-5-2843-2012

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., ... Wood, E. F. (2014, 4). The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. Retrieved from `https://journals.ametsoc.org/view/journals/bams/95/4/bams-d-12-00050.1.xml` doi: 10.1175/BAMS-D-12-00050.1

Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C. J., & Russell, L. M. (2012, 12). Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the Community Atmosphere Model 5. *Journal of Geophysical Research: Atmospheres*, *117*(D23), 23204. doi: 10.1029/2012JD018588

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., ... Slater, A. G. (2011, 1). Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *3*(1), n/a-n/a. doi: 10.1029/2011MS00045

Lohmann, K., Drange, H., Bentsen, M., Helge, A. E., Ae, D., & Bentsen, M. (2009). Response of the North Atlantic subpolar gyre to persistent North Atlantic oscillation like forcing. *Climate Dynamics*, *32*(2), 273–285. doi: 10.1007/s00382-008-0467-6

Lu, F., Harrison, M. J., Rosati, A., Delworth, T. L., Yang, X., Cooke, W. F., ... Adcroft, A. (2020). GFDL's SPEAR Seasonal Prediction System: Initialization and Ocean Tendency Adjustment (OTA) for Coupled Model Predictions. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2020MS002149. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002149` doi: https://doi.org/10.1029/2020MS002149

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013, 11). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, *41*(9-10), 2393–2409. doi: 10.1007/s00382-012-1599-2

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., ... Albers, J. (2020, 5). Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond. *Bulletin of the American Meteorological Society*, *101*(5), E608-E625. doi: 10.1175/BAMS-D-18-0326.1

Mariotti, A., Ruti, P. M., & Rixen, M. (2018, 3). Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate and Atmospheric Science 2018 1:1*, *1*(1), 1–4. doi: 10.1038/s41612-018-0014-z

Massonnet, F., Bellprat, O., Guemas, V., & Doblas-Reyes, F. J. (2016). Using climate models to estimate the quality of global observational data sets. *Science*, *354*(6311), 452–455. doi: 10.1126/science.aaf6369

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., ... Stockdale, T. (2009, 10). Decadal prediction: Can it be skillful? *Bulletin of the American Meteorological Society*, *90*(10), 1467–1485. doi: 10.1175/2009BAMS2778.1

Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., ... Xie, S. P. (2021, 4). Initialized Earth System prediction from subseasonal to decadal timescales. *Nature Reviews Earth & Environment 2021 2:5*, *2*(5), 340–357. doi: 10.1038/s43017-021-00155-x

Mochizuki, T., Ishii, M., Kimoto, M., Chikamoto, Y., Watanabe, M., Nozawa, T., ... Mori, M. (2010). Pacific decadal oscillation hindcasts relevant to near-term climate prediction. *Proceedings of the National Academy of Sciences*, *107*(5), 1833-1837. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.0906531107` doi: 10.1073/pnas.0906531107

Neale, R., Richter, J., Conley, A., Park, S., Lauritzen, P., Gettelman, A., … Lin, S.-J. (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. *Proceedings of the National Academy of Sciences*, *116*(49), 24390–24395. doi: 10.1073/pnas.1906691116

Pohlmann, H., Jungclaus, J. H., Köhl, A., Stammer, D., & Marotzke, J. (2009, 7). Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic. *Journal of Climate*, *22*(14), 3926–3938. doi: 10.1175/2009JCLI2535.1

Polkova, I., Brune, S., Kadow, C., Romanova, V., Gollan, G., Baehr, J., … Stammer, D. (2019, 1). Initialization and Ensemble Generation for Decadal Climate Predictions: A Comparison of Different Methods. *Journal of Advances in Modeling Earth Systems*, *11*(1), 149–172. doi: 10.1029/2018MS001439

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., … Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(14). doi: 10.1029/2002JD002670

Robson, J. (2010). *Understanding the performance of a decadal prediction system* (Doctoral dissertation). doi: 10.13140/RG.2.1.2183.2560

Robson, J. I., Sutton, R. T., & Smith, D. M. (2012, 10). Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. *Geophysical Research Letters*, *39*(19). doi: 10.1029/2012GL053370

Rodwell, M. J., Lang, S. T. K., Ingleby, N. B., Bormann, N., Hólm, E., Rabier, F., … Yamaguchi, M. (2016). Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 443–454. doi: 10.1002/qj.2663

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999, 9). A dipole mode in the tropical Indian Ocean. *Nature 1999 401:6751*, *401*(6751), 360–363. doi: 10.1038/43854

Sakov, P., Counillon, F., Bertino, L., Lister, K. A., Oke, P. R., & Korablev, A. (2012). TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, *8*(4), 633–656. doi: 10.5194/os-8-633-2012

Sakov, P., & Oke, P. R. (2008, 3). A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *60 A*(2), 361–371. doi: 10.1111/j.1600-0870.2007.00299.x

Schevenhoven, F., Keenlyside, N., Counillon, F., Carrassi, A., Chapman, W. E., Devilliers, M., … Duane, G. S. (2023). Supermodeling: improving predictions with an ensemble of interacting models. *Bulletin of the American Meteorological Society*. Retrieved from `https://journals.ametsoc.org/view/journals/bams/aop/BAMS-D-22-0070.1/BAMS-D-22-0070.1.xml` doi: https://doi.org/10.1175/BAMS-D-22-0070.1

Schevenhoven, F. J., & Carrassi, A. (2021). Training a supermodel with noisy and sparse observations: a case study with cpt and the synch rule on speedo-v. 1. *Geoscientific Model Development Discussions*, *2021*, 1–23.

Singh, T., Counillon, F., Tjiputra, J., Wang, Y., & Gharamti, M. E. (2022). Estimation of Ocean Biogeochemical Parameters in an Earth System Model Using the Dual One Step Ahead Smoother: A Twin Experiment. *Frontiers in Marine Science*, *9*. doi: 10.3389/fmars.2022.775394

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007, 8). Improved surface temperature prediction for the coming decade from a global climate model. *Science*, *317*(5839), 796–799. doi: 10.1126/science.1139540

Smith, D. M., Eade, R., & Pohlmann, H. (2013, 12). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, *41*(11-12), 3325–3338. doi: 10.1007/s00382-013-1683-2

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012, 4). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. doi: 10.1175/BAMS-D-11-00094.1

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., ... Hibbard, K. (2011). The representative concentration pathways: an overview. *Climatic Change*, *109*, 5–31. doi: 10.1007/s10584-011-0148-z

Volpi, D., Guemas, V., & Doblas-Reyes, F. J. (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. *Climate Dynamics*, *49*(4), 1181–1195. doi: 10.1007/s00382-016-3373-3

Wang, Y., Counillon, F., Bertino, L., & Wang, . Y. (2016). Alleviating the bias induced by the linear analysis update with an isopycnal ocean model. *Quarterly Journal of the Royal Meteorological Society Q. J. R. Meteorol. Soc*, *142*, 1064–1074. doi: 10.1002/qj.2709

Wang, Y., Counillon, F., Bethke, I., Keenlyside, N., Bocquet, M., & Shen, M. l. (2017, 6). Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. *Ocean Modelling*, *114*, 33–44. doi: 10.1016/j.ocemod.2017.04.007

Wang, Y., Counillon, F., Keenlyside, N., Svendsen, L., Gleixner, S., Kimmritz, M., ... Yongqi Gao (2019). Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF. *Climate Dynamics*, *53*, 5777–5797. doi: 10.1007/s00382-019-04897-9

Weber, R. J., Carrassi, A., & Doblas-Reyes, F. J. (2015, 11). Linking the Anomaly Initialization Approach to the Mapping Paradigm: A Proof-of-Concept Study. *Monthly Weather Review*, *143*(11), 4695–4713. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/143/11/mwr-d-14-00398.1.xml` doi: 10.1175/MWR-D-14-00398.1

Webster, P. J., Moore, A. M., Loschnigg, J. P., & Leben, R. R. (1999, 9). Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. *Nature 1999 401:6751*, *401*(6751), 356–360. doi: 10.1038/43848

Wilks, Daniel. (2019). *Statistical Methods in the Atmospheric Sciences - 4th Edition.* Retrieved from `https://www.elsevier.com/books/statistical-methods-in-the-atmospheric-sciences/wilks/978-0-12-815823-4`

Yeager, S. G., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012). A decadal prediction case study: Late twentieth-century North Atlantic Ocean heat content. *Journal of Climate*, *25*(15), 5173–5189. doi: 10.1175/JCLI-D-11-00595.1

Yeager, S. G., & Robson, J. I. (2017, 6). Recent Progress in Understanding and Predicting Atlantic Decadal Climate Variability. *Current Climate Change Reports*, *3*(2), 112–127. doi: 10.1007/s40641-017-0064-z

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., ... Lohmann, U. (2014, 8). Technical note: On the use of nudging for aerosol-climate model intercomparison studies. *Atmospheric Chemistry and Physics*, *14*(16), 8631–8645. doi: 10.5194/ACP-14-8631-2014

Zhang, S., Rosati, A., & Delworth, T. (2010, 10). The adequacy of observing systems in monitoring the Atlantic meridional overturning circulation and North Atlantic climate. *Journal of Climate*, *23*(19), 5311–5324. doi: 10.1175/2010JCLI3677.1

Zhang, S., Rosati, A., & Harrison, M. J. (2009). Detection of multidecadal oceanic variability by ocean data assimilation in the context of a "perfect" coupled model. *Journal of Geophysical Research: Oceans*, *114*(12), 12018. doi: 10.1029/2008JC005261

Intercomparison of initialization methods for Seasonal-to-Decadal Climate Predictions with the NorCPM

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

[1]Geophysical Institute, University of Bergen, Bjerknes Centre for Climate Research, 5007 Bergen, Norway
[2]Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, 5006 Bergen, Norway

Key Points:

- Constraining the ocean state to observations produces more skillful predictions than constraining the atmospheric state
- Full-field performs better than anomaly initialization at short-lead times in specific regions, but drift degrades the skill rapidly
- Anomaly nudging of atmospheric momentum can achieve skillful decadal prediction and minimizes hindcast drift

Corresponding author: Lilian Garcia-Oliva, `lilian.garcia@uib.no`

Abstract

Initialization is essential for accurate seasonal-to-decadal (S2D) climate predictions. The initialization schemes used differ on the component initialized, the Data Assimilation (DA) method, or the technique. We compare five popular schemes within NorCPM following the same experimental protocol: reanalysis from 1980–2010 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 scheme ODA+NudA-UV provides a first attempt at joint initialization of the ocean and atmospheric components. During the reanalysis, atmospheric nudging leads to atmosphere and land components best synchronized with observations. Conversely, ODA best synchronizes the ocean component with observations. The atmospheric nudging schemes are better at reproducing specific events, such as the rapid North Atlantic subpolar gyre (SPG) 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 one-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.

Plain Language Summary

This study explores the impact of a wide range of standard initialization schemes on the performance of coupled reanalysis and seasonal-to-decadal predictions produced with the same Earth System Model. We compare atmospherically-driven initialization versus ocean initialization. We also compare full-field initialization —meaning where the observations are used as are—versus anomaly initialization —when the climatological difference between the model and observations is removed. All schemes have strengths and weaknesses. As expected, ocean initialization works best in the ocean, while atmospherically driven initialization works best in the atmosphere and land. Ocean initialization has the best performance overall for seasonal and decadal predictions. Still, the atmospherically driven initialization works better for some specific regions and events—for example, the strong North Atlantic subpolar gyre shift in 1995. Full-field initialization performs better than anomaly initialization at short lead times, and it improves performance in regions where the mean state is important for representing the variability, such as the Tropical Atlantic. Constraining atmospheric temperature is important for reanalysis and seasonal prediction while constraining only the winds works better for decadal prediction.

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; N. S. Keenlyside et al., 2008; Meehl et al., 2009; Hawkins & 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 & 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).

There are various schemes for accurately initializing S2D predictions. One common practice is to initialize each component of the Earth System Models (ESMs) individually, 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 & 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 & 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 initialization shocks and prediction drift. 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 (J. 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) illustrating 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 & 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., N. S. 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., S. Zhang et al., 2009; Pohlmann et al., 2009; Karspeck et al., 2018; Counillon et al., 2016; Brune & 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 S. 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) indicate 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 (NorCPM, Counillon et al., 2014, 2016) that combines the Norwegian Earth System Model (NorESM, Bentsen et al., 2013) and the Ensemble Kalman Filter (EnKF, Evensen, 2003) data assimilation method. 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 summary and conclusions are presented in Section 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 & Smith, 1990; Bleck et al., 1992), using density as its vertical coordinate.

We use the medium-resolution version of NorESM. The atmosphere and land components use a $1.9°\times2.5°$ 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.

2.2 Ocean data assimilation with the EnKF

The Ensemble Kalman Filter (EnKF, Evensen, 2003) 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 & 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:

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

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:

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

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 & Reseghetti, 2010). The observation error for the hydrographic profiles and the localization radius varies with latitude as described in Wang et al. (2017). 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 (equation 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 & 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 equation (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 et al., 2014).

One can also apply anomaly nudging (Zhang et al., 2014), where the right-hand side of equation (6) is replaced by the anomaly terms, i.e., $X \to A$. 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 equation (7) can be expressed as:

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

With the formulations of equations (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 et al. (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 below 60 km height with tapering between 50 km to 60 km, while in the land and ocean surfaces, the model is constrained towards the prescribed data.

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 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 Section 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 Section 2.2 for details on the practical implementation). Finally, the scheme ODA+NudA-UV combines the ODA and NudA-UV (EF) experiments. We did 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 (Section 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.

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.

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 15^{th} of January, April, July, and October each year during 1985–2010 and run for a year. The decadal hindcasts start on the 15^{th} of October every other year and run for 11 years each. Each hindcast runs nine realizations (ensemble members). Initial conditions are taken from the first nine members of the 30-member ensemble reanalyses. Note that this choice does not influence the results because all members are equally likely.

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 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. All climatologies are computed using the 1980-2010 period.

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:

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

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

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, Daniel, 2019).

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:

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

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

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

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 equation (11).

For the global (or regional indices) statistics, we use grid cell area weighting:

$$\mathrm{RMSE}_u = \sum_i a_i \mathrm{RMSE}_{ui} \left(\sum_j a_j\right)^{-1}, \tag{15}$$

and

$$\mathrm{ACC} = \sum_i a_i \mathrm{ACC}_i \left(\sum_j a_j\right)^{-1}. \tag{16}$$

where a_i is the area of the corresponding i-th grid cell.

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. For the ocean surface temperature, we take SST observations from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST2, Rayner et al., 2003). We interpolate our ocean outputs towards HadISST2 horizontal grid. We obtain subsurface temperature and salinity data from the EN4.2.1 objective analysis (EN4.2.1, Gouretski & Reseghetti, 2010). We re-grid and interpolate our ocean subsurface output to EN4.2.1 dataset resolution for the comparisons. Furthermore, we consider the heat and salinity content in the first 500 m, named HC500 and SC500 respectively. We define them as the ocean depth's average temperature (and salinity).

For the verification of the decadal hindcasts, we also use the Atlantic meridional overturning circulation (AMOC) at 26° North from the RAPID dataset (RAPID, Johns et al., 2011).

3 Results

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

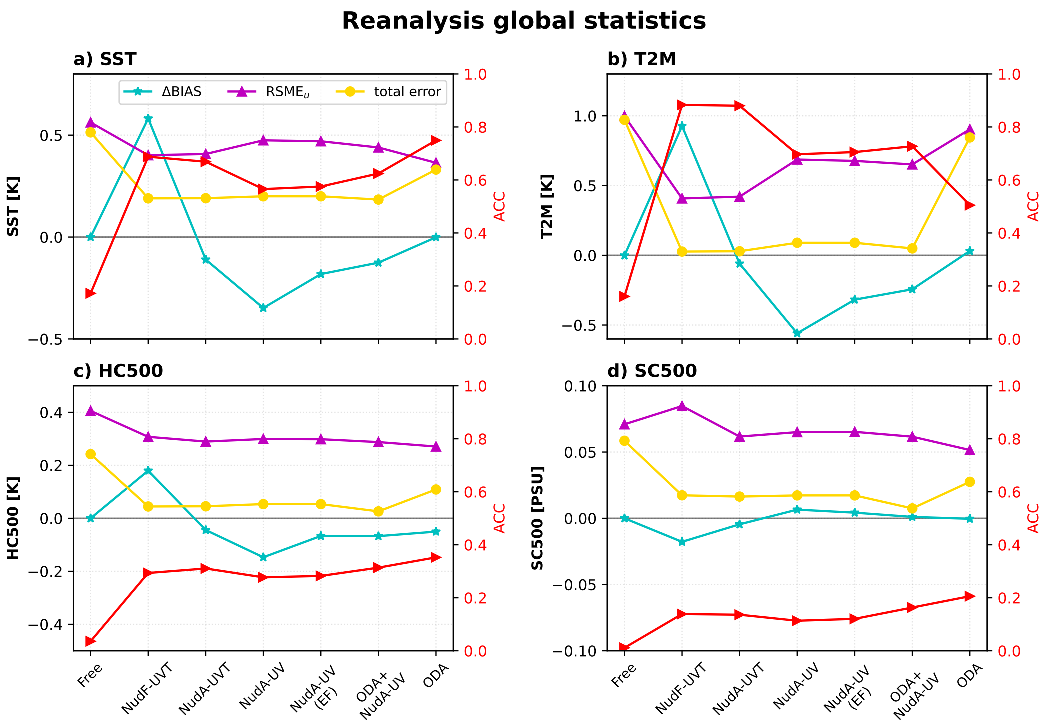

Figure 1. Global statistics of the reanalyses computed over 1980–2010, for a) SST, b) T2M, c) HC500, and d) SC500. The left-hand y-axis (in black) displays units for RMSE$_u$ (magenta), ΔBIAS (cyan), and total error (yellow), while the red right-hand y-axis is for ACC (red). The reanalyses are said to be reliable when the total error (yellow) and RMSE$_u$ (magenta) overlap. The black horizontal line marks zero.

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 (Figure 1). Globally, the reanalysis from NudF-UVT is marginally better for SST and T2M (Figures 1a and 1b), but yields a degradation for HC500 (Figure 1c) and SC500 (Figure 1d). Most of this degradation occurs in the SPG, the tropical and South Atlantic, and the Southern Ocean (Figure 2d). Furthermore, NudF-UVT exhibits a substantial RMSE$_u$ drift of HC500 and SC500 (Figure 3). Such RMSE$_u$ drift follows a parabolic shape, as the mean climatology (used for computing the metric, equation 11) is reached 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 Figure 1). 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 (equation 14) being much smaller than the RMSE$_u$ (Figures 1a and 1b). This implies that the ensemble spread (not shown) collapses during the reanalyses. The reliability for HC500 and SC500 is also poor (Figures 1c and 1d). It should be acknowledged that the HC500 (and to a minor extent SC500) reliability of Free is already too low, although it should, by construction, be satisfied by the experiment. This suggests 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 (Figures 4a and 4b), 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 & Rhines, 2004; Yeager & Robson, 2017), which induces a preconditioning of the ocean circulation state (Lohmann et al., 2009; J. I. Robson et al., 2012). Moreover, both schemes fail 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 (Figure 5b) that affects the poleward heat transport. The verification period with the RAPID (RAPID, Johns et al., 2011) 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.

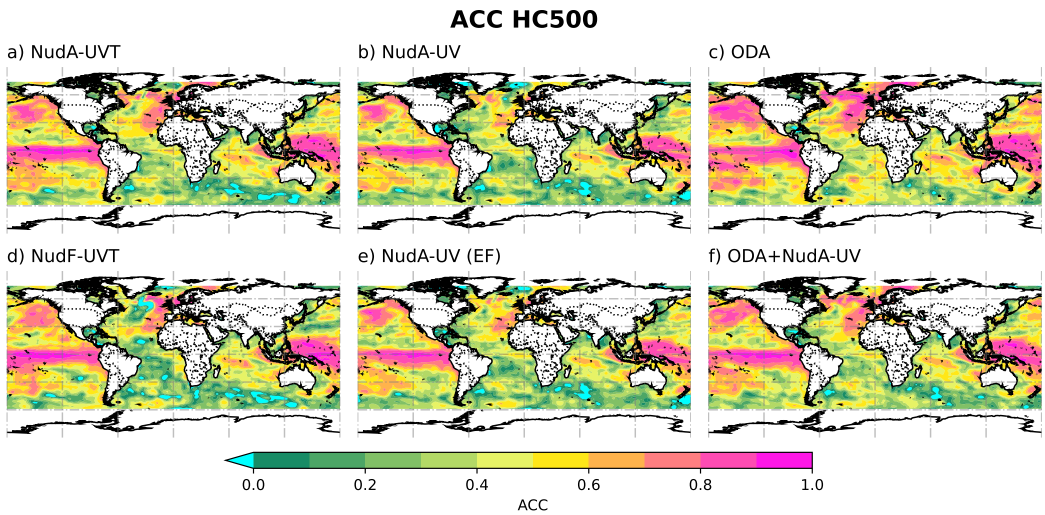

Figure 2. ACC of monthly HC500 anomalies a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV reanalysis computed against EN4 objective analysis for the period 1980–2010. Green-to-magenta colors indicate positive ACCs, and the cyan color indicates all the negative ACCs.

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 T2M), nudging only horizontal winds degrades performance (Figures 1a and 1b). For T2M, for example, NudA-UV reduces error by 0.3 K compared to Free, whereas NudA-UVT reduces it by 0.6 K. The degraded performance of NudA-UV is largest over the tropical band and is less pronounced at mid-to-high latitudes (Figures 6a and 6b). The reliability for T2M is slightly improved in NudA-UV compared to NudA-UVT (see also Table S1). In NudA-UV, there is a significant increase in climatological change ΔBIAS for SST and T2M. On the other hand, NudA-UVT sustains ΔBIAS near 0 K due to temperature nudging. Below the surface, the global skill performance of NudA-UV and NudA-UVT are similar for HC500 and SC500

(Figures 1c and 1d), 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 (Figures S6c and S6e), North East Atlantic, and South Pacific (Figures 2a and 2b). The performance for the SPG (Figure 4) and AMOC (Figure 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.

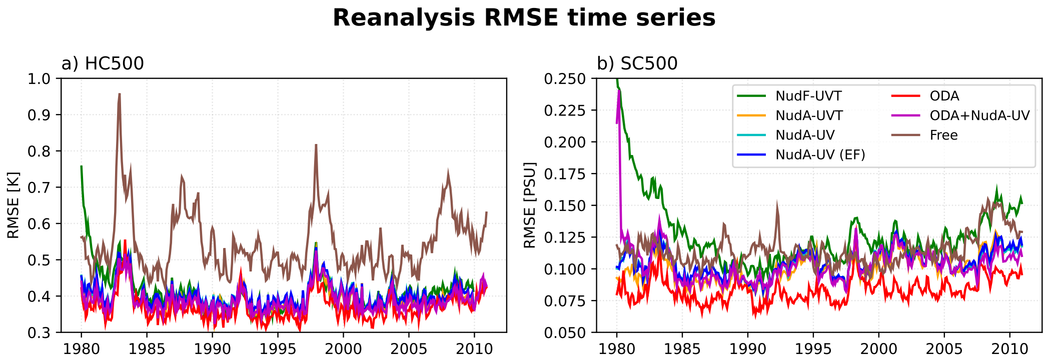

Figure 3. Time series of RMSE$_u$ for a) HC500 and b)SC500 in the different reanalyses computed against EN4 objective analysis. Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, magenta to ODA+NudA-UV, and brown is Free.

The default implementation of nudging in CAM4 deactivates the energy conservation fix in the atmospheric component (see section 2.3). Here, 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 (Figures 1a and 1d). 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 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 (Figure S6). And comparing NudA-UV and NudA-UV (EF), we find that the energy fix is very effective in improving the representation of the trend in the Iceland basin ($R = 0.31$ and 0.61, respectively, in Figures S6e and S6g).

We now compare atmospheric constraints versus ocean constraints for coupled reanalysis. The skill for T2M (Figure 1b) using atmospheric nudging is substantially better 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 (Figures 6a and 6c). 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 HadISST2 analysis, which is assimilated in the ODA system. Meanwhile, the verification of T2M is done against ERA5, 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 (Figures 1c-1d). This is also clear from Figure 3, where ODA has a consis-

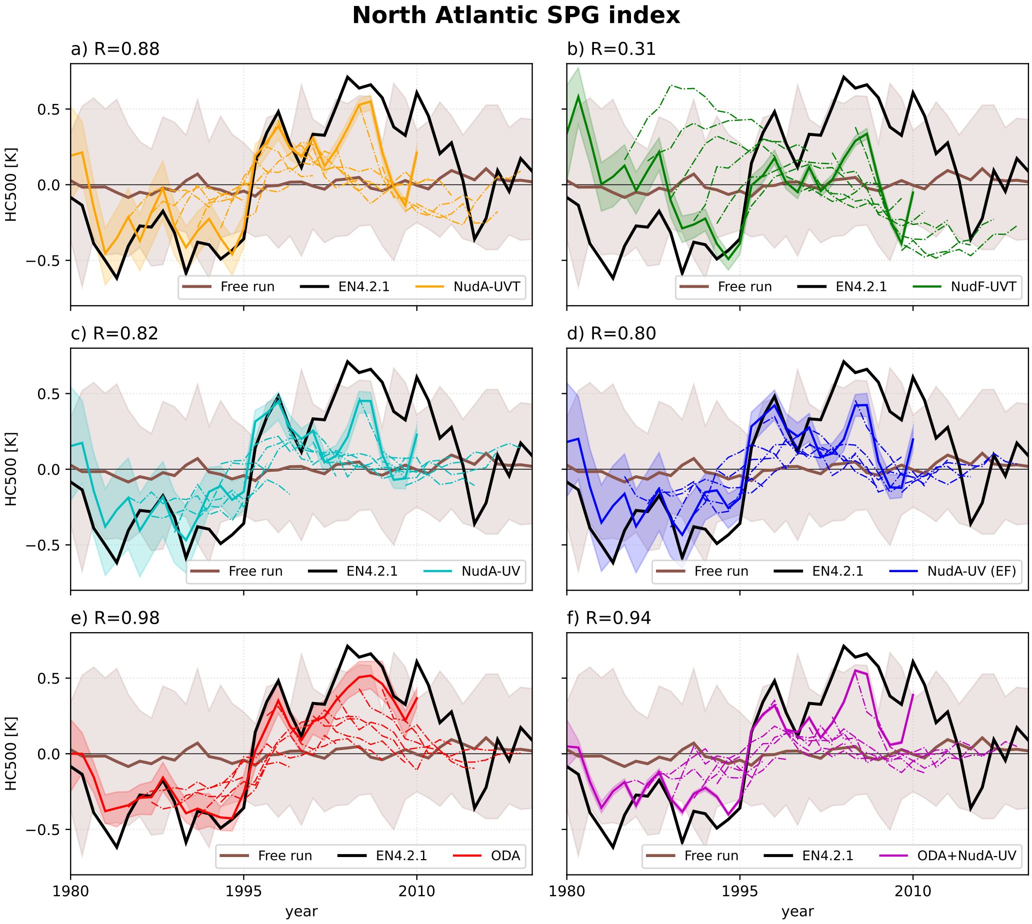

Figure 4. HC500 anomalies in the SPG box (48°-65°N, 60°-15°S) for a) NudA-UVT, b) NudF-UVT, c) NudA-UV, d) NudA-UV (EF), e) ODA and f) ODA+NudA-UV reanalyses. Solid-colored lines represent the ensemble mean of reanalysis, dash-dotted lines correspond to hindcast schemes, and the solid brown line is Free. 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

tently 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 (Figure 2c), where atmospheric nudging introduces a patch of low-skill in the Irminger and Icelandic Seas (see, for example, Figures 2a and 2b). The reliability of the system is also better preserved as we see a closer match between RMSE_u and total error σ (Figures 1c and 1d, magenta and yellow lines). In the ODA system the reliability is only marginally degraded from Free and much less than atmospheric nudging. In the case of regional indexes, ODA achieves overall the best correlation for the SPG index ($R = 0.98$, Figure 4e), and it is the only system that sustains the weak 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, Figure 4a). This is because the NAO constraint is very weak in the ODA system, and the system only adjusts a-posteriori for errors in the atmospheric forcing. Finally, for the AMOC at 26.5°N, there is a long term weakening with a stronger weak anomaly from 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 (Figures 5c and 5e), suggesting that this feature is better constrained with atmospheric variability.

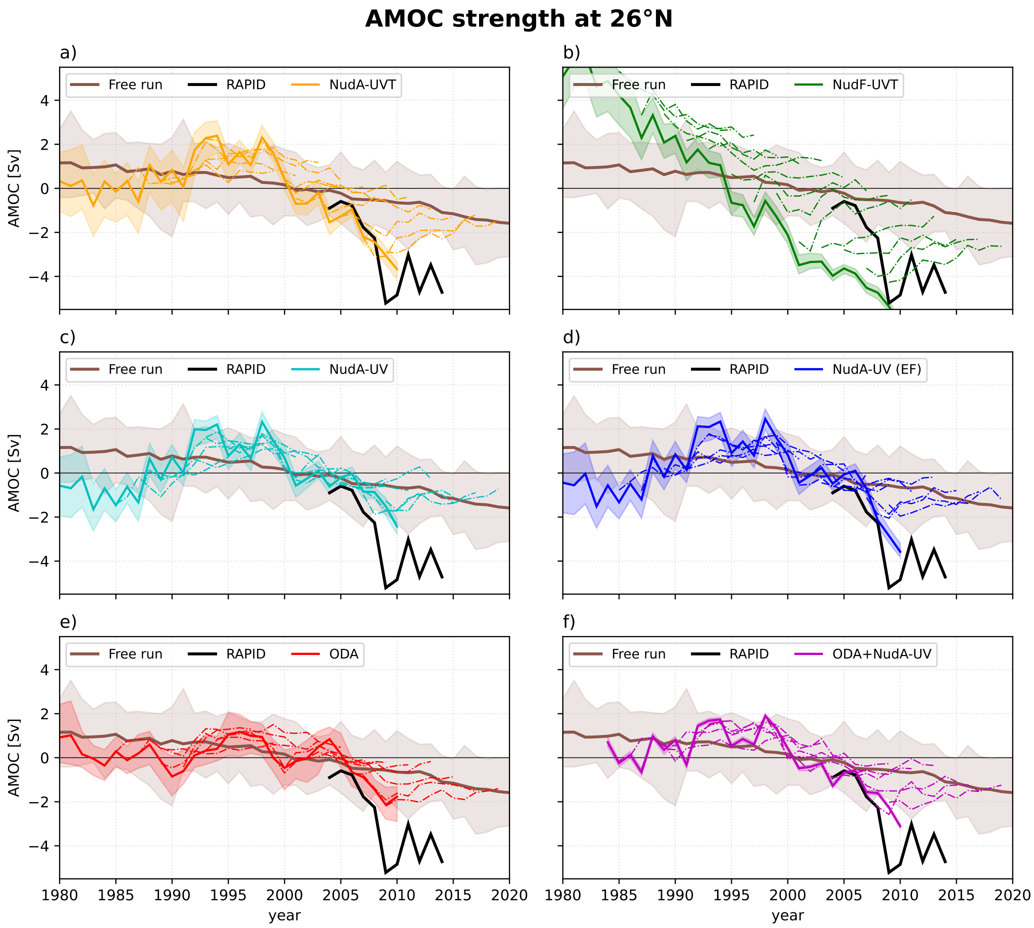

Figure 5. AMOC transport anomalies at 26.5°N with respect to the 2005–2010 period for a) NudA-UVT, b) NudF-UVT, c) NudA-UV, d) NudA-UV (EF), e) ODA and f) ODA+NudA-UV reanalyses. Solid-colored lines represent the ensemble mean of reanalysis, dash-dotted lines correspond to hindcast schemes, and the solid brown line is Free. Shading denotes ensemble minima and maxima. The solid black line is the RAPID observations.

Given the complementary skills of atmospheric nudging and the ODA systems, one would expect their combination to work best. However, comparing the global statistics of ODA and ODA+NudA-UV (Figure 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 Figures 6b, 6c and 6f). 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 collapses the ocean's ensemble spread. 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 (Figure 4f) and AMOC (Figure 5f), in good agreement with Brune et al. (2018), indicating that ODA yields improvements.

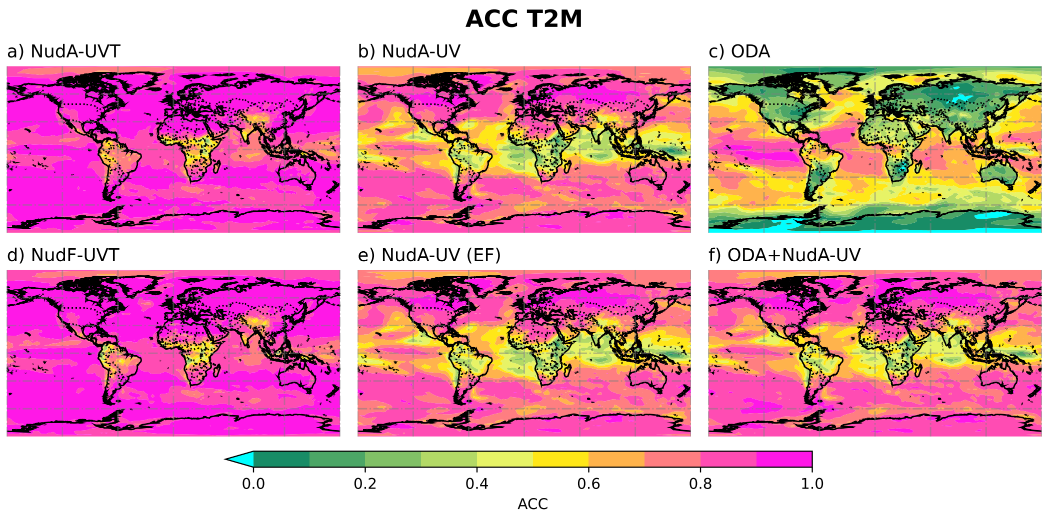

Figure 6. ACC of de-seasoned monthly T2M for a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV reanalyses computed against ERA5 for the period 1980–2010. Green-to-magenta colors indicate positive ACC values, and the cyan color indicates all the negative ACCs

3.2 Predictions

In this section, we evaluate the quality (skill) of the seasonal and decadal hindcasts initialized from the reanalysis (see section 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 (Figures 7a and 7b). On the other hand, the prediction skill for HC500 is low and only beats persistence after the sixth month; while SC500 never outperforms persistence (Figures 7c and 7d). 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 (Figure 7). This highlights the importance of ocean initialization in the prediction skill achieved.

While the globally averaged skill is low (ACCs below 0.4 in T2M and HC500, in Figures 7b and 7c), some regions show enhanced skill (Figures 8 and 9). Skill is most significant over the ocean and most notably in the tropical band driven by the El Niño–Southern Oscillation (ENSO) (Balmaseda & 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 (N. Keenlyside et al., 2020). There is also a region of significant skill in the

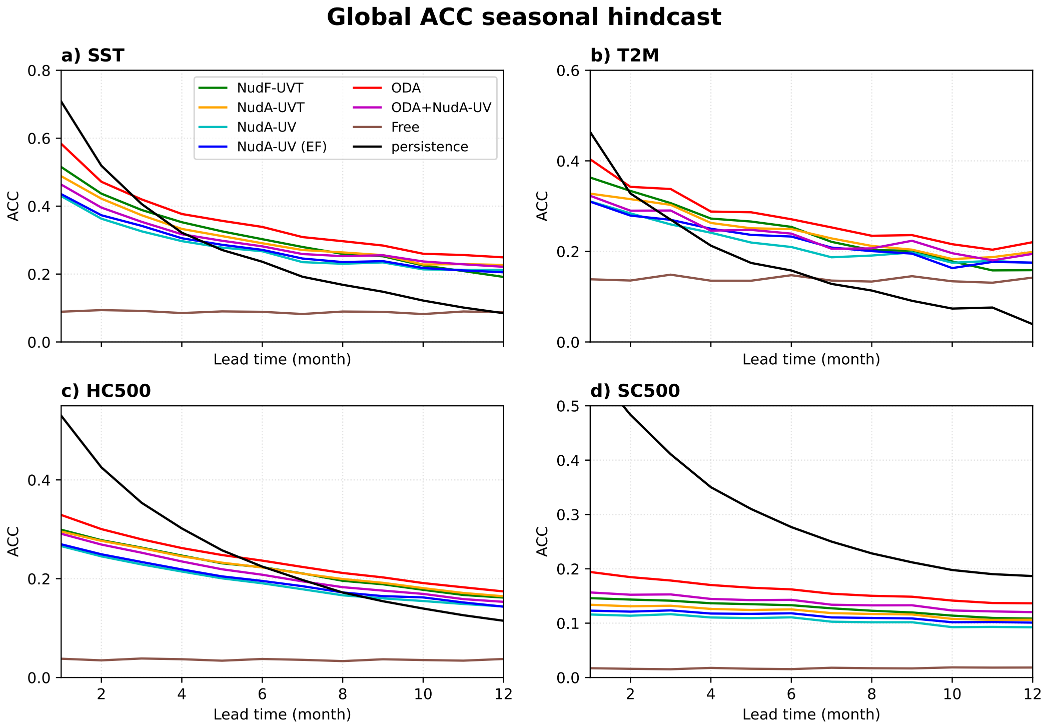

Figure 7. Global average ACC of the seasonal hindcast with lead month, for: a) sea surface temperature (SST), b) 2 m air temperature (T2M), c) 500 m heat content (HC500), and d) 500 m salinity content (SC500). Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV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

northern North Atlantic, the SPG, and the Iceland Sea, in agreement with other climate systems (e.g., Kirtman et al., 2014; Wang et al., 2019).

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 (Figure 10). All prediction systems outperform persistence, with ODA performing best. 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 S2 and S3), 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 (Figure 11). NudF-UVT performs better than all other systems, but it does not beat persistence until month six. Breaking down the analysis by start season (Figures S4b and S5b), we see that NudF-UVT performs best for the hindcast starting in May, slightly beating persistence at lead month 2 (ACC and RMSE_u), i.e., at the peak of the Atlantic Niño. Skillfully predict-

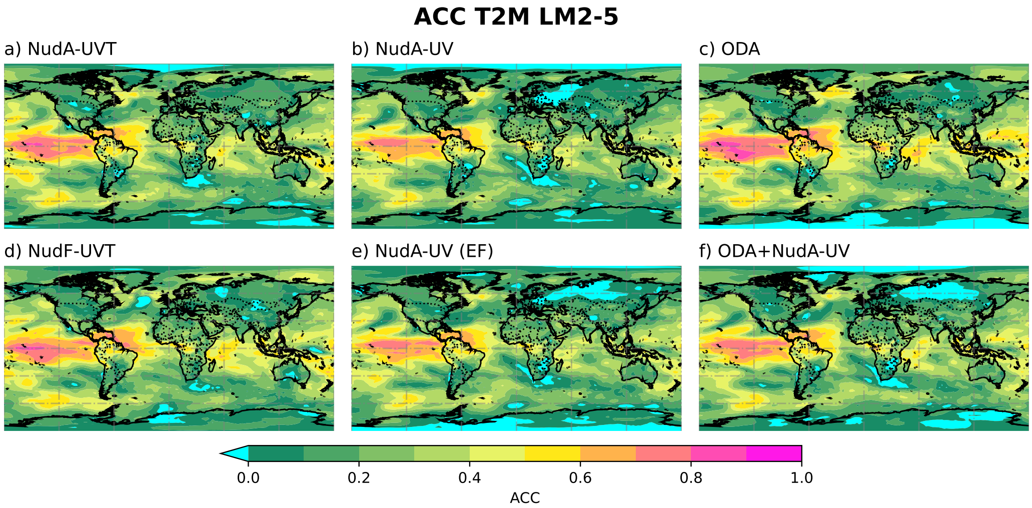

Figure 8. Seasonal hindcast 2-5 lead-month T2M ACC for a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV. Green-to-magenta colors indicate positive ACCs and cyan colour indicates all negative ACCs.

ing 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 S4 and S5), in agreement with those shown in Counillon et al. (2021). Overall, the skill remains poor in predicting Atlantic Niño variability.

Most of our experiments show good skill in predicting T2M and HC500 in the SPG at lead month 2-5. The best skill is achieved with ODA and, of all the nudging schemes, NudA-UVT performs best (Figures 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 Section 3.1) in this region is important for predictive skill at seasonal scales.

3.2.2 Decadal predictions

We assess our decadal predictions skill with ACC and RMSE$_u$ as a function of lead years. Figure 12 shows the global average skill with lead years for HC500, and Figure 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. 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 short period of our decadal hindcast, the ACCs pattern is relatively noisy, and even negative in some regions (cyan-to-blue colors in Figure 13). However, compared to the skill of a non-initialized hind-

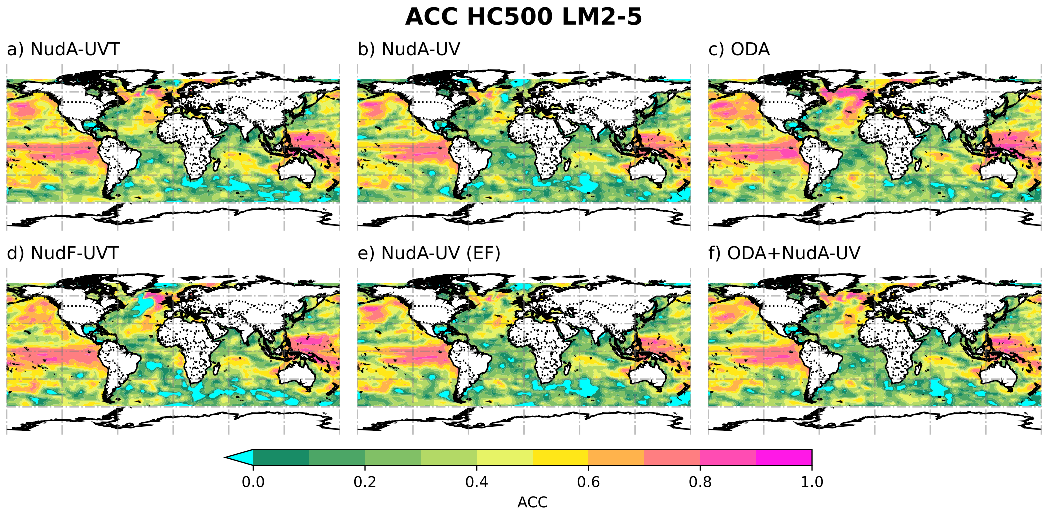

Figure 9. ACC of the seasonal hindcasts at lead-month 2-5 for HC500 with: a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) NudA-UV (EF) and f) ODA+NudA-UV computed against EN4 objective analysis. Green-to-magenta colors indicate positive ACCs and cyan colour indicates all negative ACCs.

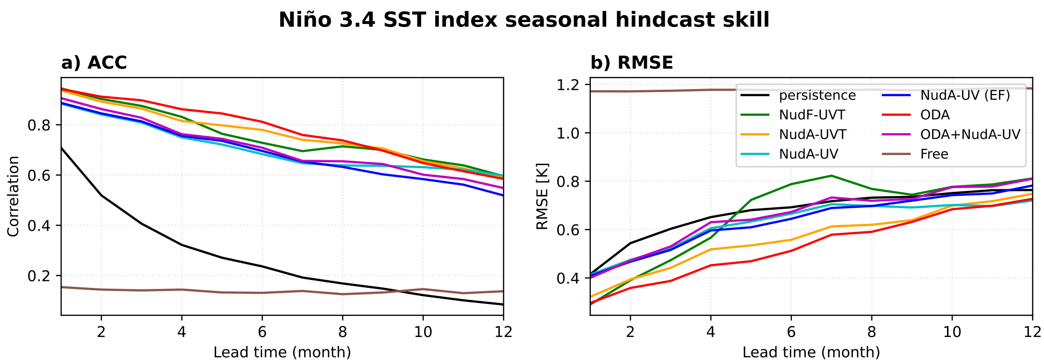

Figure 10. a) ACC of Niño 3.4 SST as a function of the lead month and b) is the same for RMSE$_u$ in K. Line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, magenta to ODA+NudA-UV, brown to Free, and persistence is the solid black line.

cast (Figure 13e), all of our schemes show regions of improved skill. These regions are the North Atlantic, the Western Pacific Ocean, and the Indian Ocean. The regions for which skill is improved when compared to Free agree with the NorCPM experiment for CMIP6 DCPP carried for 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 & Son, 2022). The skill is negative in Free at the western coasts of North and South America 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

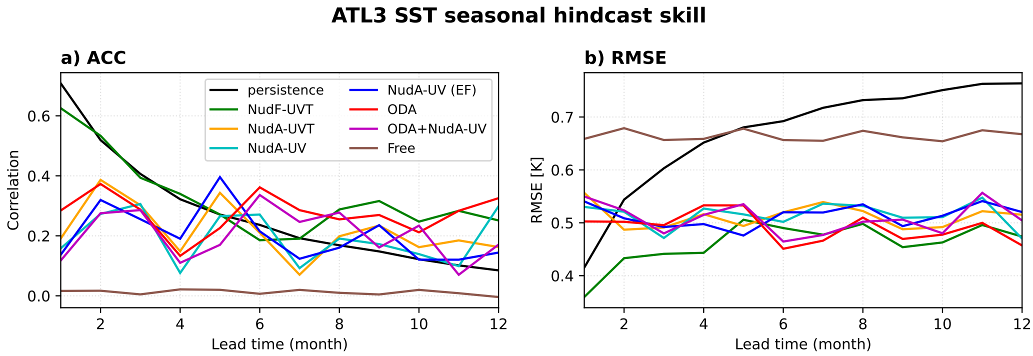

Figure 11. Same as Figure 10 for ATL3 SST.

PDV predictions (e.g., Guemas et al., 2012; Bilbao et al., 2021). Finally, ODA has the largest skill improvement in the SPG region, highlighting the importance of constraining the ocean to initialize decadal variability within the sub-polar North Atlantic.

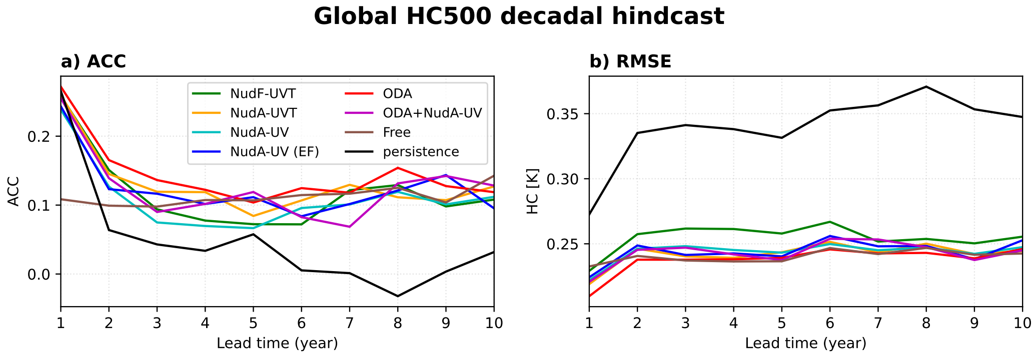

Figure 12. Global a) ACC and b) RMSE_u as a function of lead year for HC500. The line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, and magenta to ODA+NudA-UV, brown to Free, and the black line is persistence.

To further analyze the SPG variability, we evaluate the performance of the SPG index based on HC500 with lead-year (Figure 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) (Figure 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 & 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 Figure 4), showing that the mem-

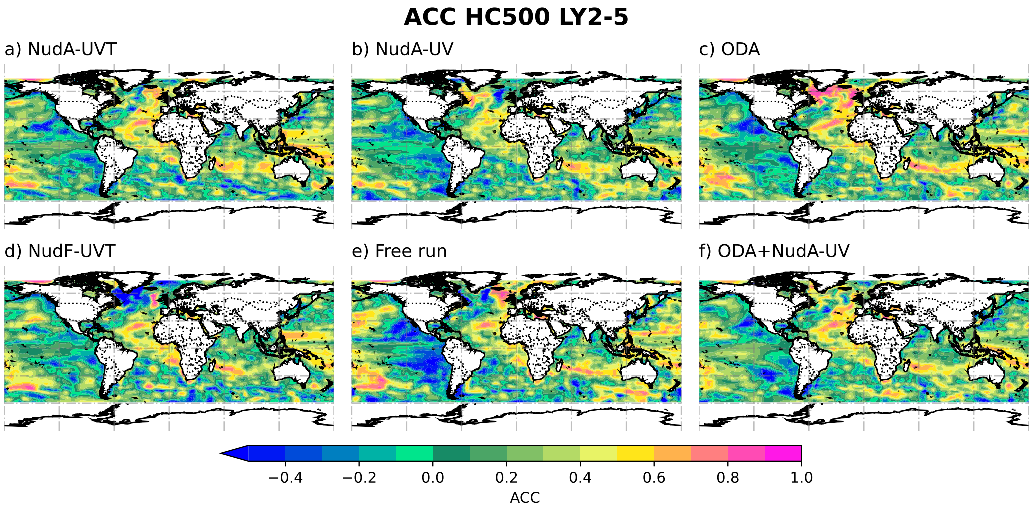

Figure 13. ACC for the decadal hindcast at lead year 2-5 of HC500 a) NudA-UVT, b) NudA-UV, c) ODA, d) NudF-UVT, e) Free and f) ODA+NudA-UV computed against EN4 objective analysis. Green-to-magenta colors indicate positive ACCs, while cyan-to-blue colors indicate negative ACCs.

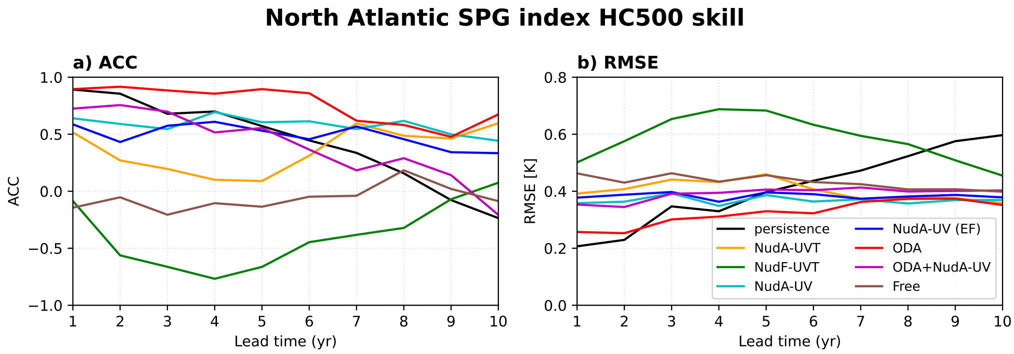

Figure 14. a) ACC and b) RMSE of the SPG index (computed from HC500 versus EN4 objective analysis) as a function of lead year. The line color green corresponds to NudF-UVT, orange to NudA-UVT, cyan to NudA-UV, blue to NudA-UV (EF), red to ODA, and magenta to ODA+NudA-UV, the brown line is Free, and the black line is persistence.

ory 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 26.5°N is shown in Figure 5 and compared to the RAPID observation program (RAPID, Johns et al., 2011) 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 Summary and Conclusions

In this study, we compared the potential of a large set of initialization schemes to constrain climate variability in an ESM 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, and 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. Still, anomaly initialization is performing overall best beyond short lead time.
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 the atmospheric nudging of horizontal winds limits the climatological change during the reanalysis, but very few differences are found during the seasonal hindcasts.
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. 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.

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). 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 & Derome, 1995) and consider using a weaker nudging.

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 & 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; F. J. Schevenhoven & Carrassi, 2021; F. Schevenhoven et al., 2023).

5 Open Research

The reanalysis and seasonal and decadal hindcasts data presented in this article are being organized and archived at `https://ns9039k.web.sigma2.no/lgarcia/initializations/`. The data is organized following the naming convention used in Table 1. Each directory contains the reanalysis and hindcasts monthly ensemble mean for 2 m temperature (T2M), sea surface temperature (SST), and temperature (T) and salinity (S). We also include the AMOC transport at 26.5°N, from annual averages. We provide the data on model grid and using netcdf format. The full simulations will be available on `https://archive.sigma2.no`, with a specific doi upon acceptance of the manuscript.

The code of the Norwegian Earth System Model (NorESM) and the Norwegian Climate Prediction Model (NorCPM version1) are available online on the Norwegian Earth System Modeling hub (`https://github.com/NorESMhub`). Specific details about NorCPM can be found in the website (`https://wiki.app.uib.no/norcpm/index.php/Norwegian_Climate_Prediction_Model`). The temperature and salinity (T, S) vertical profiles from EN4.2.1 objective analysis (Good et al., 2013) can be obtained from the Met Office Hadley Centre observations datasets website (`https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html`). And the sea surface temperature (SST) observations, HADISST2 (Rayner et al., 2003), are available at `https://www.metoffice.gov.uk/hadobs/hadisst2/data/download.html`. The reference data used for 2 m temperature (T2M), from ERA5 (Hersbach et al., 2020), can be obtained the Copernicus web services (`https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form`). The AMOC measurements used are available in the RAPID-AMOC website (`https://rapid.ac.uk`).

Acknowledgments

This study was partly funded by the Trond Mohn Foundation, under project number: BFS2018TMT01, the NFR INES ((INES; 270061), and Climate Futures (309562). This work has also received a grant for computer time from the Norwegian Program for supercomputing (NOTUR2, project number nn9039k) and a storage grant (NORSTORE, NS9039k).

References

Anderson, J. L. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. doi: 10.1175/1520-0493(2001)129⟨2884:AEAKFF⟩2.0.CO;2

Balmaseda, M., Alves, O., Arribas, A., Awaji, T., Behringer, D., Ferry, N., … Stammer, D. (2009, 9). Ocean Initialization for Seasonal Forecasts. *Oceanography*, *22*(3), 154–159. Retrieved from `http://www.jstor.org/stable/24860997`

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

Barthélémy, S., Brajard, J., Bertino, L., & Counillon, F. (2022). Super-resolution data assimilation. *Ocean Dynamics*, *72*(8), 661–678. Retrieved from `https://doi.org/10.1007/s10236-022-01523-x` doi: 10.1007/s10236-022-01523-x

Bellprat, O., Massonnet, F., Siegert, S., Prodhomme, C., Macias-Gómez, D., Guemas, V., & Doblas-Reyes, F. (2017). Uncertainty propagation in observational references to climate model scales. *Remote Sensing of Environment*, *203*, 101–108. doi: 10.1016/J.RSE.2017.06.034

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., ... Kristjánsson, J. E. (2013, 5). The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, *6*(3), 687–720. doi: 10.5194/gmd-6-687-2013

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., ... Eldevik, T. (2021, 11). NorCPM1 and its contribution to CMIP6 DCPP. *Geosci. Model Dev*, *14*(11), 7073–7116. doi: 10.5194/gmd-14-7073-2021

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., ... Vegas-Regidor, J. (2021, 2). Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth System Dynamics*, *12*(1), 173–196. Retrieved from `https://esd.copernicus.org/articles/12/173/2021/` doi: 10.5194/esd-12-173-2021

Bitz, C. M., Shell, K. M., Gent, P. R., Bailey, D. A., Danabasoglu, G., Armour, K. C., ... Kiehl, J. T. (2012, 5). Climate Sensitivity of the Community Climate System Model, Version 4. *Journal of Climate*, *25*(9), 3053–3070. doi: 10.1175/JCLI-D-11-00290.1

Bleck, R., Rooth, C., Hu, D., & Smith, L. T. (1992). Salinity-driven Thermocline Transients in a Wind- and Thermohaline-forced Isopycnic Coordinate Model of the North Atlantic. *Journal of Physical Oceanography*, *22*(12), 1486–1505. doi: 10.1175/1520-0485(1992)022⟨1486:SDTTIA⟩2.0.CO;2

Bleck, R., & Smith, L. T. (1990, 3). A wind-driven isopycnic coordinate model of the north and equatorial Atlantic Ocean: 1. Model development and supporting experiments. *Journal of Geophysical Research: Oceans*, *95*(C3), 3273–3285. doi: 10.1029/JC095IC03P03273

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., ... Eade, R. (2016, 10). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3751–3777. doi: 10.5194/gmd-9-3751-2016

Brune, S., & Baehr, J. (2020, 5). Preserving the coupled atmosphere–ocean feedback in initializations of decadal climate predictions. *Wiley Interdisciplinary Reviews: Climate Change*, *11*(3). doi: 10.1002/WCC.637

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J. (2018). Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts. *Climate Dynamics*, *51*, 1947–1970. doi: 10.1007/s00382-017-3991-4

Carrassi, A., Weber, R. J., Guemas, V., Doblas-Reyes, F. J., Asif, M., & Volpi, D. (2014, 4). Full-field and anomaly initialization using a low-order climate model: A comparison and proposals for advanced formulations. *Nonlinear Processes in Geophysics*, *21*(2), 521–537. doi: 10.5194/npg-21-521-2014

Choi, J., & Son, S. W. (2022, 4). Seasonal-to-decadal prediction of El Niño–Southern Oscillation and Pacific Decadal Oscillation. *npj Climate and Atmospheric Science 2022 5:1*, *5*(1), 1–8. doi: 10.1038/s41612-022-00251-9

Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., & Zheng, F. (2014). Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: A twin experiment. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *66*(1). doi: 10.3402/tellusa.v66.21074

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *68*(1), 32437. doi: 10.3402/tellusa.v68.32437

Counillon, F., Keenlyside, N., Toniazzo, T., Koseki, S., Teferi, D., Bethke, I., & Wang, Y. (2021). Relating model bias and prediction skill in the equatorial Atlantic. *Climate Dynamics*, *56*, 2617–2630. Retrieved from https://doi.org/10.1007/s00382-020-05605-8 doi: 10.1007/s00382-020-05605-8

Counillon, F., Keenlyside, N., Wang, S., Devilliers, M., Gupta, A., Koseki, S., & Shen, M.-L. (2023). Framework for an Ocean-Connected Supermodel of the Earth System. *Journal of Advances in Modeling Earth Systems*, *15*(3), e2022MS003310. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003310 doi: https://doi.org/10.1029/2022MS003310

Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., ... Wang, Q. (2014). North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modelling*, *73*, 76–107. doi: 10.1016/J.OCEMOD.2013.10.005

Dee, D. P. (2006, 1). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3323–3343. doi: 10.1256/qj.05.137

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Rosnay, d. P. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society Q. J. R. Meteorol. Soc*, *137*, 553–597. doi: 10.1002/qj.828

Ding, H., Greatbatch, R. J., Latif, M., & Park, W. (2015, 7). The impact of sea surface temperature bias on equatorial Atlantic interannual variability in partially coupled model experiments. *Geophysical Research Letters*, *42*(13), 5540–5546. doi: 10.1002/2015GL064799

Dippe, T., Greatbatch, R. J., & Ding, H. (2018, 7). On the relationship between Atlantic Niño variability and ocean dynamics. *Climate Dynamics*, *51*(1-2), 597–612. doi: 10.1007/S00382-017-3943-Z/FIGURES/12

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., ... Van Oldenborgh, G. J. (2013, 4). Initialized near-term regional climate change prediction. *Nature Communications 2013 4:1*, *4*(1), 1–9. doi: 10.1038/ncomms2704

Dunstone, N. J., & Smith, D. M. (2010, 1). Impact of atmosphere and sub-surface ocean data on decadal climate prediction. *Geophysical Research Letters*, *37*(2), 2709. doi: 10.1029/2009GL041609

El Gharamti, M., McCreight, J. L., Noh, S. J., Hoar, T. J., Rafieeinasab, A., & Johnson, B. K. (2021, 9). Ensemble streamflow data assimilation using WRF-Hydro and DART: Novel localization and inflation techniques applied to Hurricane Florence flooding. *Hydrology and Earth System Sciences*, *25*(9), 5315–5336. doi: 10.5194/hess-25-5315-2021

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

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014, 8). Why Should Ensemble Spread Match the RMSE of the Ensemble Mean? *Journal of Hydrometeorology*, *15*(4), 1708–1713. doi: 10.1175/JHM-D-14-0008.1

García-Serrano, J., Guemas, V., & Doblas-Reyes, F. J. (2015, 5). Added-value from initialization in predictions of Atlantic multi-decadal variability. *Climate Dynamics*, *44*(9-10), 2539–2555. doi: 10.1007/S00382-014-2370-7/FIGURES/9

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), 6704–6716. doi: 10.1002/2013JC009067

Gouretski, V., & Reseghetti, F. (2010, 6). 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. doi: 10.1016/j.dsr.2010.03.011

Guemas, V., Doblas-Reyes, F. J., Lienert, F., Soufflet, Y., & Du, H. (2012, 10).

Identifying the causes of the poor decadal climate prediction skill over the North Pacific. *Journal of Geophysical Research: Atmospheres*, *117*(D20). Retrieved from `http://doi.wiley.com/10.1029/2012JD018004` doi: 10.1029/2012JD018004

Häkkinen, S., & Rhines, P. B. (2004, 4). Decline of Subpolar North Atlantic Circulation during the 1990s. *Science*, *304*(5670), 555–559. doi: 10.1126/SCIENCE.1094917/SUPPL/FILE/HAKKINEN.SOM.PDF

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

Hawkins, E., & Sutton, R. (2009, 8). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. doi: 10.1175/2009BAMS2607.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. N. (2020, 7). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/QJ.3803

Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M., ... Roberts, M. J. (2017, 12). Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales? *Ocean Modelling*, *120*, 120–136. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S1463500317301774` doi: 10.1016/j.ocemod.2017.11.002

Hoke, J. E., & Anthes, R. A. (1976, 12). The Initialization of Numerical Models by a Dynamic-Initialization Technique. *Monthly Weather Review*, *104*(12), 1551–1556. doi: https://doi.org/10.1175/1520-0493(1976)104⟨1551:TIONMB⟩2.0.CO;2

Houtekamer, P. L., & Derome, J. (1995, 7). Methods for Ensemble Prediction. *Monthly Weather Review*, *123*(7), 2181–2196. doi: 10.1175/1520-0493(1995)123⟨2181:mfep⟩2.0.co;2

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., ... Marshall, S. (2013, 9). The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. doi: 10.1175/BAMS-D-12-00121.1

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., ... Weston, P. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3130` doi: https://doi.org/10.1002/qj.3130

Johns, W. E., Baringer, M. O., Beal, L. M., Cunningham, S. A., Kanzow, T., Bryden, H. L., ... Curry, R. (2011). Continuous, Array-Based Estimates of Atlantic Ocean Heat Transport at 26.5°N. *Journal of Climate*, *24*(10), 2429–2449. doi: 10.1175/2010JCLI3997.1

Karspeck, A. R., Danabasoglu, G., Anderson, J., Karol, S., Collins, N., Vertenstein, M., ... Craig, A. (2018). A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed. *Quarterly Journal of the Royal Meteorological Society*, *144*(717), 2404–2430. doi: 10.1002/qj.3308

Keenlyside, N., Kosaka, Y., Vigaud, N., Robertson, A. W., Wang, Y., Dommenget, D., ... Matei, D. (2020). Basin interactions and predictability. In C. R. Mechoso (Ed.), *Interacting climates of ocean basins: Observations, mechanisms, predictability, and impacts* (p. 258–292). Cambridge University Press. doi: 10.1017/9781108610995.009

Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L., & Roeckner, E. (2008, 5). Advancing decadal-scale climate prediction in the North Atlantic sector. *Na-*

ture, *453*(7191), 84–88. doi: 10.1038/nature06921

Kirkevåg, A., Iversen, T., Seland, Ø., Hoose, C., Kristjánsson, J. E., Struthers, H., ... Schulz, M. (2012). Aerosol-climate interactions in the Norwegian Earth System Model – NorESM. *Geosci. Model Dev. Discuss.*, *5*, 2599–2685. doi: 10.5194/gmdd-5-2843-2012

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., ... Wood, E. F. (2014, 4). The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. Retrieved from `https://journals.ametsoc.org/view/journals/bams/95/4/bams-d-12-00050.1.xml` doi: 10.1175/BAMS-D-12-00050.1

Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C. J., & Russell, L. M. (2012, 12). Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the Community Atmosphere Model 5. *Journal of Geophysical Research: Atmospheres*, *117*(D23), 23204. doi: 10.1029/2012JD018588

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., ... Slater, A. G. (2011, 1). Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *3*(1), n/a-n/a. doi: 10.1029/2011MS00045

Lohmann, K., Drange, H., Bentsen, M., Helge, A. E., Ae, D., & Bentsen, M. (2009). Response of the North Atlantic subpolar gyre to persistent North Atlantic oscillation like forcing. *Climate Dynamics*, *32*(2), 273–285. doi: 10.1007/s00382-008-0467-6

Lu, F., Harrison, M. J., Rosati, A., Delworth, T. L., Yang, X., Cooke, W. F., ... Adcroft, A. (2020). GFDL's SPEAR Seasonal Prediction System: Initialization and Ocean Tendency Adjustment (OTA) for Coupled Model Predictions. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2020MS002149. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002149` doi: https://doi.org/10.1029/2020MS002149

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013, 11). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, *41*(9-10), 2393–2409. doi: 10.1007/s00382-012-1599-2

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., ... Albers, J. (2020, 5). Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond. *Bulletin of the American Meteorological Society*, *101*(5), E608-E625. doi: 10.1175/BAMS-D-18-0326.1

Mariotti, A., Ruti, P. M., & Rixen, M. (2018, 3). Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate and Atmospheric Science 2018 1:1*, *1*(1), 1–4. doi: 10.1038/s41612-018-0014-z

Massonnet, F., Bellprat, O., Guemas, V., & Doblas-Reyes, F. J. (2016). Using climate models to estimate the quality of global observational data sets. *Science*, *354*(6311), 452–455. doi: 10.1126/science.aaf6369

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., ... Stockdale, T. (2009, 10). Decadal prediction: Can it be skillful? *Bulletin of the American Meteorological Society*, *90*(10), 1467–1485. doi: 10.1175/2009BAMS2778.1

Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., ... Xie, S. P. (2021, 4). Initialized Earth System prediction from subseasonal to decadal timescales. *Nature Reviews Earth & Environment 2021 2:5*, *2*(5), 340–357. doi: 10.1038/s43017-021-00155-x

Mochizuki, T., Ishii, M., Kimoto, M., Chikamoto, Y., Watanabe, M., Nozawa, T., ... Mori, M. (2010). Pacific decadal oscillation hindcasts relevant to near-term climate prediction. *Proceedings of the National Academy of Sciences*, *107*(5), 1833-1837. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.0906531107` doi: 10.1073/pnas.0906531107

Neale, R., Richter, J., Conley, A., Park, S., Lauritzen, P., Gettelman, A., ... Lin, S.-J. (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. *Proceedings of the National Academy of Sciences*, *116*(49), 24390–24395. doi: 10.1073/pnas.1906691116

Pohlmann, H., Jungclaus, J. H., Köhl, A., Stammer, D., & Marotzke, J. (2009, 7). Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic. *Journal of Climate*, *22*(14), 3926–3938. doi: 10.1175/2009JCLI2535.1

Polkova, I., Brune, S., Kadow, C., Romanova, V., Gollan, G., Baehr, J., ... Stammer, D. (2019, 1). Initialization and Ensemble Generation for Decadal Climate Predictions: A Comparison of Different Methods. *Journal of Advances in Modeling Earth Systems*, *11*(1), 149–172. doi: 10.1029/2018MS001439

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., ... Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(14). doi: 10.1029/2002JD002670

Robson, J. (2010). *Understanding the performance of a decadal prediction system* (Doctoral dissertation). doi: 10.13140/RG.2.1.2183.2560

Robson, J. I., Sutton, R. T., & Smith, D. M. (2012, 10). Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. *Geophysical Research Letters*, *39*(19). doi: 10.1029/2012GL053370

Rodwell, M. J., Lang, S. T. K., Ingleby, N. B., Bormann, N., Hólm, E., Rabier, F., ... Yamaguchi, M. (2016). Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 443–454. doi: 10.1002/qj.2663

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999, 9). A dipole mode in the tropical Indian Ocean. *Nature 1999 401:6751*, *401*(6751), 360–363. doi: 10.1038/43854

Sakov, P., Counillon, F., Bertino, L., Lister, K. A., Oke, P. R., & Korablev, A. (2012). TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, *8*(4), 633–656. doi: 10.5194/os-8-633-2012

Sakov, P., & Oke, P. R. (2008, 3). A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *60 A*(2), 361–371. doi: 10.1111/j.1600-0870.2007.00299.x

Schevenhoven, F., Keenlyside, N., Counillon, F., Carrassi, A., Chapman, W. E., Devilliers, M., ... Duane, G. S. (2023). Supermodeling: improving predictions with an ensemble of interacting models. *Bulletin of the American Meteorological Society*. Retrieved from `https://journals.ametsoc.org/view/journals/bams/aop/BAMS-D-22-0070.1/BAMS-D-22-0070.1.xml` doi: https://doi.org/10.1175/BAMS-D-22-0070.1

Schevenhoven, F. J., & Carrassi, A. (2021). Training a supermodel with noisy and sparse observations: a case study with cpt and the synch rule on speedo-v. 1. *Geoscientific Model Development Discussions*, *2021*, 1–23.

Singh, T., Counillon, F., Tjiputra, J., Wang, Y., & Gharamti, M. E. (2022). Estimation of Ocean Biogeochemical Parameters in an Earth System Model Using the Dual One Step Ahead Smoother: A Twin Experiment. *Frontiers in Marine Science*, *9*. doi: 10.3389/fmars.2022.775394

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007, 8). Improved surface temperature prediction for the coming decade from a global climate model. *Science*, *317*(5839), 796–799. doi: 10.1126/science.1139540

Smith, D. M., Eade, R., & Pohlmann, H. (2013, 12). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, *41*(11-12), 3325–3338. doi: 10.1007/s00382-013-1683-2

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012, 4). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. doi: 10.1175/BAMS-D-11-00094.1

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., … Hibbard, K. (2011). The representative concentration pathways: an overview. *Climatic Change*, *109*, 5–31. doi: 10.1007/s10584-011-0148-z

Volpi, D., Guemas, V., & Doblas-Reyes, F. J. (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. *Climate Dynamics*, *49*(4), 1181–1195. doi: 10.1007/s00382-016-3373-3

Wang, Y., Counillon, F., Bertino, L., & Wang, . Y. (2016). Alleviating the bias induced by the linear analysis update with an isopycnal ocean model. *Quarterly Journal of the Royal Meteorological Society Q. J. R. Meteorol. Soc*, *142*, 1064–1074. doi: 10.1002/qj.2709

Wang, Y., Counillon, F., Bethke, I., Keenlyside, N., Bocquet, M., & Shen, M. l. (2017, 6). Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. *Ocean Modelling*, *114*, 33–44. doi: 10.1016/j.ocemod.2017.04.007

Wang, Y., Counillon, F., Keenlyside, N., Svendsen, L., Gleixner, S., Kimmritz, M., … Yongqi Gao (2019). Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF. *Climate Dynamics*, *53*, 5777–5797. doi: 10.1007/s00382-019-04897-9

Weber, R. J., Carrassi, A., & Doblas-Reyes, F. J. (2015, 11). Linking the Anomaly Initialization Approach to the Mapping Paradigm: A Proof-of-Concept Study. *Monthly Weather Review*, *143*(11), 4695–4713. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/143/11/mwr-d-14-00398.1.xml` doi: 10.1175/MWR-D-14-00398.1

Webster, P. J., Moore, A. M., Loschnigg, J. P., & Leben, R. R. (1999, 9). Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. *Nature 1999 401:6751*, *401*(6751), 356–360. doi: 10.1038/43848

Wilks, Daniel. (2019). *Statistical Methods in the Atmospheric Sciences - 4th Edition.* Retrieved from `https://www.elsevier.com/books/statistical-methods-in-the-atmospheric-sciences/wilks/978-0-12-815823-4`

Yeager, S. G., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012). A decadal prediction case study: Late twentieth-century North Atlantic Ocean heat content. *Journal of Climate*, *25*(15), 5173–5189. doi: 10.1175/JCLI-D-11-00595.1

Yeager, S. G., & Robson, J. I. (2017, 6). Recent Progress in Understanding and Predicting Atlantic Decadal Climate Variability. *Current Climate Change Reports*, *3*(2), 112–127. doi: 10.1007/s40641-017-0064-z

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., … Lohmann, U. (2014, 8). Technical note: On the use of nudging for aerosol-climate model intercomparison studies. *Atmospheric Chemistry and Physics*, *14*(16), 8631–8645. doi: 10.5194/ACP-14-8631-2014

Zhang, S., Rosati, A., & Delworth, T. (2010, 10). The adequacy of observing systems in monitoring the Atlantic meridional overturning circulation and North Atlantic climate. *Journal of Climate*, *23*(19), 5311–5324. doi: 10.1175/2010JCLI3677.1

Zhang, S., Rosati, A., & Harrison, M. J. (2009). Detection of multidecadal oceanic variability by ocean data assimilation in the context of a "perfect" coupled model. *Journal of Geophysical Research: Oceans*, *114*(12), 12018. doi: 10.1029/2008JC005261

Supporting Information for Intercomparison of initialization methods for Seasonal-to-Decadal Climate Predictions with the NorCPM

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

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

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

Contents of this file

Introduction We have included figures and extra information in tables to provide additional evidence for the topics discussed in the main text.

Corresponding author: L. Garcia-Oliva, Geophysical Institute, University of Bergen, Bjerknes Centre for Climate Research, 5007 Bergen, Norway. (lilian.garcia@uib.no)

Table S1. Reanalysis reliability[a].

Configuration	SST	T2M	HC500	SC500
Free	1.09	1.03	1.67	1.21
NudF-UVT	2.11	15.25	6.90	4.88
NudA-UVT	2.14	14.50	6.39	3.76
NudA-UV	2.37	7.65	5.59	3.77
NudA-UV (EF)	2.35	7.55	5.58	3.78
ODA	1.10	1.07	2.49	1.87
ODA+NudA-UV	2.39	13.22	11.03	8.25

[a] $RMSE_u/\sigma$, see equation (14) in main text.

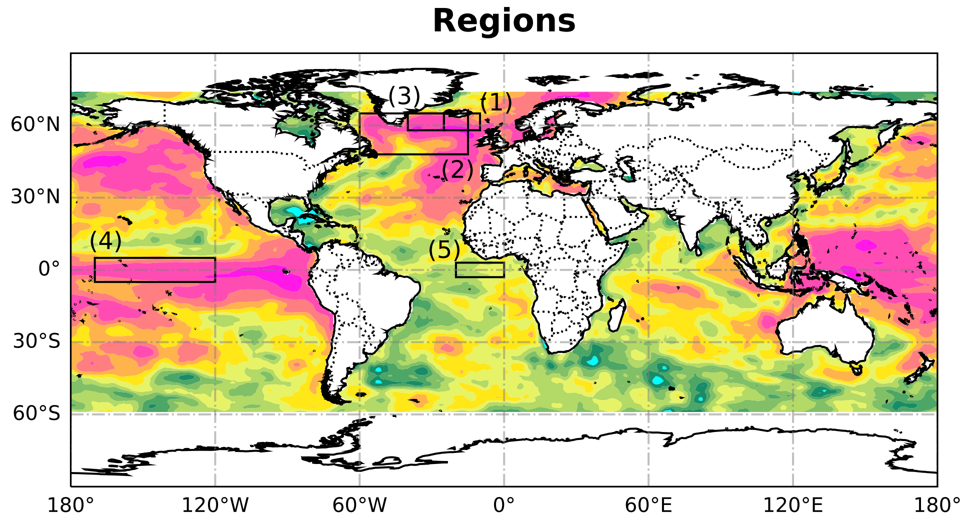

Figure S1. Regions studied. Numbered regions are (1) Iceland Sea (ICS), (2) North Atlantic Subpolar Gyre (SPG), (3) Irminger Sea (IRS), (4) El Niño 3.4 (ENSO), and (5) Atlantic 3 (ATL3). In colors: HC500 reanalysis of ODA experiment.

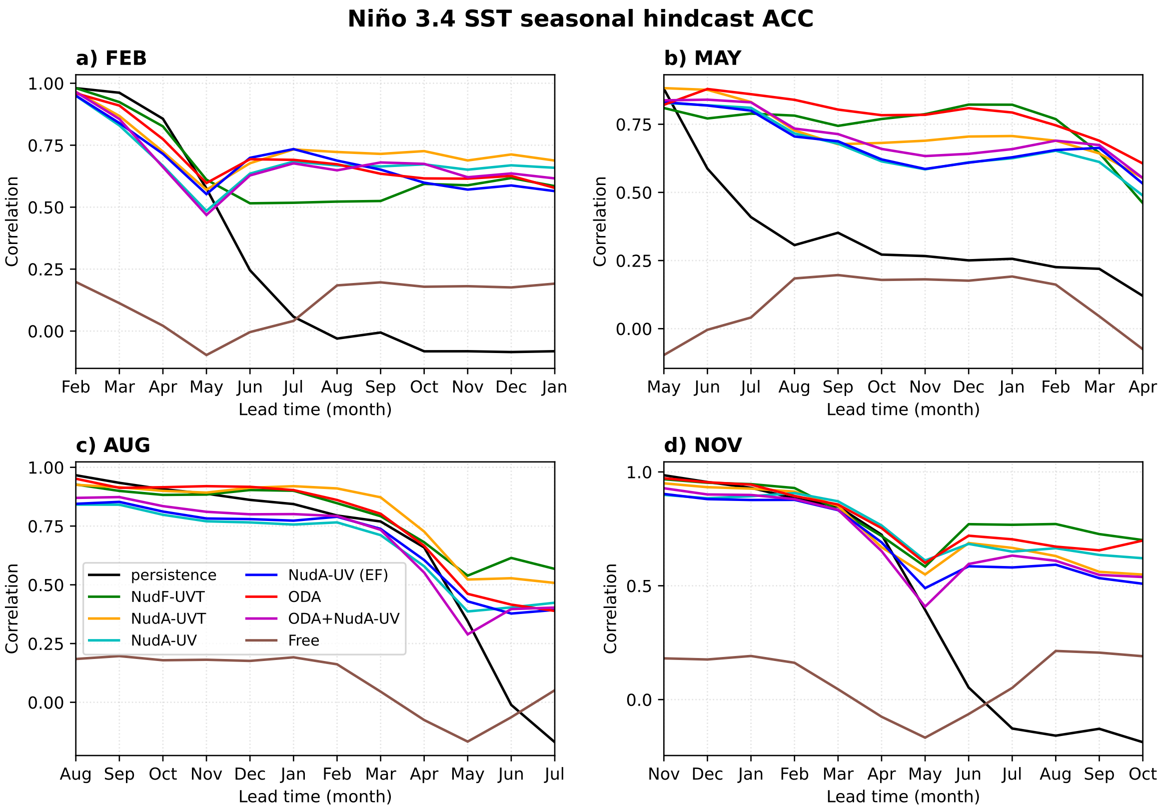

Figure S2. ACC of Niño 3.4 SST with lead-month computed against HadISST2 decomposed by starting month of the hindcasts.

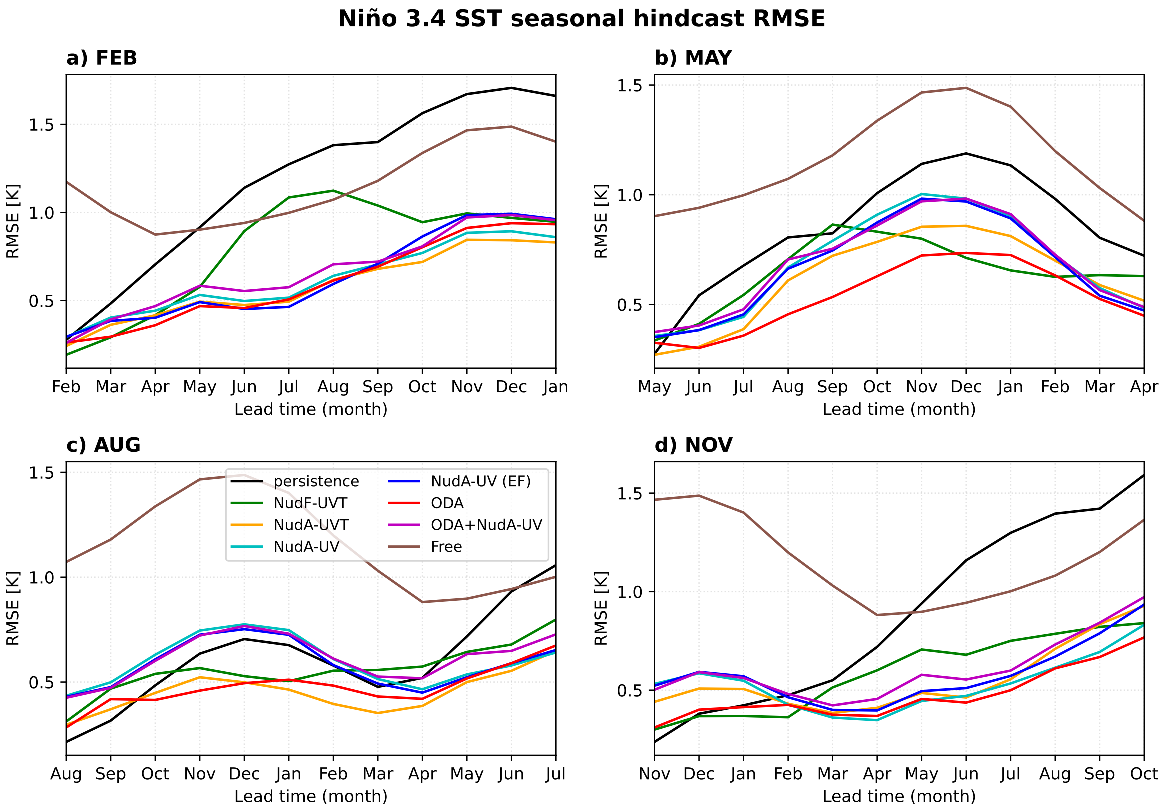

Figure S3. RMSE of Niño 3.4 SST with lead-month computed against HadISST2 and decomposed by starting month of the hindcasts.

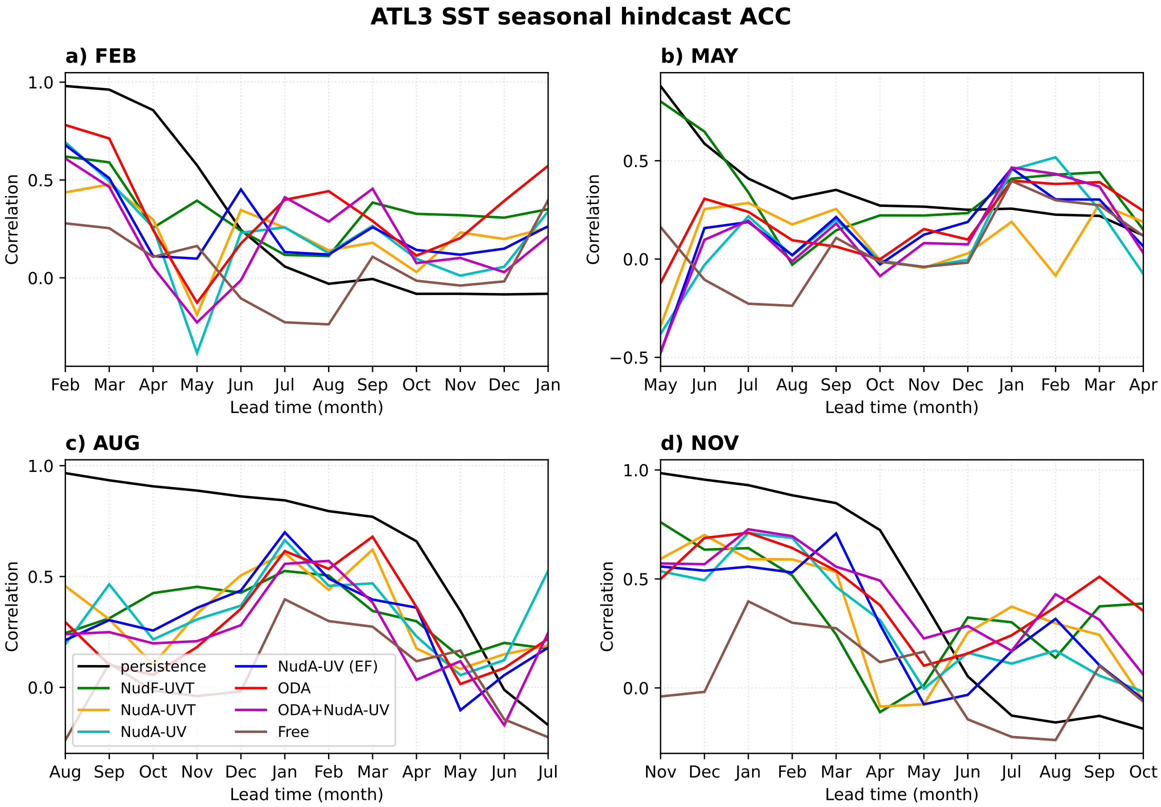

Figure S4. Same as Fig. S2, but for ATL3 SST.

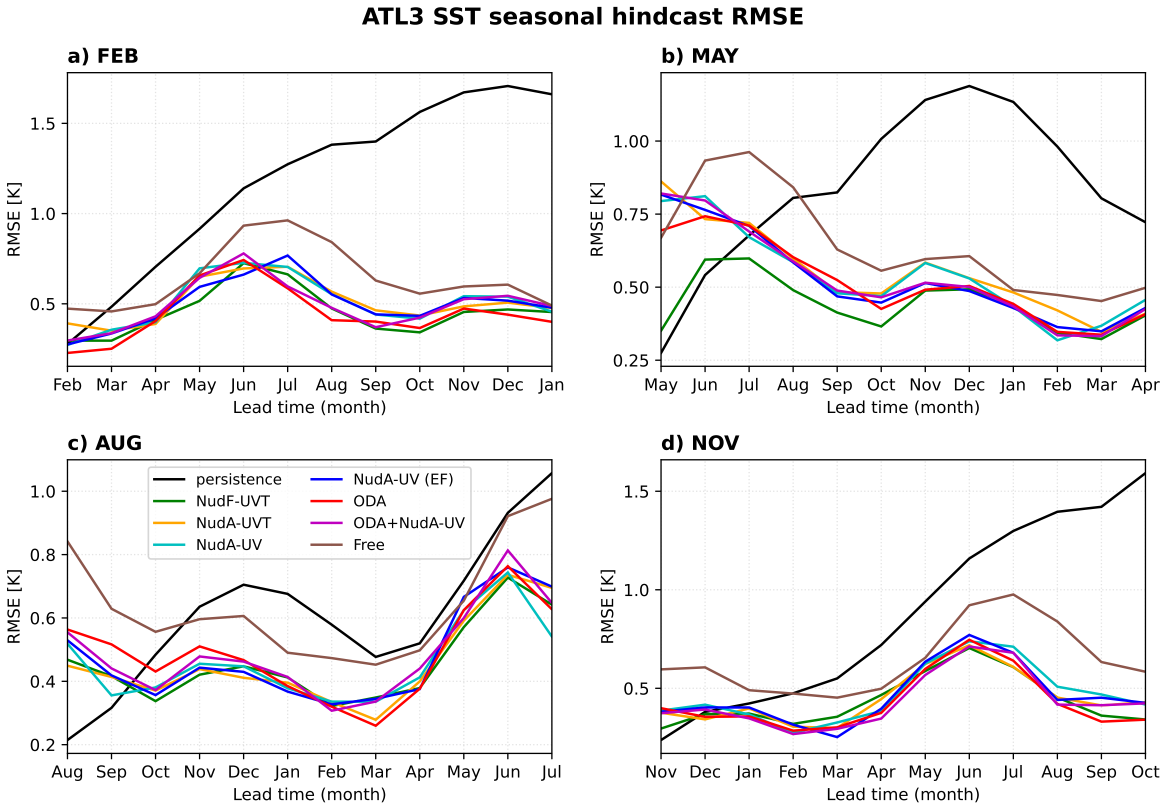

Figure S5. Same as Fig. S3, but for ATL3 SST.

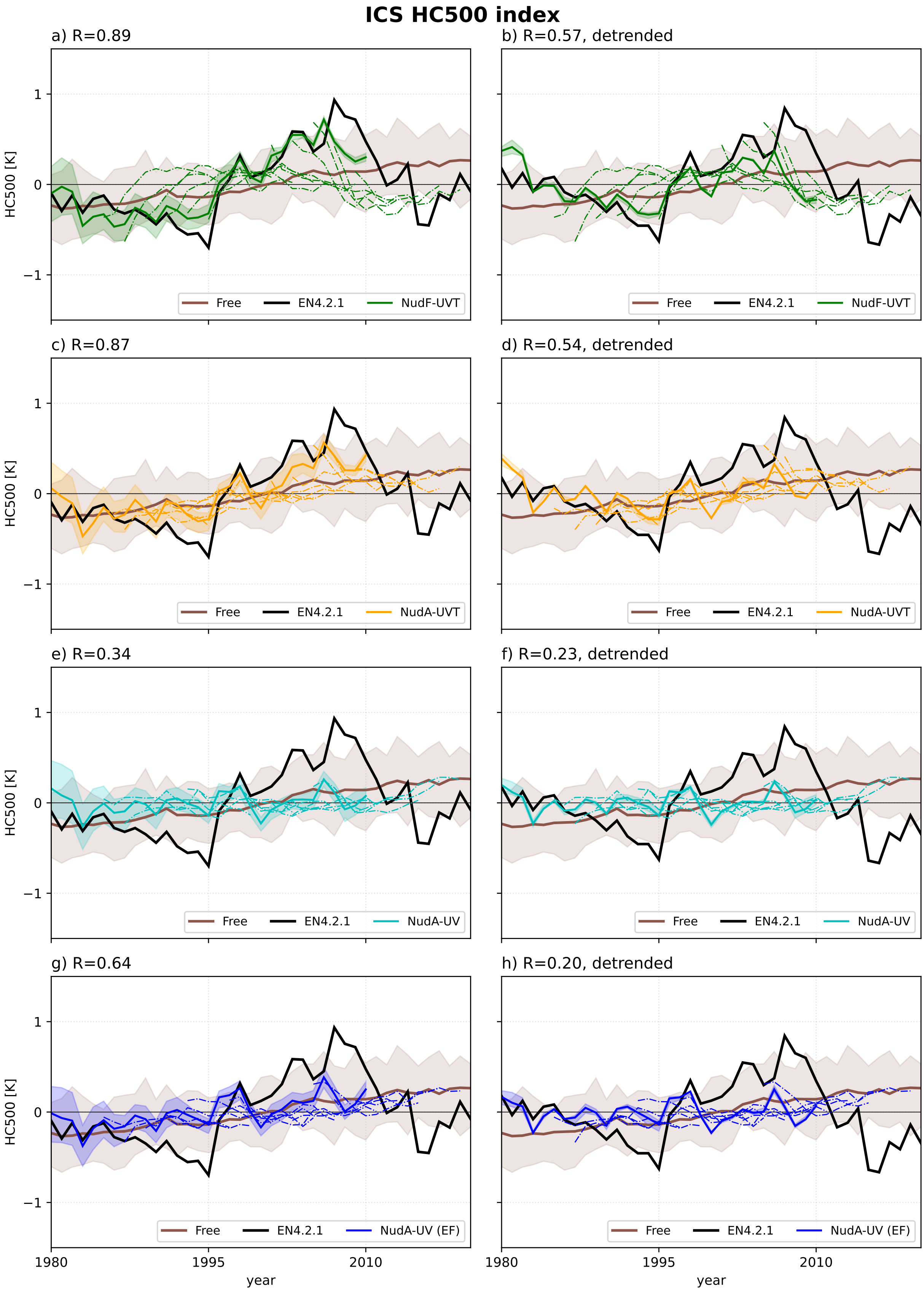

Figure S6. HC500 in the Iceland Sea basin box for NudF-UVT (a,b), NudA-UVT (c,d), NudA-UV (e,f) and NudA-UV EF (g,h). Panels on the left are with the trend and panels on the right are detrended. The correlation coefficient with EN4 objective analysis estimate is included.