Investigating the predictability of North Atlantic sea surface height

Robert Fraser[1] · Matthew Palmer[2] · Christopher Roberts[3] · Chris Wilson[4] · Dan Copsey[2] · Laure Zanna[1]

Abstract

Interannual sea surface height (SSH) forecasts are subject to several sources of uncertainty. Methods relying on statistical forecasts have proven useful in assessing predictability and associated uncertainty due to both initial conditions and boundary conditions. In this study, the interannual predictability of SSH dynamics in the North Atlantic is investigated using the output from a 150 year long control simulation based on HadGEM3, a coupled climate model at eddy-permitting resolution. Linear inverse modeling (LIM) is used to create a statistical model for the evolution of monthly-mean SSH anomalies. The forecasts based on the LIM model demonstrate skill on interannanual timescales $\mathcal{O}(1–2\ \text{years})$. Forecast skill is found to be largest in both the subtropical and subpolar gyres, with decreased skill in the Gulf Stream extension region. The SSH initial conditions involving a tripolar anomaly off Cape Hatteras lead to a maximum growth in SSH about 20 months later. At this time, there is a meridional shift in the 0 m-SSH contour on the order of 0.5°–1.5°-latitude, coupled with a change in SSH along the US East Coast. To complement the LIM-based study, interannual SSH predictability is also quantified using the system's average predictability time (APT). The APT analysis extracted large-scale SSH patterns which displayed predictability on timescales longer than 2 years. These patterns are responsible for changes in SSH on the order of 10 cm along the US East Coast, driven by variations in Ekman velocity. Our results shed light on the timescales of SSH predictability in the North Atlantic. In addition, the diagnosed optimal initial conditions and predictable patterns could improve interannual forecasts of the Gulf Stream's characteristics and coastal SSH.

1 Introduction

Forecasts of sea surface height (SSH) on interannual time-scales are affected by several sources of uncertainty. Assessing such uncertainty and associated mechanisms are crucial to provide reliable forecasts and potentially mitigate the

effects of coastal flooding in certain regions. Furthermore, this can have implications for devising strategies on how best to design ocean observing systems and initialise climate models (Zanna et al. 2018). Finally, more skilful SSH predictions could help improve predictions of other parts of the climate system, for example increased skill in predicting the Gulf Stream's meridional position could impact predictions of air–sea heat fluxes and atmospheric blocking (Scaife et al. 2011).

Increased understanding of the ocean-atmosphere processes which modulate interannual sea level variability could aid in improving SSH forecasts. Cabanes et al. (2006) found that the mechanisms which contribute to observed interannual SSH variability are regionally dependent, with the majority of the interannual variability controlled by both the local steric response to heat fluxes, and the large-scale oceanic adjustments to variations in the wind stress. In modelling studies (e.g., Roberts et al. 2016), the SSH interannual variability in subpolar gyre has been found to be mostly

✉ Robert Fraser
robbiefraser70@hotmail.com

1 Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford, UK

2 Hadley Centre for Climate Change, Met Office, Exeter, UK

3 European Centre for Medium-Range Weather Forecasts, Reading, UK

4 National Oceanography Centre, Liverpool, UK

buoyancy driven through variations in both thermosteric and halosteric SSH components. Whereas, in the subtropical gyre, variability is primarily driven by the momentum forcing and variations in the thermosteric SSH component (e.g., Roberts et al. 2016). In the Gulf Stream region, intrinsic ocean processes (which are defined as ocean processes generated in the absence of atmospheric forcing variability) are responsible for the majority of the SSH interannual variability (Penduff et al. 2011). Such intrinsic ocean processes include mesoscale eddies.

Several studies have investigated seasonal to interannual SSH predictability (Chowdhury et al. 2007; Wang et al. 2013; Qiu et al. 2014) with some key mechanisms identified in General circulation models (GCMs) (e.g., Roberts et al. 2016). Nonaka et al. (2016) and Roberts et al. (2016) focused on interannual SSH variability in eddy-resolving and eddy-permitting ocean models, respectively. Nonaka et al. (2016) investigated the predictability of mid-latitude ocean currents, using an ensemble of eddy-resolving ocean GCM (OGCM) experiments with horizontal resolution of ∼ 1/10°. They find a lack of predictability in the jet regions due to the contribution from the mesoscale eddy field. Roberts et al. (2016) examined interannual SSH predictability using the eddy-permitting (∼ 1/4°) Hadley Centre Global Environment Model version 3 (HadGEM3) and found predictive skill in the Tropics on timescales of several years with a lack of skill in jet regions. Ensemble forecasts of SSH did not exhibit skill on 2–5 years that could beat persistence forecasts.

Other predictability studies, have used low resolution OGCMs (1°–2°) to investigate dynamic and steric SSH predictability (Schneider and Griffies 1999; Miles et al. 2014; Polkova et al. 2015). Polkova et al. (2015) found predictive skill in interannual steric SSH predictions in the subtropics on timescales of 2–5 years. Such skill was related to adjustments due to baroclinic Rossby waves. Skill was also found in the North Atlantic subpolar gyre on timescales of 2–5 years, which was related to changes in spiciness along isopycnals. Schneider and Griffies (1999) found SSH predictability in the North Atlantic on times of up to 17 years, using an ensemble of coupled climate model runs, related to the large-scale ocean circulation.

The work presented here focuses on the predictability of dynamic sea level, i.e. variations which arise from ocean processes. These variations are linked to changes in ocean circulation, large scale heat transports, position of gyre boundaries and thus air–sea interactions.

The aims of this study are to

– Establish the timescales of predictability due to internal variability of SSH in the North Atlantic;
– Identify the regions where forecasts are most sensitive to perturbations in the initial conditions;
– Evaluate the spatio-temporal characteristics relevant to assessing North Atlantic SSH predictability.
– Relate any internally generated predictability to large-scale ocean characteristics, with an emphasis on both mid-latitude jets and the gyre circulations;

We focus our analysis on statistical methods, which are used to evaluate predictability generated via internal variability in a fully coupled climate model. We will make use of an extended model control run, without interannual variations in the model's boundary conditions. This should enable us to isolate any interannual predictability related to the model's internal variability. A perfect model approach and a combination of Linear Inverse Modeling (LIM), non-normal mode analysis, and average predictability time (APT) are used to evaluate how initial conditions influence error growth (Penland 1989; Farrell and Ioannou 1996; DelSole and Tippett 2009a).

This paper is organised as follows, Sect. 2 details the model set up and the interannual SSH variability present. Section 3 contains information on the statistical methods used to evaluate predictability and an investigation into the influence of eddy-mean flow interactions on forecast skill. Section 4 examines the predictability related to the initial conditions of the ocean model through both the optimal initial conditions of SSH and the predictable components diagnosed by evaluation of the APT of the system. The final section contains a discussion of the results.

2 Characterising interannual sea surface height variability in the North Atlantic in HadGEM3

2.1 HadGEM3 model description

We use the output from a 150-year free-running control simulation of a coupled climate model, HadGEM3 GC2.0 (Williams et al. 2015). This control simulation has repeated-year radiative forcings (e.g., aerosols and greenhouse gases) taken from the year 2000 (identical to experiment 2 in the Coupled Model Inter-comparison Project Phase 3, CMIP3). The ocean component of HadGEM3, GO5.0 is based on version 3.4 of NEMO (Nucleus for European Modelling of the Ocean) and is described in detail in Megann et al. (2014). The model is on an ORCA025 horizontal grid, which uses an eddy-permitting 1/4° horizontal resolution and 75 vertical levels, with the vertical level spacings increasing from 1 m at the surface to 200 m at depth. The vertical level spacing provides high resolution near the surface for short to mid-range forecasting. It uses both a linear free surface and an energy conserving momentum advection scheme. The vertical mixing of tracers and momentum is parameterised by a

turbulent closure scheme (Gaspar et al. 1990; Madec 2008). The horizontal viscosity used is bi-Laplacian and the bottom friction is quadratic (Megann et al. 2014).

Monthly mean SSH anomalies are defined as departures from a time-mean, constructed from the entire model run. These anomalies are then used to create the statistical forecast models in Sects. 3 and 4. In addition, monthly mean fields of surface net heat fluxes and both the zonal and meridional wind stress are used in Sect. 4.4 in the analysis of the mechanisms responsible for the predictable components. All the fields are linearly-detrended and have their seasonal cycles removed. Two independent methods were trialed to remove the seasonal variability. In the first method, the output seasonal cycle is by removed through Fourier-filtering, by fitting cosine and sine waves to the annual and semiannual harmonics and removing variability at these frequencies. The second method deseasoned the model output by subtracting the monthly climatology from each month. The results are insensitive to the method used and so the second method is chosen for simplicity.

2.2 Evaluation of HadGEM3 against observational references

Previous model verification experiments by Williams et al. (2015) have shown that there are some small model biases in SST, primarily located just to the south of Greenland. However, HadGEM3 fields are significantly improved relative to its predecessor HadGEM2-ES. HadGEM3 has a more accurate Gulf Stream path, which which leads to improved atmospheric blocking statistics for the UK and Europe (Scaife et al. 2011; Williams et al. 2015).

The time-mean SSH and its variability in the HadGEM3 model are compared to observations to ensure the model demonstrates a sufficient level of realism. An AVISO mean dynamic topography estimate for the period 1993–2012 (Rio et al. 2014), is used to evaluate the HadGEM3 time-mean SSH for the full model simulation. A uniform offset is applied to the AVISO data to give a similar domain average to the HadGEM3 simulation. These mean SSH profiles of the observation-based data and the model output are shown in Fig. 1a, b. There is a reasonable agreement between the model and the observations in both the magnitude and spatial pattern of the mean field. However, there are some slight discrepancies between the two: the observations have more prominent recirculation gyres flanking the Gulf Stream near to its detachment point; and the mean SSH is slightly lower in the west of the subpolar gyre in the model.

The observational estimate of sea level variability comes from the the European Space Agency (ESA) Climate Change Initiative (CCI) version 2.0 monthly gridded fields of surface height anomaly (Legeais et al. 2018). Monthly data for the period 1993–2012 are aggregated into annual mean values

before computing the standard deviation for each grid box. For comparison, we select the last 20-years of the HadGEM3 simulation (representative of the simulation as a whole) to compute the standard deviation of annual mean values. The standard deviation of the interannual SSH anomalies of both the observations and the HadGEM3 output are shown in Fig. 1c, d. The patterns are in good agreement both in magnitude and spatially. In the observations there is a higher amount of interannual variability located at the Gulf Stream's detachment point (5 cm), this is unsurprising as the 1/4° resolution model will be missing some variability due to eddy-mean flow interactions in the most turbulent regions.

2.3 Interannual sea surface height variability

Figure 2a shows the time-mean SSH of the control run. The characteristic double-gyre structure is evident, and the strong SSH gradient is indicative of the location of the Gulf Stream and its extension. The power spectra of SSH anomalies at several locations in the domain are shown in Fig. 2b. The power spectral density measured along the Gulf Stream (black and blue lines) is larger at all timescales than that within the gyre regions (red and green lines). The largest spectral power at interannual frequencies is near Gulf Stream's detachment point (blue line). As a reference, the spectra are compared to the Zang and Wunsch (2001) canonical frequency-wavenumber spectrum (gray dashed line). As a function of frequency (σ), this spectrum is proportional to $\sigma^{-1/2}$ on periods longer than 100 days, whereas, for periods shorter than 100 days it is proportional to σ^{-2} (i.e. red noise). Hughes and Williams (2010) also highlight regional deviations from this canonical spectrum. The spectra taken in the vicinity of the Gulf Stream, display approximately red noise profiles up to timescales of a year with whiter noise profiles on longer timescales. This whitening is indicative that the predictability of SSH in the Gulf Stream may be limited on interannual timescales. In contrast, the profiles taken in the subpolar and subtropical gyres (red and green lines), are closer to being red noise like in nature for all time periods. This is indicative that skillful interannual SSH forecasts can potentially be made in these regions.

Figure 3a shows the standard deviation of annual mean SSH anomalies. This again shows that most of the interannual variability is located in the vicinity of the Gulf Stream's extension and in the subpolar gyre. There are several potential mechanisms for such interannual variability in the Gulf Stream's extension, including: baroclinic Rossby waves directly modulating the jet extension (Sasaki and Schneider 2011; Qiu et al. 2014); variations in the western boundary currents due to changes in wind forcing (Andres 2016); and modulation by the mesoscale eddy field (Spall 1996; Berloff et al. 2007).
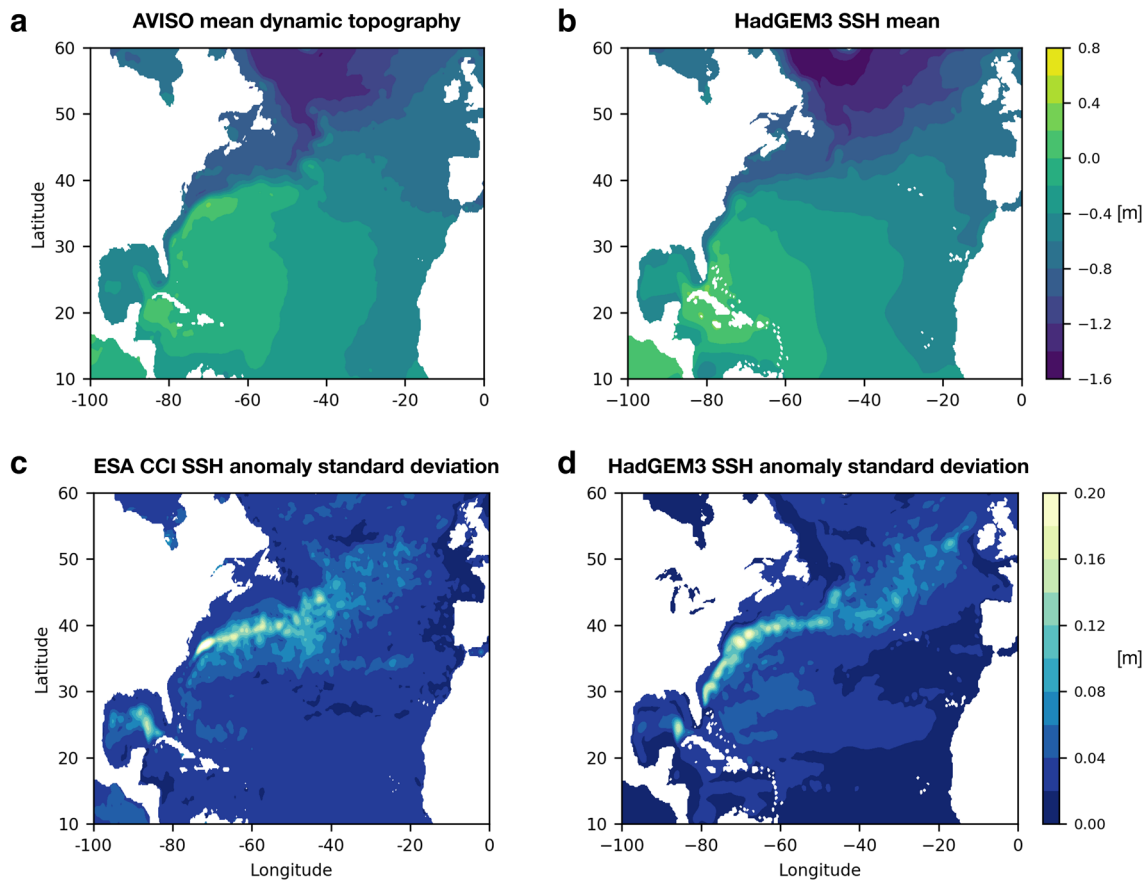
Fig. 1 A comparison of mean SSH from **a** AVISO mean dynamic topography and **b** the HadGEM3 model output. The standard deviation of the interannual SSH anomalies calculated from **c** the ESA CCI version 2.0 monthly gridded fields of surface height anomaly and **d** the HadGEM3 model output. Both observation based plots are based on the the 20 year period of 1993–2012 inclusive. The mean HadGEM3 SSH is calculated from the entire 150 year model run. Whereas, the model SSH anomaly standard deviation plot is calculated from a representative 20-year period of the model run

The signal to noise ratio of interannual variability in the North Atlantic is investigated by diagnosing $\frac{\sigma_N}{\sigma_1}$, where σ_N represents the standard deviation of N-year means of SSH. This measure is often referred to as potential predictability (Boer 2004; Hawkins et al. 2011). Figure 3b, c show the potential predictability associated with σ_3 and σ_5. Regions with weak interannual variability (standard deviations < 0.02m) are masked (white regions in Fig. 3b, c). These regions of low variability are located along the eastern edge of the basin. The US east coast south of 45°N stands out as the only portion of coastline where any potential interannual predictability is present (Fig. 3b, c). The largest potential predictability is located in the subpolar gyre. The dynamics of this region are likely to be relatively linear, dominated by mixed layer's response to forcing and mean advection, and not heavily influenced by the effects of turbulent mesoscale eddies (Sérazin et al. 2015). However, such relatively linear dynamics in the model may also be because the 1/4° model resolution will not fully resolve the small internal deformation radius in the subtropical gyre and therefore will likely underestimate effects related to the baroclinic instability. Figure 3a shows large values of interannual variability in the Gulf Stream extension, however, Fig. 3b, c demonstrate this pattern of large variability does not translate into a comparable pattern of large potential predictability. Nevertheless, although the potential predictability is lower in the Gulf Stream region it is still non zero, in agreement with the power spectra on interannual timescales in Fig. 2b (black and blue profiles). Therefore even in the eddy-active Gulf Stream region, there appears to be some potentially predictable interannual variability.

3 LIM forecast analysis: influence of eddy field initialisation on interannual forecasts

Although the SSH variability analysis hints at the presence of interannual timescales, an investigation of SSH forecasts is needed to evaluate any interannual SSH predictability present. Traditionally the statistics needed to evaluate
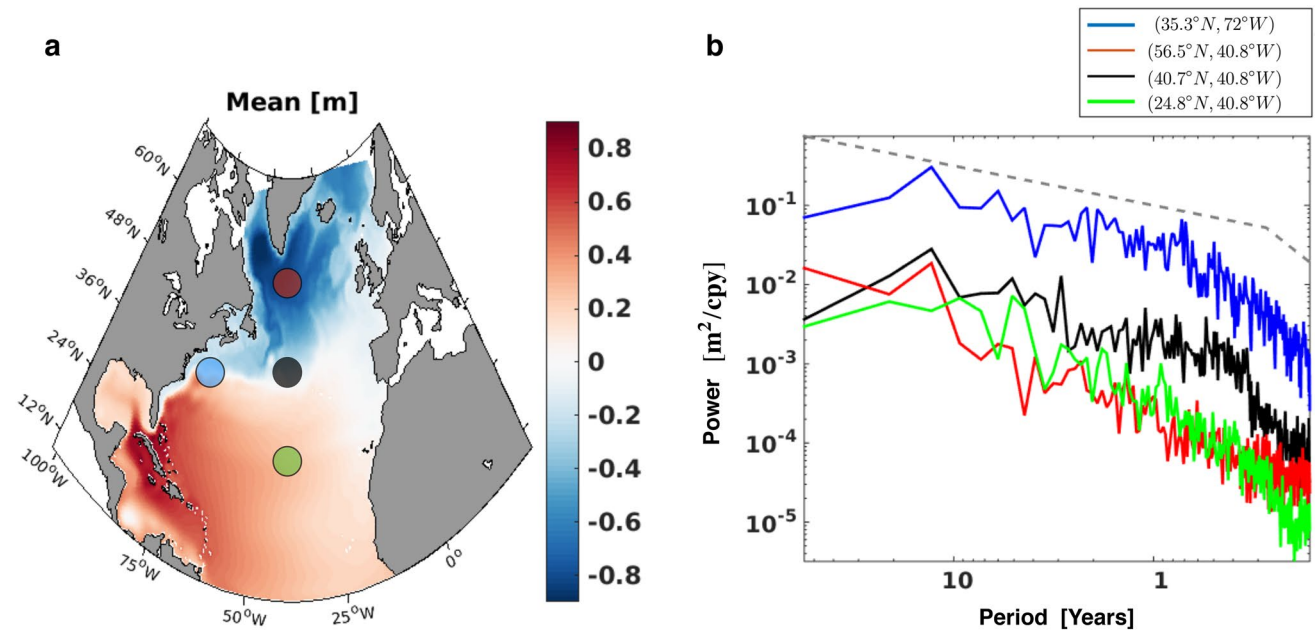
a

b

Fig. 2 HadGEM3 150-year control run: **a** Time mean SSH; **b** Power spectra of SSH anomalies taken at the four locations in the North Atlantic as indicated in **a**: near the detachment point of the Gulf Stream (blue, 35.3°N, 72.0°W), in the subtropical gyre (green, 24.8°N, 40.8°W), in the Gulf Stream extension region (black, 40.7°N, 40.8°W) and in the subpolar gyre (red, 56.5°N, 40.8°W). The gray dashed line represents the Zang and Wunsch (2001) spectrum at an arbitrary chosen amplitude. The power spectral density is in units of (m²/cpy) where cpy denotes cycles per year

predictability in a GCM are generated by creating an ensemble of model simulations. Depending on the model used, and the size of the ensemble, this process can be computationally expensive (Collins 2007). In this paper we use a contrasting approach whereby the methods used to evaluate predictability are based on statistical models created from one long dynamical model run. The statistical models used here have the benefit of being computationally cheap to run and can still provide insights into the dynamics of the system due to their simplicity (e.g., Sonnewald et al. 2018).

Linear Inverse Modeling (LIM; Penland 1989) has previously been used to evaluate predictability in sea surface temperature, in both models and observations (e.g., Penland 1989; Hawkins et al. 2011; Zanna 2012; Huddart et al. 2017; Dias et al. 2018). The method models the evolution of the desired fields as a linear process forced by white noise. In doing so, the linear inverse model gives information about the predictability of the fluctuations in the system. To enable this calculation, the SSH anomalies are decomposed into empirical orthogonal functions (EOFs) and their related principal components (PCs),

$$SSH(x, y, t) = \sum_i EOF_i(x, y) PC_i(t). \tag{1}$$

The EOFs are constructed using monthly-mean SSH model output. The EOFs are also weighted by the area of their grid boxes as the NEMO grid is irregular. In the calculation

of the EOFs, SSH in the Gulf of Mexico is not used as we wished to focus on the SSH predictability in the main ocean basin.

In the following analysis we use 25 EOFs explaining 63% of the variance; the leading three EOFs (responsible for 9.9%, 5.2% and 5.0% of the variance, respectively) are shown in Fig. 4. The leading EOF has a spatial structure reminiscent to that calculated using observations (Häkkinen and Rhines 2004; Häkkinen et al. 2013). Häkkinen et al. (2011) attributed this pattern of SSH variability to variability in the wind stress curl. As the wind stress curl varies, there are associated variations in the strength and sizes of the subpolar and subtropical gyres and a resultant change in SSH (Häkkinen et al. 2013). This 'gyre mode' varies between a state with a small subpolar gyre with a large eastward extended subtropical gyre and a state with a large eastward extended subpolar gyre with a small contracted subtropical gyre. A similar variation and associated dependence on the wind stress curl has been identified in this model (shown in the supplementary material). Moreover, there is a lagged response of the first principal component of SSH to a leading principal component of the wind stress, again in agreement with Häkkinen et al. (2011).

The evolution of the PCs of SSH anomalies is approximated by a linear stochastic model (Penland and Sardeshmukh 1995)
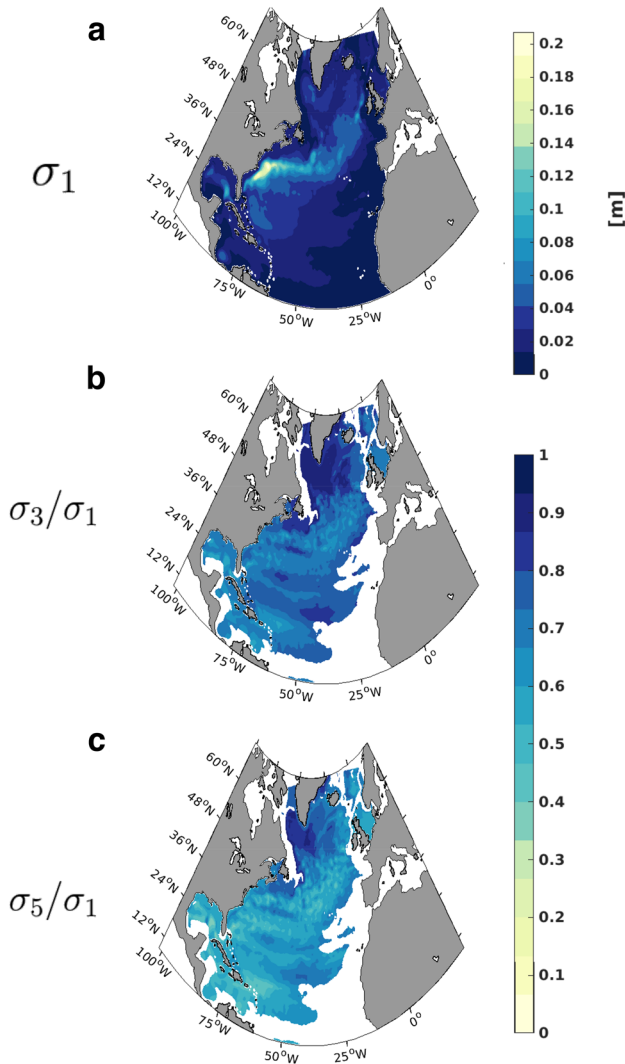
Fig. 3 **a** Standard deviation of SSH anomalies of the control run created with 1 year means. Ratio of standard deviations of the **b** 3 and **c** 5 year means to standard deviation based on 1-year means

$$\frac{d\mathbf{P}}{dt} = \mathbf{A}P(t) + \xi, \tag{2}$$

where **P** is the vector of n-PCs, with dimensions of n by 1, ξ is a stochastic forcing term, and **A** is a linear n by n matrix which controls the temporal evolution of the n-PCs. The linear operator

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

contains dynamical information about the variability of the system as it is constructed using the, n by n, covariance matrices at lag-τ_0 and lag-0, which are calculated from the PCs as

$$\mathbf{C}(\tau_0) = \langle \mathbf{P}(t + \tau_0)\mathbf{P}^T(t) \rangle,$$
$$\mathbf{C}(0) = \langle \mathbf{P}(t)\mathbf{P}^T(t) \rangle. \tag{4}$$

Here $\langle \rangle$ indicates an average over all times.

Forecasts, $\hat{\mathbf{P}}$, are then generated using the model such that, with respect to the initial time, t,

$$\hat{\mathbf{P}}(t + \tau) = \mathbf{B}(\tau)\mathbf{P}(t), \tag{5}$$

where τ is the forecast lead time and the n by n matrix **B** is the forecast propagator,

$$\mathbf{B}(\tau) = \exp(\mathbf{A}\tau) = \exp\left[\frac{\tau}{\tau_0}\ln[\mathbf{C}(\tau_0)\mathbf{C}(0)^{-1}]\right], \tag{6}$$

Predictability can then be evaluated by examining the difference between the probability distribution of the predictions and that of the climatology. For LIM to be applicable, the system being examined is required to possess several characteristics (Penland and Sardeshmukh 1995):

- it can be described by Gaussian statistics;
- **A** is independent of the time lag, τ_0, used to calculate it;
- all real parts of the eigenvalues of **A** must be negative and therefore decay.

Finally, to prevent overfitting of the linear models, the data used in each experiment is separated into a training and a verification data set. In the presented results the training and verification data sets are 140 years and 10 years long respectively. The linear model is constructed using the training data set. This model is then used to make predictions for the verification set, and the skill of these predictions is used to evaluate the system's predictability.

Tests to assess how well these conditions are met for SSH anomalies in the North Atlantic and are shown in Fig. 5. Panel a shows a comparison of the cumulative density function of the 150 years of SSH anomalies with that of an idealised Gaussian distribution with the mean and variance of the model output, all calculated in the area used to calculate the EOFs (shown in Fig. 4e). The agreement between the two profiles demonstrates that the system is well described by Gaussian statistics. To examine the influence of mesoscale eddies on the skill of the forecasts two different linear operators are constructed. In each experiment, the same reduced basis, consisting of 25 EOFs and PCs, is used to create each propagator. The first propagator contains all available frequencies. A second temporally smoothed propagator is constructed by applying an 18-month running mean filter to the PCs. Figure 5b shows the Frobenius norm of **A** as a function of different lag times, τ_0, for these two operators. Both operators possess regions where the norm of **A** only varies by a small amount as a function of τ_0. However, for the 18 month filter there is strong variation in the Frobenius
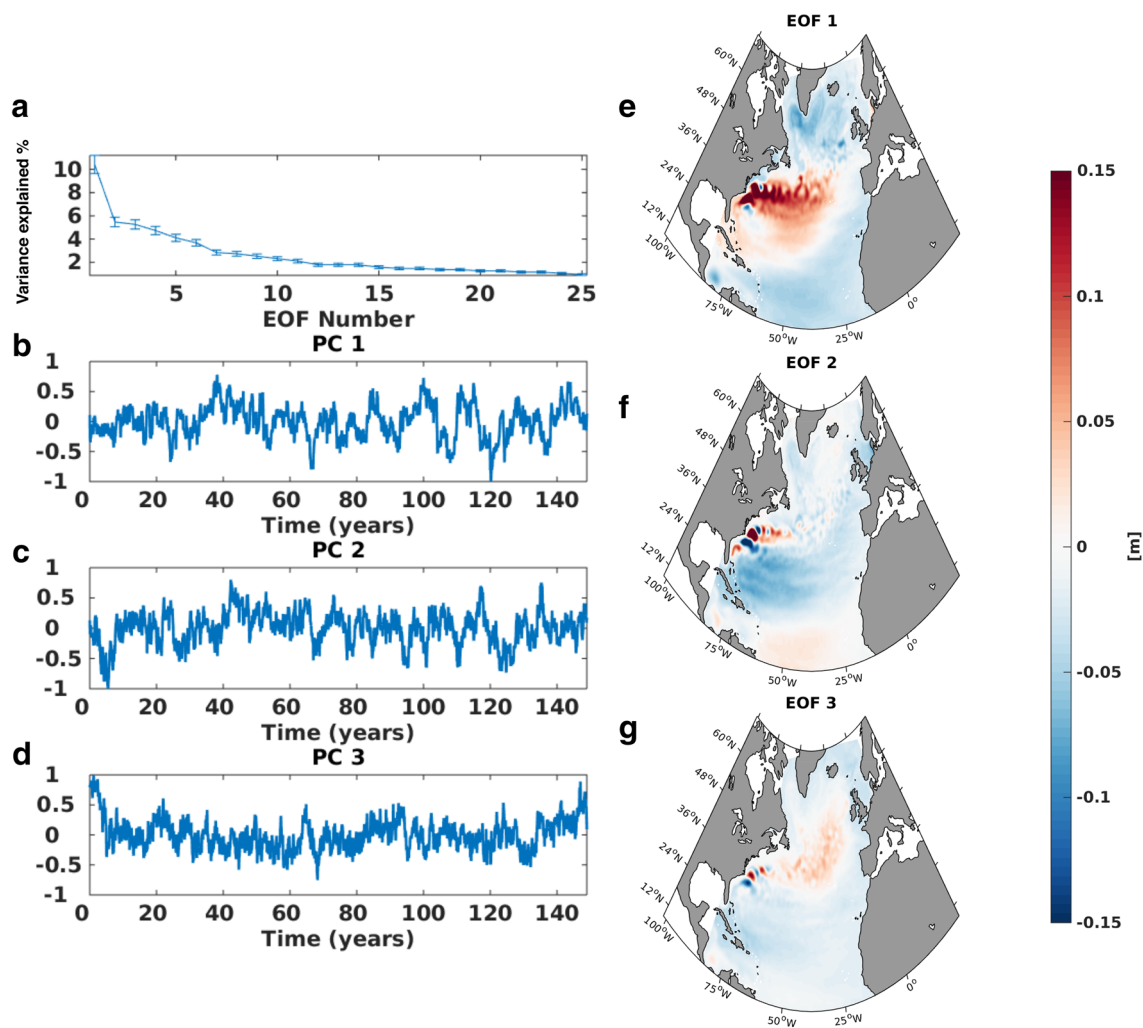
Fig. 4 **a** The fraction of variance each of the leading 25 EOFs explains, calculated from 150 years of monthly mean SSH model output. The error bars represent the one standard deviation error related to sampling (calculated using equation 24 in North et al. (1982)). Timeseries of the: **b** first, **c** second and **d** third principal components. The spatial components of the: **e** first, **f** second and **g** third EOFs

norm of **A**, when τ_0 is 15 months. This large variation indicates that in this parameter range the model output may not be well represented by a system of the form shown in Eq. 2. When using the operator constructed with monthly means a τ_0 of 6 months is chosen and when using the smoothed operator a τ_0 of 2 months is used. In both cases, all the real parts of the eigenvalues of **A** are negative and thus satisfy the final necessary condition.

We can now create forecasts of SSH anomalies using the viable LIMs. Forecasts are initialised every 6 months throughout the 150 years of model output, creating 300 forecasts in total. When creating these forecasts it is crucial to ensure a difference between the training and test datasets. At each forecast initalisation date the LIM propagator used to generate the forecasts is trained on 140 years of model data. In each case these 140 years are comprised of the the data which is not within a ten-year window centered on the

forecast initialisation date. In order to create a benchmark for the forecasts made using the LIM models, lagged correlation forecasts are also made (Lorenz 1963),

$$x(t_0 + \tau) = \beta(\tau)x(t_0), \tag{7}$$

where β is the auto-correlation of the time series at a point in space, x is time series of the quantity being predicted and τ is the lag time of the forecast. This is a type of 'damped persistence' forecast which may provide forecasts better than climatology ($\beta = 0$) and persistence ($\beta = 1$) forecasts. The model output used to construct these damped persistence forecasts is reconstructed from the same EOFs and principal components used to construct the LIM models, to allow for a fair comparison.

To evaluate the skill of the statistical models relative to climatological forecasts, we use a root mean square error metric ($RMSE_{Relative}$) (Hawkins et al. 2011):
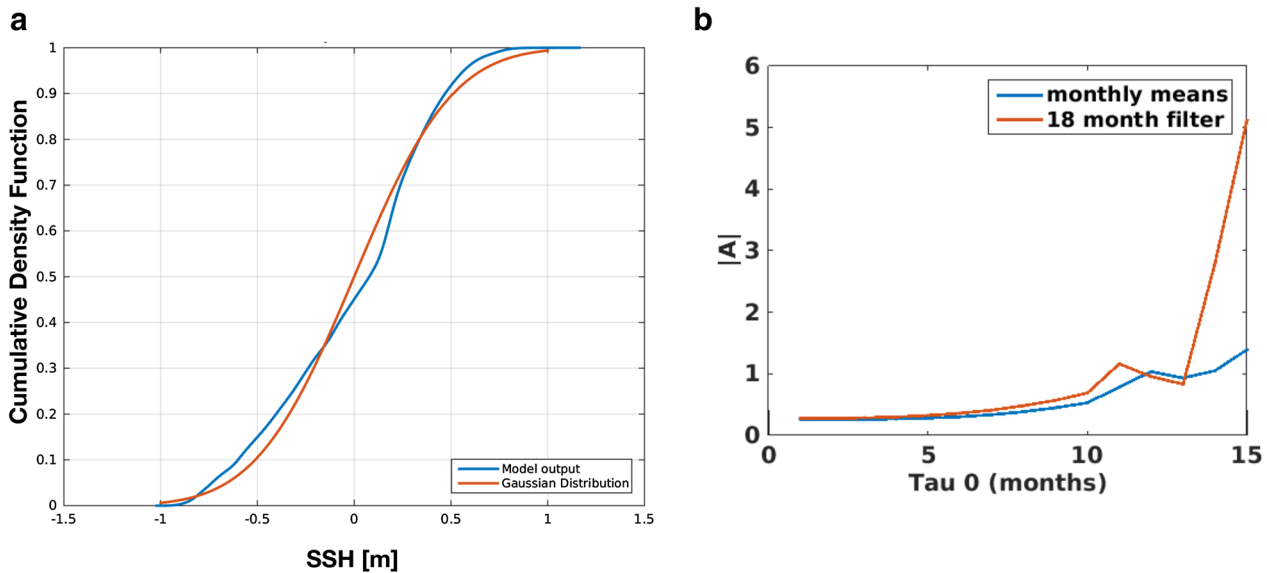
a

b

Fig. 5 **a** A comparison of the cumulative distribution functions of the monthly mean SSH model output and an idealised Gaussian distribution constructed using the mean and standard deviation of the SSH model output. **b** The Frobenius norm of the operator **A** as a func-

tion of different lag times, τ_0. The two operators are constructed with monthly mean and 18 month filtered principal components respectively

$$RMSE_{Relative} = \frac{RMSE_{pred}}{RMSE_{clim}} \qquad (8)$$

where $RMSE_{clim}$ is the root mean square error of a climatological forecast over the forecast period and $RMSE_{pred}$ is the root mean square error of the predicted field. The climatological forecast assumes the SSH anomalies keep their initial values, i.e. $\hat{\mathbf{P}}(t + \tau) = \mathbf{P}(t)$ and is constructed using the same EOFs and PCs used to construct the LIM models (i.e the time filtered forecasts are verified against the time-filtered truth). The RMSEs are calculated relative to the control run solution, which has been reconstructed using the same EOFs and PCs used to construct the LIM models. A value greater than unity indicates that the model's forecasts are inferior to those generated using the climatology, whereas, a value less than unity demonstrates forecast skill superior to climatology.

The forecast error maps for the LIM model trained on monthly data are shown in Fig. 6d–f. Errors emerge rapidly in the subpolar and subtropical gyres (seen in panels a and b). Only in the Gulf Stream region and southern part of the domain are any areas of skillful forecasts seen (confirmed in panel c). The damped persistence forecasts created with the monthly mean model output (panels j, k, and l) exhibit small errors in the subpolar gyre and parts of the subtropical gyre (panels a and b), coinciding with regions of large potential predictabilities (Fig. 3). The forecasts created with the LIM model trained on monthly mean SSH anomalies are less skillful than those produced with the damped persistence

model. The inclusion of the high-frequency components of the SSH in the construction of the LIM model means that predictably is not exhibited on timescales longer than a year.

The error maps which are subject to 18-month filtering produce smaller errors in all regions (Fig. 6a–c, g–i, and m–o). These models again display the smallest relative RMSEs in the subpolar and subtropical gyres, with more substantial errors in the Gulf Stream region. The LIM model outperforms the damped persistence forecasts in the majority of areas and timescales [the exception being in the subtropical gyre (panel b)]. The subpolar gyre emerges as the region with the largest amount of predictability, on timescales longer than a year (panels a and i). Small errors are also exhibited in the tropics, and extending eastwards towards the Iberian Peninsula. The US east coast stands out as the only section of coastline which borders a region with forecast errors of less than 0.8 on interannual timescales. These results are in agreement to those found by Nonaka et al. (2016), where a lack of any predictability, on timescales longer than a few months, is also found in the Kuroshio.

4 Predictable patterns: optimal initial conditions and average predictability time

The spatio-temporal structure of the predictability can also be analysed by explicitly identifying any patterns which are predictable on interannual timescales. Two methods are now used: (1) an examination of the growth of optimal initial

conditions leading to a maximum increase of variance and (2) a decomposition of the system into predictable components, ranked by their relative contributions to the total average predictability time present.

4.1 Non-normal mode analysis and optimal initial conditions

The characteristics of the trained linear model can be used to infer information about a system's sensitivity to initial conditions. In a series of papers, Farrell and co-authors developed a methodology, generalised linear stability theory, to investigate the transient behaviour resulting from initial perturbations to its mean state (Farrell and Ioannou 1996). This methodology has been used to examine a range of geophysical problems including: Couette flow (Farrell 1982), atmospheric forecast error growth (Farrell 1990), quasi-geostrophic turbulence (Farrell and Ioannou 1995), the El Niño Southern Oscillation (Penland and Sardeshmukh 1995), Gulf Stream dynamics (Farrell and Moore 1992) and the Atlantic meridional overturning circulation (Zanna and Tziperman 2005, 2008; Hawkins and Sutton 2009).

This analysis investigates the transient growth in linearly-stable fluid dynamical systems. It may appear counter-intuitive that there can exist disturbances which lead to growth in a stable system. However, when the operator $\mathbf{A}$ is non-normal, i.e. $\mathbf{A}\mathbf{A}^{\mathbf{T}} \neq \mathbf{A}^{\mathbf{T}}\mathbf{A}$, it is possible for the eigenmodes of the system to interact and give a large amplification of variance at a finite-time (Farrell and Ioannou 1996). The solutions to

$$\frac{d\mathbf{P}}{dt} = \mathbf{A}P(t), \tag{9}$$

can be written in terms of the eigenvectors, $\mathbf{e}_i$ as

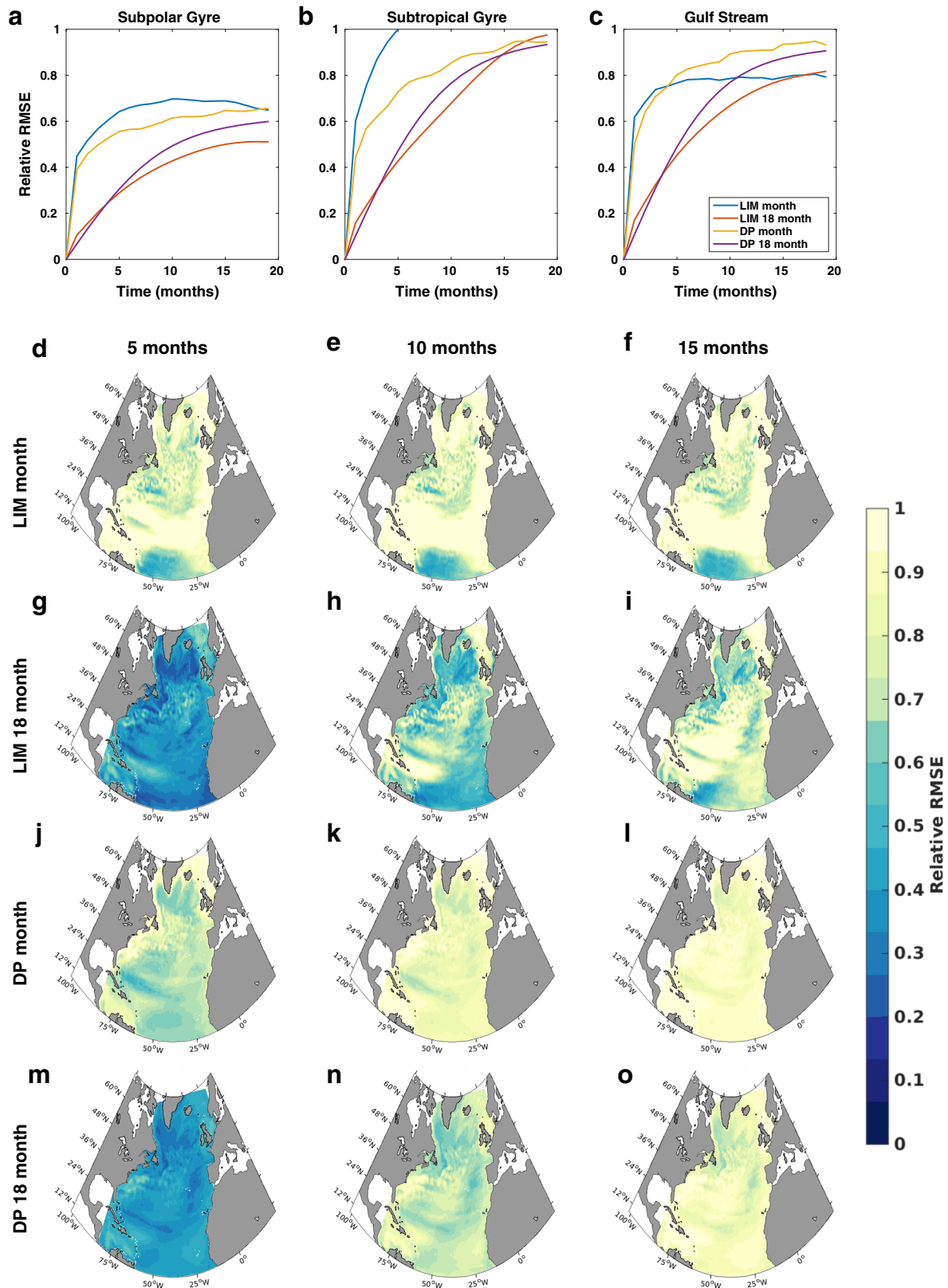$$\mathbf{P}(t) = \sum_i \mathbf{e}_i a_i \exp \lambda_i t, \tag{10}$$

where λ_i are the eigenvalues of $\mathbf{A}$ and a_i is a complex constant. The SSH anomaly growth at time τ by non-normal eigenmode interference is given by

$$\mu(\tau) = \frac{\mathbf{P}(\tau)^{\mathbf{T}}\mathbf{P}(\tau)}{\mathbf{P}(0)^{\mathbf{T}}\mathbf{P}(0)}. \tag{11}$$

The longest timescale on which this growth occurs can be thought of as an optimistic upper bound on the predictability of linear events without forcing. The corresponding spatial patterns, which lead to a maximum growth at a time τ, are called the optimal initial conditions and are given by calculating the leading singular vector of $\mathbf{B}(\tau)$. In this section, the LIM constructed using 18 month temporally smoothed principal components is used as it exhibits skillful forecasts on interannual timescales.

The curve depicting the growth of SSH anomalies, $\mu(\tau)$, is shown in Fig. 7a. The perturbations can grow through non-normal interactions on time scales of up to 100 months, with the maximum growth occurring at 20 months. The optimal initial condition pattern in SSH, which leads to the largest growth in SSH anomalies after 20 months, is shown in Fig. 7b. This pattern has a very weak gyre scale tripolar pattern, reminiscent of EOF 1 (shown in Fig. 4e). The pattern has two main notable smaller scale features, a tripolar structure off Cape Hatteras (situated at 32.5°–42.5°N, 67° –74.55°W, shown by the green ellipse in panel e) and a single sign SSH anomaly along the US east coast (black ellipse, panel e). The propagated optimal initial condition is shown in Fig. 7, panels c and d, at 10 and 20 months, respectively. After 10 months, the SSH anomaly along the boundary no longer has a single sign. There is an increase in SSH along the path of the Gulf Stream and in the subtropical recirculation gyre, and a decrease in SSH in the subpolar gyre. After 20 months, an SSH anomaly grows along the Gulf Stream path, and its magnitude is seen to double. The magnitude of SSH in the subpolar and subtropical gyres is also seen to increase significantly. One interpretation of these optimal initial conditions is that it is especially important to constrain the position of the Gulf Stream separation in the initial conditions as initial errors in this region lead to gyre-scale errors within 10–20 months. However, it is also possible that it is the weaker gyre-scale pattern present in the optimal initial conditions which leads to this growth, as SSH anomalies can be integrated by the gyre circulation on interannual timescales.

Figure 8 shows the positive optimal initial condition (of the same magnitude as that shown in Fig. 7b) and its evolution after 20 months when it is added to the mean SSH field. It also shows the negative version of the optimal initial condition added to mean field, which is an equally valid solution since the evolution is linear. The initial and propagated version of the positive optimal initial condition, Fig. 8, panels a and b, demonstrates an increase in strength of the subpolar gyre, as well as an increase in the SSH gradient across the Gulf Stream. The change in the SSH gradient is linked to variations in the geostrophic transport along the Gulf Stream path, shown in Fig. 9. The resultant geostrophic velocity anomalies act in different directions in the two gyres and are particularly evident in the subtropical gyre. The SSH 0m contour is also seen to be shifted to a higher latitude. However, this is a marginal effect as shown by the contours in Fig. 8's panel b (less than a degree in latitude, for an initial perturbation with double the magnitude of that shown in Fig. 7a). The evolution of the negative optimal initial conditions, shown in Fig. 8c, d, demonstrates an increase in SSH along the US east coast North of Cape Hatteras, as well as a southward shifted Gulf Stream detachment point. The SSH gradient across the Gulf Stream is also lower, indicating a

a Subpolar Gyre
b Subtropical Gyre
c Gulf Stream
d 5 months
e 10 months
f 15 months

LIM month — LIM 18 month — DP month — DP 18 month

decrease in Gulf Stream transport. Panel d shows that the SSH 0m contour's position can move significantly southward (approximately 5° in latitude, for an initial perturbation with double the magnitude of that shown in Fig. 7a and that the subtropical gyre contracts to the west of the basin. The initial conditions associated with timescales (τ) ranging from

10 to 30 months are also calculated and compared to the optimal calculated at the maximum amplification time. The spatial correlation between these initial patterns are found to be at least 0.8, and the patterns behave in a qualitatively similar manner when propagated in time. The optimal initial conditions, calculated from similar models with differing numbers of EOFs, exhibited small-scale (1/2°) spatial differences in the Gulf Stream's extension, however both the signal along the US east coast and the tripolar pattern appear robust. Moreover, the propagated optimals all resemble that shown in Fig. 7d.

4.2 Optimal initial conditions occurring in the model output

It is important to determine how often the optimal initial conditions and their evolved patterns are realised in the model output. Figure 10a, shows the projections of the initial states on the model output (at each time t the projection is the product of the $\mathbf{P}(t)$ and the optimal initial condition, i.e. the leading singular vector of $\mathbf{B}$), as well as the projections of the evolved initial conditions 20 months later. The growth in these projections is seen to be close to that predicted by the maximum amplification curve (Fig. 7a). The occurrences of the the tripolar SSH pattern, seen in the optimal initial condition, are detected by using an algorithm which calculates the 2D correlation coefficients between the monthly mean SSH anomalies and optimal initial condition in the relevant area (the green ellipse in Fig. 7e). Spatial correlations which are greater than 0.8 are retained. Out of the 1800 monthly-mean SSH anomalies comprising the model output, 404 are found to display a tripolar anomaly structure off Cape Hatteras. After 15–20 months from those 404 tripolar anomaly patterns, 310 (77%) lead to SSH anomaly growth along the US east coast (as in Fig. 7f, green ellipse). This growth is detected by evaluating the change in sign of the SSH anomalies (in the the green ellipse in Fig. 7f). About 140 (out of 310) also display a change in sign of SSH along the coast (as in Fig. 7f, black ellipse). The occurrence of these optimal evolutions in the model output indicates that changes in the SSH in the Gulf Stream near its detachment

point are potentially important in predicting SSH variations along the US east coast.

4.3 Average predictability time

We complement the analysis of the optimal patterns, which depends on the target timescale, by examining predictable patterns that persist over all timescales, and are therefore the most predictable over a range of target times (DelSole and Tippett 2007). This is done by calculating the average predictability time (APT) (DelSole and Tippett 2009a). This index of predictability is based on the Mahalanobis signal (DelSole and Tippett 2007),

$$S(\tau) = \frac{1}{k} tr[(\boldsymbol{\Sigma}_\infty - \boldsymbol{\Sigma}_\tau)\boldsymbol{\Sigma}_\infty^{-1}], \tag{12}$$

where k is a constant related to the number of principal components used in the analysis, tr is the trace of the matrix, $\boldsymbol{\Sigma}_\tau$ is the covariance matrix of the forecast error at lead time τ, $\boldsymbol{\Sigma}_\infty$ is covariance matrix of the forecast distribution at long lead times. Here, $S(\tau)$ has a value of 1 when the system is completely predictable, and a value of 0 when the forecast covariance matrix is the same as the climatological covariance matrix, meaning the system is unpredictable. This method has been used before to examine the predictability of several geophysical fields, including the upper ocean temperature and the AMOC (Branstator et al. 2012; Branstator and Teng 2014).

The APT can be defined by integrating the Mahalanobis signal over all lead times (DelSole and Tippett 2009a), leading to

$$APT = 2 \sum_{\tau=1}^{\infty} S(\tau). \tag{13}$$

The factor of two makes APT agree with the e-folding time in the univariate case. In one dimension, APT resembles a root mean square error and is given by DelSole and Tippett (2009b)

$$APT = 2 \sum_{\tau=1}^{\infty} \frac{\sigma_\infty^2 - \sigma_\tau^2}{\sigma_\infty^2} = 2 \sum_{\tau=1}^{\infty} \left(1 - \frac{\sigma_\tau^2}{\sigma_\infty^2}\right). \tag{14}$$

Since APT is the integral of predictability over all times, it is independent of the chosen lead time. This measure can also be used to define predictable components by finding the projection vectors $\mathbf{q}$ that maximize APT. In which case, the component $\mathbf{q^T P}$, with $\mathbf{P}$ being the principal component state vector, has forecast and climatological variances given by $\sigma_\tau^2 = \mathbf{q^T} \boldsymbol{\Sigma}_\tau \mathbf{q}$ and $\sigma_\infty^2 = \mathbf{q^T} \boldsymbol{\Sigma}_\infty \mathbf{q}$, respectively.

In this study, the APT of the whole system and of the leading predictable components are calculated using the
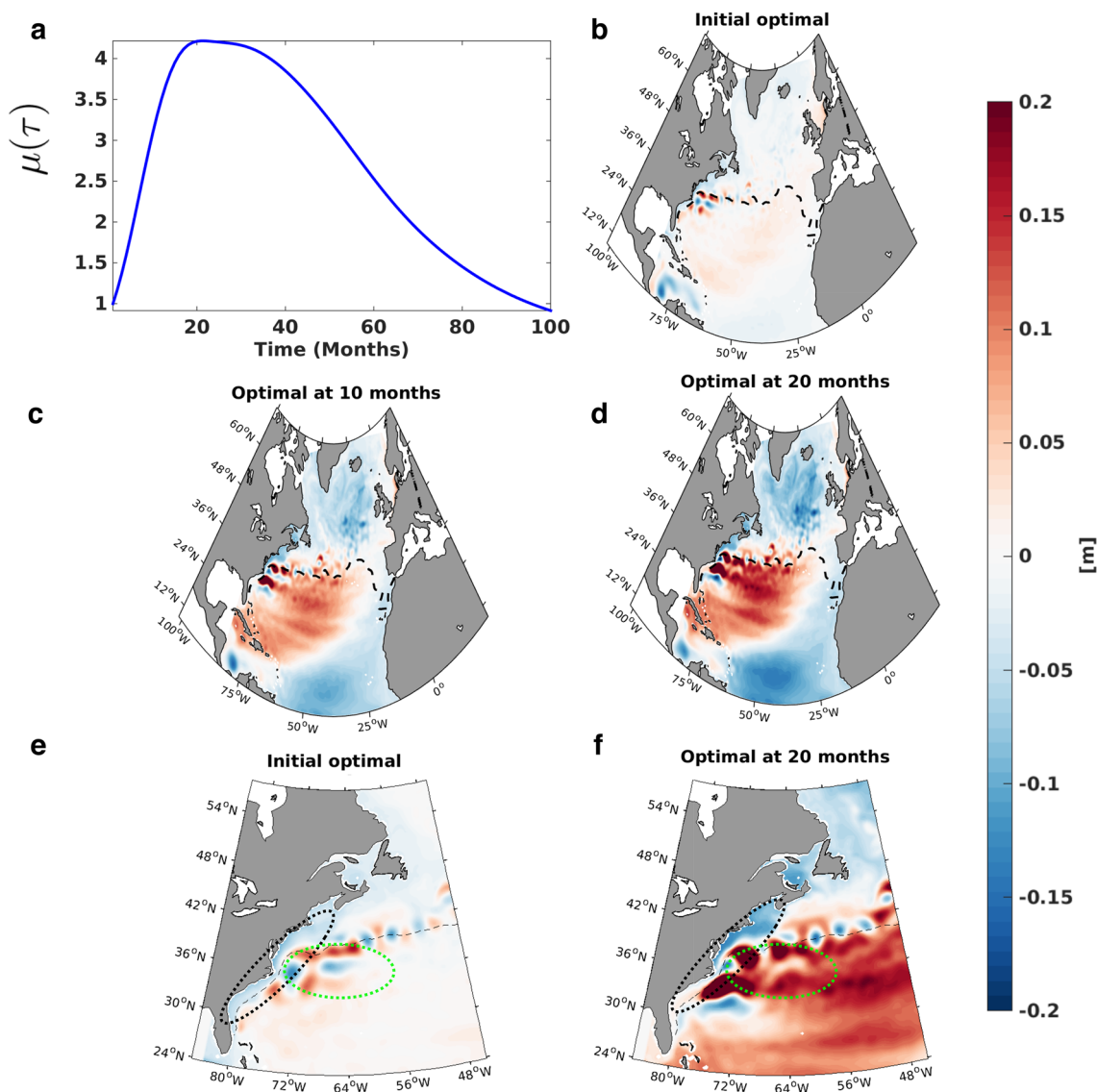
Fig. 7 a The maximum amplification curve (Eq. 11). **b** The optimal initial condition with its tripolar pattern. **c** The optimal at 10 months. **d** The optimal at 20 months, its state of maximal growth. **e** The initial optimal in just the area bordering the US east coast. **f** The optimal at 20 months in just the area bordering the US east coast. The black and green dotted ellipses indicate the regions which are correlated with monthly means from the model output. The ellipses are characteristic anomalies described in the text. The black dotted lines indicate the 0m contour in the time mean SSH

method contained in DelSole and Tippett (2009b). Firstly, to prevent overfitting, the data is separated into training and verification data sets, in the same manner as described previously. Forecasts are then generated by forming linear regression models from the training data (DelSole and Tippett 2009b), i.e. the projections $\hat{\mathbf{P}}_{\mathbf{L}}(t + \tau)$, are given by

$$\hat{\mathbf{P}}_{\mathbf{L}}(t + \tau) = \mathbf{C}(\tau)\mathbf{C}(0)^{-1}\mathbf{P}(t). \tag{15}$$

Using such models and in the case of a zero-mean stationary process, meaning $\mathbf{C}(0) = \boldsymbol{\Sigma}_\infty$, the forecast error covariance matrix is given by

$$\boldsymbol{\Sigma}_\tau = \mathbf{C}(0) - \mathbf{C}(\tau)\mathbf{C}(0)^{-1}\mathbf{C}(\boldsymbol{\tau})^T. \tag{16}$$

These values for $\boldsymbol{\Sigma}_\infty$ and $\boldsymbol{\Sigma}_\tau$ can be substituted into Eq. 13 to calculate the APT of the entire system. In order to maximize APT in Eq. 14, the problem reduces to solving the generalized eigenvalue problem (See DelSole and Tippett (2009b) for a full derivation),

$$\mathbf{Gq} = \lambda \mathbf{C}(0)\mathbf{q} \tag{17}$$

where

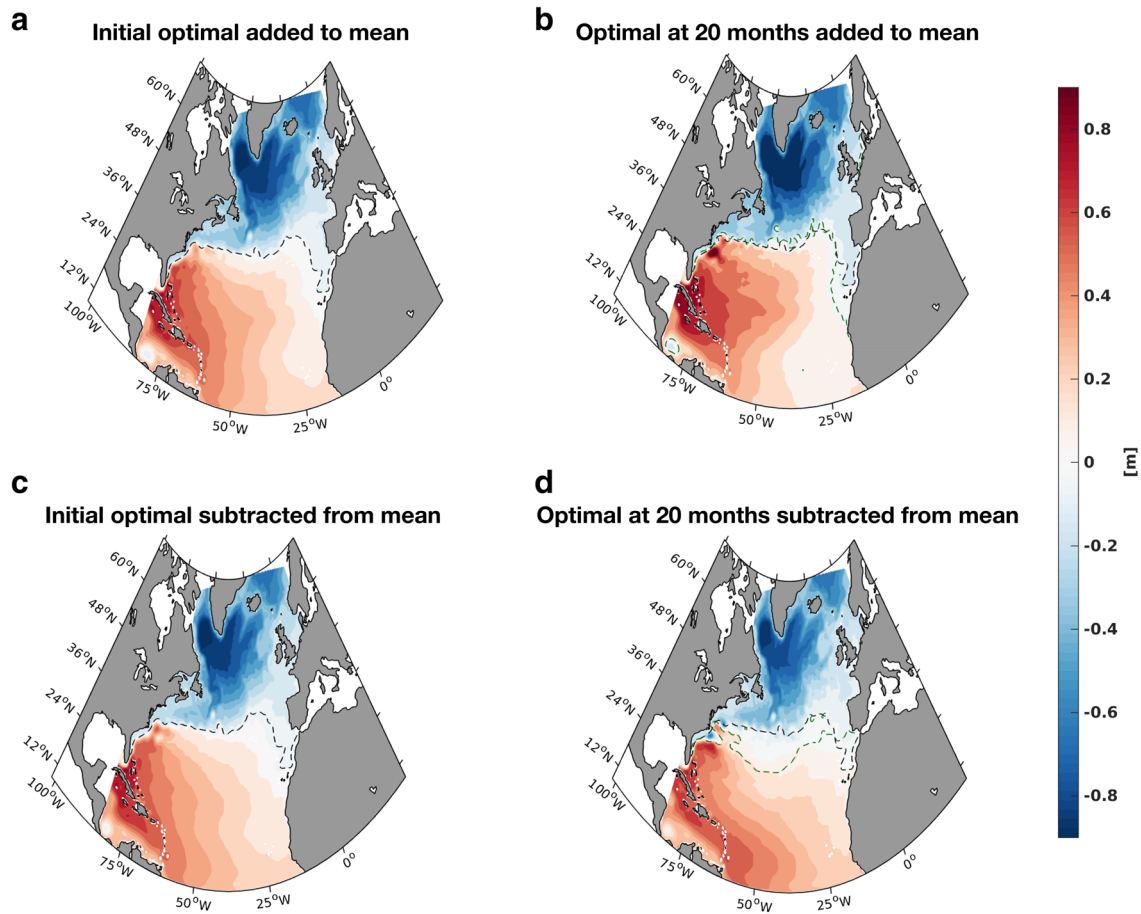$$\mathbf{G} = \sum_{\tau=1}^{\infty} \mathbf{C}(\tau)\mathbf{C}(0)^{-1}\mathbf{C}(\tau)^T. \tag{18}$$

Fig. 8 **a** The initial optimal added to the mean field. **b** The initial optimal propagated forward in time by 20 months added to the mean field. **c** The negative version of the initial optimal added to the mean field. **d** The negative version of the initial optimal propagated forward in time by 20 months added to the mean field. The black dot-ted lines indicate the 0m contour in the time mean SSH. The green lines denote the SSH 0m contour when the optimal initial condition of double the magnitude of that shown in Fig. 7b evolves, after 20 months

The projection vectors **q** are uncorrelated with each other because **G** and Σ_∞ are symmetric. The spatial patterns, **p**, associated with the projection vectors **q** are found by using

$$\mathbf{p} = \langle \mathbf{PP^T q} \rangle = \Sigma_\infty \mathbf{q}; \tag{19}$$

these spatial patterns can be projected back onto the EOFs and are referred to as the predictable components. The predictable components of the system are calculated using only the training data set. To prevent overfitting and calculate APT of each predictable component, the projection vector, **q**, calculated from the training data is applied to verification data set. Thus, the squared multiple correlation between the component time series and the verification data is

$$\mathbf{R}_\tau^2 = \frac{\mathbf{q}^T \mathbf{C}(\tau) \mathbf{C}(0)^{-1} \mathbf{C}(\tau)^T \mathbf{q}}{\mathbf{q}^T \mathbf{C}(0) \mathbf{q}}, \tag{20}$$

where **q** is calculated from the training data set and the correlations are calculated from the verification set. Therefore,

$\mathbf{R}_\tau^2$ can be interpreted as the variance of the predictable component time series, which is explained by a linear regression prediction at time lag τ. The predictability time of each component, APT_p, is then calculated as,

$$APT_p = 2 \sum_{\tau=1}^{\infty} \mathbf{R}_\tau^2. \tag{21}$$

Figure 11a shows that the predictability of the SSH of the whole system, measured by the Mahalanobis signal (solid blue line), diminishes rapidly, reaching a value of approximately 0.25 after 10 months. This is in approximate agreement with the timescale found in the LIM study. However, Fig. 11a (blue shaded region) also shows that several of the individual components of the system have Mahalanobis signals which decay on longer timescales. The APT of the
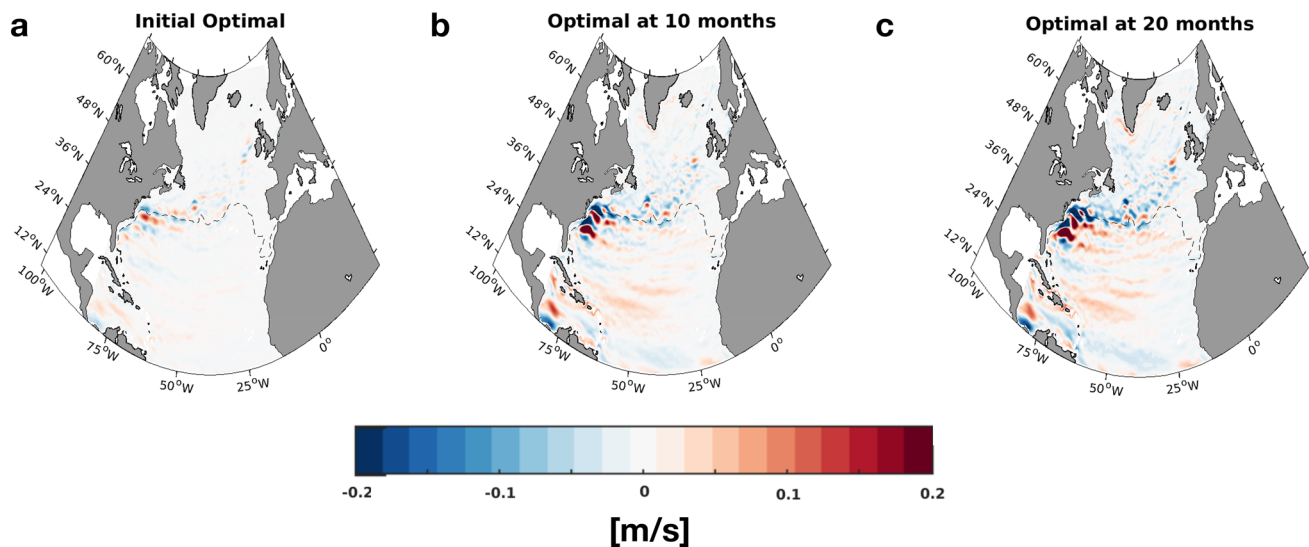
Fig. 9 Anomalies in the zonal geostophic velocities calculated from the SSH of: **a** the initial optimal, **b** the initial optimal propagated 10 months forward in time, **c** the initial optimal propagated 20 months forward in time (All with the same magnitudes as the patterns shown in Fig. 7)
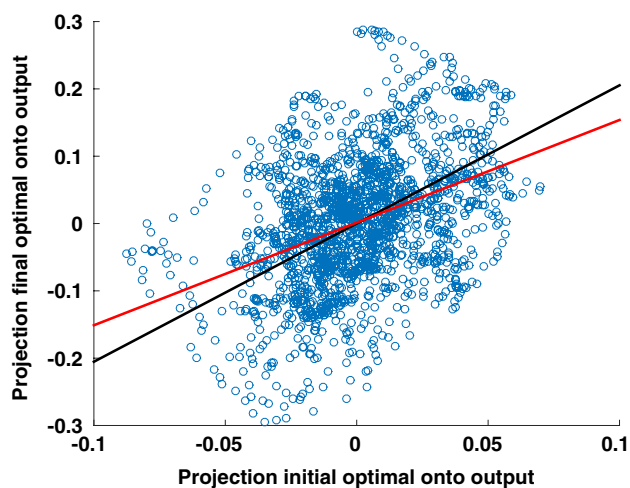

Fig. 10 Projections of the initial and final states on the model output. The black line is a linear fit to these projections, and the red line is a line with a slope corresponding to $\mu^{1/2}$ ($\tau = 20$ months)

leading 25 predictable components shown in panel b, confirms that several components demonstrate predictability on timescales longer than 2 years. The three leading predictable components (those which have the largest values of APT) have average predictability times of 26–28 months. The corresponding spatial patterns of the leading three components are shown in panels c, d, and e. The pattern which is associated with the largest value of APT (panel c) has large SSH magnitudes in the jet extension region. The second pattern (panel d) is localised mainly to the US east coast and Gulf Stream extension region, whereas, the third pattern (panel e) is similar to the evolved optimal initial conditions (Fig. 7),

and EOF1. The time series related to each of these spatial patterns, shown in panels f, g and h, all display interannual variability and have autocorrelation times of 30–60 months.

The similarities between the third leading predictable component, the evolved optimal initial conditions, and EOF1 indicate some robustness of the constructed predictability patterns. The time series associated with the third predictable component correlates strongly with the leading principal component at zero lag. The leading predictable patterns (1 and 2) are not merely EOF1, highlighting that the mode capturing most of the variance is not necessarily the most predictable.

4.4 The influence of atmospheric forcings on the predictable components

The steric component dominates interannual variability in SSH in the North Atlantic. Roberts et al. (2016) confirmed that this is also the case in an ocean only component of HadGEM3 (forced NEMO simulation) and that it is the thermosteric and wind-driven components which contribute most to the interannual variability in the subtropical gyre. In the subpolar gyre, the variability is caused by both the thermosteric and halosteric components and is dominated by the response of the ocean to variations in the buoyancy forcings. Furthermore, in the Gulf Stream region, the variations due to intrinsic ocean processes are also important (Penduff et al. 2011; Sérazin et al. 2015).

Attempts are now made to establish the dynamical origin of the predictable patterns by examining their relationships with fields relating to the wind and buoyancy driven circulations, namely the Ekman components of SSH and
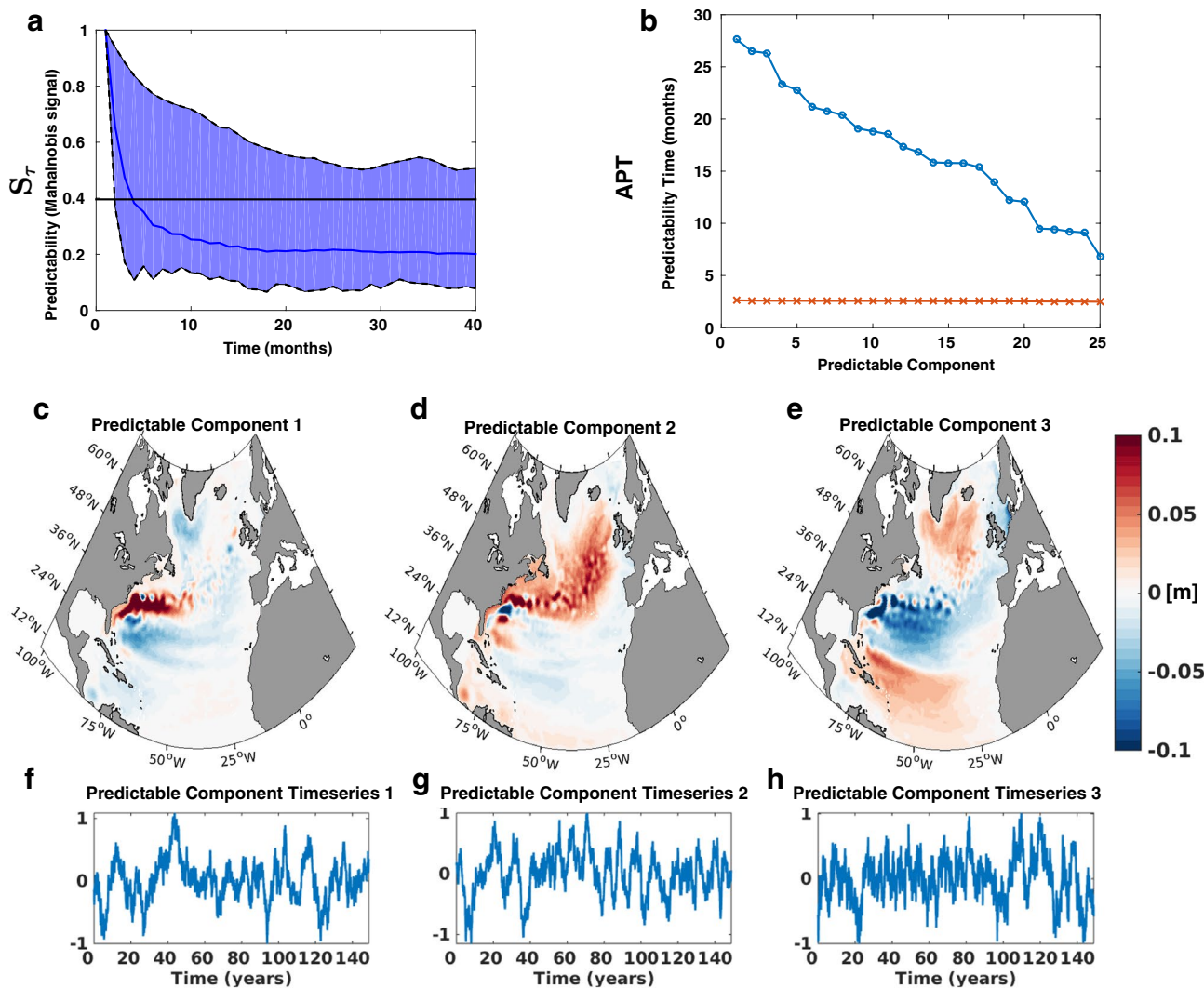
Fig. 11 a Predictability measured by the Mahalanobis signal (S_τ as defined in Fig. 12) of the whole system (blue line), created using linear regression models and the leading 25 principal components. The envelope denotes the spread in predictability measured by the Mahalanobis signal (S_τ) for each of the leading predictable components. The solid black line is representative of 5% significance level calculated using a student's t-test. **b** The average predictability time (APT) for the first 25 predictable components. The orange line indicates the 5 % significance level estimated using the Monte Carlo method discussed in the Appendix. The spatial patterns of the **c** first, **d** second and **e** third predictable components (**p**), ranked in order of their values of APT. The associated time series for the **f** first, **g** second and **h** third predictable components ($\mathbf{q^T P}$)

the net heat fluxes. The interannual variability detected is likely related to the oceanic adjustment to variations in these forcings. The net heat fluxes contribute to the thermosteric buoyancy forced component of SSH, and the wind stresses contribute to the steric advective components.

The SSH, meridional and zonal wind stress fields are used to decompose the ocean currents into the associated geostrophic $\mathbf{u_g} = (u_g, v_g)$ and Ekman components $\mathbf{u_e} = (u_e, v_e)$,

$$\mathbf{u} = \mathbf{u_g} + \mathbf{u_e}. \tag{22}$$

The Ekman components are calculated from,

$$v_e = -\frac{\tau_s^x}{f \rho_0 d_{Ek}} \quad \text{and} \quad u_e = \frac{\tau_s^y}{f \rho_0 d_{Ek}}, \tag{23}$$

where f is the Coriolis parameter, τ_s^x and τ_s^y are the zonal and meridional components of the wind stress, τ_s, at the ocean's surface. The density, ρ_0, and the Ekman depth, d_{Ek}, are taken to be constants of 1025 kg/m^3 (a typical value at the surface of the North Atlantic (Wang et al. 2010)) and 100m (a typical value in the Subtropical gyre in the winter (Stommel 1979)), as most of the variability in the Ekman

velocity component is due to variations in the wind stress. The Ekman pumping velocity is also calculated as

$$w_e = \frac{1}{f\rho_0}(\nabla \times \boldsymbol{\tau}_s), \tag{24}$$

and the geostrophic currents are calculated as

$$v_g = \frac{g}{af\cos\phi}\frac{\partial \eta}{\partial \lambda} \quad \text{and} \quad u_g = -\frac{g}{af}\frac{\partial \eta}{\partial \phi}, \tag{25}$$

where η is the monthly mean SSH, a is the radius of Earth, ϕ is latitude, λ is longitude and g is gravity.

These fields and the net heat fluxes are then regressed against the normalized time series of the three leading predictable components (Fig. 11). The fields are smoothed with a 6-month running mean before regression, to focus on the interannual variability. The regression coefficients of the geostrophic currents and the first predictable component are shown in Fig. 12. The regression coefficients with SSH are also shown as contours. These show a westward propagation of SSH in the subtropical gyre. The meridional geostrophic velocities have large regression coefficients with the leading predictable component, in the subpolar gyre along the Canadian east coast. At times where the current is leading the predictable component time series, there is also a positive signal at the Gulf Stream's detachment point. The lack of a clear lead–lag relationship here makes causality hard to distinguish. However, from these strong correlations, it is apparent that there is significant interannual predictability present in the western boundary current in the subpolar gyre. There is a lack of any apparent changes in the large-scale patterns of the zonal geostrophic current regression
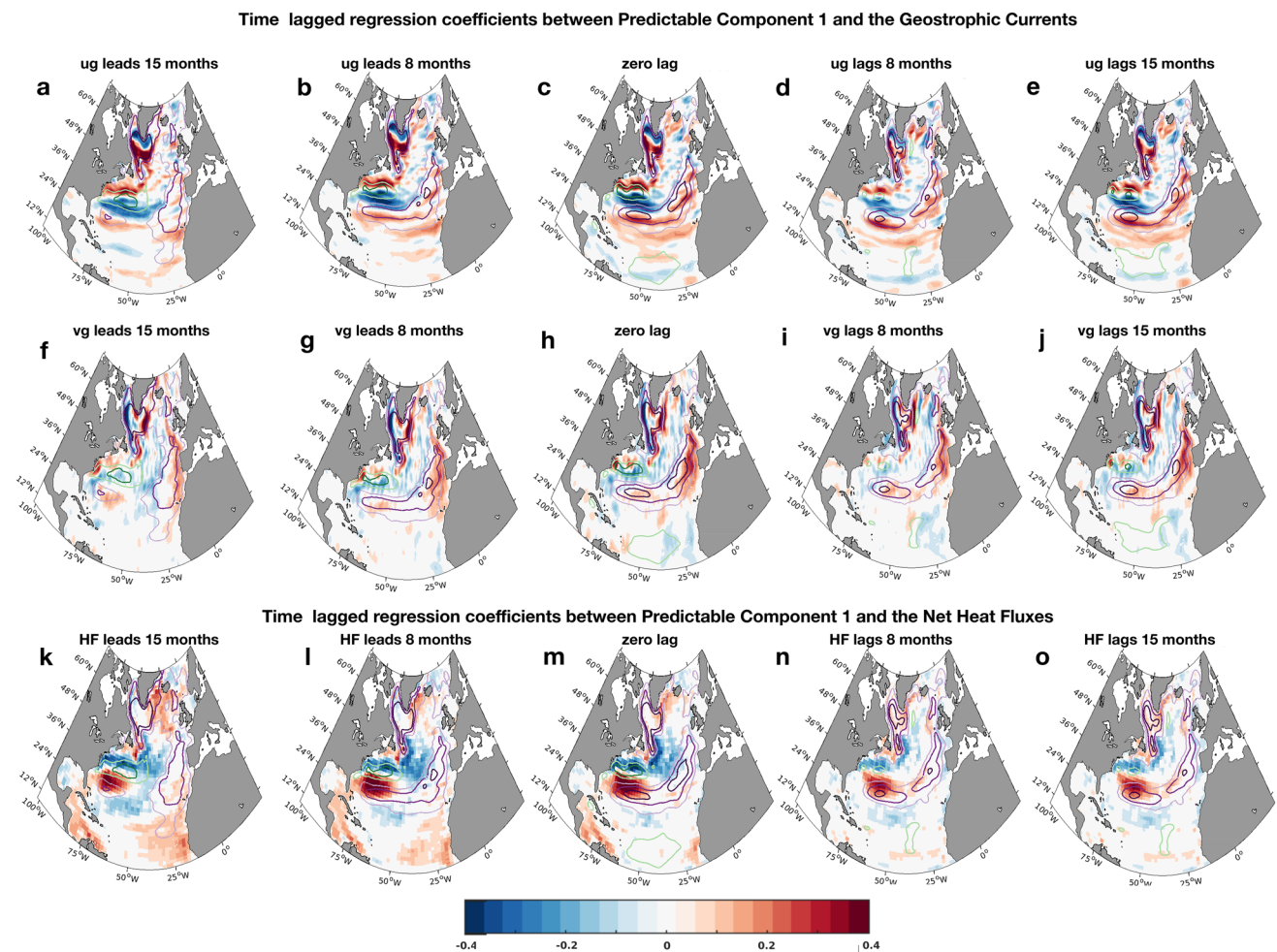

Fig. 12 Time lagged linear regression coefficients between the time series of the first predictable component and monthly mean zonal and meridional geostrophic currents, created using 150 years of control run output. Only regression coefficients with p values of < 0.05 are retained. Units of the regression are m/s for those involving the geostrophic components and W/m^2 for those investigating the net heat fluxes. The contours show the time-lagged regression coefficients of the predictable component and SSH at 0.1 m intervals. The maximum regression coefficient magnitude is 0.3 m. The green contours denote positive regression coefficients, whereas, the purple contours denote negative coefficients

coefficients. These coefficients are large in the subtropical gyre and in the Gulf Stream region. The regression coefficients relating to the net heat fluxes have a sizeable dipolar pattern in the Gulf Stream region at all times. However, in the east of the Subpolar gyre, a strong signal appears at lead times of 15 months.

The regression coefficients of the leading predictable component with the Ekman currents are shown in Fig. 13. There is an apparent time-lagged relationship present, with variations in the Ekman currents leading the predictable component strongly on timescales of up to 15 months. The Ekman currents are associated with the large-scale Sverdrup transport, within the wind-driven gyres. These regression coefficients imply that wind-driven variations in the gyre circulations lead to a predictable change in SSH on interannual timescales. The associated changes in gyre scale variations of the Ekman currents translate to variations in SSH in the

Gulf Stream extension and subpolar gyre regions. This result is also indicative that interannual forecasts of SSH can be improved by better representing the zonal and meridional wind stress fields, on longer than monthly timescales.

It is difficult to discern anything about the variability of the time-lagged geostrophic regression coefficients and predictable component 2 (the figures relating to the regression analysis for the second and third predictable components are contained in the supplementary information). However, there is a signal in the regression coefficients of the net heat fluxes which leads the predictable component by 8–15 months. This signal is located in the Gulf Stream extension region. The regression coefficients relating to the zonal and meridional Ekman components also strongly lead the predictable component and are related to changes in the wind stress in the east of the North Atlantic and the subpolar gyre.
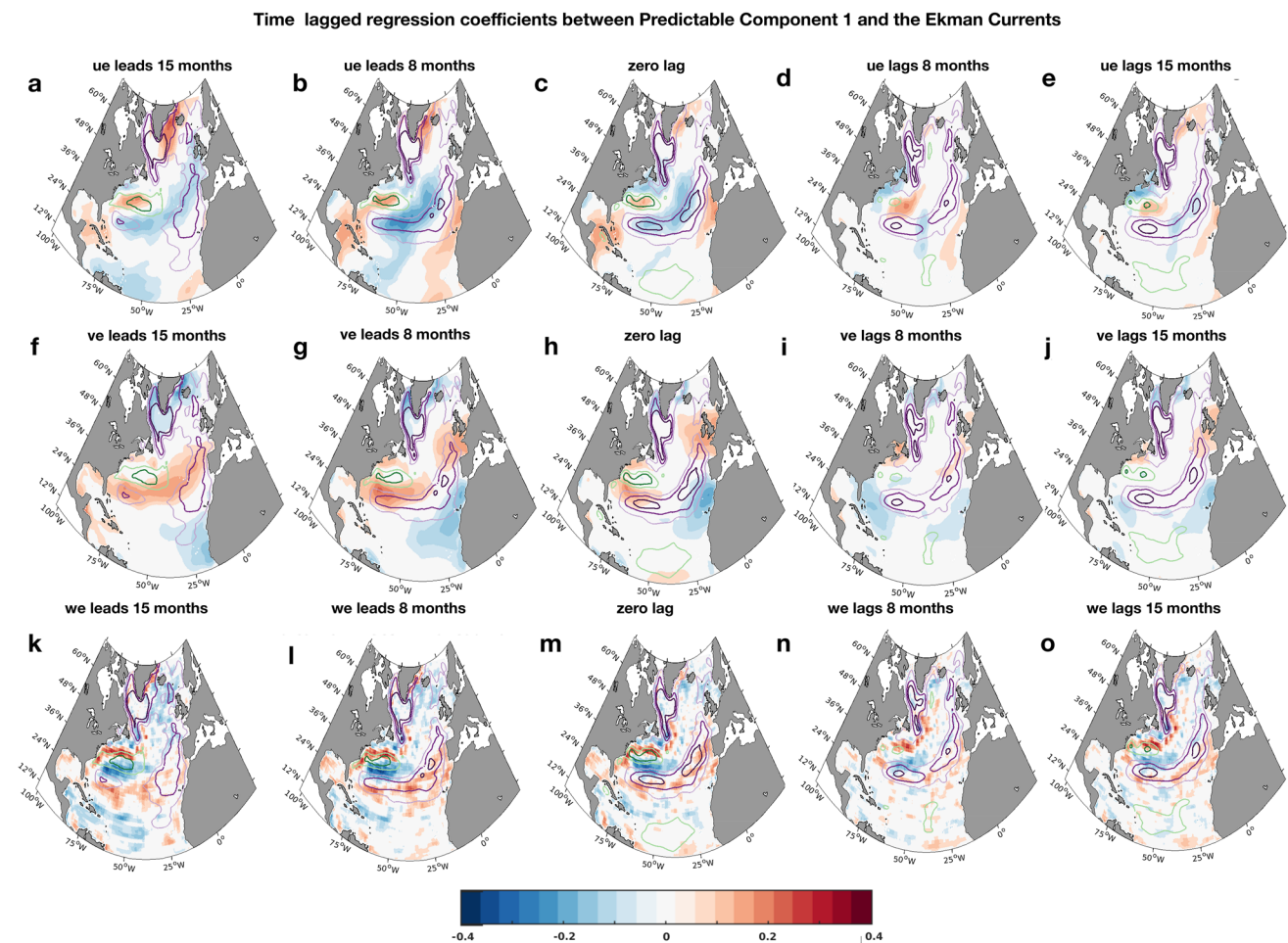

Fig. 13 Time lagged linear regression coefficients between the time series of the first predictable component and monthly mean zonal and meridional Ekman currents and the Ekman pumping/suction velocity, created using 150 years of control run output. Only regression coefficients with p values of < 0.05 are retained. Units of the regression are m/s. The contours show the time-lagged regression coefficients of the predictable component and SSH at 0.1 m intervals. The maximum regression coefficient magnitude is 0.3 m. The green contours denote positive regression coefficients, whereas, the purple contours denote negative coefficients

The third predictable component's regression with the Ekman components demonstrate a clear lead–lag relationship. The Ekman components are seen to lead variations in the predictable component. The patterns of the regression coefficients are large scale and sizeable in the west of the basin. The meridional components of the Ekman currents can be interpreted as causing a convergence or divergence of SSH in the Gulf Stream region as the gyres react to changes in the wind stress. There are also variations in the net heat fluxes in the subpolar gyre, which lead a change in the predictable component.

Therefore it is concluded that predictable component one is largely a response to variations in the wind at the latitudes of the Gulf Stream. Predictable component two is likely due to variations in both the net heat fluxes and the wind stress in the subpolar gyre and in the Gulf Stream extension region. Finally, predictable component three is likely due to the oceanic adjustment resulting from a combination of both variations in the wind stress in the subpolar gyre and east of the ocean basin, and the response to variations in the net heat fluxes in the subpolar gyre. All three patterns show that changes in the atmospheric forcings lead large-scale predictable patterns of SSH. Even though the variations in the wind stress and net heat fluxes are unpredictable at times longer than a few months, the oceans adjustment to them appears to be predictable on timescales of approximately 1–2 years. However, further experiments and analysis are needed to determine how the dynamical processes present generate the diagnosed predictable components.

5 Summary and discussion

The predictability of SSH in the North Atlantic in a control run of a fully coupled model (HadGEM3) was evaluated using methods based on linear inverse modeling and average predictability time. The key findings from this study include:

– predictability of SSH in the subpolar gyre and along the west coast of the Atlantic basin on timescales of up to 20 months (using LIM).
– Short predictability times in the Gulf Stream extension region (5–10 months).
– Optimal initial conditions resulting in regional SSH changes in the subpolar and subtropical gyres and a change in SSH gradient along the Gulf Stream's extension over a timescales of approximately 20 months. The optimals consist of a weak large-scale SSH tripole, with a stronger signal at the Gulf Stream's detachment point.
– large-scale predictable patterns on timescales of 26–28 months, characterized by SSH variations of order 5–10 cm along the US east coast, extending to the gyre scale, which are predictable on timescales of 26–28 months.

These predictable patterns are calculated using a complementary method to LIM, namely APT, which is independent of target time.
– these predictable components correlate significantly with persistent, large-scale, evolving features in SSH, and which appear to be induced by wind and heat flux forcing in the preceding 8–15 months.

To our knowledge, this is the first analysis of SSH using such a comprehensive set of linear predictability methods. These linear methods provide a computationally inexpensive alternative to ensemble modelling techniques (Hawkins and Sutton 2009). There is an expected trade-off between a linear approximation of the dynamical system and computational savings. However, the use of temporal filtering or averaging appears to improve interannual predictions, most likely because the direct influence on the variability from the strongly nonlinear ocean mesoscale is removed. For example, optimal initial conditions of SSH identified by the non-normal mode analysis, when pattern matched in the full, nonlinear forward model, evolve at 15–20 months as predicted by the linear method for almost 80 % of events.

The interannual ocean variability in mid-latitude jet extensions is dominated by the intrinsic component (Sérazin et al. 2015). Our study shows that the variability in the Gulf Stream extension, is not generally predictable on interannual timescales. As the predictability in SSH in the turbulent jet extension regions is limited to less than order 5–10 months (in agreement with Nonaka et al. (2016) and Roberts et al. (2016)). However, in the subpolar gyre and in areas of the subtropical gyre significant predictive skill was found on timescales over a year. As, predictability in SSH and ocean dynamics might be enhanced via atmospheric forcing integrated over large-scale regions (Cabanes et al. 2006). In addition, SSH predictability along the US east coast might be influenced by the position and strength of the Gulf Stream. The timescales and patterns of predictability of SSH in the North Atlantic derived from statistical forecasts trained on model output are comparable to those in Roberts et al. (2016), diagnosed using multiple runs of a fully dynamical model indicating that the results are robust to the method chosen.

The maximum amplification of the optimal initial conditions occurs 20 months after initialisation, which can be used as a predictability timescale and results in several different effects. Firstly, the amplification can lead to a doubling in the magnitude of the initial SSH anomalies in the Gulf Stream region (e.g., an optimal initial condition perturbation with such a tripolar structure and amplitude 3 cm can propagate to give anomalies of 6 cm along the Gulf Stream path). Secondly, the amplification acts to increase (or decrease, depending on the sign) the SSH gradient across the Gulf Stream, leading to a geostrophic velocity anomaly of the

order of 10 cm/s (for an optimal perturbation, P, with pattern and magnitude as in Fig. 7b). Moreover, these changes can lead to a meridional shift of an SSH contour of 0 m by several degrees in latitude (a southward shift of approximately 5° for a perturbation $-2P$; see also Fig. 8b, d). The optimal perturbation, P, results in a 5 cm increase in SSH along the US east coast at latitudes of 30°–40°N (5 cm decrease for a perturbation $-P$). This change is large compared to the recent (1993–2009) rate of global mean SSH rise from satellite altimetry, which is 3.2 ± 0.4 mm year^{-1} (Church and White 2011).

In order to investigate the dynamical evolution of the optimal initial condition the climate model could be restarted with the optimal initial conditions, and with parts of the optimal initial conditions masked (however, restarting the climate model with optimal initial conditions would require multivariate 3D restarts). Such experiments would help examine the role of the oceanic dynamical processes which lead to the growth of the initial conditions, and the relative importance of atmospheric noise and model error.

The optimal initial conditions also have implications for observations. The initial conditions found are indicative that to better constrain interannual predictions of SSH, in the North Atlantic, it would be beneficial to incorporate a higher number of ocean observations (SSH, temperature and velocity fields) in the region near to the Gulf Stream's detachment point. This area has already been the subject of many observational studies (e.g., Line W, Toole et al. (2011)) and is well observed by altimetry (see Lillibridge and Mariano (2013) and references within). Furthermore, initialisation of GCM ensemble simulations with the optimal initial conditions could provide a better estimate of initial condition uncertainty in SSH prediction (Zanna et al. 2018).

This study did not entirely decouple the effects of applying interannual external forcings from the intrinsic variability due to the mesoscale eddy fields. It also made use of a single model, and therefore the optimal initial conditions presented may be model specific. It would therefore be beneficial to examine the SSH predictability with a more extensive ensemble of model simulations, including those which isolate the effects of intrinsic processes (e.g., Sérazin et al. (2015) and Zanna et al. (2018)). Moreover, a comparison with altimetry or higher resolution model output may further elucidate the effects of the eddy field on interannual variability. It would be interesting to assess how a change in spatial resolution affect the diagnosed optimal initial conditions. Such a resolution change might impact the rectification and behavior of the jets, and therefore the diagnosed mode of Gulf Stream variability. Finally, a series of idealised simulations, with selective timescales of the wind and buoyancy forcings, may aid in explaining the dynamical origin of the predictable components. Such ensembles already exist for a range of applications (Gregory et al. 2016; Roberts et al.

2016; Meyssignac et al. 2017). Alternatively, a probabilistic approach as described by Bessières et al. (2017) could be used to disentangle the forced and intrinsic variability components, thus, better explaining the dynamical origin of the predictable patterns.

Numerical APT calculation and significance tests

Two significant difficulties arise when calculating APT using Eq. 17 and a finite time series; (1) the time lag co-variances cannot be estimated at large lags, as the series is finite; (2) there is the danger of over-fitting if the dimensions of state space exceed the number of samples. The first issue is addressed by calculating the APT using a truncated weighted sum, with a lag window, known as the Parzen window, (DelSole and Tippett 2009b)

$$APT = 2 \sum_{\tau=1}^{M} y_\tau S(\tau), \tag{26}$$

where

$$y_\tau = \begin{cases} 1 - 6\left(\frac{\tau}{M}\right)^2 + 6\left(\frac{\tau}{M}\right)^3, & \text{if } 0 \le \tau \le M/2, \\ 2\left(1 - \frac{\tau}{M}\right)^2, & \text{if } M/2 \le \tau \le M, \end{cases} \tag{27}$$

and M is the defined truncation time. To overcome the second difficulty the APT is calculated in EOF space.

Two significance tests are employed to validate the predictable components and ensure the statistical significance of the calculated APT. Firstly, as $\mathbf{R}_\tau^2$ can be interpreted as a multivariate generalization of the correlation coefficient, and therefore its statistical significance can be calculated using standard methods such as the student's t-test (Jia and DelSole 2011). Secondly, the statistical significance of the APT is evaluated relative to the null hypothesis that the principal component time series are unpredictable, as in Jia and

DelSole ([2011](#)). This is done by carrying out several Monte Carlo experiments: for N_s spatial distributions and N_t time steps, we generate $2N_sN_t$ random numbers from a Gaussian distribution with zero mean and unit variance. These distributions are substituted into Eq. [17](#) and the eigenvalues of the problem are calculated. This is repeated 100 times, and the 95th percentile of each eigenvalue is selected. If the APT value calculated from the training data for each component exceeds the eigenvalue's 95th percentile, the hypothesis that there is no predictability is rejected (Jia and DelSole [2011](#)).

References

Andres M (2016) On the recent destabilization of the gulf stream path downstream of cape hatteras. Geophys Res Lett 43(18):9836–9842

Berloff P, Hogg AMC, Dewar W (2007) The turbulent oscillator: a mechanism of low-frequency variability of the wind-driven ocean Gyres. J Phys Oceanogr 37(9):2363–2386. https://doi.org/10.1175/JPO3118.1

Bessières L, Leroux S, Brankart J-M, Molines J-M, Moine M-P, Bouttier P-A, Penduff T, Terray L, Barnier B, Sérazin G (2017) Development of a probabilistic ocean modelling system based on nemo 3.5: application at eddying resolution. Geosci Model Dev 10(3):1091–1106. https://www.geosci-model-dev.net/10/1091/2017/. Accessed 9 Apr 2018

Boer GJ (2004) Long time-scale potential predictability in an ensemble of coupled climate models. Clim Dyn 23(1):29–44

Branstator G, Teng H (2014) Is AMOC more predictable than North Atlantic heat content? J Clim 27(10):3537–3550. https://doi.org/10.1175/JCLI-D-13-00274.1

Branstator G, Teng H, Meehl GA, Kimoto M, Knight JR, Latif M, Rosati A (2012) Systematic estimates of initial-value decadal predictability for six AOGCMs. J Clim 25:1827–1846

Cabanes C, Huck T, Colin de Verdière A (2006) Contributions of wind forcing and surface heating to interannual sea level variations in the Atlantic Ocean. J Phys Oceanogr 36:1739–1750

Chowdhury M, Chu P-S, Schroeder T, Colasacco N (2007) Seasonal sea-level forecasts by canonical correlation analysis–an operational scheme for the us-affiliated pacific islands. Int J climatol 27(10):1389–1402

Church JA, White NJ (2011) Sea-level rise from the late 19th to the early 21st century. Surv Geophys 32(4):585–602. https://doi.org/10.1007/s10712-011-9119-1

Collins M (2007) Ensembles and probabilities: a new era in the prediction of climate change. Philos Trans R Soci A Math Phys Eng Sci 365:1957–1970

DelSole T, Tippett MK (2007) Predictability: recent insights from information theory. Rev Geophys 45(4):1188–1204

DelSole T, Tippett MK (2009a) Average predictability time. Part I: theory. J Atmos Sci 66:1172–1187

DelSole T, Tippett MK (2009b) Average predictability time. Part II: seamless diagnoses of predictability on multiple time scales. J Atmos Sci 66(5):1188–1204

Dias DF, Subramanian A, Zanna L, Miller AJ (2018) Remote and local influences in forecasting pacific sst: a linear inverse model and a multimodel ensemble study. Clim Dyn. https://doi.org/10.1007/s00382-018-4323-z

Farrell B, Ioannou PJ (1996) Generalized stability theory. Part I: autonomous operators. J Atmos Sci 53:2025–2040

Farrell BF (1982) The initial growth of disturbances in a baroclinic flow. J Atmos Sci 39(8):1663–1686

Farrell BF (1990) Small error dynamics and the predictability of atmospheric flows. J Atmos Sci 47(20):2409–2416

Farrell BF, Ioannou PJ (1995) Stochastic dynamics of the midlatitude atmospheric jet. J Atmos Sci 52(10):1642–1656

Farrell BF, Moore AM (1992) An adjoint method for obtaining the most rapidly growing perturbation to oceanic flows. J Phys Oceanogr 22(4):338–349

Gaspar P, Grégoris Y, Lefevre J-M (1990) A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: tests at station papa and long-term upper ocean study site. J Geophys Res Oceans 95(C9):16179–16193. https://doi.org/10.1029/JC095iC09p16179

Gregory JM, Bouttes N, Griffies SM, Haak H, Hurlin WJ, Jungclaus J, Kelley M, Lee WG, Marshall J, Romanou A, Saenko OA, Stammer D, Winton M (2016) The flux-anomaly-forced model intercomparison project (fafmip) contribution to cmip6: investigation of sea-level and ocean climate change in response to co_2 forcing. Geosci Model Dev 9(11):3993–4017. https://www.geosci-model-dev.net/9/3993/2016/. Accessed 14 Mar 2018

Häkkinen S, Rhines PB (2004) Decline of subpolar north atlantic circulation during the 1990s. Science 304(5670):555–559. http://science.sciencemag.org/content/304/5670/555. Accessed 21 Mar 2018

Häkkinen S, Rhines PB, Worthen DL (2011) Warm and saline events embedded in the meridional circulation of the northern north atlantic. J Geophys Res Oceans 116(C3):0148–0227

Häkkinen S, Rhines PB, Worthen DL (2013) Northern North Atlantic sea surface height and ocean heat content variability. J Geophys Res Oceans 118(7):3670–3678. https://doi.org/10.1002/jgrc.20268

Hawkins E, Robson J, Sutton R, Smith D, Keenlyside N (2011) Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach. Clim Dyn 37(2495):2509

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 22:3960–3978

Huddart B, Subramanian A, Zanna L, Palmer T (2017) Seasonal and decadal forecasts of atlantic sea surface temperatures using a linear inverse model. Clim Dyn 49(5):1833–1845. https://doi.org/10.1007/s00382-016-3375-1

Hughes CW, Williams SDP (2010) The color of sea level: importance of spatial variations in spectral shape for assessing the significance of trends. J Geophys Res (Oceans) 115:C10048. https://doi.org/10.1029/2010JC006102

Jia L, DelSole T (2011) Diagnosis of multiyear predictability on continental scales. J Clim 24(19):5108–5124

Legeais J-F, Ablain M, Zawadzki L, Zuo H, Johannessen JA, Scharffenberg MG, Fenoglio-Marc L, Fernandes MJ, Andersen OB, Rudenko S, Cipollini P, Quartly GD, Passaro M, Cazenave A, Benveniste J (2018) An improved and homogeneous altimeter sea level record from the esa climate change initiative. Earth Syst Sci Data 10(1):281–301. https://www.earth-syst-sci-data.net/10/281/2018/. Accessed 3 Aug 2018

Lillibridge JL, Mariano AJ (2013) A statistical analysis of gulf stream variability from 18+ years of altimetry data. Deep sea research part II: topical studies in oceanography modern physical oceanography and Professor H.T. Rossby. 85:127–146

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20(130):141

Madec G (2008) Nemo ocean engine, technical document. https://eprints.soton.ac.uk/64324/. Accessed 17 Jan 2019

Megann A, Storkey D, Aksenov Y, Alderson S, Calvert D, Graham T, Hyder P, Siddorn J, Sinha B (2014) Go5. 0: the joint nerc-met office nemo global ocean model for use in coupled and forced applications. Geosci Model Dev 7(3):1069–1092

Meyssignac B, Piecuch CG, Merchant CJ, Racault M-F, Palanisamy H, MacIntosh C, Sathyendranath S, Brewin R (2017) Causes of the regional variability in observed sea level, sea surface temperature and ocean colour over the period 1993–2011. Surv Geophys 38(1):187–215. https://doi.org/10.1007/s10712-016-9383-1

Miles ER, Spillman CM, Church JA, McIntosh PC (2014) Seasonal prediction of global sea level anomalies using an ocean-atmosphere dynamical model. Clim Dyn 43(7):2131–2145. https://doi.org/10.1007/s00382-013-2039-7

Nonaka M, Sasai Y, Sasaki H, Taguchi B, Nakamura H (2016) How potentially predictable are midlatitude ocean currents? Sci Rep 6(10):20153

North GR, Bell TL, Cahalan RF, Moeng FJ (1982) Sampling errors in the estimation of empirical orthogonal functions. Mon Weather Rev 110(7):699–706

Penduff T, Juza M, Barnier B, Zika J, Dewar WK, Treguier A-M, Molines J-M, Audiffren N (2011) Sea level expression of intrinsic and forced ocean variabilities at interannual time scales. J Clim 24(21):5652–5670. https://doi.org/10.1175/JCLI-D-11-00077.1

Penland C (1989) Random forcing and forecasting using principal oscillation pattern analysis. Mon Weather Rev 117:10

Penland C, Sardeshmukh PD (1995) The optimal growth of tropical sea surface temperature anomalies. J Clim 8:1999–2024

Polkova I, Köhl A, Stammer D (2015) Predictive skill for regional interannual steric sea level and mechanisms for predictability. J Clim 28(18):7407–7419. https://doi.org/10.1175/JCLI-D-14-00811.1

Qiu B, Chen S, Schneider N, Taguchi B (2014) A coupled decadal prediction of the dynamic state of the Kuroshio extension system. J Clim 27(4):1751–1764. https://doi.org/10.1175/JCLI-D-13-00318.1

Rio M-H, Mulet S, Picot N (2014) Beyond goce for the ocean circulation estimate: synergetic use of altimetry, gravimetry, and in situ data provides new insight into geostrophic and ekman currents. Geophys Res Lett 41(24):8918–8925. https://doi.org/10.1002/2014GL061773

Roberts CD, Calvert D, Dunstone N, Hermanson L, Palmer MD, Smith D (2016) On the drivers and predictability of seasonal-to-interannual variations in regional sea level. J Clim 29(21):7565–7585. https://doi.org/10.1175/JCLI-D-15-0886.1

Sasaki YN, Schneider N (2011) Interannual to decadal gulf stream variability in an eddy-resolving ocean model. Ocean Model 39(3–4):209–219. http://www.sciencedirect.com/science/article/pii/S1463500311000771. Accessed 21 Mar 2018

Scaife AA, Copsey D, Gordon C, Harris C, Hinton T, Keeley S, O'Neill A, Roberts M, Williams K (2011) Improved Atlantic winter blocking in a climate model. Geophys Res Lett. https://doi.org/10.1029/2011GL049573

Schneider T, Griffies SM (1999) A conceptual framework for predictability studies. J Clim 12:3133–3155

Sérazin G, Penduff T, Grégorio S, Barnier B, Molines J-M, Terray L (2015) Intrinsic variability of sea level from global ocean simulations: spatiotemporal scales. J Clim 28(10):4279–4292. https://doi.org/10.1175/JCLI-D-14-00554.1

Sonnewald M, Wunsch C, Heimbach P (2018) Linear predictability: a sea surface height case study. J Clim 31(7):2599–2611. https://doi.org/10.1175/JCLI-D-17-0142.1

MA Spall (1996) Dynamics of the Gulf Stream/deep western boundary current crossover. Part II: low-frequency internal oscillations. J Phys Oceanogr 26:2169–2182

Stommel H (1979) Determination of water mass properties of water pumped down from the ekman layer to the geostrophic flow below. Proc Natl Acad Sci 76(7):3051–3055. http://www.pnas.org/content/76/7/3051. Accessed 14 Aug 2018

Toole J, Curry R, Joyce T, McCartney M, Pena-Molino B (2011) Transport of the north atlantic deep western boundary current about 39 n, 70 w: 2004–2008. Deep sea research part II. Topical Studies in Oceanography climate and the Atlantic Meridional Overturning Circulation. 58(17):1768–1780. http://www.sciencedirect.com/science/article/pii/S096706451100021X

Wang C, Dong S, Munoz E (2010) Seawater density variations in the north atlantic and the atlantic meridional overturning circulation. Clim Dyn 34(7):953–968. https://doi.org/10.1007/s00382-009-0560-5

Wang Q, Mu M, Dijkstra AH (2013) Effects of nonlinear physical processes on optimal error growth in predictability experiments of the Kuroshio Large Meander. J Geophys Res Oceans 118(12):6425–6436. https://doi.org/10.1002/2013JC009276

Williams KD, Harris CM, Bodas-Salcedo A, Camp J, Comer RE, Copsey D, Fereday D, Graham T, Hill R, Hinton T, Hyder P, Ineson S, Masato G, Milton SF, Roberts MJ, Rowell DP, Sanchez C, Shelly A, Sinha B, Walters DN, West A, Woollings T, Xavier PK (2015) The met office global coupled model 2.0 (gc2) configuration. Geosci Model Dev 8(5):1509–1524. http://www.geosci-model-dev.net/8/1509/2015/. Accessed 16 Jan 2019

Zang X, Wunsch C (2001) Spectral description of low-frequency oceanic variability. J Phys Oceanogr 31(10):3073–3095. https://doi.org/10.1175/1520-0485(2001) 031<3073:SDOLFO>2.0.CO;2

Zanna L (2012) Forecast skill and predictability of observed Atlantic Sea surface temperatures. J Clim 25(14):5047–5056. https://doi.org/10.1175/JCLI-D-11-00539.1

Zanna L, Brankart JM, Huber M, Leroux S, Penduff T, Williams PD (2018) Uncertainty and scale interactions in ocean ensembles: from seasonal forecasts to multidecadal climate predictions. Q J R Meteorol Soc. https://doi.org/10.1002/qj.3397

Zanna L, Tziperman E (2005) Nonnormal amplification of the thermohaline circulation. J Phys Oceanogr 35(9):1593–1605

Zanna L, Tziperman E (2008) Optimal surface excitation of the thermohaline circulation. J Phys Oceanogr 38(8):1820–1830. https://doi.org/10.1175/2008JPO3752.1