Climate-mode initialization for decadal climate predictions

Iuliia Polkova[1] · Armin Köhl[1] · Detlef Stammer[1]

Abstract

In the context of decadal climate predictions, a climate-mode initialization method is being tested by which ocean ORAS4 reanalysis is projected onto dominant modes of variability of the Earth System Model from the Max Planck Institute for Meteorology (MPI-ESM). The method aims to improve the prediction skill of the model by filtering out dynamically unbalanced noise during the initialization step. Used climate modes are calculated as statistical 3-D modes based on the bivariate empirical orthogonal function (EOF) analysis applied to temperature and salinity anomalies from an ensemble of historical simulations from the MPI-ESM. The climate-mode initialization method shows improved surface temperature skill, particularly over the tropical Pacific Ocean at seasonal-to-interannual timescales associated with the improved zonal momentum balance. There, the new initialization somewhat outperforms the surface temperature skill of the anomaly initialization also for lead years 2–5. In other parts of the world ocean, both initialization methods currently are equivalent in skill. However, only 44% of variance in the original ORAS4 reconstruction remains after the projection on model modes, suggesting that the ORAS4 modes are not fully compatible with the model modes. Moreover, we cannot dismiss the possibility that model modes are not sufficiently sampled with the data set underlying the EOF analysis. The full potential of the climate-mode initialization method for future decadal prediction systems therefore still needs to be quantified based on improved modal representation.

Keywords Decadal prediction · Initialization · Climate mode

1 Introduction

How to optimally use available climate observations to generate initial conditions for climate predictions and how to best insert them into a coupled climate model with minimum loss of prediction skill remain active areas of research with many open questions to be answered (Balmaseda and Anderson 2009; Meehl et al. 2009, 2014; Kirtman et al. 2013; Boer et al. 2016; Penny et al. 2017). Among them are questions concerned with prediction errors, ensemble generation, predictability and processes involved in generating prediction skill.

With respect to prediction errors, they encompass: (i) model biases—the errors that are attributed to uncertainties

✉ Iuliia Polkova
  iuliia.polkova@uni-hamburg.de

[1] Institut für Meereskunde, Universität Hamburg, CEN, Bundesstrasse 53, 20146 Hamburg, Germany

in representation of physical processes in a model, to simplifications and assumptions made for numerical modeling and resolution limitations, and (ii) drifts or initial shocks—the errors that occur when predictions are started from unbalanced initial conditions imposed during assimilation or initialization procedure (Malanotte-Rizzoli et al. 1989). Having experienced initial shock, the model begins adjusting to dynamical imbalances in the region by spurious transports and exchange of spurious fluxes with the adjacent model components until the dynamical balance is reached. The associated errors can thus lead to a rapid degradation of prediction skill. Mulholland et al. (2015) distinguishes three major sources of initial shock in a coupled model imposed by: (i) imperfect initialization procedure from off-line assimilation efforts in which there is a mismatch in communicated fluxes between the ocean and the atmosphere, (ii) using different model set-ups for generating initial states and predictions, and (iii) switching off bias correction at the start of predictions resulting in a change in the dynamics of the model components. The latter is similar to using full-field initialization approach, when in the assimilation the model is

brought as close as possible to the observed attractor eliminating model biases and then, in the prediction mode, is let to run freely without bias control.

Some decadal prediction studies report that a large-scale decrease of skill and dynamical adjustments that last several lead years can happen due to initial shock (Pohlmann et al. 2017; Kröger et al. 2018). These few examples show that the degree of the associated decadal prediction skill reduction could be considerable. However, to what extent different sources of initial shock (also as compared to model biases) can effect decadal prediction skill remains rather unknown.

Based on the long-standing experience gained in short-term climate predictions (Rosati et al. 1997; Sugiura et al. 2008; Balmaseda and Anderson 2009), it can be expected that initial conditions respecting the model's dynamics should lead to the best prediction skill also in decadal predictions (Counillon et al. 2014; Liu et al. 2017; Mochizuki et al. 2016; Polkova et al. 2019). From a theoretical point of view, it appears obvious that only a dynamically consistent assimilation approach for generating initial conditions applied to the same coupled model that is being used to perform the predictions can lead to a best prediction skill through a reduced initial model adjustment shock. This would involve using the same assimilation approach to remove model biases by adjusting uncertain model parameters and using this improved system during prediction effort. Although first emerging steps in this direction appear promising (Stammer et al. 2016; Penny et al. 2017), no functioning system exists at the time of writing and might not do so for some time to come.

Due to the absence of such a sophisticated assimilation and initialization approach, previous studies tested various practical initialization methods (i.e., full-field initialization, anomaly initialization and flux correction) for decadal predictions aiming to find a method which can best handle prediction errors and yield high prediction skill (Magnusson et al. 2013; Hazeleger et al. 2013; Smith et al. 2013; Polkova et al. 2014; Volpi et al. 2016). A refinement of the anomaly initialization for decadal predictions was proposed recently (Volpi et al. 2017), where the initial states were weighted with the ratio between the modeled and the observed variability to avoid initialization which goes beyond of the range of the model variability. Still, all these initialization methods remain suboptimal when dealing with non-stationary errors including initial shocks (Goddard et al. 2013; Magnusson et al. 2013).

The idea that the initial state contains predictable and non-predictable components, and that filtering out non-predictable ones can yield more long-lasting skill, was previously tested in the context of numerical weather predictions. Along these lines, different filtering approaches were applied to initial conditions to improve forecasts by minimizing noise from the internal-gravity waves (Williamson 1976; Ballish 1981) and remove random components from initial conditions, which limit predictability, and retain those that are potentially more predictable (Branstator et al. 1993). In the context of seasonal El-Niño Southern Oscillation (ENSO) predictions, an idea of initialization of coupled climate modes of variability was tested, where observed coupled modes of variability were remapped onto modeled ones (Hurrell et al. 2009). This remapping procedure also filters out components from coupled initial states that do not match model variability. Such approaches of selectively initializing model variability have not been fully explored for decadal predictions.

In the current study, we build on these ideas of filtered initialization and design the initialization method, which aims to bring ocean initial conditions in consistency with the dynamics of the decadal prediction system by initializing the climate modes of variability of the prediction system and filtering out components from the initial conditions that cannot be predicted by the prediction system. For this, we project the ocean reanalysis onto model variability modes and subsequently test the sensitivity of the prediction skill to the filtered versus non-filtered initial conditions. Climate modes of variability are presented in terms of the statistical modes from the empirical orthogonal function (EOF) analysis applied to an ensemble of the twentieth century simulations. The EOF modes do not necessarily correspond to physical modes of variability. The ocean reanalysis anomalies are then projected onto a truncated set of the EOF-modes. In this mapping step, the reanalysis variability that is not compatible with the climate modes from the model is filtered out, retaining the ocean states that serve as initial conditions for ensembles of decadal predictions. A possible problem that we envision from such initialization is the removal of part of the observed variability by the projection step, which can at initialization time reduce the skill of predictions, in contrast to using the complete information about initial conditions. The expectation is that this part of the skill is quickly lost anyway such that the skill remains more persistent when initialized by climate modes. The climate-mode initialization method addresses initial shocks that arise from using different model set-ups for generating initial states and predictions. The method is implemented with anomaly initialization omitting the drift which is present when the model is initialized from the full-field state. Since we only project the ocean state in this initialization approach, some imbalance between the ocean and the atmosphere is still possible. Ideally, we would aim for deriving coupled climate modes and projecting the ocean and the atmosphere states on them.

The remaining paper describes, in Sect. 2, the prediction system (based on the Max Planck Institute for Meteorology Earth System Model—MPI-ESM), climate-mode initialization method and experiments. This section also describes details of the calculation of the EOF-modes and compares variability of the model and the ORAS4 ocean reanalysis (Balmaseda et al. 2013) that is used as a source of initial conditions in this study. Section 3 deals with prediction skill resulting from climate-mode initialization and compares it

with a reference approach based on anomaly initialization. A discussion and concluding remarks are given in Sect. 4.

2 Methodology of the climate-mode initialization

The implementation of the climate-mode initialization requires the following elements and steps:

1. An ensemble of the twentieth-century simulations (hereafter historical simulations) that are carried out with the MPI-ESM.
2. Derivation of climate modes through a bivariate EOF analysis of the historical simulations.
3. Establishment of filtered initial conditions by projecting the ORAS4 ocean reanalysis onto the truncated set of EOF-modes.
4. Nudging run and ensembles of decadal predictions (hereafter initialized hindcasts) started from the climate-mode initialization and carried out with the MPI-ESM.

2.1 Model and input for the EOF analysis

During this study all model simulations were performed with the MPI-ESM1.2. The model consists of the atmospheric component ECHAM6.3 with a resolution of T63L47 and the oceanic component MPIOM1.6.3 with 1.5° horizontal resolution and 40 vertical levels (Jungclaus et al. 2013). The resolution used in this study corresponds to a low-resolution (LR) configuration. For an analysis of the MPI-ESM performance with respect to its resolution see (Müller et al. 2018). The MPI-ESM in similar configuration was recently used for decadal predictions (Marotzke et al. 2016; Kröger et al. 2018; Polkova et al. 2019).

The historical simulations that are used as input for the EOF analysis were forced using the CMIP5 solar irradiance data, aerosol and greenhouse gas concentrations (Taylor et al. 2012). The CMIP5 historical simulations cover the period 1850–2005. For the current analysis, we use only the

time slice 1958–2005 and 15 members of historical-simulations ensemble (see Table 1).

2.2 Derivation of climate modes

Applying the bivariate EOF analysis (as used by e.g., Nardelli and Santoleri 2005; Hawkins and Sutton 2007; Bretherton et al. 1999), modes of climate variability are derived from the bivariate model state vector, X, composed out of monthly potential temperature (θ) and salinity (S) anomalies from each depth level at every latitude and longitude of the model grid. The anomalies are sampled from a 15-member ensemble of historical simulations. Because the resulting EOFs should capture interannual variability, we consider only October monthly mean anomalies for: (i) the EOF analysis, (ii) the reconstruction of the ORAS4 anomalies and (iii) the nudging of these fields into the MPI-ESM. October is selected because it by one month precedes the initialization of the preoperational MiKlip decadal prediction experiments used as a reference for comparison here (Marotzke et al. 2016; Polkova et al. 2019). For both parameters, θ and S, monthly October anomalies are calculated with respect to the period 1958–2005, overlapping with historical simulations and the ORAS4 ocean reanalysis. The total number of temperature and salinity data points entering the state vector are $m \times n$, where m represents the sum of all vertical and horizontal grid points and n represents the number of October monthly fields available from 15 ensemble members each being 48 years long. Hence, the dimension of X is $2m \times n = 2 \cdot 220 \cdot 256 \cdot 40 \times 15 \cdot 48 = 4,505,600 \times 720$. The anomalies from the ensemble of historical simulations are arranged in the following form:

$$X = \begin{pmatrix} \theta(z_1,t_1) & \theta(z_1,t_2) & \cdots & \theta(z_1,t_n) \\ \theta(z_2,t_1) & \theta(z_2,t_2) & \cdots & \theta(z_2,t_n) \\ \vdots & \vdots & \ddots & \vdots \\ \theta(z_m,t_1) & \theta(z_m,t_2) & \cdots & \theta(z_m,t_n) \\ S(z_1,t_1) & S(z_1,t_2) & \cdots & S(z_1,t_n) \\ S(z_2,t_1) & S(z_2,t_2) & \cdots & S(z_2,t_n) \\ \vdots & \vdots & \ddots & \vdots \\ S(z_m,t_1) & S(z_m,t_2) & \cdots & S(z_m,t_n) \end{pmatrix}, \qquad (1)$$

Table 1 Summary of experiments

Experiment	Ocean nudging	Simulation length	Ensemble members	References
Historical		1958–2005 used for the EOF analysis	15	CMIP5 experiment (Giorgetta et al. 2013)
ANOM-INIT	ORAS4 anomalies	Yearly started over 1960–2015, 10-years long	10	MiKlip Preop-LR
ANOM-1m-INIT	October ORAS4 anomalies	Yearly started over 1960–2015, 10-years long	10	
MODE-INIT	Filtered October ORAS4 anomalies	Yearly started over 1960–2015, 10-years long	10	FAI experiment (Polkova et al. 2019)

where z and t stand for space and time dimensions, respectively. Each of the monthly mean three-dimensional θ and S fields is normalized prior to the compilation of the X state vector. Ensuring parity in variance at each grid cell and aiming for capturing deep-ocean variability or regions of strongest density changes, several combinations of normalization and weighting are possible. We tested several schemes including commonly used normalization by the standard deviation (as in e.g., Hawkins and Sutton 2007); for more details see Supplementary Fig. S1. Eventually, we picked the scheme which led to the highest variance explained after ORAS4 was projected onto the EOF-modes that were constructed based on this weighting/normalization scheme. Thus, we apply weighting the EOF input fields by their contribution to density changes (the thermal expansion (α) and haline contraction (β) coefficients) and by the square root of the grid-cell area ($\sqrt{w_A}$): $w_\theta = \alpha \cdot \sqrt{w_A}$ and $w_S = \beta \cdot \sqrt{w_A}$, for temperature and salinity, respectively. The grid-cell area weighting, $\sqrt{w_A}$, represents the ratio of a grid-cell area to the total area; the $\sqrt{\cdot}$ is used here, to assure that during the EOF decomposition the variance is weighted by the grid-cell area. Coefficients α and β are calculated as: $\alpha = -\rho_0^{-1} \partial \bar{\rho} / \partial \bar{\theta}$ and $\beta = \rho_0^{-1} \partial \bar{\rho} / \partial \bar{S}$, where ρ_0 is the reference density and the overbar refers to long-term October monthly mean for potential temperature, salinity and density fields (ρ).

Starting from the state vector, X, as given above, the EOF-modes are computed by solving the eigenvalue problem:

$$X^T X E = EL, \tag{2}$$

using a SVD decomposition. In Eq. 2, $E = [e_k]$ represents the $k = 1, ..., K$ number of eigen modes describing patterns of variability $e_k = (e_k^\theta, e_k^S)$ for temperature and salinity anomalies, respectively. The vector $L = \lambda I$ contains the associated eigenvalues on its diagonal. The fraction of total variance explained by each eigenvector is expressed as $\sigma^2 = \frac{\lambda_k}{\sum_1^K \lambda_k}$.

The cumulative eigenvalue spectrum and the first bivariate full-depth EOF-mode are shown in Figs. 1 and 2, respectively. The first EOF-mode is shown for the surface layer, and along the equator as function of depth. This first-mode explains 11.4% of the total model variability (Fig. 1). At the ocean surface, the EOF structure for temperature reminds of the El-Niño Southern Oscillation joint by an Indian Ocean Dipole pattern. In the North Atlantic, there is some indication of a gyre pattern; however it is weaker in amplitude than what we see in the Pacific and the Indian Oceans due to the fact that, globally, the temperature and salinity variances are dominated by the changes in the Pacific Ocean. To represent each basin by its own modal structure, properly merged basin-scale EOFs might be required.

The vertical structure of the mode for both temperature and salinity fields along the equator is also shown in Fig. 2. The first mode reflects ocean dynamics mostly in the upper 1000 m. Further description of the vertical modal structure is provided in Fig. 3 in terms of the first four vertical temperature modes plotted for the three dynamically distinct locations marked in Fig. 2. In all three locations the EOFs are dominated by near-surface variability as represented by mode 1. Typically amplitudes decay with increasing mode number; however, in some locations, mode 3 can be larger than mode 2, e.g., in the North Pacific. The shape of some gravest vertical EOF-modes reminds that of vertical dynamic modes (Supplementary Fig. S2). With respect to amplitudes,
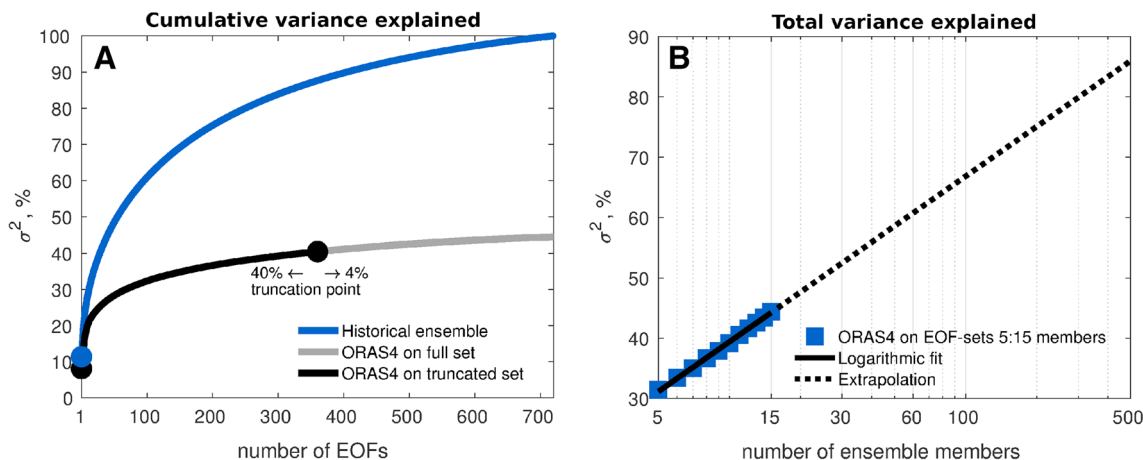

Fig. 1 a Cumulative variance explained by the EOF-modes (blue) and the ORAS4 reconstruction (black and grey). Here in black is ORAS4 projected onto truncated set of EOFs and in grey—onto full set of EOF-modes. The variance explained by the reconstruction is calculated as the difference of total variance of the ORAS4 weighted anomalies, $var(X)$, and the variance of the difference between the reconstructed and ORAS4 weighted anomalies, $var(\hat{X} - X)$, and further expressed as a fraction of total variance: $\sigma^2 = \left(1 - \frac{var(\hat{X}-X)}{var(X)}\right) \cdot 100\%$. b Logarithmic fit (black solid) to σ^2 for the reconstructed ORAS4 using 5- to 15-members to derive EOF-modes (blue squares) and extrapolation (black dashed) of σ^2 if using more ensemble members to derive the EOF-modes
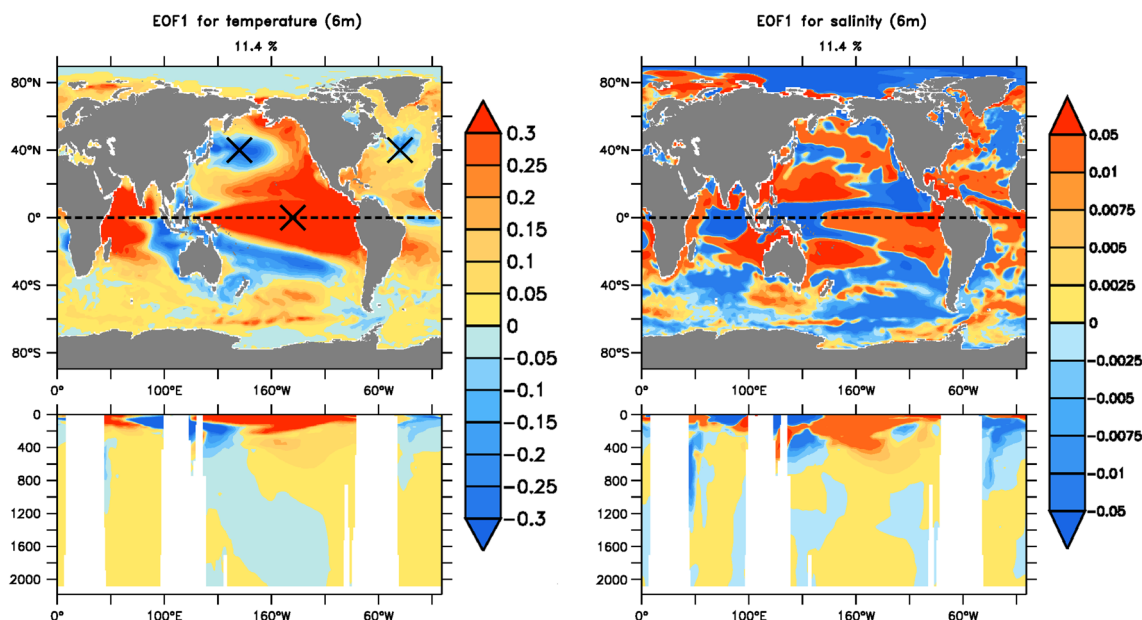
Fig. 2 The first bivariate 3-D EOF-mode for temperature (left, °C) and salinity (right, psu) for the ocean surface layer (upper panels) and for the equatorial section (lower panels)

for the equatorial Pacific Ocean, enhanced amplitudes are limited to the upper 800 m. In the North Pacific amplitudes different from zero reach to 1000 m, while in the central North Atlantic enhanced amplitudes are also distributed over the top 1600 m but with a complex vertical structure, as it would result from a different tendency in the near-surface water and the upper North Atlantic Deep Water. The different tendencies in the near-surface and deep North Atlantic is reminiscent of Atlantic Multidecadal Variability signatures noted by Polyakov et al. (2005) and Kim et al. (2018).

2.3 ORAS4 reconstruction

Projecting ORAS4 ocean reanalysis onto the set of eigenvectors e_k provides the principal component (PC) time-series:

$$a_k^{ORAS4}(\tau) = \sum_{i=1}^{2m} e_{ik} X_i^{ORAS4}(\tau), \qquad (3)$$

where τ represents the time dimension (1958–2016) of the ocean reanalysis anomalies, $X^{ORAS4}(z, \tau)$. The truncation level for index $k = 1...n^*$ is picked at the arbitrary point, where the reconstruction losses about 4% of variance explained, i.e., the first 360 EOFs explain 40% of variance and the further 360 only—4% (Fig. 1, grey curve). The time evolution of the filtered 3-D ORAS4 temperature, $\hat{\theta}^{ORAS4}(z, \tau)$, and salinity, $\hat{S}^{ORAS4}(z, \tau)$, anomalies are then obtained according to:

$$\hat{\theta}^{ORAS4}(z, \tau) \equiv \sum_{k=1}^{n^*} a_k^{ORAS4}(\tau) e_k^{\theta}(z),$$

$$\hat{S}^{ORAS4}(z, \tau) \equiv \sum_{k=1}^{n^*} a_k^{ORAS4}(\tau) e_k^{S}(z). \qquad (4)$$

Figure 4 shows examples of original ORAS4 temperature and salinity fields for October 2010 from 6 to 150 m depth, respectively, together with reconstructed fields based on Eq. 4. In agreement with the visual impression, the reconstructed fields are smoothed version of the original fields; they represent only 40% of the total variance in the ORAS4 reanalysis. The patterns of variance explained for several depth layers are provided in Supplementary Fig. S3. Overall, the filtering approach seems to strongly reduce the small-scale (in space and time) temperature and salinity signal, especially in the Atlantic Ocean. By contrast, the Pacific Ocean signal in general is well captured and represented.

These findings are further supported by Fig. 5 showing time series of October temperature and salinity anomalies averaged over the Niño 3.4 and the North Atlantic subpolar gyre regions. While the tropical Pacific ORAS4 time series are well represented by the reconstruction, the interannual signal in the North Atlantic is largely muted in both temperature and salinity. However, the reconstruction reproduces the multi-annual to decadal variability in the Atlantic Ocean. Good representation of the initial signal in the North Atlantic Ocean is considered to be important for enhanced
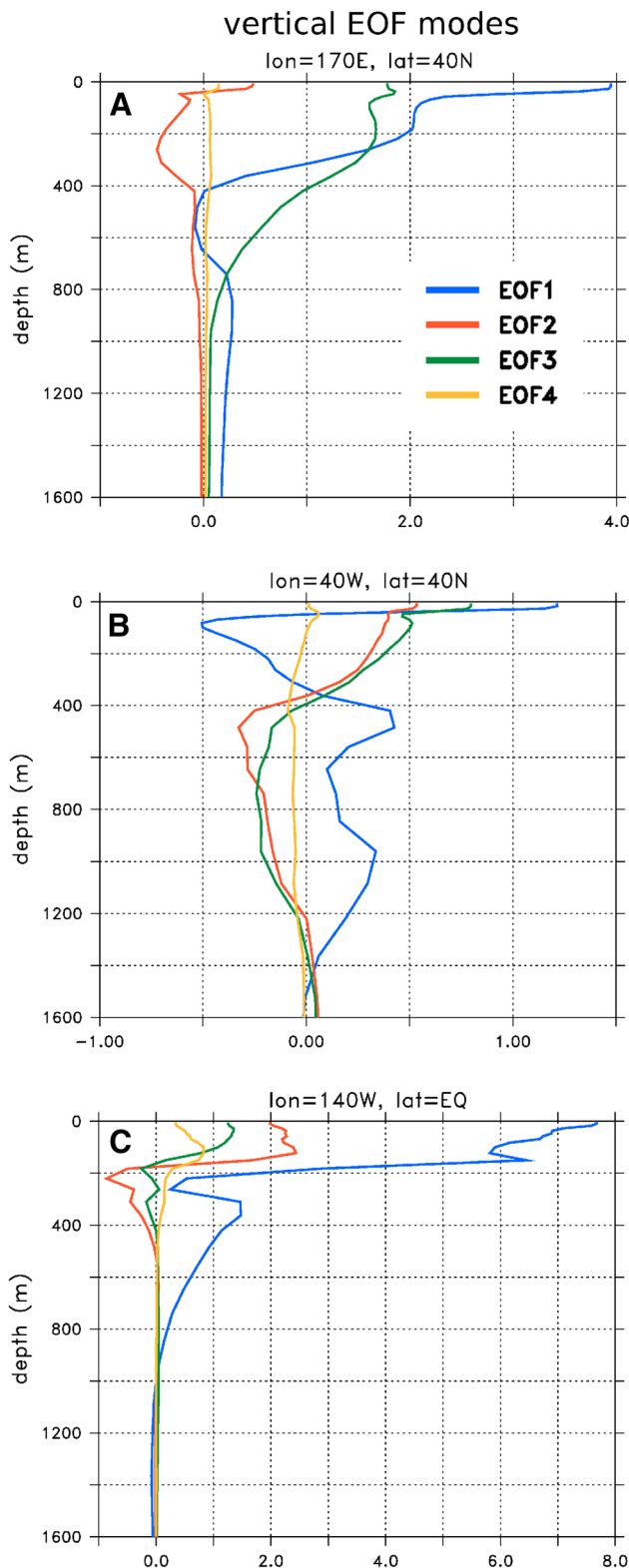
vertical EOF modes

Fig. 3 First four vertical EOF-modes for temperature scaled by their explained variances at three locations depicted in Fig. 2: in the North Pacific (upper panel, 170°E, 40°N), the North Atlantic (middle panel, 40°W, 40°N) and the equatorial Pacific (lower panel, 140°W)

prediction skill in decadal predictions (Yeager and Robson 2017). We therefore have to expect that diminished anomalies will lead to less skill in those regions in comparison to non-filtered anomaly initialization method.

The value of 40% of variance explained in the reconstruction is rather low and raises the questions: will the remaining signal be sufficient for initializing decadal hindcasts, and why does the filtering process eliminate so much of ORAS4 variance? By fitting a logarithmic curve to the variance explained plotted versus number of EOFs considered, we find that using more members of historical simulations for the EOF analysis could increase the value of explained variance in the reconstructed data (Fig. 1b). In addition, further understanding how compatible ORAS4 is with the model variability is important for the design of the initialization method and understanding its performance.

To discuss the latter, we show in Fig. 6 the model and reanalysis standard deviation (STD) patterns for linearly detrended temperature and salinity fields averaged for different depth levels. Model STD is calculated from one ensemble member of the historical simulations. The model and the reanalysis STD patterns for temperature resemble relatively well in the Northern Hemisphere near the ocean surface and in the upper-ocean layer. The figure reveals that near the surface the model simulates slightly stronger variability in the Southern Hemisphere, except near the equator, where the simulated variability is slightly weaker in comparison to the ORAS4 reanalysis. Below 300 m, the model simulates about 3 times stronger STD over the subtropical gyres than what is shown by the ORAS4 reanalysis. In the layer between 700 and 2000 m (not shown), the strongest STD is localized in the North Atlantic in ORAS4; in the historical simulation, the largest values of STD are along the Antarctic Circumpolar Current and the western sides of the subtropical gyres. In terms of the STD for salinity, the resemblance is worse than for temperature fields already in the surface layer. In general, the salinity STD from the model is stronger than what is shown by the reanalysis for all depth layers.

The discrepancy in the STD patterns between the historical simulations and the ORAS4 reanalysis that holds for both temperature and salinity and increases with depth, might very well indicate that ORAS4 variability modes are not entirely compatible with model modes.

3 Mode-initialized predictions

3.1 Initialized experiments

In this study, anomaly initialization (ANOM-INIT) is used as a reference against which the climate-mode initialization (MODE-INIT) is compared (see Table 1). The
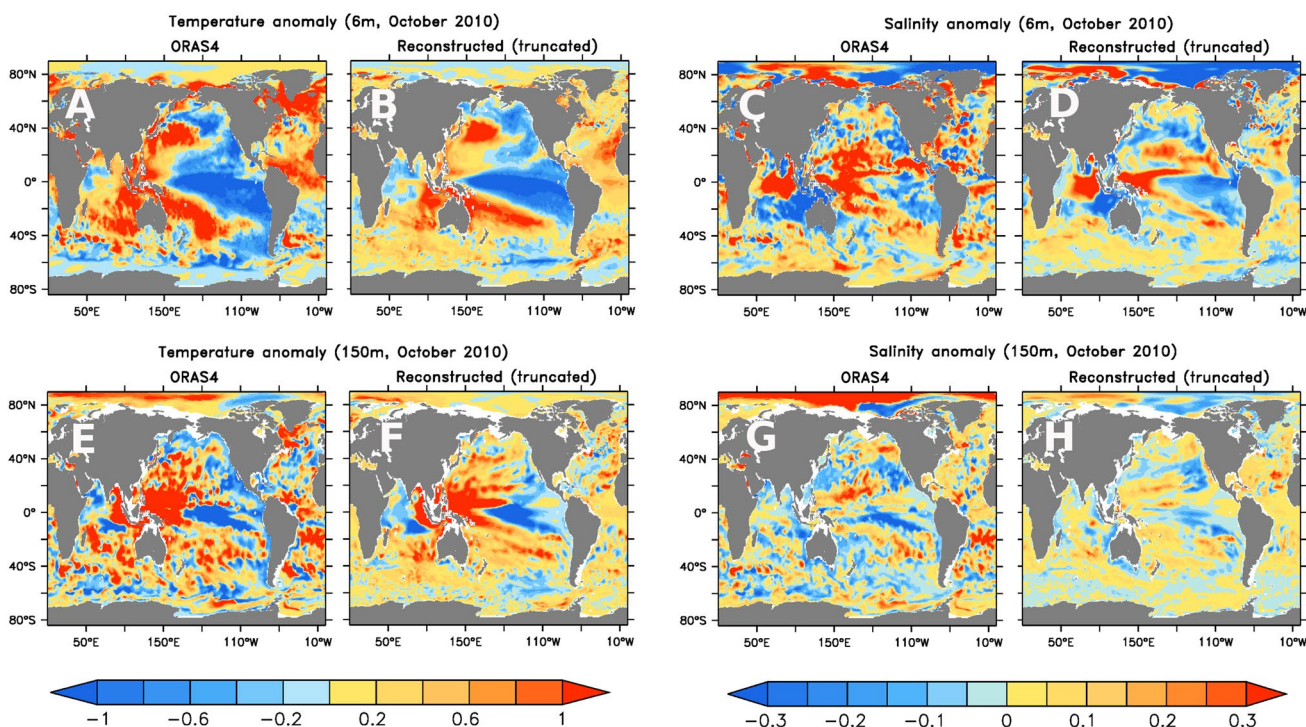
Fig. 4 An example of ORAS4 (**a**, **c**, **e** and **g**) and filtered ORAS4 (**b**, **d**, **f** and **h**) temperature anomalies (in °C; **a**, **b**, **e** and **f**) and salinity anomalies (in *psu*; C, D, G and H) for October 2010 at 6 m depth (upper panels) and 150 m depth (lower panels)
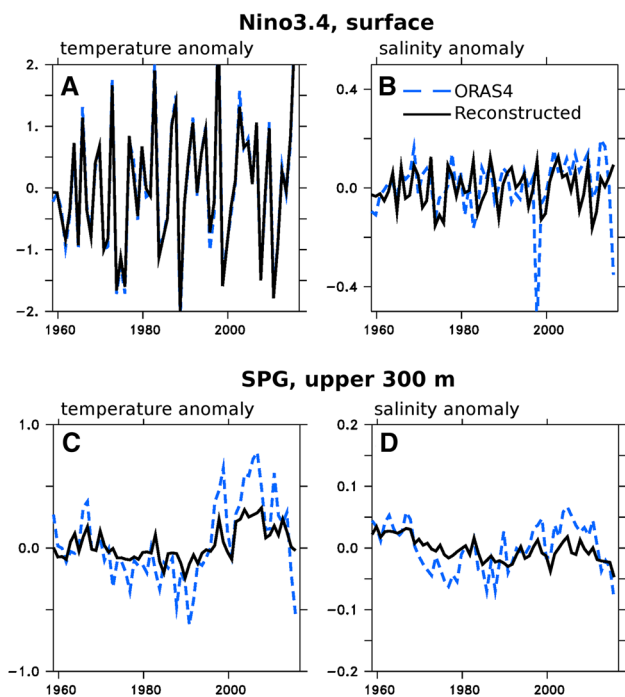

Fig. 5 Time series of October temperature (left, °C) and salinity (right, *psu*) anomalies from ORAS4 (blue) and the reconstruction (black): for the Niño 3.4 region (**a** and **b**) and the North Atlantic sub-polar-gyre region (SPG; **c** and **d**). The time series for the Niño 3.4 region are calculated at the ocean surface, 5°S–5°N and 170–120°W. For the SPG region, temperature and salinity anomalies are averaged over the upper-ocean 300 m, 50°–60°N and 65°W–10°E

anomaly initialization is common to many decadal prediction efforts (Kirtman et al. 2013; Meehl et al. 2014). It is concerned with an attempt to retain the model close to its attractor during the initialization step and further during prediction itself. An initial condition is constructed by adding observed (or reanalysis) monthly anomalies to the model's climatology. Predictions initialized in such a way eliminate initial drift (lead-time dependent bias) as compared to predictions initialized close to the observed climate state as in full-field initialization.

ANOM-INIT and MODE-INIT are started from nudging runs, ANOM-ASSIM and MODE-ASSIM, respectively. In ANOM-ASSIM, the ocean component of the MPI-ESM is nudged toward ORAS4 temperature and salinity (T&S) anomalies added to model's climatology. In MODE-ASSIM, the ocean component is nudged toward the filtered ORAS4 anomalies added to model's climatology. The nudging runs are started from the historical simulation and are carried out over the period 1960–2015. Different to ANOM-ASSIM, MODE-ASSIM starts every year from a historical simulation on the 30th of September. This is because we only consider October monthly mean anomalies for the EOF analysis, for the reconstruction of ORAS4 and now for nudging. October is selected because it precedes the initialization dates (November 1) in the reference ANOM-INIT experiments (Polkova et al. 2019). To estimate the effect of one-month nudging, we carried
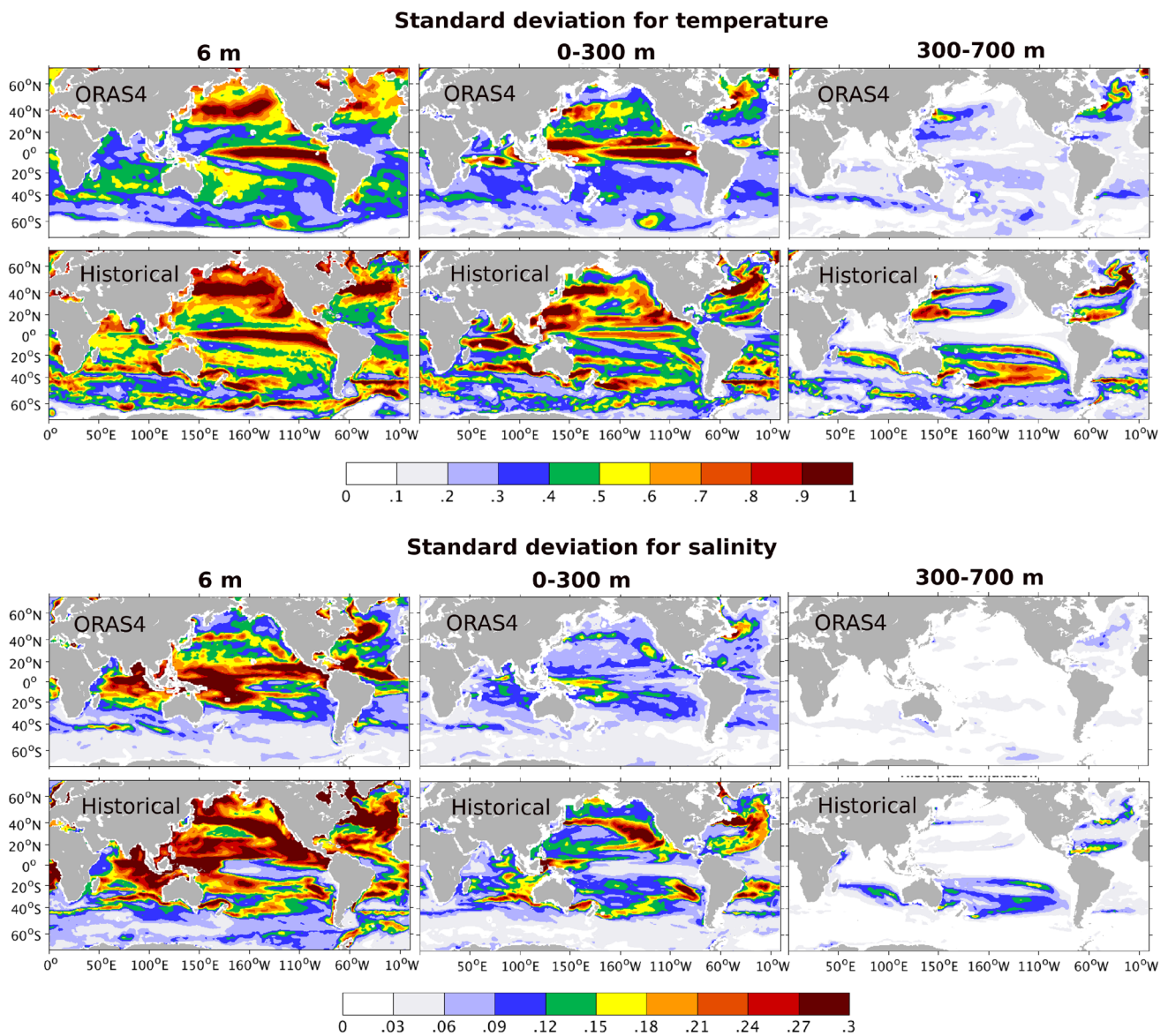
Standard deviation for temperature

Standard deviation for salinity

Fig. 6 STD from a historical simulation (one member) and ORAS4 for October temperature (°C) and salinity (*psu*) fields averaged for different depth layers and linearly detrended at each grid-point: surface layer (left panels), upper 300 m (middle panels) and 300–700 m (right panels)

out anomaly nudging over one month with non-filtered ORAS4 states (ANOM-1m-ASSIM) and compared the results against ANOM-ASSIM.

Figure 7 demonstrates STD for the temperature snapshots at the beginning of initialized hindcasts from three nudging runs. The difference in terms of STD between the two experiments ANOM-1m-ASSIM and ANOM-ASSIM appears to be very marginal. Overall, the STD at different depth levels in MODE-ASSIM is about 1.5–2 times lower than in ANOM-1m-ASSIM and ANOM-ASSIM. The variability at the ocean surface over the Pacific basin is well represented in MODE-ASSIM as was also suggested by previous Figs. 4 and 5. While in the Atlantic Ocean, the amplitude of the

initial anomalies in MODE-ASSIM is lower than in the non-filtered nudging experiments (see for instance, the Labrador Sea). The STD from MODE-ASSIM is also lower than what the model simulates in the historical simulations, apart from the tropics (Fig. S4 in the Supplementary material). To sum-up, regions of high STD are mostly captured in MODE-ASSIM, however in many regions with lower amplitudes than in other nudging runs and the historical simulations.

The initialized hindcasts use ERA-40/ERA-Interim (Uppala et al. 2005; Dee et al. 2011) temperature, vorticity, divergence and surface pressure full-field values for nudging the atmosphere. The relaxation time is 11 days in the ocean; in the atmosphere it is 6 h for vorticity, 24 h for temperature
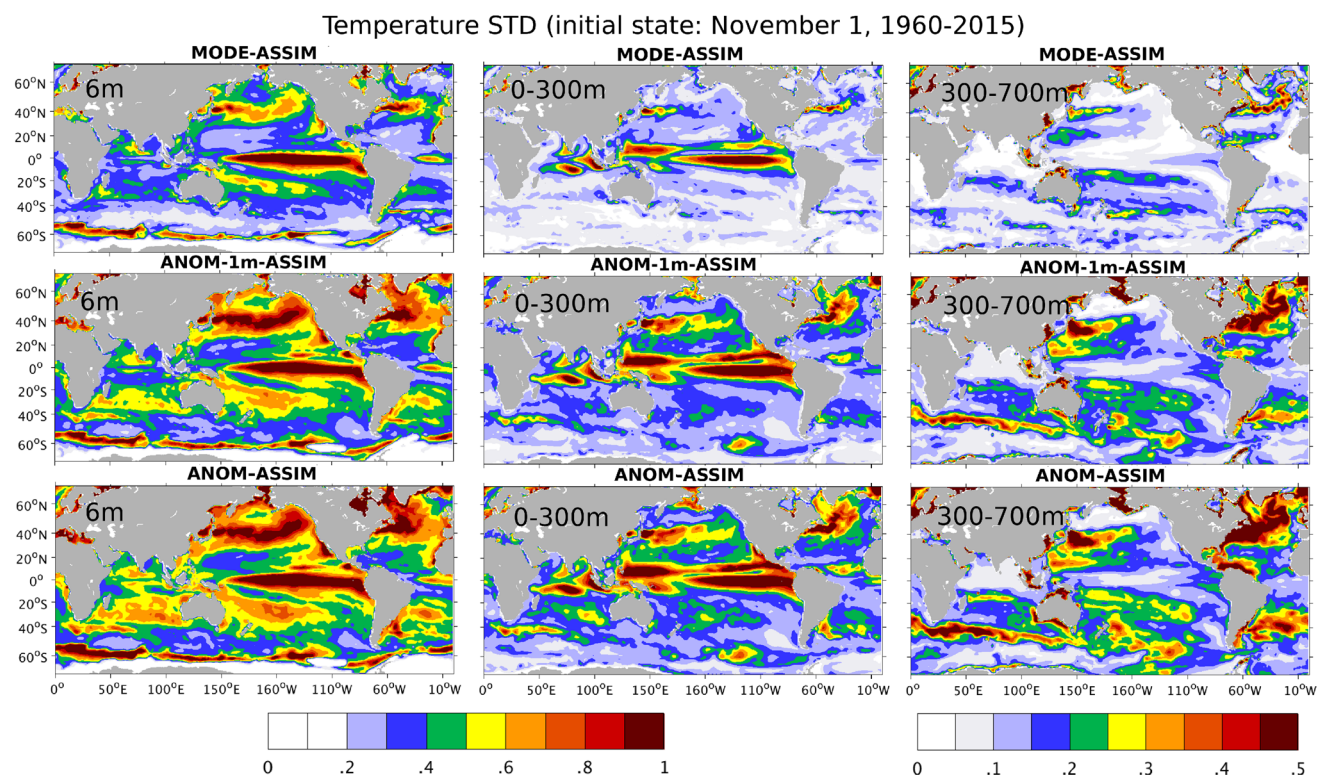
Temperature STD (initial state: November 1, 1960-2015)

Fig. 7 STD for temperature snapshots (°C) from three nudging runs: MODE-ASSIM (upper panels), ANOM-1m-ASSIM (middle panels) and ANOM-ASSIM (lower panels) for different depth layers: 6 m (left column), 0–300 m (middle column) and 300–700 m (right column). Snapshots correspond to initialization dates (November 1) over the period 1960–2015

and surface pressure, and 48 h for divergence. For sea-ice concentration, the nudging toward the NSIDC data (Fetterer et al. 2016) is applied using a 11-days relaxation time scale.

All initialized hindcasts are yearly started from the corresponding nudging run from November 1 over the period 1960–2015, and are 10 years and 2 months long. As with ANOM-INIT, MODE-INIT also uses lagged initialization to generate an ensemble of predictions, however with some differences in detail: because MODE-ASSIM and ANOM-1m-INIT are 1-month long, initial conditions are sampled from 9-days long free runs following each assimilation. Similar to historical simulations, the initialized experiments until 2005 use the same external forcing and the RCP4.5 pathway over 2006–2025. A summary of all experiments is given in Table 1. When analyzing hindcast anomalies, a long-term mean is removed from the time-series of ensemble mean hindcast and the verification data set for a particular lead time at each grid-point. This procedure also removes the mean bias from the hindcasts. The MODE-INIT hindcasts do not show a lead-time dependent bias.

3.2 Prediction skill

To demonstrate performance of climate-mode initialization, we provide in Figs. 8 and 9 the correlation skill patterns

for sea surface temperature (SST) with respect to HadISST (Rayner et al. 2003) for seasons of the first lead year and multi-year means, respectively. The correlation skill differences with respect to a reference ANOM-1m-INIT (both Figures) and historical simulations (Fig. 9) are also shown. ANOM-1m-INIT correlation skill patters are provided in the supplementary; ANOM-INIT skill was reported earlier by Polkova et al. (2019).

Shortly after initialization (Fig. 8), in DJF and MAM, the ANOM-1m-INIT slightly outperforms MODE-INIT in the subpolar regions and tropical Atlantic, while elsewhere the correlation skill is comparable. However, in JJA and SON, the results suggest that MODE-INIT outperforms ANOM-1m-INIT in the tropics. The percentage of the area where MODE-INIT significantly improves (reduces) the correlation skill is estimated as a number of grid-cells of positive (negative) significant correlation difference to the total number of grid-cells of significant correlation difference values. Statistically significant values are estimated based on the t-test at 90% confidence level (Weaver and Wuensch 2013). Thus, the area percentage of the significantly improved skill is 52% in JJA and 64% in SON as estimated globally, and 67% in JJA and 92% in SON as estimated for the tropical oceans. Improvements in skill on seasonal timescales might
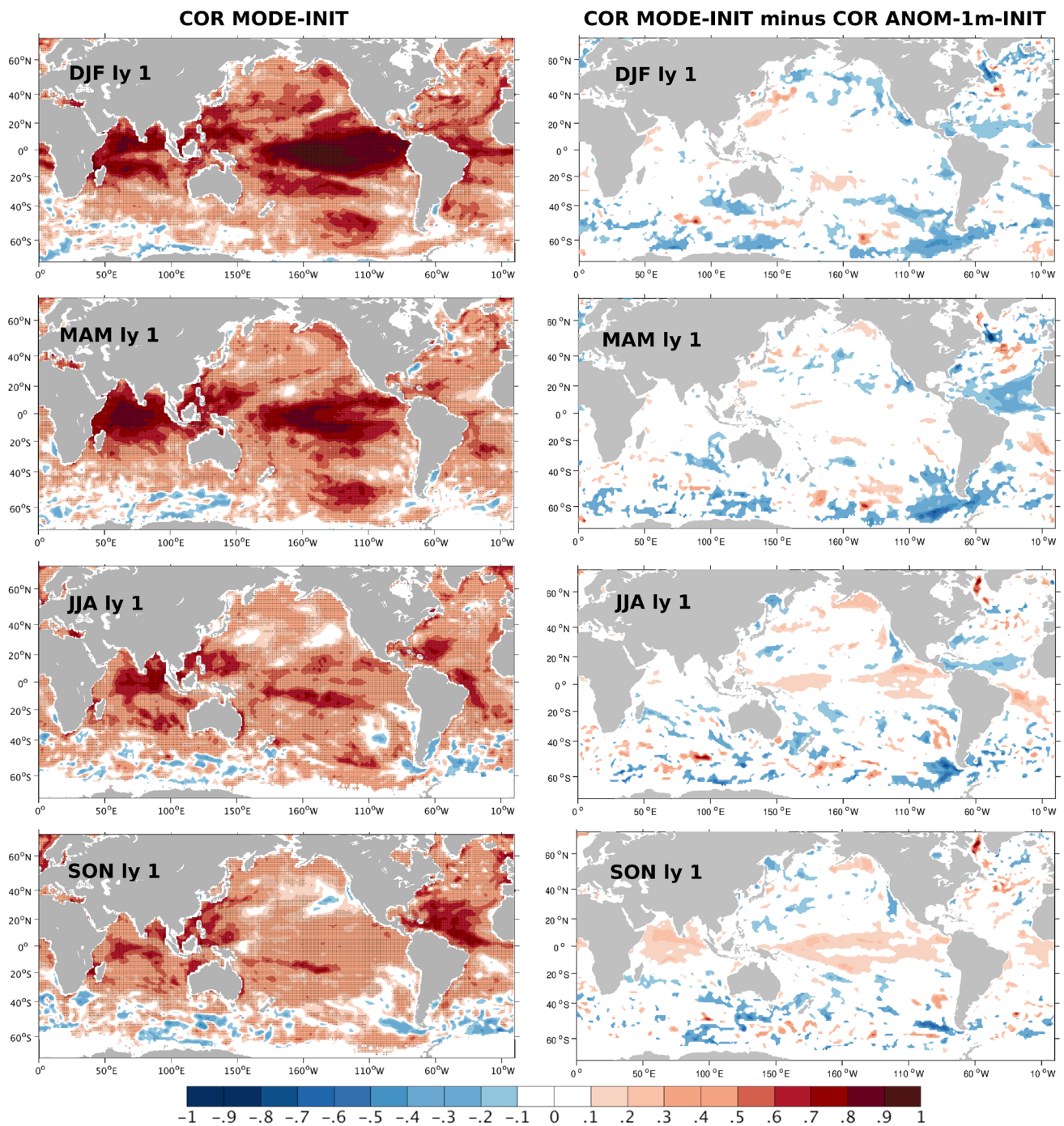
COR MODE-INIT

COR MODE-INIT minus COR ANOM-1m-INIT

Fig. 8 Correlation skill of the ensemble mean SST for DJF (the first row), MAM (the second row), JJA (the third row) and SON (the fourth row) seasons of the first lead year from MODE-INIT (left) and the significant correlation skill difference between MODE-INIT and ANOM-1m-INIT (right). Verification dataset is HadISST. Hatch-ing (left) and colored regions (right) indicate statistically significant values estimated based on the *t*-test at 90% confidence level Weaver and Wuensch (2013). Skill is estimated for the experiments initialized over the period 1961–2015

reflect improved initialization of the El-Niño Southern Oscillation (ENSO).

In terms of SST skill for multi-year averages (Fig. 9), MODE-INIT outperforms ANOM-1m-INIT in the northern

and tropical Pacific in lead years 2–5. The percentage of grid-cells with significantly improved skill amounts to 63 % globally and 89% over the tropics. At lead years 6–9, the skill from both initialization methods is comparable.
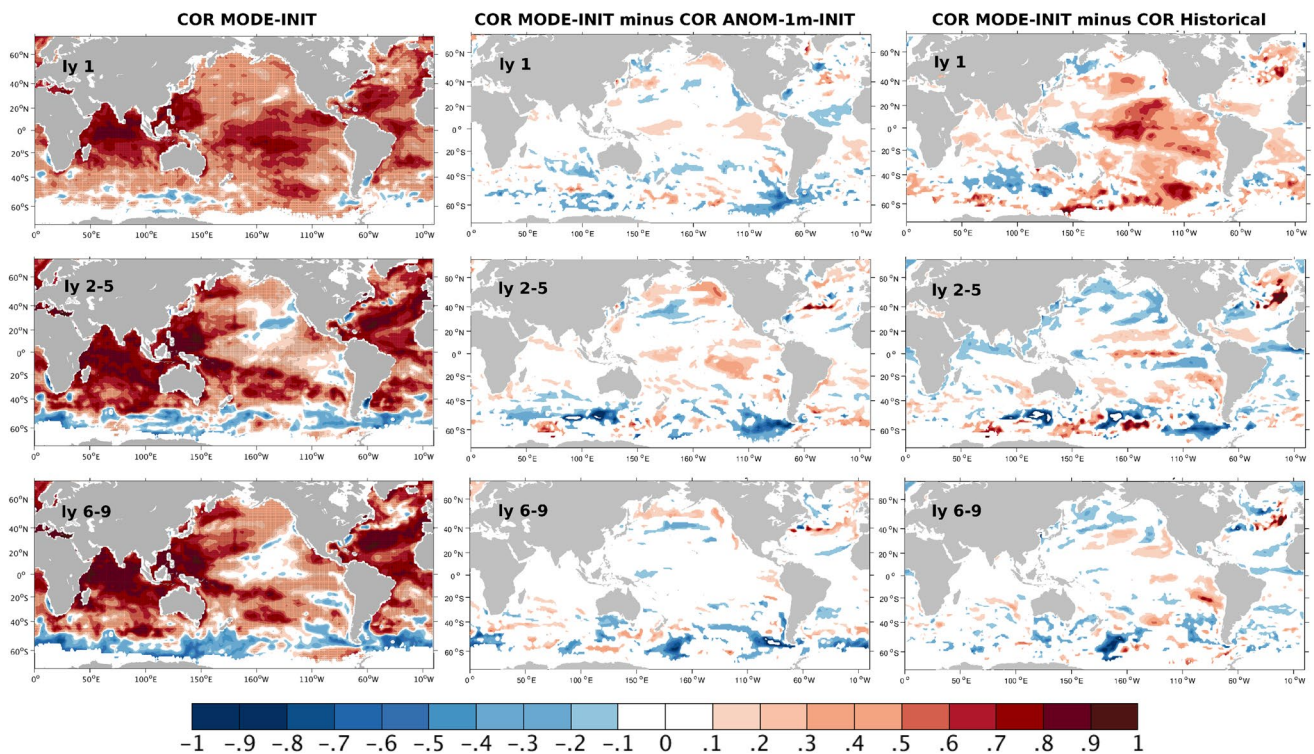
Fig. 9 Correlation skill of the ensemble mean SST for lead year 1 (the first row), 2–5 (the second row) and 6–9 (the third row) from MODE-INIT (the first column) and the significant correlation skill difference between MODE-INIT and ANOM-1m-INIT (the second column) and the significant correlation skill difference between MODE-INIT and historical simulations (the third column). Verification dataset is HadISST. Hatching (left column) and colored regions (middle and right columns) indicate statistically significant values estimated based on the *t*-test at 90% confidence level (Weaver and Wuensch 2013). Skill is estimated for the experiments initialized over the period 1961–2015

To assess prediction skill due to external forcing changes, the skill of MODE-INIT is compared to that of historical simulations. In the first lead year, initialization brings significant improvement in 86% of the areas globally and 95% in the tropics. Because historical simulations by design are not synchronized with observations, on short time-scales (seasonal-to-interannual), it is unlikely that historical simulations can outperform any initialized predictions. For longer lead years and multi-year averages, the externally forced response dominates the prediction skill for surface temperature and, apart from the subpolar North Atlantic, the improvements from initialization are less obvious.

Interestingly, MODE-INIT for multi-year averages shows some indication of slightly improved skill over ANOM-1m-INIT along the Gulf Stream path. This also holds when comparing skill from MODE-INIT and ANOM-INIT (Polkova et al. 2019). An analysis of the Gulf Stream paths in the two experiments shows some differences in the hindcast performance, however does not allow to explain the reasons of the skill improvements. We think that the problem of low skill along the Gulf Stream path arises due to the existing mismatch between mean and variability when composing the fields for the anomaly initializations. Namely, it is well

known that after the Gulf Stream separation at Cape Hatteras, the flow path evolves more zonally in most of the climate models than in reality. The path from the ocean reanalysis that we use for initialization also somewhat deviates from that of the model. Thus, in the superposition of reanalysis anomalies and model climatology, meridional shifts of the observed Gulf Stream are not realized as shifts but materialize as local minima or maxima that cannot be dynamically sustained in the same way as shifts could be but creates unphysical anomalies. As the model has low variability in regions of high observed Gulf Stream variability, the MODE-INIT method will essentially filter out the observed variability rather than placing it at the wrong position. Thus, this issue of wrongly placed anomalies might not be as strong as in ANOM-1m-INIT or ANOM-INIT.

To understand what causes the improved skill for MODE-INIT in the equatorial and tropical Pacific Ocean, we analyze the Niño 3.4 index (Fig. 10) and the momentum balance in the equatorial Pacific (Fig. 11). That initialized hindcasts carried out with the MPI-ESM sometimes lag the observed ENSO events by one year was reported recently (Polkova et al. 2019) and is also apparent here: The hindcasts initialized in November are not able to predict El-Niño events
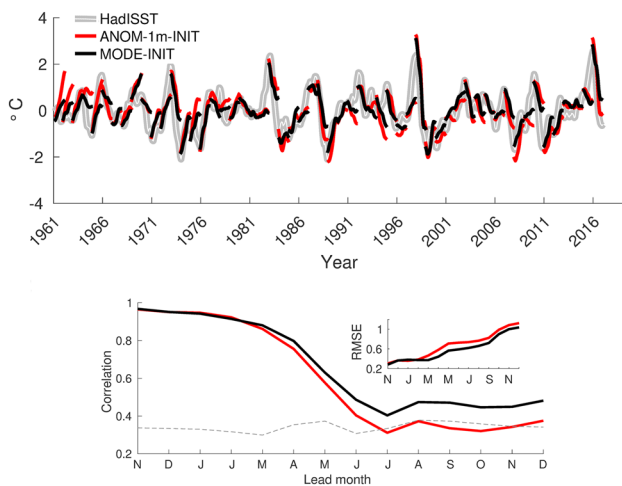
Fig. 10 Monthly Niño 3.4 Index of lead year 0–1 (14 months) from HadISST (grey) and the hindcasts: MODE-INIT (black) and ANOM-1m-INIT (red). A 3-months running mean is applied to the time series. Correlation coefficients, root-mean-square errors (RMSE) and the 95% confidence level according to the *t*-test (gray dashed) are shown
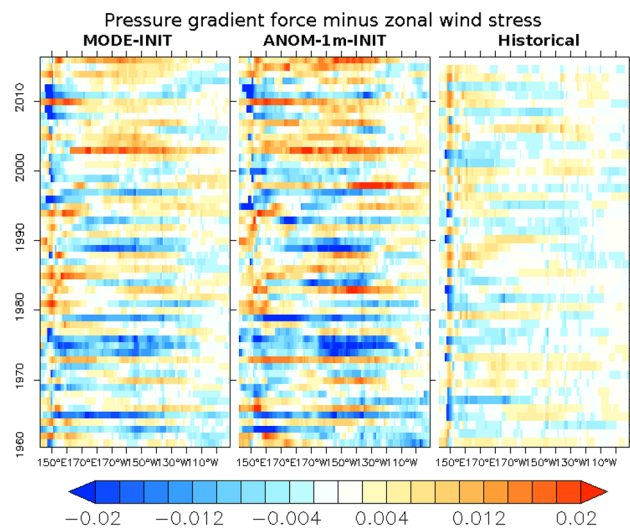

Fig. 11 Zonal momentum balance (N/m^2) of the upper equatorial Pacific calculated as a difference between integrated pressure gradient force and zonal wind stress. The balance is plotted for the MODE-INIT (left) and ANOM-1m-INIT (middle) in the first month after initialization, and for the historical simulation (right) over the counterpart period

developing in fall of the following year (e.g., 1982 and 1997). However, once the hindcasts are initialized from the strong anomaly (e.g., 1983 and 1998) they usually show good skill by the right duration of persistence. That the predictability of the El-Niño is limited to less than a year is known from previous studies as "spring-barrier". In terms of differences in the performance, ANOM-1m-INIT tends

to simulate stronger ENSO events, for instance, at the beginning of the 1960s. Also the initial point is slightly offset in the two experiments.

Overall, MODE-INIT shows higher skill for the Niño 3.4 region in terms of both correlation and the root-mean-squared error (RMSE) calculated with respect to HadISST (Fig. 10). Both experiments start from the same level of skill, but it seems that ANOM-1m-INIT losses skill faster than MODE-INIT, which is consistent with our hypothesis that filtering out noise component from the initial state might lead to a more persistent prediction skill.

The momentum balance between the zonal pressure gradient and the zonal wind stress forcing was shown previously to be a crucial element for balanced initial conditions (Liu et al. 2017; Thoma et al. 2015). Offsetting the balance can lead to an artificially increased number of El-Niño events and explain reduced skill in the tropical Pacific (Liu et al. 2017). Figure 11 shows the residual of the zonal momentum balance along the equator for the first month of initialized hindcasts. The residual is derived from the momentum balance equation in the upper equatorial Pacific in the zonal direction:

$$A_v \frac{\partial u}{\partial z} = \frac{1}{\rho_0}\Big(\tau_x - \int_0^{zo} \frac{\partial P}{\partial x} dz\Big), \qquad (5)$$

where A_v is the vertical eddy viscosity, u is the zonal velocity. The wind stress, τ_x, is compensated by the pressure gradient force, $\int_0^{zo} \frac{\partial P}{\partial x} dz$, integrated down to the depth, zo, where the vertical shear of zonal velocity $\frac{\partial u}{\partial z}$ becomes zero (Bryden and Brady 1985). From this follows that for the dynamical balance in the equatorial Pacific, the residual between the integrated pressure gradient force and the zonal wind stress should be zero.

Figure 11 suggests that MODE-INIT has indeed a better balance of the ocean and the atmosphere states at the beginning of the hindcasts as compared to ANOM-1m-INIT. The balance from the historical simulations is used as a reference, demonstrating that a certain amount of imbalance is expected to allow for variability, in particular, at the western side of the Pacific Ocean, where for instance westerly windburst may trigger El-Niño events. The MODE-INIT zonal momentum balance is between that of historical simulations and ANOM-1m-INIT. Further analysis shows a better match between the terms of the momentum balance equation in the eastern equatorial Pacific basin (120–90°W) and lower amplitudes of the terms in the western and central basins (160°E–120°W) for MODE-INIT as compared to ANOM-1m-INIT (Supplementary Fig. S7).

Nevertheless, reduced imbalances alone may not be indicative of an improved initialization, as balanced states may still miss the required variability. In fact, the STD of the upper 300 m temperature of the historical simulation

shown in Fig. 6 demonstrates much smaller variability in comparison to ORAS4. A further factor is a mismatch of the ocean and atmosphere mean fields. The initialization procedure, which we employ in this study, uses full-field ERA40/ERA-Interim initialization in the atmosphere and anomaly ORAS4 initialization in the ocean. Analyzing wind-stress from the first lead year of the hindcasts and the counter-part from the ERA-40/ERA-Interim reanalyses indicates that the reanalyses have larger zonal wind-stress values than the initialized hindcasts (not shown). The thermocline from the hindcasts is also slightly weaker than from the ORAS4 reanalysis (not shown). Differences in mean states were in fact responsible for large number of artificial El-Niño events and reduced skill in comparison to more consistent initialization reported earlier (Liu et al. 2017). However, as both hindcasts use anomaly initialization, difference in the mean between MODE-INIT and ANOM-1m-INIT remain small and is unlikely to explain the difference in prediction skill.

4 Discussion and concluding remarks

Dealing with observational and model errors is one of the major challenges for decadal climate predictions and poses questions as how to best initialize internal variability and initialization of which regions and variables matters for the prediction skill at decadal time scale? For optimal use of observations, it thus remains important to identify predictable components of the climate system that have to be initialized. In terms of the skill, decadal predictions might benefit from filtering out unpredictable elements in the observations. This task involves identifying spatial and temporal characteristics of internal variability that should be reflected in initial states. In the current study, we address this problem by reshaping reanalysis variability onto model variability modes to filter out unpredictable signal. Though, the current initialization method best addresses variability in the upper ocean, it achieves comparably high prediction skill and improves the reference anomaly initialization method, particularly in the tropics and on time scales larger than 6 months. This suggests that initializing the ocean subsurface might be sufficient for short-term predictions. With respect to the results in the North Atlantic, deep ocean observations are needed to better define the initial state.

We anticipate that further improvements can be expected once improved EOFs are being used during the climate-mode initialization. In the following, we offer several suggestions to improve the method:

– To derive the EOFs, 15-member ensemble of historical simulations was used. Using a large ensemble, such as now available 100-member ensembles (Bittner et al.

2016), could significantly boost the variance explained in the reconstruction data set, not only at the subsurface but also in the deeper ocean layers. This step is also connected to a decision on truncation level and calls further testing a boundary that separates dominant modes from the noise. Further normalization/weighting used for the EOF analysis can be tested to improve scheme for representing inhomogeneity of variance in different data sets.

– In the EOF analyses, we did not exclude linear trends associated with the external forcing response, since this is one of the major sources of prediction skill at decadal time scales (Boer et al. 2013). The trends in MODE-ASSIM and MODE-INIT for some variables (e.g., the North Atlantic sub-polar gyre index) resemble historical simulations rather than ORAS4. It is thus a further question of whether to include the trend in the EOF analysis and if not, how to alternatively represent the trend in the initial fields.

– For future studies, we suggest to use merged regional rather than global EOF-modes to better represent regional anomalies "carrying" prediction skill, e.g., in the North Atlantic Ocean. It has been shown previously that regional modes can perform significantly better in reconstructing Atlantic sea level variability (Meyssignac et al. 2012) in comparison to using global EOFs (Carson et al. 2017).

– Near the equator, initializing velocity field in addition to temperature and salinity can carry extra benefits to better represent meridional transports and pressure balances. To provide better-balanced states anomaly assimilation should also be employed for the atmosphere. Ultimately, this might call for coupled-mode initialization.

Although used here only in a pilot setting, the climate-mode initialization method shows encouraging results and suggests potential for future application in climate predictions. The current initialization efforts either use the initial state from the "native" data assimilation systems built specifically for the prediction system or they introduce external data assimilation products into the prediction system. The latter is cheaper but external reanalyses might not be consistent with the prediction system. If the method that is proposed in this study is further improved it could be useful for producing centers which cannot afford advanced data assimilation systems.

The important question that still remains to be answered by future studies on initialization of decadal predictions is: to what extent can dynamically consistent initialization improve skill? Seasonal prediction studies showed that dynamically consistent initialization pays off for the ENSO skill (Zhang et al. 2005; Chen and Cane 2008; Liu et al. 2017). For decadal prediction, the associated benefit is not clear. This may call into question the need for advanced data

assimilation schemes that take better care of imbalances. Thus, more evidence is needed to demonstrate that advanced data assimilation and initialization schemes pay off in terms of better prediction skill on decadal time-scales.

The results of this study reveal improved prediction skill in the equatorial Pacific on seasonal time scales resulting from an improved dynamical balance in the filtered initial conditions. Also on longer time scales, there are some modest skill improvements over anomaly initialization. Since, the only difference in the compared experiments comes from filtering initial states, we believe that the method proved to have realized its intention to improve skill by improving dynamical consistency between the prediction system and the initial states. However, as we suggested above, there is room for improving the method. The climate-mode initialization method shows a shortcoming in the North Atlantic, where the EOF basis over-smoothes the climate signal in the initial conditions responsible for prediction skill. A comparison with other initialization methods shows an added value in terms of the prediction skill from the climate-mode initialization method (Polkova et al. 2019). We therefore anticipate further benefits over the current setting once the improvements suggested above were implemented.

References

Ballish B (1981) A simple test of the initialization of gravity modes. Monthly Weather Rev 109(6):1318–1321

Balmaseda M, Anderson D (2009) Impact of initialization strategies and observations on seasonal forecast skill. Geophys Res Lett 36(1):L01701

Balmaseda M et al (2013) Evaluation of the ECMWF ocean reanalysis system ORAS4. Q J R Meteorol Soc 139(674):1132–1161

Bittner M et al (2016) Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty. Geophy Res Lett 43(17):9324–9332

Boer G et al (2013) Decadal predictability and forecast skill. Clim Dyn 41:1817–1833. https://doi.org/10.1007/s00382-013-1705-0

Boer G et al (2016) The decadal climate prediction project (DCPP) contribution to CMIP6. Geosci Model Dev 9(10):3751

Branstator G et al (1993) Identification of highly predictable flow elements for spatial filtering of medium-and extended-range numerical forecasts. Monthly Weather Rev 121(6):1786–1802

Bretherton C et al (1999) The effective number of spatial degrees of freedom of a time-varying field. J Clim 12(7):1990–2009

Bryden HL, Brady EC (1985) Diagnostic model of the three-dimensional circulation in the upper equatorial Pacific Ocean. J Phys Oceanogr 15(10):1255–1273

Carson M et al (2017) Regional sea level variability and trends, 1960–2007: a comparison of sea level reconstructions and ocean syntheses. J Geophy Res Oceans 122(11):9068–9091

Chen D, Cane M (2008) El-Niño prediction and predictability. J Comput Phys 227(7):3625–3640

Counillon F et al (2014) Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment. Tellus A Dy Meteorol Oceanogr 66(1):21074

Dee D et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597

Fetterer F et al (2016) Sea ice index, version 2. National Snow and Ice Data Center (NSIDC), Boulder

Giorgetta M et al (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. J Adv Model Earth Syst 5(3):572–597

Goddard L et al (2013) A verification framework for interannual-to-decadal predictions experiments. Clim Dyn 40(1–2):245–272

Hawkins E, Sutton R (2007) Variability of the Atlantic thermohaline circulation described by three-dimensional empirical orthogonal functions. Clim Dyn 29(7–8):745–762

Hazeleger W et al (2013) Multiyear climate predictions using two initialisation strategies. Geophys Res Lett 40(9):1794–1798

Hurrell J et al (2009) A unified modeling approach to climate system prediction. Bull Am Meteorol Soc 90(12):1819–1832

Jungclaus J et al (2013) Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth System Model. J Adv Model Earth Syst 5(2):422–446

Kim WM et al (2018) Key role of internal ocean dynamics in Atlantic multidecadal variability during the last half century. Geophys Res Lett 45(24):13–449

Kirtman B et al (2013) Near-term climate change: projections and predictability. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, pp 953–1028

Kröger J et al (2018) Full-field initialized decadal predictions with the MPI Earth System Model: an initial shock in the North Atlantic. Clim Dyn 51(7–8):2593–2608

Liu X et al (2017) Impact of in-consistency between the climate model and its initial conditions on climate prediction. Clim Dyn 49(3):1061–1075

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

Malanotte-Rizzoli P et al (1989) Initialization and data assimilation experiments with a primitive equation model. Dyn Atmos Oceans 13(3–4):349–378

Marotzke J et al (2016) MiKlip-a national research project on decadal climate prediction. Bull Am Meteorol Soc 97(12):2379–2394

Meehl G et al (2009) Decadal prediction: can it be skillful? Bull Am Meteorol Soc 90(10):1467–1485. https://doi.org/10.1175/2009BAMS2778.1

Meehl G et al (2014) Decadal climate prediction: an update from the trenches. Bull Am Meteorol Soc 95(2):243–267

Meyssignac B et al (2012) An assessment of two-dimensional past sea level reconstructions over 1950–2009 based on tide-gauge data and different input sea level grids. Surv Geophys 33(5):945–972

Mochizuki T et al (2016) Multiyear climate prediction with initialization based on 4d-var data assimilation. Geophys Res Lett 43(8):3903–3910

Mulholland D et al (2015) Origin and impact of initialization shocks in coupled atmosphere-ocean forecasts. Monthly Weather Rev 143(11):4631–4644

Müller W et al (2018) A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1. 2-HR). J Adv Model Earth Syst 10(7):1383–1413

Nardelli B, Santoleri R (2005) Methods for the reconstruction of vertical profiles from surface data: multivariate analyses, residual GEM, and variable temporal signals in the North Pacific Ocean. J Atmos Ocean Technol 22(11):1762–1781

Penny S et al (2017) Coupled data assimilation for integrated earth system analysis and prediction. Bull Am Meteorol Soc 97(7):ES169–ES172

Pohlmann H et al (2017) Initialization shock in decadal hindcasts due to errors in wind stress over the tropical Pacific. Clim Dyn 49(7–8):2685–2693

Polkova I et al (2014) Impact of initialization procedures on the predictive skill of a coupled ocean-atmosphere model. Clim Dyn 42:3151–3169. https://doi.org/10.1007/s00382-013-1969-4

Polkova I et al (2019) Initialization and ensemble generation for decadal climate predictions: a comparison of different methods. J Adv Model Earth Syst 11(1):149–172

Polyakov I et al (2005) Multidecadal variability of North Atlantic temperature and salinity during the twentieth century. J Clim 18(21):4562–4581

Rayner N et al (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res Atmos 108(D14):4407

Rosati A et al (1997) The impact of ocean initial conditions on ENSO forecasting with a coupled model. Monthly Weather Rev 125(5):754–772

Smith D et al (2013) A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. Clim Dyn 41(11–12):3325–3338

Stammer D et al (2016) Ocean data assimilation in support of climate applications: status and perspectives. Ann Rev Mar Sci 8:491–518

Sugiura N et al (2008) Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. J Geophys Res Oceans 113(C10):C10017

Taylor K et al (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93(4):485–498

Thoma M et al (2015) Decadal hindcasts initialized using observed surface wind stress: evaluation and prediction out to 2024. Geophys Res Lett 42(15):6454–6461

Uppala S et al (2005) The ERA-40 re-analysis. Q J R Meteorol Soc 131(612):2961–3012

Volpi D et al (2016) Comparison of full field and anomaly initialisation for decadal climate prediction: towards an optimal consistency between the ocean and sea-ice anomaly initialisation state. Clim Dyn 49(4):1181–1195

Volpi D et al (2017) Decadal climate prediction with a refined anomaly initialisation approach. Clim Dyn 48(5–6):1841–1853

Weaver B, Wuensch KL (2013) SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. Behav Res Methods 45(3):880–895

Williamson D (1976) Normal mode initialization procedure applied to forecasts with the global shallow water equations. Monthly Weather Rev 104(2):195–206

Yeager S, Robson J (2017) Recent progress in understanding and predicting Atlantic decadal climate variability. Curr Clim Change Rep 3(2):112–127

Zhang S et al (2005) Initialization of an ENSO forecast system using a parallelized ensemble filter. Monthly Weather Rev 133(11):3176–3201