



Seasonal and decadal forecasts of Atlantic Sea surface temperatures using a linear inverse model

Benjamin Huddart¹ · Aneesh Subramanian² · Laure Zanna² · Tim Palmer²

Received: 11 March 2016 / Accepted: 23 September 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Predictability of Atlantic Ocean sea surface temperatures (SST) on seasonal and decadal timescales is investigated using a suite of statistical linear inverse models (LIM). Observed monthly SST anomalies in the Atlantic sector (between 22°S and 66°N) are used to construct the LIMs for seasonal and decadal prediction. The forecast skills of the LIMs are then compared to that from two current operational forecast systems. Results indicate that the LIM has good forecast skill for time periods of 3–4 months on the seasonal timescale with enhanced predictability in the spring season. On decadal timescales, the impact of inter-annual and intra-annual variability on the predictability is also investigated. The results show that the suite of LIMs have forecast skill for about 3–4 years over most of the domain when we use only the decadal variability for the construction of the LIM. Including higher frequency variability helps improve the forecast skill and maintains the correlation of LIM predictions with the observed SST anomalies for longer periods. These results indicate the importance of temporal scale interactions in improving predictability on decadal timescales. Hence, LIMs can not only be used as benchmarks for estimates of statistical skill but also to isolate contributions to the forecast skills

from different timescales, spatial scales or even model components.

Keywords Sea surface temperature · Predictability · Linear inverse model · Atlantic Ocean

1 Introduction

The Earth's climate is a highly nonlinear system in which variability on different temporal and spatial scales interact (e.g., Palmer 1999). These interactions from spatial scales of tens of meters to thousands of kilometers and temporal scales of hours to years give rise to the very rich atmospheric and oceanic dynamics (Sasaki et al. 2014) with upscale and downscale turbulent transfer of energy and momentum. Previous studies have shown evidence that the large-scale, slow variability in sea surface temperature (SST) anomalies can be explained as the ocean's response to high-frequency stochastic atmospheric forcing (Frankignoul and Hasselmann 1977; Frankignoul 1985). In addition, the low frequency (multi-decadal) variability in surface temperature or in the Atlantic Meridional Overturning circulation can be modulated by the higher frequency variability in the ocean such as mesoscale and frontal dynamics (Latif et al. 2004; Thomas and Zhai 2013). Hence, studying the interactions amongst scales of variability can help improve our understanding of the role of such interactions in the predictability on different timescales.

The SST in the Atlantic sector has significant variability on seasonal to multi-decadal timescales. Atlantic SST anomalies are believed to have a large impact on the weather and climate over Europe, Eastern United States and Africa (e.g., Sutton and Hodson 2005; Meehl et al. 2009; Stockdale et al. 2011; Jung et al. 2011; O'Reilly

Electronic supplementary material The online version of this article (doi:[10.1007/s00382-016-3375-1](https://doi.org/10.1007/s00382-016-3375-1)) contains supplementary material, which is available to authorized users.

✉ Aneesh Subramanian
subramanian@atm.ox.ac.uk

¹ Downing College, University of Cambridge, Regent Street, Cambridge CB2 1DQ, UK

² Clarendon Laboratory, Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

et al. 2016). Therefore, accurate predictions of SST anomalies in the Atlantic sector can translate into substantial improvement in atmospheric weather prediction over many continental regions, which has many beneficial implications for the economies and societies (Mathieu et al. 2004; Stockdale et al. 2006; Robertson et al. 2015). However, the mechanisms that govern Atlantic SST predictability on seasonal to decadal timescales are not well understood (Stocker et al. 2013).

Operational forecasting centers around the globe spend a considerable amount of effort to improve predictions on sub-seasonal to decadal timescales. Yet, most coupled modeling systems have poor forecast skill (Stockdale et al. 2011; Meehl et al. 2013) in predicting events on these timescales such as tropical cyclones (Hodges and Emerton 2015), the MJO (Klingaman et al. 2015) and atmospheric blocking (Matsueda 2009) events. Many of these events are significantly influenced by the SST, which can help keep memory longer than a couple of weeks in the system. Skillful forecasts on seasonal timescales have been made in the Tropics (e.g., Barnston et al. 2012). However, the skill and use of seasonal forecasts in the extratropics is more questionable for events such as heat waves, blocking and extreme storms. Yet, the demand for such forecasts over the past few decades for various uses from the energy sector to water resource management is increasing and providing skillful forecasts for these events would be very useful for many applications (Robertson et al. 2015).

Much of the variability in the extratropics can be directly attributed to the internal variability and nonlinearity of the system, unlike in the Tropics where the internal chaotic variability of the large-scale flow is relatively weak. Hence, a theoretical assessment of the potential predictability on seasonal timescales in the extra tropics is harder than for the tropical large-scale flow (Palmer and Anderson 1994). Previous studies of predictability in the extra-tropics suggest that the summer season should have enhanced predictability compared to winter months due to reduced internal variability in summer (Branković et al. 1994; Palmer et al. 2004). Yet, much of the extratropical predictability originates from the tropically forced Rossby wave teleconnections. These teleconnections are strongest in the winter season due to the enhanced meridional potential vorticity gradients and hence imply that winter seasons may have enhanced predictability associated with the tropical predictability (Palmer and Anderson 1994). Recent studies show that multi-model ensemble forecast systems have varied levels of skill on seasonal timescales for different decades in the North Atlantic (Shi et al. 2015; Kumar 2009) with no evidence of ensemble overdispersion. On the other hand, results from Eade et al. (2014) claim that a single model ensemble forecasting system is over-dispersive and the real world is more predictable than predictability estimates from perfect model

experiments in this region. The potential predictability of Atlantic SST and ocean circulation on interannual to decadal timescales studied extensively using modeling experiments (e.g., Meehl et al. 2009; Jung et al. 2011) does not necessarily translate into equivalent forecast skill on these timescales. They show that the many sources of uncertainties in the forecast system, such as hindcast length, sampling size and model inaccuracies can influence the assessment of the true potential predictability in the system (Palmer and Hagedorn 2006 and references therein). Hence, being able to find the true sources of predictability in the system and improving the modeling of these components in forecast systems will lead to improvement in the forecast skill.

Current Earth system models (ESMs) being developed for seamless predictions in forecasting centers around the world are highly complex and imperfect. These models are being tested extensively for their forecast skill and are continuously being improved in their formulation. Along with these comprehensive ESMs, we also have extensive long-term observations which can be used further for improving our understanding of predictability and forecast skill on different time scales (Rodwell et al. 2013). Our main goal in this study is to use one of the longest observed records of SST in the Atlantic (Rayner et al. 2003) and study its statistical properties using an idealized stochastic-dynamic model. We further investigate issues related to predictability and forecast skill over the region and compare the stochastic-dynamic model to the state-of-the-art seasonal and decadal forecasting systems. We use similar configurations of the linear inverse model for evaluating both the seasonal and decadal forecasts to emulate the concept of a “seamless forecasting” linear inverse modeling system.

We use a linear inverse model (LIM, Penland and Magorian 1993) in a forecast framework. LIM is a reduced system of equations of a complete stochastic-dynamic model for the climate (Majda et al. 1999, 2008). LIM has been used for forecast skill evaluation of global, tropical and extratropical weather and climate (Pegion and Sardeshmukh 2011; Vimont 2012; Zanna 2012; Newman 2013). These studies have shown that LIM has a comparable forecast skill from medium-range to decadal timescales to that of global circulation models, despite their reduced order.

LIM has been shown to make skillful predictions of decadal variability in SST using annually averaged fields in the North Atlantic sector (Hawkins and Sutton 2009; Zanna 2012). We use monthly mean SST data to construct a suite of LIMs (using temporal filters) to investigate their forecast skill on seasonal and decadal timescales. We study the modal interactions on different spatio-temporal scales and how they contribute to the predictability of Atlantic SST anomalies using this framework. We conclude by comparing the forecast skill of these models to some of the climate models currently employed: decadal prediction using

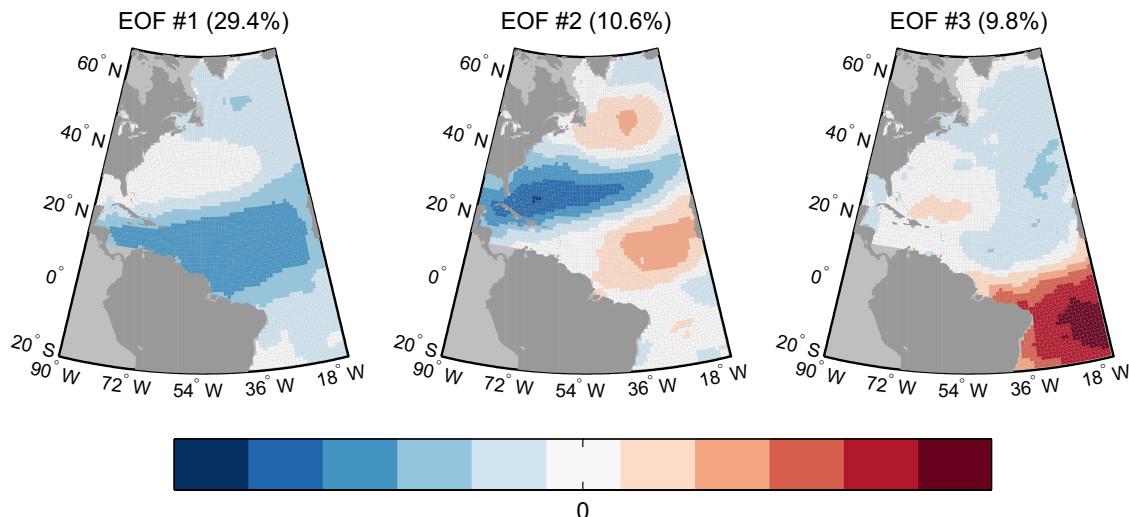


Fig. 1 First three EOFs for the monthly SST anomalies over the North Atlantic are shown. The contour intervals are equal for each of the plots. The fraction of variance associated with each mode is indicated in the parentheses

DePreSys and the ECMWF Seasonal Forecast System 4 (ECMWF S4).

This paper is organized as follows. In Sect. 2 we explain how SST anomalies were constructed from the data available and how these can be used to train our LIM. Here we also address how the SST anomaly is filtered for the construction of temporally filtered LIM. These models are tested and a comparison of their forecast skill is made with current climate models in Sect. 3. A discussion of the results as well as an evaluation of the model is contained in Sect. 4.

2 Methods and model description

2.1 Data

We use the Met Office Hadley Centre's SST interpolated dataset (HadISST1) on a 1° by 1° latitude-longitude grid from 1870 to 2015 (Rayner et al. 2003). We use an interpolated monthly SST data set to have continuous records throughout the time period of interest for making predictions on various timescales. Data collected before 1900 was discarded, due to poor data coverage and less accurate measurement in the nineteenth century (Rayner et al. 2003).

The domain considered is the North Atlantic basin between latitudes 22°S and 66°N and longitudes of 90°W and 25°W . Monthly SST anomalies (SSTA) in this region were constructed by first removing the monthly climatology at each grid-point. The data is then detrended and the time mean anomaly at each grid cell is set to zero. We define this to be our monthly SSTA, which will further be used for comparison with operational model forecasts.

We normalize the anomalies at each grid cell by their standard deviations to compare models of biased variability with the forecast fields made from observations. We then further normalize the temperature anomaly at each location by the surface area of the grid cell to weight equally tropical and extratropical one degree grid cells. The forecast fields are de-normalized prior to analysis and comparison with other model forecasts as to make the SSTA consistent with the definition above.

2.2 Linear inverse modelling

The SSTA field is decomposed into empirical orthogonal functions (EOFs) and associated principal components (PCs) that describe their spatial patterns and time evolution, respectively using singular value decomposition (SVD). The first three EOFs are shown in Fig. 1 and discussed further in the following section. The total SSTA field, $T(x, y, t)$, as a function of space and time can be decomposed as follows:

$$T(x, y, t) = \sum_i \lambda_i EOF_i(x, y) PC_i(t). \quad (1)$$

Only those EOFs/PCs contributing a significant fraction of the variance, λ_i , to the SSTA are retained, generating a reduced space representation of the data. In linear inverse modelling (LIM) we assume that the time evolution of the PC state vector, \mathbf{P} , can be written as a linear dynamical system forced with white stochastic forcing \mathcal{N} such that

$$\frac{d\mathbf{P}}{dt} = \mathbf{A}\mathbf{P} + \mathcal{N}. \quad (2)$$

The matrix \mathbf{A} , a time-independent and stable dynamics matrix (with all negative eigenvalues), describes the time evolution as estimated from linear regression (Penland and Sardeshmukh 1995) as

$$\mathbf{A} = \frac{\ln[\mathbf{C}(\tau_0)\mathbf{C}(0)^{-1}]}{\tau_0}. \quad (3)$$

We use a lead time of $\tau_0 = 1$ month, ensuring that the propagator (see later) is real, allowing month-to-month predictions to be made. The time-lagged and zero-lag covariance matrices are given by $\mathbf{C}(\tau_0) = \langle \mathbf{P}(t + \tau_0)\mathbf{P}^T(t) \rangle$ and $\mathbf{C}(0) = \langle \mathbf{P}(t)\mathbf{P}^T(t) \rangle$ respectively, where $\langle \dots \rangle$ denotes the average over t . We use 30 EOFs/PCs to determine \mathbf{A} , explaining 92% of the total variance (further properties of the EOFs and PCs are discussed in the Supplementary Material).

Evaluating the logarithm of a matrix can result in computational errors, particularly if some of the matrix elements are small. We therefore work exclusively with the exponential of \mathbf{A} , the propagator \mathbf{B} , which is given as:

$$\mathbf{B}(\tau) = \exp(\mathbf{A}\tau) = \left[\frac{\mathbf{C}(\tau_0)}{\mathbf{C}(0)} \right]^{\frac{\tau}{\tau_0}}. \quad (4)$$

We can then use the propagator to make forecasts of \mathbf{P} , given by $\hat{\mathbf{P}}(t + \tau) = \mathbf{B}(\tau)\mathbf{P}(t)$. Projection of these principal components onto the EOFs then gives anomaly predictions for the region.

We will later refer to a “perfect reconstruction” of the SSTa field when assessing the skill of forecasts made using the LIM. We reconstruct the data set using (1), including the same number of EOFs as are used in training the model. These reconstructions coincide with the forecast SSTa in the case where the model captures the dynamics of the principal components exactly.

2.3 Filtered linear inverse model

In this study we use a third order Butterworth filter to filter the SSTa time series into three different timescales: decadal with periods greater than 10 years, interannual with periods between 1 and 10 years and intraannual with periods of less than 1 year. We therefore end up with three data sets, each containing only variability within a certain frequency band. EOFs and PCs are constructed separately from each of these subsets and the EOFs corresponding to each of time scales are shown in Fig. 2. The fraction of variance contained within each timescale is estimated from the SSTa power spectrum (Fig. 3) and the fraction of this belonging to each mode (within a given timescale) is deduced from the singular value decomposition, as before.

The most significant PCs from each timescale form the PC state vector $\mathbf{P}(t) = (d_1, \dots, d_{N_d}, e_1, \dots, e_{N_e}, a_1, \dots, a_{N_a})^T$, where d, e and a label the modes with decadal, interannual and intraannual variability, respectively. The time lagged and zero lag covariance matrices corresponding to this state vector are used to determine the propagator, as described in the previous section.

This propagator is used to predict $\hat{\mathbf{P}}(t + \tau)$ from the combined PC state vector. We then isolate the components of $\hat{\mathbf{P}}(t + \tau)$ corresponding to the decadal variability and project these onto the decadal EOFs to yield predictions for the decadal component of future SSTa. Though we only consider the decadal parts of the resulting state vector, the propagator will, in general, include coupling between modes at different timescales and hence the values of decadal PCs could depend on those of interannual and intraannual PCs at earlier times. This model therefore allows us to study the influence of higher frequency variability on decadal forecasts.

2.4 Forecasts made using GCMs

The forecast fields obtained using the LIM are compared with those from two coupled GCMs for two different timescales. The shorter time scale predictions made using the unfiltered LIM are compared with forecasts made using ECMWF seasonal forecast system 4. Details for the ECMWF System 4 can be found in Molteni et al. (2011) and <http://www.ecmwf.int/products/forecasts/seasonal/documentation/system4>. The decadal component of SSTa predicted by the filtered LIM is compared with results from DePreSys. DePreSys is the UK Met Office decadal climate prediction system (Smith et al. 2007).

For the ECMWF model we have sets of 7-month long forecasts of SST (with monthly data), initialized in November or May of the years 1991–2009. There are 41 ensemble members for each initialization date; here we work solely with the ensemble mean, which is usually the best estimate of the ensemble forecast. We do not use the probabilistic information of the ensemble, as the LIM forecasts are deterministic, with no uncertainty information in the forecasts. The data has a 1.25° by 1.25° resolution so linear interpolation was used to regrid the data to allow comparison with HadISST1. Anomalies were calculated for both observations and model fields by subtracting the respective monthly climatologies.

From DePreSys we have 10-year long forecasts of SST initialized yearly in November from 1980 to 2004. The model forecast fields, originally at 1.25° by 1.25° were interpolated linearly to the same grid as HadISST for comparison. We discuss the results of forecast experiments and their implications in the following sections.

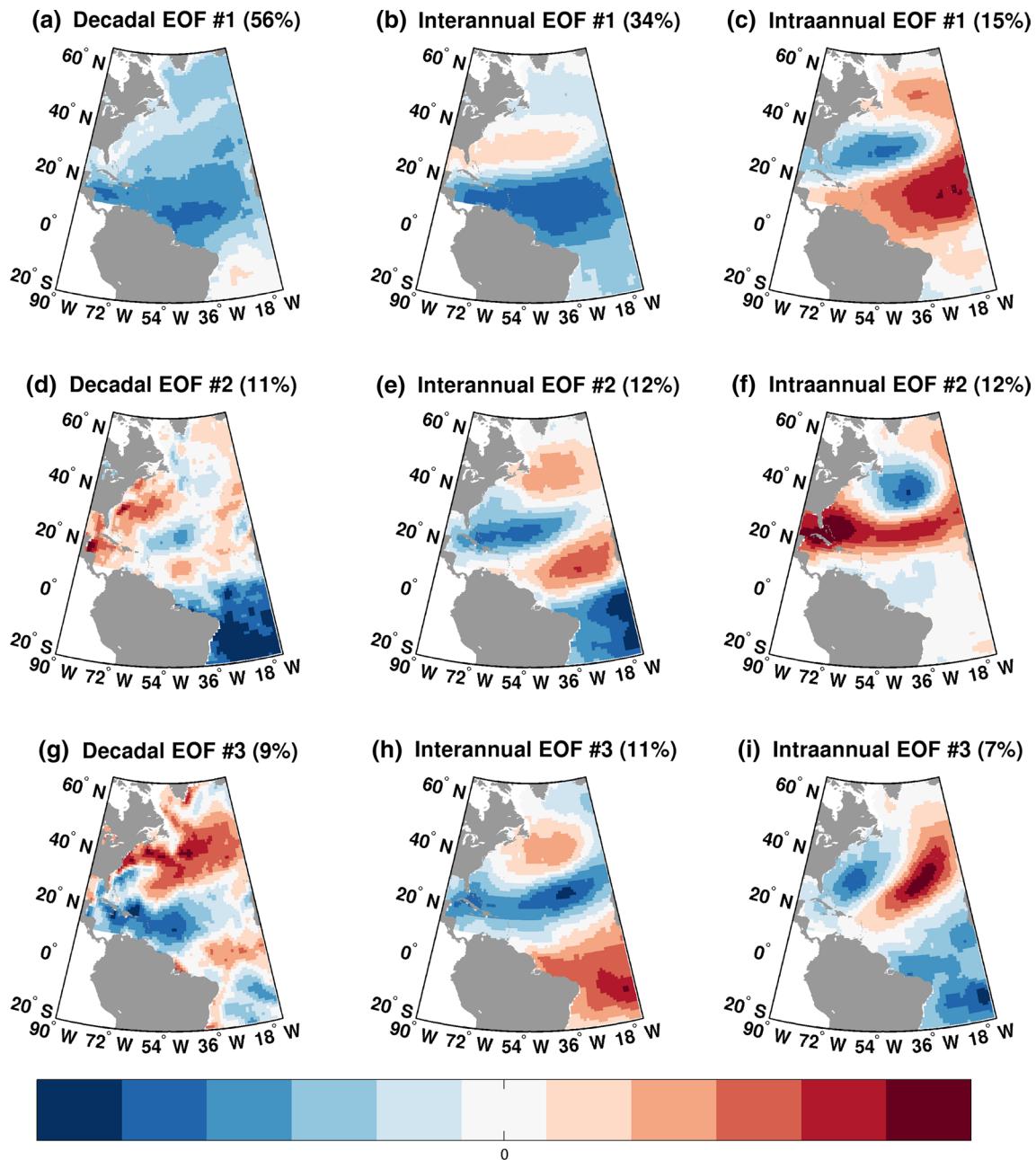


Fig. 2 The leading EOFs for each of the three time scales (decadal, interannual and intraannual). The fraction of variance represented by each EOF, within its particular timescale, is given in parenthesis

3 Results

3.1 Principal components and power spectra

The three most significant EOFs for the North Atlantic basin from 1900 to the present are shown in Fig. 1, which correspondingly explain 29.4, 10.6 and 9.8% of the total variance. These three primary modes of variability are similar in spatial pattern and fraction of variance explained to

the modes identified in other previous studies (Deser and Blackmon 1993; Marshall et al. 2001). The SSTa power spectrum (Fig. 3) was constructed using the time series from 1900 onwards for both the raw data and reconstructions of SSTa in a truncated EOF space. The reconstruction used to train and verify the unfiltered LIM includes only the 30 most significant EOFs, while the filtered LIM contains the 20 most significant EOFs corresponding to each of the three timescales.

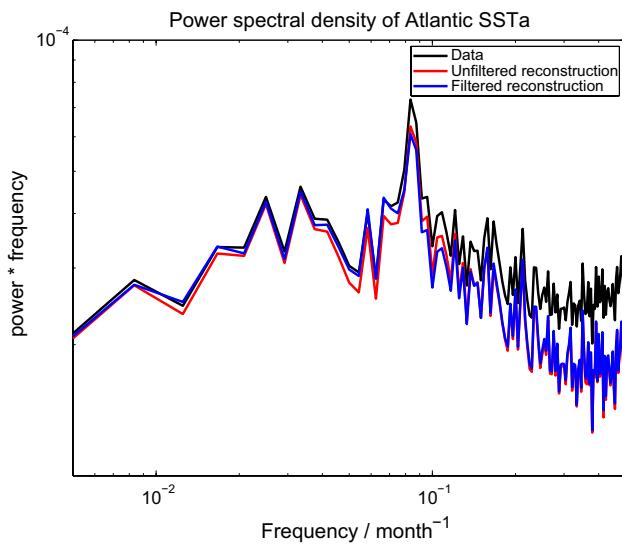


Fig. 3 The power spectrum for the raw SST fields (*black line*) and the reconstructions in a reduced EOF space used for training both the filtered (*blue*) and unfiltered LIM (*red*)

The power spectra shows the expected behaviour for SST: the slope is pink at lower frequencies with a transition to red noise spectra at higher frequencies as the power increases with period. The EOF decomposition does not split the variability into distinct frequencies, the three leading PCs have power at all frequencies, though all three have a large proportion of their power in the low frequencies (not shown). The first EOF has the largest power at low frequencies. This corresponds to a monopole oscillation of the SSTa which varies on decadal timescales. The tripolar oscillation seen in the second EOF is characteristic of the North Atlantic Oscillation imprint on the SSTs (NAO, Hurrell 1995) with power on interannual timescales. The gap in power between the raw data and the reconstructions widens at higher frequencies which means the reconstruction and therefore the LIM will not capture the variability as well as for sub-seasonal variability. Integration of the power spectral density gave the contribution to the total power from decadal, interannual and intraannual variability as 30.82, 41.44 and 27.74% respectively.

3.2 Seasonal forecasts

In order to assess the forecast skill of the model, it is useful to compare the forecasts made with those obtained from two commonly used reference values: climatology and persistence. A climatological forecast implies that the sea surface temperature evolves according to their monthly mean climatology with the anomalies assumed to be zero, $\hat{\mathbf{P}}(t + \tau) = 0$. These forecasts tend to perform better (relative to other models) at longer lead times, after which the

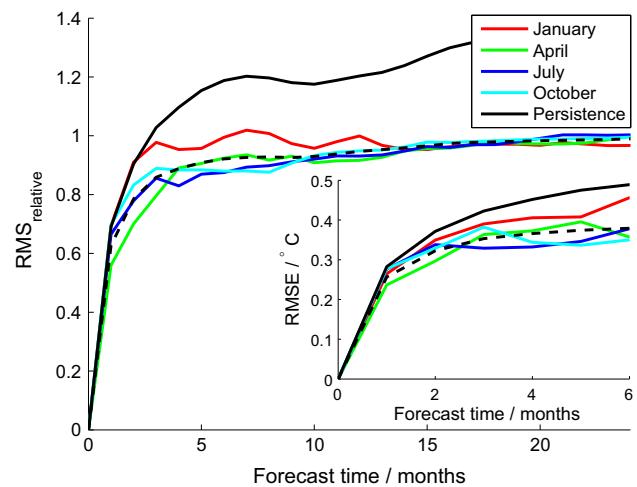


Fig. 4 RMS error relative to climatology for 2-year forecasts initialised at different months as shown in the *color legend*. The *dotted line* indicates the average RMS error over all forecasts initialised over all months in the year and the *solid black line* indicates the average RMS error in persistence forecasts. *Inset* Initial growth rate of raw RMSE

initial conditions are not as important. Persistence assumes that the SSTa keep their initial values, $\hat{\mathbf{P}}(t + \tau) = \mathbf{P}(t)$. It is generally more accurate than climatology, especially for shorter lead times or for more slowly varying fields, as in these cases the temperatures will not have deviated too far from their initial values.

In order to compare the LIM forecast skill to climatological forecasts we define a relative RMS (Newman et al. 2009) by

$$\text{RMS}_{\text{relative}} = \frac{\text{RMS}_{\text{pred}}}{\text{RMS}_{\text{clim}}}, \quad (5)$$

where the root mean square error (RMS) for the predicted field is RMS_{pred} and the RMS for the climatology over the same period is RMS_{clim} . A relative RMS of less (more) than one means the model is performing better (worse) than merely predicting climatology. The model is initialized at four different months (January, April, July and October) in each of the years from 1980 to 1989, with the LIM being trained using the data from the 80 years previous to each initialization date. EOFs and PCs are calculated for the 2-year forecast period following each initialization date. The LIM is initialized with the zero-lead time PCs and the model generates the PC time series for the remainder of the forecast period. Projecting the forecast PCs onto the corresponding EOFs gives the forecast SSTa, which are then compared to the reconstruction obtained using the true PC time series.

The ensemble average relative RMS (averaged over the basin) obtained from these 10 forecasts for each month is shown in Fig. 4. Forecasts show skill for lead times of up

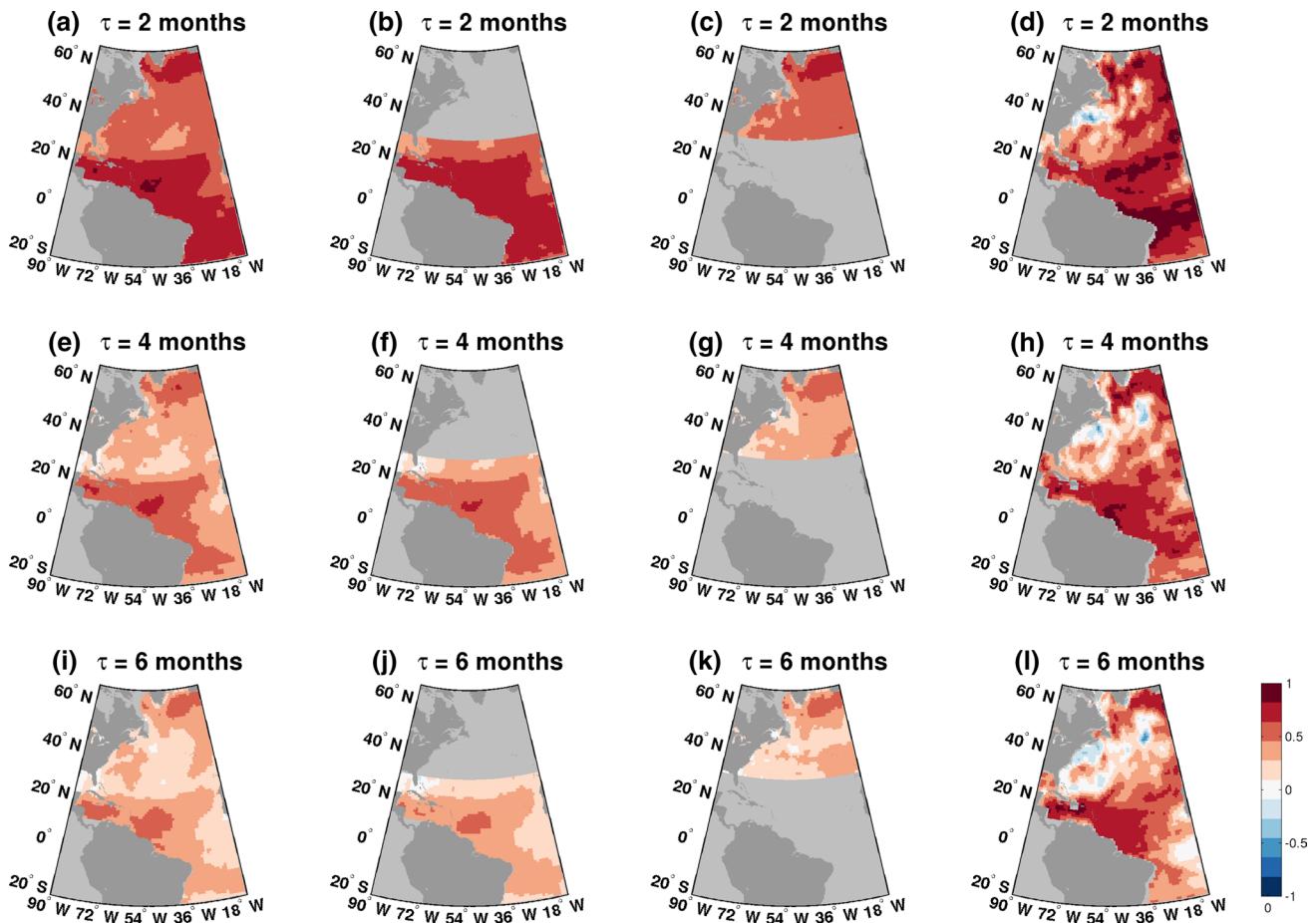


Fig. 5 Maps of temporal correlations for the SSTa time series forecasted at a given lead time. **a, e, i** Forecasts were made using LIM trained on data for the whole domain. **b, f, j** Forecasts were made using LIM trained only on the SSTa in the tropical region. **c, g,**

k Forecasts were made using LIM trained only on the SSTa in the extratropical region. **d, h, l** Forecasts were made from the ECMWF S4 seasonal forecasting system

to 5–6 months, after which the model forecast has similar skill to using climatological predictions. Predictions initialized in January or October appear to be the least skillful with errors exceeding climatology between 2 and 3 months. April forecasts have better skill and outperform climatology until about 5 months, which agrees with other GCM studies showing improved skill in summer months (Palmer et al. 2004). The forecasts later than 6 months are comparable to climatology. The model is also shown to outperform persistence at all lead times. The inset in Fig. 4 shows the raw RMS_{pred} for each of the predictions. For the unfiltered LIM the error at short lead time is largest for October forecasts. January and October forecasts have a slightly faster error growth (indicated by a steeper gradient) and have similar RMS_{pred} until 2 months after which their errors diverge. April forecasts also perform the best according to this metric, with a much flatter error growth rate.

Another metric for quantifying forecast skill is anomaly correlation. Here we construct time series from the model

output such that each time series consists of anomalies calculated at a given lead time. These time series are then compared with ones composed of anomalies calculated from the perfect reconstruction described in Sect. 2.2 (i.e., what one would obtain if the PCs evolved exactly according to LIM). The reconstructions are based on the EOF/PC decomposition of the next 12 months of SSTa following the initialization date. The model is initialized using the initial values of these PCs and the forecast PCs projected on the EOFs based on this period to allow fair comparison with the perfect reconstruction. Carrying out this test at each grid cell allows us to quantify the skill of the forecast at a given lead time at different parts of the region of interest, thereby allowing us to see the areas in which SSTa evolution is well approximated by linear dynamics.

From the first column of Fig. 5 we see that the correlation deteriorates most rapidly in mid-latitudes and the sub-polar gyre, with significant loss of correlation between latitudes 20°N and 40°N between 4 and 6 months. The

forecast skill is generally better in the tropics, though there is a patch of high correlation at the highest latitudes for a period of 4–6 months and longer. We see a band of high correlation between the equator and 20°N, lasting for up to 12 months (not shown). Moderate correlation is retained off the east coast of Brazil for up to 4 months; this area has a large contribution from EOF #3 which could explain the success of linear dynamics in the forecasts. Increased predictability in the Tropics is expected due to the smaller variability in SST on seasonal timescales and a lot of the variability in the Tropical dynamics can be expected to be related to linear waves propagating zonally across the basin, although strong air-sea interaction and coupled processes can disrupt these linear dynamics during various periods.

In addition to determining the areas in which skill is highest we also investigated possible links between the SSTa behaviour in the mid and high-latitude Atlantic with those around the tropical region. This was done by dividing the domain into two regions: the tropical region here defined as 20°S to 30°N and the extratropical region North of 30°N. The LIM was trained separately on each region by masking out the data from the rest of the domain; the 30 leading EOFs calculated from the remaining data are retained in each case. We then evaluated the skill at predicting SST in each region separately (see the second two columns in Fig. 5, panels b, c, f, g, j, k). The correlation coefficients in different regions are qualitatively similar for forecasts made using each region separately and those made using the whole domain (first column in Fig. 5); the regions with greater skill are the same in each case. However, small differences in forecast skill do arise when the regions are subset for forecasts. The forecast skill in the sub-tropical gyre tends to be better for forecasts with data from the North Atlantic only (Fig. 5k); when the tropical SST is included in the model formulation, it is likely that the large variability in the tropics could reduce the strength of North Atlantic variability captured in the EOFs over the entire domain (Fig. 5i). On the other hand, tropical forecasts that exclude the North Atlantic variability tend to be worse in some regions (Fig. 5j), which suggests that there may be some form of coupling between extratropical and tropical variability which is a source of predictability for the tropical SST anomalies. The coupling could be a manifestation of the interaction between the subpolar and subtropical gyres via Ekman transport or an atmospheric teleconnection pathway between high and low latitudes. This would need targeted GCM experiments to explore mechanisms that lead to this coupling for enhanced predictability.

As an attempt to contrast the skill achieved by this simple statistical model with an advanced forecast model, we compare the results of the LIM with the ECMWF S4 seasonal forecasts (Molteni et al. 2011). For each forecast,

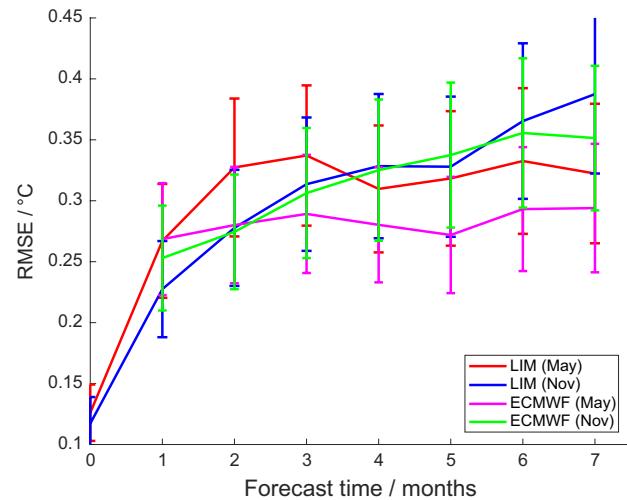


Fig. 6 Domain averaged root mean square error between the SSTa predictions made using the LIM (red, blue) and ECMWF seasonal forecast System 4 model (pink and green) and the observations from HadISST1 as a function of forecast time. The 95% confidence interval error bars are computed and plotted as *vertical bars*

the observations from HadISST corresponding to the same time period as the ECMWF S4 forecasts are removed and a LIM is trained on the remaining record. The LIM is then used to predict the SSTa over this forecast period. Repeating this over all forecasts results in a data set with the same format as the ECMWF data. We then calculate the RMS error between the two (LIM, ECMWF S4) model predictions and raw observations. The LIM is now compared to the raw SSTa rather than reconstructions in a reduced EOF space as before in order to allow a fairer comparison in skill with ECMWF S4. Note that observations are linearly interpolated in time for comparison with the ECMWF S4 results as the ECMWF model gives values centered on the start of each month whereas the monthly means in HadISST are centered on the middle of each month. The result of averaging over all grid-cells, ensembles and years resulted in Fig. 6. This shows us how the RMSE grows with lead time.

The RMSE in predictions using LIM are comparable, though slightly worse than those in forecasts made using ECMWF S4, for forecasts initialized in both May and November. We did not have access to the value of the SSTa at the time of initialization for the ECMWF forecasts. Though it is expected that there would be some initial error, as ECMWF S4 forecasts are initialized from a coupled reanalysis state different from HadISST. For the LIM forecasts, the non-zero error at zero lead time is due to errors in initial conditions resulting from truncation of the number of EOFs/PCs. The error growth rate in ECMWF S4 is slightly slower than for the LIM forecasts. For both models, May forecasts exhibit a faster initial error growth, with the RMSE levelling off after around two months; November

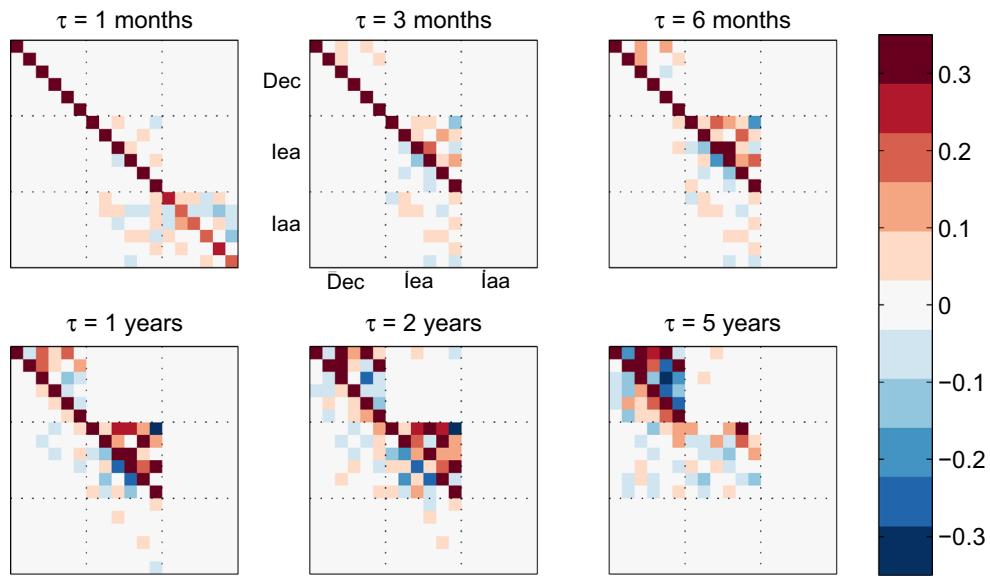


Fig. 7 The propagator matrix constructed from the six leading modes of each time scale, at various lead times. The *dotted lines* indicate the matrix boundaries for the different blocks of modes for the three timescales. The block diagonal (6×6) elements are the modes of one timescale interacting amongst themselves while the off-diagonal

block elements are modes from two timescales interacting with each other. The legend ‘Dec’ indicates the row/column in each matrix corresponding to the modes of decadal timescales. The ‘Iea’ legend indicates the same for inter-annual timescales and the ‘Iaa’ indicates the modes of intra-annual variability

forecasts show a steadier error growth. In the May initializations we see a dip in the error in both models at lead times of about 4 months (corresponding to September). However, it was found that this is a property of the RMS of SSTa itself rather than a return of skill due to some kind of re-emergence phenomena.

Figure 5d, h, l shows the temporal correlation at each grid point between the ECMWF S4 seasonal forecast at lead time τ for each panel. Contrasting the rate at which correlation reduces as a function of lead time between the ECMWF S4 and that of the LIM, we note that the ECMWF S4 maintains a higher correlation in the subpolar gyre region for a longer time, while the correlation decay in the Tropics is at a similar rate for both the LIM and the ECMWF S4. The temporal correlation in the Gulf Stream region falls off faster in the ECMWF S4, as likely the Gulf Stream dynamics is poorly resolved in the ocean component of the S4 system (Molteni et al. 2011; Balmaseda et al. 2013). The LIM lacks any nonlinear feedbacks in terms of eddy-mean flow interactions and hence, the lack of such dynamics helps maintain the predictability in this region for longer than a misrepresentation of such dynamics as in ECMWF S4.

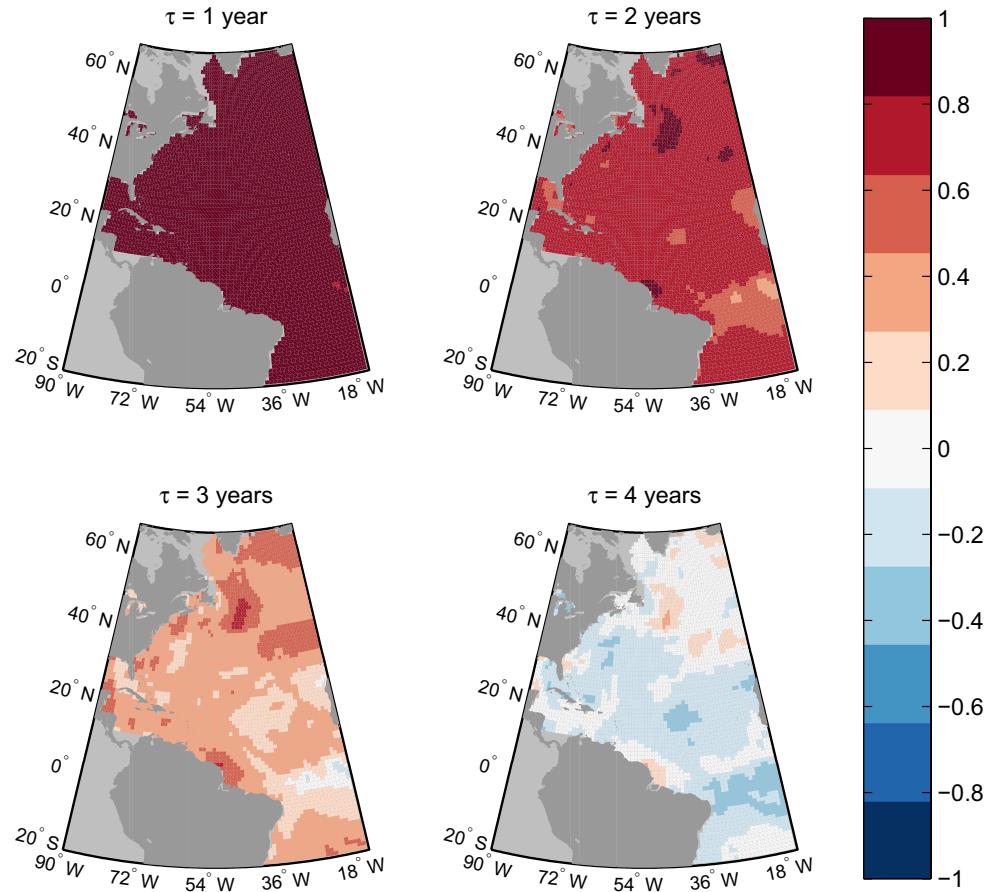
3.3 Decadal forecasts

We now use the filtering outlined in Sect. 2 to explore the predictability of the decadal component of the SSTa. We take the number of modes for each to be

$N_d = N_e = N_a = 6$. The propagator constructed using these PCs (as in Eq. 1) is shown in Fig. 7, for various values of τ . The interaction between decadal modes is as expected more significant on longer time scales. Coupling between intraannual modes is significant at lead times of one month, but this variability decays extremely quickly. Dependence of intraannual modes on interannual variability is seen at short lead times but is no longer significant beyond 3 months. Two-way coupling between interannual and decadal modes increases with lead time up to about 1–2 years, after which this coupling begins to diminish. The sub-matrix describing the evolution of interannual modes becomes less diagonal over time with the magnitudes of the off-diagonal elements increases for times up to 6 months. For lead times of more than a couple of years, the evolution of the interannual variability depends almost entirely on the decadal part of the signal. Significant coupling between intraannual and decadal modes is not present at any lead times.

In order to facilitate multiple forecasts of several years in length we changed the way in which we divide the data into training and verification periods as compared to the seasonal forecasting experiment. We now remove 10 years worth of data from the time series and train the model on the remaining data. The model output is then compared to a reconstruction of the 10-year verification period using the same number of EOFs/PCs as the model. This method results in small errors when computing the elements of the time-lag correlation matrix at the point where the data from

Fig. 8 Temporal correlation over the region for a range of lead times for the decadal component of SSTa for the forecast fields with HadISST1 field as the verification. The LIM was constructed using only the decadal component of the signal after filtering the raw time series



either side of the removed period meet. However, as the lag chosen is only one month, this only affects two elements from a very long time series. It has the advantage that the model is trained using a large amount of the time series so the EOFs/PCs used in the model should very closely resemble those presented earlier (which were calculated for the entire time series). This may be helpful when trying to justify features of the predictability in terms of the shapes of the EOFs.

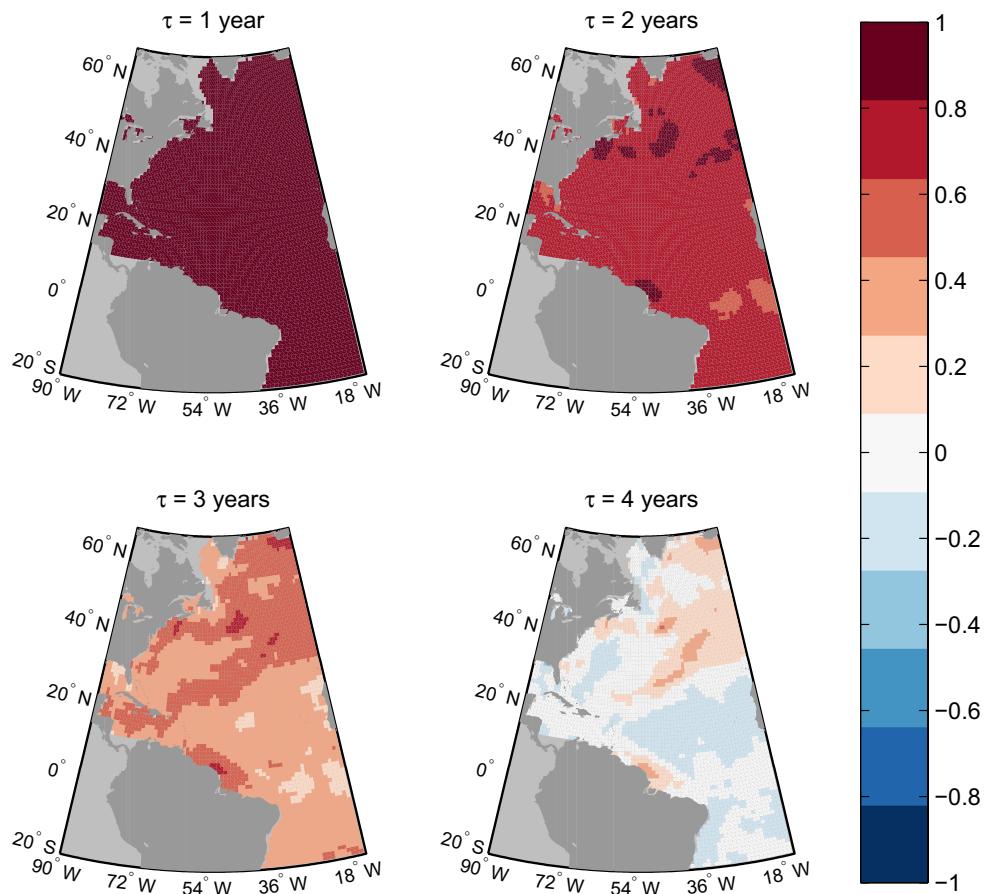
As a first test we build a propagator using only the PCs describing evolution on decadal timescales; $N_d = 20$ components were used. Forecasts were initialized at subsequent months from Jan 1951–Jan 2001, resulting in 600 forecasts. The maps of temporal correlation at a range of lead times can be seen in Fig. 8. Forecast skill is lost between 3 and 4 years in most of the domain where the correlation drops below 0.2. Differences in skill between different regions are not as sharp; skill seems to be more even across the basin as compared to the seasonal forecasting skill. This is likely to be due to the shape of the EOFs for the decadal variability. The leading decadal EOF (Fig. 2a) is close to uniform across the basin and contributes a very large fraction of the variance (56%). The skill in the sub-polar gyre is comparable to the rest of the basin and forecasts in this

region now tend to be the most skillful (whereas on seasonal timescales the predictability was the worst in this region). This suggests that the non-linear dynamics which proved problematic for LIM on monthly timescales are not seen in the decadal component of the variability and that the decadal timescale variability is more linear than on seasonal timescales.

We now properly utilise the filtered LIM to test if knowledge of the shorter timescale variability can improve the forecasting of the decadal variability. To do this we initialize the filtered LIM with $N_d = N_e = N_a = 20$ and a PC state vector containing PCs for each of the time scales. Only the PCs corresponding to the decadal part of the signal from the LIM forecast fields are retained and these are projected onto the decadal EOFs to give the forecast field. As seen in Fig. 7 the decadal PCs interact with the interannual and the intraannual, so these decadal PCs will vary slightly from those calculated using the decadal modes only. The temporal correlations at each lag can be seen in Fig. 9.

Including the higher frequency components in this manner improves the forecasts everywhere in the region of interest. Areas that have improved the most notably are along the equator where the region of negative correlation

Fig. 9 Temporal correlation over the region for a range of lead times for the decadal component of SSTa. The LIM was constructed using both the decadal and the higher frequency components



at $\tau = 3$ years is no longer seen and in the sub-tropical gyre where modest correlation is now seen even at leads of 4 years. We have performed further experiments to isolate the impact of including interannual and intraannual variability separately and find that the interannual variability adds to the predictability of the model over having just the decadal variability signal. Adding only the intraannual variability to the model with decadal variability does not change the forecast skill of the model (not shown). This can also be inferred by noticing that the decadal modes mostly interact with the interannual modes in the LIM dynamics matrix. This result is pleasing as it demonstrates a link between short-term weather and long-term climate and hence, supports efforts to build unified seamless prediction systems for all scales.

We conclude by comparing these predictions of the decadal component of SSTa with the decadal predictions made using DePreSys. The metric used here is the predictability of the domain averaged SST, with a 3-year running mean. This is a similar quantity to the AMO index, but is more favourable for the DePreSys model in two regards. The 10-year running mean used in the calculation of the AMO index means this quantity cannot be calculated for predictions extending only to 10-year lead times as in DePreSys.

The restriction of the domain to the North Atlantic means that we would not be measuring how well the tropical SSTa were predicted. The shorter running mean used in our index means that we can report the value of this index at lead times of up to 8 years. The RMSE between this index as calculated by each of the models and observations can be seen in Fig. 10. The RMSE is quite flat (particularly in the case of DePreSys) and is qualitatively similar (in both magnitude and shape) to the predictability of the AMO index reported by Newman (2013) with regards to similar models.

4 Discussion

In this study, the hindcast skill for a suite of linear inverse models (LIMs) on both seasonal and decadal timescales has been presented and compared with two current operational forecast systems. Our results reveal that the LIM has good forecast skill for time periods of 3–5 months on the seasonal timescale, with increased predictability in the Spring season compared to the other seasons. On the decadal timescale, the LIM has skill for about 3–4 years over most of the North Atlantic domain as found in Zanna (2012) and Newman (2013).

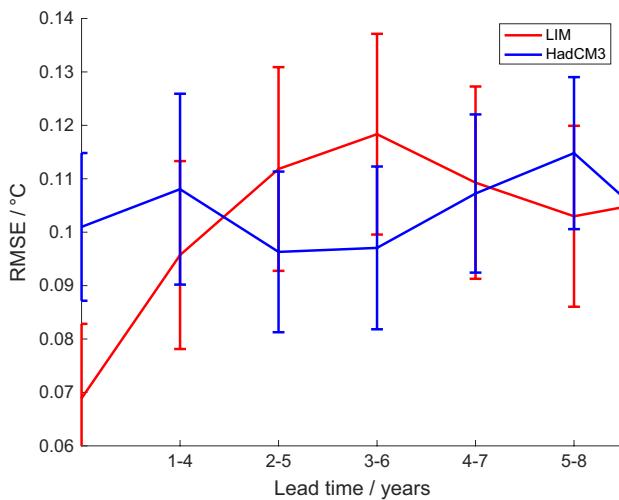


Fig. 10 RMSE between model and observations of the index defined as domain averaged SSTa, averaged over 3-year periods. The 95% confidence interval *error bars* are computed and plotted as *vertical bars*

We also evaluate the forecast skill for a LIM formulated by using the leading modes of variability from filtered time series to study the time scale interactions of SST components on longer term predictability. A linear inverse model configured with only the decadal leading modes of variability has good forecast skill up to 4 years in most of the North Atlantic domain. Including variability at frequencies higher than decadal scales helps improve the LIM forecast skill especially in the North Atlantic sub-polar gyre region, which indicates that modal interactions of higher frequencies with the low frequency variability is a source of longer term predictability in this region.

There is enhanced predictability on seasonal timescales over the tropical oceans and subtropical gyre region compared to the sub-polar gyre region for the LIM forecasts. On decadal timescales, the sub-polar gyre region tends to show increased predictability over the sub-tropical gyre region using the LIM.

The LIM forecast skill on both the seasonal and decadal timescales were similar to the global operational forecast systems ECMWF S4 and DePreSys respectively. This suggests that statistical models such as LIM are a useful tool to not only add to our suite of forecast methods but can also provide valuable information about regional predictability and variability. However, we note that they cannot be used as an alternative to global operational forecast systems as they are of less value to perform ensemble forecasts and evaluate probabilistic forecasts due to their inherent linear decay nature. Also, the value of operational ensemble forecasting systems are the probabilistic forecasts they provide, along with uncertainty estimates on the forecasts. This is one of the distinct strengths of a fully nonlinear ensemble

forecast system, which is not possible to reproduce in a LIM with an inherent decay timescale.

The LIM forecasts provide a more skillful baseline than using persistence or climatology to evaluate GCM forecasts. The knowledge gained from using LIMs can then be transferred to evaluate what modes are under represented in models that can be improved for extracting predictability. Furthermore, we encourage the use of suite of models as presented here, including multiple datasets and multiple timescales, to emulate the concept of a “seamless forecasting linear inverse modeling system.

Acknowledgements We gratefully acknowledge funding from the Met Office Academic Partnership program for supporting BH. LZ was partially supported by NERC Grant NE/K013548/1. TP and AS thank the ERC Grant 291406 and AS thanks NSF support under Grant OCE-1419306.

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

References

- Balmaseda MA, Mogensen K, Weaver AT (2013) Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q J R Meteorol Soc* 139(674):1132–1161
- Barnston A, Tippett M, L'Heureux M, Li S, DeWitt D (2012) Skill of real-time seasonal ENSO model predictions during 2002–2011: is our capability increasing? *Bull Am Meteorol Soc* 93:631–651
- Branković Ć, Palmer TN, Ferranti L (1994) Predictability of seasonal atmospheric variations. *J Clim* 7(2):217–237
- Deser C, Blackmon ML (1993) Surface climate variations over the North Atlantic Ocean during winter: 1900–1989. *J Clim* 6(9):1743–1753
- Eade R, Smith D, Scaife A, Wallace E, Dunstone N, Hermanson L, Robinson N (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys Res Lett* 41(15):5620–5628
- Frankignoul C (1985) Sea-surface temperature anomalies, planetary-waves, and air-sea feedback in the middle latitudes. *Rev Geophys* 23(4):357–390
- Frankignoul C, Hasselmann K (1977) Stochastic climate models, part II. Application to sea-surface temperature anomalies and thermocline variability. *Tellus* 29(4):289–305
- Hawkins E, Sutton R (2009) Decadal predictability of the Atlantic Ocean in a coupled GCM: forecast skill and optimal perturbations using linear inverse modeling. *J Clim* 22(14):3960–3978
- Hodges K, Emerton R (2015) The prediction of northern hemisphere tropical cyclone extended life cycles by the ECMWF ensemble and deterministic prediction systems. Part I: tropical cyclone stage*. *Mon Weather Rev* 143(12):5091–5114
- Hurrell JW (1995) Decadal trends in the North-Atlantic oscillation—regional temperatures and precipitation. *Science* 269(5224):676–679
- Jung T, Miller M, Palmer T, Towers P, Wedi N, Achuthavarier D, Adams J, Altshuler E, Cash B, Kinter J, Marx L, Stan C, Hodges

- K (2011) High-resolution global climate simulations with the ECMWF model in Project Athena: experimental design, model climate and seasonal forecast skill. *J Clim.* doi:[10.1175/JCLI-D-11-00265.1](https://doi.org/10.1175/JCLI-D-11-00265.1)
- Klingaman NP, Woolnough SJ, Jiang X, Waliser D, Xavier PK, Petch J, Caian M, Hannay C, Kim D, Ma H, Merryfield WJ, Miyakawa T, Pritchard M, Ridout JA, Roehrig R, Shindo E, Vitart F, Wang H, Cavanaugh NR, Mapes BE, Shelly A, Zhang GJ (2015) Vertical structure and physical processes of the Madden-Julian oscillation: linking hindcast fidelity to simulated diabatic heating and moistening. *J Geophys Res* 120(10):4690–4717
- Kumar A (2009) Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Mon Weather Rev* 137(8):2622–2631
- Latif M, Roeckner E, Botzet M, Esch M, Haak H, Hagemann S, Jungclaus J, Legutke S, Marsland S, Mikolajewicz U, Mitchell J (2004) Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with sea surface temperature. *J Clim* 17(7):1605–1614
- Majda AJ, Timofeyev I, Vanden-Eijnden E (1999) Models for stochastic climate prediction. *Proc Natl Acad Sci USA* 96(26):14687–14691
- Majda AJ, Franzke C, Khouider B (2008) An applied mathematics perspective on stochastic modelling for climate. *Phil Trans R Soc* 366(1875):2427–2453
- Marshall J, Kushnir Y, Battisti D, Chang P, Czaja A, Dickson R, Hurrell J, McCartney M, Saravanan R, Visbeck M (2001) North Atlantic climate variability: phenomena, impacts and mechanisms. *Int J Climateatol* 21(15):1863–1898. doi:[10.1002/joc.693](https://doi.org/10.1002/joc.693)
- Mathieu PP, Sutton RT, Dong B, Collins M (2004) Predictability of winter climate over the North Atlantic European Region during ENSO Events. *J Clim* 17(10):1953–1974
- Matsueda M (2009) Blocking predictability in operational medium-range ensemble forecasts. *SOLA* 5:113–116
- Meehl GA, Goddard L, Murphy J, Stouffer RJ, Boer G, Danabasoglu G, Dixon K, Giorgetta MA, Greene AM, Hawkins E, Hegerl G, Karoly D, Keenlyside N, Kimoto M, Kirtman B, Navarra A, Pulwarty R, Smith D, Stammer D, Stockdale T (2009) Decadal prediction: can it be skillful? *Bull Am Meteorol Soc* 90(10):1467–1485
- Meehl GA, Goddard L, Boer G, Burgman R, Branstator G, Cassou C, Corti S, Danabasoglu G, Doblas-Reyes F, Hawkins E, Karspeck A, Kimoto M, Kumar A, Matei D, Mignot J, Msadek R, Pohlmann H, Riener M, Rosati T, Schneider E, Smith D, Sutton R, Teng H, van Oldenborgh GJ, Vecchi G, Yeager S (2013) Decadal climate prediction: an update from the trenches. *Bull Am Meteorol Soc* 95:243–267
- Molteni F, Stockdale TN, Balmaseda M, Balsamo G, Buizza R, Ferranti L, Magnusson L, Mogensen K, Palmer T, Vitart F (2011) The new ECMWF seasonal forecast system (System 4). Technical report, ECMWF, Sheffield Road
- Newman M, Sardeshmukh PD, Penland C (2009) How important is air-sea coupling in ENSO and MJO evolution? *J Clim* 22(11):2958–2977
- Newman M (2013) An empirical benchmark for decadal forecasts of global surface temperature anomalies. *J Clim* 26(14):5260–5269
- O'Reilly CH, Huber M, Woollings T, Zanna L (2016) The signature of low frequency oceanic forcing in the atlantic multidecadal oscillation. *Geophys Rev Lett.* doi:[10.1002/2016GL067925](https://doi.org/10.1002/2016GL067925)
- Palmer TN (1999) Climate change from a nonlinear dynamical perspective. *J Clim* 12:575–591
- Palmer TN, Anderson D (1994) The prospects for seasonal forecasting—a review paper. *Q J R Meteorol Soc* 120(518):755–793
- Palmer TN, Alessandri A, Andersen U, Cantelaube P, Davey M, Delecluse P, Déqué M, Diez E, Doblas-Reyes FJ, Feddersen H, Graham R, Gualdi S, Gueremy JF, Hagedorn R, Hoshen M, Keenlyside N, Latif M, Lazar A, Maisonnave E, Marletto V, Morse AP, Orfila B, Rogel P, Terres JM, Thomson MC (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMIETER). *Bull Am Meteorol Soc* 85(6):853–872
- Palmer T, Hagedorn R (2006) Predictability of weather and climate. Cambridge University Press, Cambridge
- Pegion K, Sardeshmukh PD (2011) Prospects for improving subseasonal predictions. *Mon Weather Rev* 139(11):3648–3666
- Penland C, Magorian T (1993) Prediction of nino-3 sea-surface temperatures using linear inverse modeling. *J Clim* 6(6):1067–1076
- Penland C, Sardeshmukh PD (1995) The optimal growth of tropical sea surface temperature anomalies. *J Clim* 8:1999–2024. doi:[10.1175/1520-0442\(1995\)008<1999:TOGOTS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2)
- Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res* 108(D14):2156–2202
- Robertson AW, Kumar A, Pena M, Vitart F (2015) Improving and promoting subseasonal to seasonal prediction. *Bull Am Meteorol Soc* 96:ES49–ES53
- Rodwell MJ, Magnusson L, Bauer P, Bechtold P, Bonavita M, Cardinali C, Diamantakis M, Earnshaw P, Garcia-Mendez A, Isaksen L, Kaellen E, Klocke D, Lopez P, McNally T, Persson A, Prates F, Wedi N (2013) Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull Am Meteorol Soc* 94(9):1393–1405
- Sasaki H, Klein P, Qiu B, Sasai Y (2014) Impact of oceanic-scale interactions on the seasonal modulation of ocean dynamics by the atmosphere. *Nat Commun* 5:5636
- Shi W, Schaller N, MacLeod D, Palmer TN, Weisheimer A (2015) Impact of hindcast length on estimates of seasonal climate predictability. *Geophys Res Lett* 42(5):1554–1559
- Smith DM, Cusack S, Colman AW, Folland CK, Harris GR, Murphy JM (2007) Improved surface temperature prediction for the coming decade from a global climate model. *Science* 317(5839):796–799
- Stockdale TN, Balmaseda MA, Vidard A (2006) Tropical Atlantic SST prediction with coupled ocean-atmosphere GCMs. *J Clim* 19(23):6047–6061
- Stockdale TN, Anderson DLT, Balmaseda MA, Doblas-Reyes F, Ferranti L, Mogensen K, Palmer TN, Molteni F, Vitart F (2011) ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Clim Dyn* 37:455–471. doi:[10.1007/s00382-010-0947-3](https://doi.org/10.1007/s00382-010-0947-3)
- Stocker T, Qin D, Plattner G, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex B, Midgley B (eds) (2013) IPCC, 2013: 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, UK; New York
- Sutton RW, Hodson DLR (2005) Atlantic Ocean forcing of North American and European summer climate. *Science* 309:115–118
- Thomas MD, Zhai X (2013) Eddy-induced variability of the meridional overturning circulation in a model of the North Atlantic. *Geophys Res Lett* 40(11):2742–2747
- Vimont DJ (2012) Analysis of the Atlantic meridional mode using linear inverse modeling: seasonality and regional influences. *J Clim* 25:1194–1212. doi:[10.1175/JCLI-D-11-00012.1](https://doi.org/10.1175/JCLI-D-11-00012.1)
- Zanna L (2012) Forecast skill and predictability of observed Atlantic sea surface temperatures. *J Clim* 25(14):5047–5056