Signal-to-noise and predictable modes of variability in winter seasonal forecasts

Received: 26 September 2022 | Revised: 10 June 2023 | Accepted: 22 June 2023 | Published on: 7 August 2023

DOI: 10.1002/qj.4522

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Signal-to-noise and predictable modes of variability in winter seasonal forecasts

Daniel L.R. Hodson[1] | **Rowan T. Sutton**[1] | **Adam A. Scaife**[2,3]

[1]NCAS, Department of Meteorology, University of Reading, Reading, UK

[2]Hadley Centre, Met Office, Exeter, UK

[3]Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK

Correspondence
Daniel L.R. Hodson, NCAS, Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading, RG6 6BB, UK.
Email: d.l.r.hodson@reading.ac.uk

Funding information
Natural Environment Research Council; DEFRA; BEIS

Abstract

Recent studies suggest seasonal forecasts for European winters are now skilful, but they also identify a "signal-to-noise paradox", wherein models predict the real world more skilfully (higher correlation) than the evolution of their ensemble members. Here, we analyse seasonal hindcasts from the Met Office GloSea5 seasonal forecast system to identify sources of predictability and seek insight into the signal-to-noise problem. For the first time, we use an optimal detection method to identify predictable signals over the North Atlantic region within the forecast system on subseasonal time-scales. We find two primary predictable modes: a Pacific North America (PNA)-like mode and a North Atlantic oscillation (NAO)-like mode. The latter is the leading predictable mode in December–January, and its spatial pattern closely resembles the NAO. The PNA-like mode dominates in January–February. Whereas the PNA-like mode is driven by Pacific Ocean sea-surface temperatures, the NAO-like mode is driven at least partly by Indian Ocean sea-surface temperatures, not solely due to the common trend. We develop a novel method of comparing the magnitude of these modes in the forecast system and observations that complements previous approaches. This suggests that the signal-to-noise problem in GloSea5 is primarily a feature of the December–January NAO-like mode, with the observed mode being three times larger than in the model. The magnitude of the PNA-like mode is better captured by the forecasts, although there is still evidence of a weaker signal-to-noise problem. This suggests particular mechanisms may lead to the lower signal to noise seen in NAO hindcasts, rather than a global weakness of the forecast system in responding to initialization and external forcing. Our results, though specific to GloSea5, provide insights into the causes of the signal-to-noise problem in seasonal forecasts of European winters. They also imply there is significant potential for improving such forecasts and suggest how such improvements may be achieved.

KEYWORDS

forecast, NAO, north Atlantic, predictability, signal-to-noise, skill

1 | INTRODUCTION

Seasonal forecasting of climate has the potential to bring substantial benefits to society. Skilful predictions of climate variables at months, seasons, or even years ahead would improve society's ability to plan and mitigate for future variations in climate. Notable seasonal forecasting skill has been demonstrated for the evolution of the El Niño–Southern Oscillation (ENSO) (Chen *et al.* (2004); Jin *et al.* (2008)) and the consequent climatic impacts driven by ENSO (e.g., Trenberth and Caron (2000)), but much lower forecast skill has been reported at extratropical latitudes. Recently, however, significant skill has been demonstrated in predicting the surface winter North Atlantic oscillation (NAO; Walker and Bliss (1932)) several months (Scaife *et al.* (2014); Athanasiadis *et al.* (2017); Baker *et al.* (2018); Dobrynin *et al.* (2018)), to more than a year ahead (Dunstone *et al.* (2016); Smith *et al.* (2020)). The benefits of such predictive ability could be significant, particularly for the European and North American region. Further studies have demonstrated skill would be of significant value for applications in the power generation industry (Ely *et al.* (2013); Clark *et al.* (2017)), in transport (Palin *et al.*, 2016), and in anticipating seasonal river flows (Bell *et al.*, 2017).

High correlation skill for predictions of the NAO has now been demonstrated in several seasonal forecasting systems (Athanasiadis *et al.* (2017); Baker *et al.* (2018)), but this skill is a specific feature of the "ensemble mean" hindcasts and the correlation skill measure – the reasons for this are discussed extensively in Christiansen (2018). The amplitude of variations in the ensemble mean hindcasts is much less than is found in observations. Moreover, and surprisingly, it has been shown that the ensemble mean provides significantly higher skill (higher correlation) for predicting the observed NAO than it does for predicting a single ensemble member from the same forecast system. This is the "signal-to-noise paradox" (Scaife and Smith, 2018). The evidence suggests that the NAO in GloSea5 and similar seasonal forecast systems may be responding more weakly to sources of predictability than the observed NAO does. This motivates investigation of what these sources of predictability are and what the nature of the responses they drive is. If these features can be identified and examined, it may be possible to begin to understand the cause of this weaker response. To address these challenges, we need to further consider the information provided by ensemble mean hindcasts.

Consider a seasonal forecast model ensemble initialized each year from observations, with some noise added to represent the uncertainty in those initial conditions, generating an ensemble of n forecasts. Each ensemble member will produce a slightly different response, but they will cluster around a common forced or predictable response. The ensemble mean of the resulting n forecasts will be a good (unbiased) estimator of the result we would obtain if we had an infinite ensemble. However, the variance of the ensemble mean will be a biased estimator, as it will contain a contribution from the internal variability (ensemble spread; $\sigma^2_{\text{noise}}/n$).

If we are just interested in the ensemble mean, this does not present a problem. However, the ensemble mean is a superposition of a set of climate modes, and a more detailed analysis of the response of these modes in the forecast involves decomposing the ensemble mean response into the leading modes of variability, often using an empirical orthogonal function (EOF; Wilks (1995)) analysis technique.

EOF analysis requires computing covariances (over time) of the ensemble mean and, as with the variance, these will also contain a contribution from the internal variability. For an infinite ensemble, this contribution from the internal variability $\sigma^2_{\text{noise}}/n$ will vanish and we will be left with a set of modes that are a response to initialization (and external forcing), a set of forced or "predictable" modes. For a small ensemble, however, the contribution from the internal variability may be much larger than any signal present in the forecast. Hence, for a small ensemble, EOFs computed from the ensemble mean may simply be modes of internal variability in the model, rather than the predictable modes we seek. It is unclear how large an ensemble (n) needs to be before the predictable modes emerge from the background noise modes when computing EOFs in this way.

One approach to overcome this and reliably extract the predictable modes in a finite forecast ensemble is optimal detection (Allen and Smith (1997); Venzke *et al.* (1999)). This statistical technique uses estimates of the internal variability (derived from the ensemble spread) to guarantee arriving at the best "unbiased" estimates of the leading predictable modes. It has been successfully used in previous studies to estimate the predictable modes in climate model ensembles (Venzke *et al.* (1999); Sutton and Hodson (2003)), but it has not as yet been applied to a hindcast ensemble from an operation seasonal forecast model. In this article we examine an ensemble of winter (December–February, DJF) hindcasts from the GloSea5 seasonal forecast model and use the optimal detection technique applied to subseasonal (December–January and January–February) and seasonal means (DJF) to extract the leading predictable modes of North Atlantic variability across the season. We also compare these modes with observations.

The article is structured as follows. In Section 2 we discuss the hindcast ensemble produced using the GloSea5 seasonal forecast system and the observational data used

for evaluation, and we provide an outline explanation of the optimal detection method used to extract the leading predictable modes from the hindcast ensemble over the North Atlantic and how we estimate the scaling between the observed and model modes. In Section 3 we present the results, the comparison with observations, and their implications. Finally, we present our conclusions in Section 4. Further technical details concerning the scaling of predictable modes in the hindcasts and observation (defined as γ) and the relationship between the ratio of predictable components (RPC; Eade *et al.* (2014) and γ are presented in the Appendix.

2 | DATA AND METHODS

2.1 | Hindcast ensemble

We analyse data from an ensemble of winter (DJF) seasonal hindcasts/forecasts from the UK Met Office GloSea5 seasonal forecasting system (MacLachlan *et al.* (2015)) for 1993–2021 (28 years). GloSea5 is a seasonal forecasting system consisting of the UK Met Office Atmosphere (GA3.0; Walters *et al.* (2011); Brown *et al.* (2012)) coupled to the NEMO ocean (V3.0; Madec (2016)) and JULES land surface (V3.0; Best *et al.* (2011)) and CICE sea ice (V3.0; Hunke *et al.* (2015)) models. The atmosphere includes 35 levels in or above the stratosphere, with a fixed lid at 85 km. The atmosphere is initialized from observed state estimates from the operational four-dimensional data variational assimilation system for forecast members, and from European Centre for Medium-Range Weather Forecasts Reanalysis–Interim for the hindcast members. The ocean and sea ice are initialized from a short-range ocean and sea-ice data assimilation system (FOAM; Blockley *et al.* (2014)). For more details, see MacLachlan *et al.* (2015). These seasonal hindcast/forecasts were extracted from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu). Hindcasts were available for 1992–2016, and forecasts for 2017–2021.

For each winter hindcast/forecast, the model was initialized from four different start dates in October–November (October 9, 17, 25, and November 1). For the hindcasts (1992–2016), seven ensemble members were launched for each start date, with the model's stochastic physics scheme (MacLachlan *et al.* (2015)) providing perturbed initial conditions, resulting in a 28-member hindcast ensemble. Fifty ensemble members were available for the forecast years (2017–2021). We chose to randomly select 28 members from this ensemble to be consistent with the hindcast ensemble. In the subsequent analysis, we examine this hindcast/forecast ensemble for each winter subseason (DJ and JF) together with the whole winter season (DJF).

To examine how the initialized sea-surface temperatures (SSTs) may drive a DJF atmosphere response, we use the hindcast/forecast ensemble mean SST for the November immediately preceding each December, both for examining spatial correlations and for construction an El Niño (Niño 3.4) index.

2.2 | Observed data

Ensemble mean mean-sea-level pressures (MSLP) from both the European Centre for Medium-Range Weather Forecasts Reanalysis–Interim (Dee *et al.*, 2011) and the National Oceanic and Atmospheric Administration 20th century reanalysis (Compo *et al.*, 2006, https://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html) were used as the observed MSLP datasets in the model–observation comparison. NAO indices (computed from MSLP station data) were taken from the Climatic Research Unit NAO dataset (Jones *et al.*, 1997, https://crudata.uea.ac.uk/cru/data/nao/).

2.3 | Statistical significance

We test for the significance ($P < 0.05$, two tailed) of Pearson's correlation coefficients (correlation, hereafter) in Figures 2–7. We use a table of critical correlation values for ($P < 0.05$, two tailed) (e.g., https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118342978.app2). For more details see von Storch and Zwiers (1999).

Each of the data points in our analysis is separated by a year; and as the intrinsic time-scale of the atmosphere is much shorter than this (on the order of days), we may consider these data points to be statistically independent from an atmospheric point of view. However, serial correlation that arises from forcings external to the atmosphere (changing boundary conditions, multidecadal ocean variability) may lead to an overestimate of the significance of correlations. von Storch and Zwiers (1999) suggest that, unlike tests of the mean, estimates of the significance of correlation are not strongly affected by serial correlation, at least for low correlation values.

Zwiers (1990) examines the impact of serial correlation on the significance of correlation and shows that the 95% significance threshold for correlation between two time series of 20 points is almost unaffected when each time series has a lag 1 autocorrelation of <0.3. We repeated these calculation by computing Monte Carlo estimates of the 95% significance threshold for random time series of length 28 (as used in this article), with

increasing levels of serial correlation (lag 1 autocorrelation, assuming an autoregressive process of order 1), using 1 million repeats for each level. The thresholds are reported in Supporting Information Table S1. We also computed the lag 1 autocorrelation of the Niño 3.4 and NAO indices and the principal components (PCs) for each season (Supporting Information Table S2). Comparing the lag 1 autocorrelations for all the indices and seasons, and their respective adjusted correlation threshold (from Supporting Information Table S1), it is clear that the statistical significance for the indices in Figures 2–4 does not change if we correct for serial correlation in this way.

2.4 | Optimal detection

The goal of this analysis is to identify potential sources of skill in the winter (DJF) GloSea5 hindcasts (Scaife *et al.*, 2014) over the North Atlantic region (102° W–20° E, 0°–80° N). Ultimately, any such skill originates from the observations (atmosphere, ocean, and land) used to initialize the model (observations that are the time integral of the model response to the slowly varying external forcings; e.g., greenhouse gases). The atmosphere in each ensemble member will evolve from its initialized state in response to any common external forcing and forcing from the ocean and land surface (which are in turn slowly evolving from their own initialized states).

Though hindcasts diverge over time (due to growth of initial differences and stochastic physics parametrization; MacLachlan *et al.* (2015)), they will remain clustered around a "predictable" response. We can use standard EOF analysis (Wilks, 1995) to decompose this response into the leading patterns of forced variability, or "predictable modes". For a hypothetical infinite ensemble, these predictable modes are simply the leading EOFs of the ensemble mean. But for a finite ensemble, these EOFs will be contaminated by a contribution from the sampling of internal variability, or noise. For a sufficiently large ensemble, this contamination may be small enough such that a simple EOF of the ensemble mean will be enough to extract modes close enough to the true predictable modes (which indeed turns out to be the case in this study – see Supporting Information Figures S1–S3). But we would not know the extent of this contamination without comparing such an analysis with a robust unbiased estimate of the true predictable modes. However, it is still possible to obtain an unbiased estimate of these predictable modes in a finite ensemble by following the methods of Venzke *et al.* (1999), which we summarize as follows.

For an ensemble of simulations, $\mathbf{X}_j$ is a space × time matrix representing a spatial field, varying in time (forecast date), for ensemble member j. We can write

$$\mathbf{X}_j = \mathbf{X}^{\mathrm{F}} + \boldsymbol{\mu}_j, \tag{1}$$

where $\mathbf{X}^{\mathrm{F}}$ is the "predictable response" common to all ensemble members and $\boldsymbol{\mu}_j$ is internal variability, or noise ($\tilde{N}(0, \sigma_\mu^2)$).

Our goal is to determine the leading modes of variability in the predictable response $\mathbf{X}^{\mathrm{F}}$ by applying PC/EOF analysis (Wilks, 1995):

$$\mathbf{X}^{\mathrm{F}} = \mathbf{E}^{\mathrm{F}}\boldsymbol{\Lambda}^{\mathrm{F}}(\mathbf{P}^{\mathrm{F}})^{\mathrm{T}}, \tag{2}$$

where $\mathbf{E}^{\mathrm{F}}$ are the EOFs (spatial patterns), $(\mathbf{P}^{\mathrm{F}})^{\mathrm{T}}$ are the PCs (or associated time series), and $\boldsymbol{\Lambda}^{\mathrm{F}}$ are the singular values (amplitudes) associated with each mode. In order to compute these modes, we need to calculate the covariance matrix $\mathbf{C} = \mathbf{X}_j(\mathbf{X}_j)^{\mathrm{T}}$. For an infinite number of ensemble members, $\mathbf{C} = \overline{\mathbf{X}\mathbf{X}}^{\mathrm{T}}$, where $\overline{\mathbf{X}}$ is the ensemble mean of $\mathbf{X}_j$. But, for a finite number of ensemble members, j, the resulting covariance matrix will be contaminated by $\boldsymbol{\mu}_j$. An alternative approach is to suppress $\boldsymbol{\mu}_j$ before computing the covariance matrix as follows. First, we compute an estimate of the noise, $\boldsymbol{\mu}_j$:

$$\hat{\boldsymbol{\mu}}_j = \mathbf{X}_j - \overline{\mathbf{X}}, \tag{3}$$

where $\overline{\mathbf{X}}$ is $\mathbf{X}_j$ averaged over all ensemble members j. We concatenate the j noise estimates $\hat{\boldsymbol{\mu}}_j$ along the t index and then compute the EOFs of this noise estimate. We retain the first K of these EOFs, ranked by the fraction of the noise variance they represent, and construct a new basis from these K noise EOFs, each weighted by the reciprocal of the magnitude of the noise variance each represents. The ensemble mean $\overline{\mathbf{X}}$ is then projected into this new subspace. Once projected, the ensemble mean now has noise that has unit variance in each direction in the subspace, by construction. Hence, any additional variance must arise from the predictable response $\mathbf{X}^{\mathrm{F}}$. The leading modes of variability in this subspace, therefore, can be used to construct the leading modes of $\mathbf{X}^{\mathrm{F}}$.

Following Venzke *et al.* (1999), the truncation K is chosen as follows. First, we project the ensemble mean $\overline{\mathbf{X}}$ onto the Nth noise EOF and compute the resulting variance in this projection. The ratio of this variance to that of the Nth noise EOF then gives us a measure of the signal-to-noise ratio S_N in the Nth direction in the noise EOF subspace. The running mean of S_N is then computed:

$$R_K = \frac{1}{K}\sum_{N=1}^{K} S_N. \tag{4}$$

We then choose K when R_K first reaches 95% of its total value when including all noise EOFs. Examining

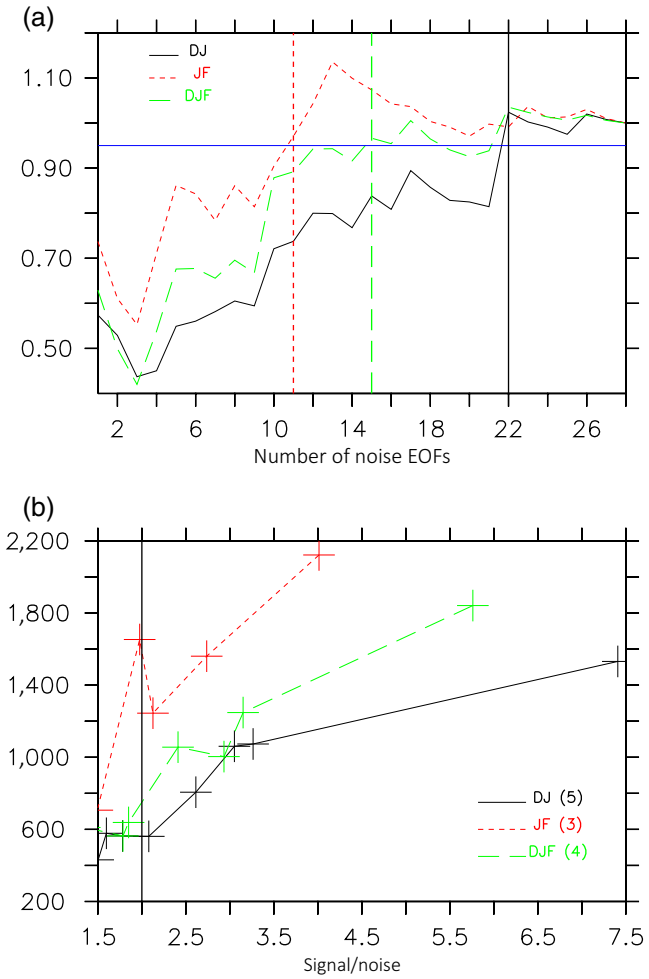

FIGURE 1 (a) Cumulative average ratio of variances (ensemble mean/noise) with inclusion of each successive noise empirical orthogonal function (EOF) for each month (December–January [DJ]: black solid; January–February [JF]: red dotted; December–February [DJF]: green dashed), divided by the ratio when using all noise EOFs (i.e., $n = 28$). The vertical lines show the chosen noise EOF cut-off for subsequent analysis (DJ: 22; JF: 11; DJF: 15), where the ratio first exceeds 95% (blue line) of the value when including all noise EOFs ($n = 28$). (b) Ensemble-mean variance versus signal/noise ratio for each optimal filter when retaining the chosen number of noise EOFs, for DJ (22, black solid), JF (11, red dotted), and DJF (15, green dashed). The vertical line marks the signal/noise ratio of 2, used for the optimal filter truncation. [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 1a, we choose to truncate at $K = 22$ (DJ), $K = 11$ (JF), and $K = 15$ (DJF), as the mean R_K exceeds 95% upon the inclusion of this noise EOF.

We now use these K noise EOFs, combined with the ensemble mean $\overline{\mathbf{X}}$, to compute a set of "optimal filters" that maximize the ratio of ensemble-mean variance to noise variance (Venzke *et al.* (1999)). We then use the first R of these optimal filters to reconstruct the unbiased estimate of $\mathbf{X}^{\mathrm{F}}(\mathbf{X}^{\mathrm{F}})^{\mathrm{T}}$ and hence the leading modes or EOFs. This second truncation (R) is chosen by comparing the variance of the ensemble mean projected onto a given optimal filter with the signal/noise ratio associated with that optimal filter. Figure 1b plots these quantities for DJ, JF, and DJF. We then choose to retain the R optimal filters that have a signal/noise >2; these represent directions that have both high signal/noise *and* capture the most variance in the ensemble mean. Examining Figure 1b, we chose R to be 5 for DJ, 3 for JF, and 4 for DJF. We then use these R optimal filters to compute our unbiased EOFs of $\mathbf{X}^{\mathrm{F}}$.

In summary, the optimal detection procedure is as follows:

- Compute the ensemble mean of the N members.
- Subtract the ensemble mean from each member producing N estimates of the noise.
- Concatenate these N noise estimates along the time axis.
- Compute the EOFs of this, giving the noise EOFs.
- Select the leading K noise EOFs that capture most of the variance in the ensemble mean.
- Project the ensemble mean onto these noise EOFs, each weighted by the reciprocal of their corresponding singular value (eigenvalue). All noise in this projection will now have unit variance.
- Compute the EOFs of this projection. The resulting EOFs will be the directions/modes that have a high signal/noise ratio.
- From these EOFs, select the R leading EOFs that have singular values >2 (signal/noise > 2).
- Project these R EOFs back onto the noise EOFs, each weighted this time by their corresponding singular value. We now have a filtered ensemble mean – a reconstruction of the ensemble mean where noise has been suppressed.
- Compute the EOFs of this filtered ensemble mean. These are the predictable modes.

This represents an outline sketch of the optimal detection method – in order to explain the purpose of the two truncation choices K and R. A more comprehensive description of this procedure can be found in Venzke *et al.* (1999).

2.5 | Estimating predictable response ratio

(Scaife and Smith, 2018) show that the NAO index in the GloSea hindcast has a lower amplitude than that seen in observations, yet both are well correlated (0.62) (Scaife

et al., 2014). This suggests that both the model and the observed NAO are responding to a common source of prescribed external forcing and hence contain a common predictable (forced) signal or mode. However, the amplitude of this forced, predictable mode is smaller in the model than in observations. If we assume that such a predictable (forced) mode exists both in the model hindcasts (X^{model}) and observations (Y^{obs}) and that these modes are identical except for a scaling γ, then we can express the relation between model and observations as:

$$Y^{\text{obs}} = \gamma X^{\text{model}} \tag{5}$$

We can estimate γ using estimates of X^{model} derived from the optimal detection analysis already described herein. This estimate of γ will be a joint property of the forecast model and the observations, so estimates of γ with other seasonal forecast models may differ. We note that γ is equivalent to $1/\beta$ as defined by Siegert *et al.* (2016), and that γ is also related to the RPC as defined by Eade *et al.* (2014); we explore this further in Appendix A.2.

For example, to estimate γ for DJ, we compare the regression patterns for observations regressed (over the time index) on the first PC (the time series associated with the first predictable mode) in Figure 6a with that for the hindcast ensemble mean regressed on the same PC in Figure 6b. Then, for each grid point (k, l), we pair the regression value (Y^{obs}_{kl}) for the observations with the regression value (X^{model}_{kl}) for the hindcast ensemble mean. We exclude any grid points where the regression value is not significant ($P < 0.05$) in either observations or the model.

We then fit a regression line through all these points (Y, X). The resulting regression slope is an estimate of the amplitude ratio γ between the observations and the model. However, this will be a biased estimate of γ, as it is dependent on the ensemble size J (see Appendix A.1). To provide a robust estimate of the limiting value for $J \to \infty$, we subsample the ensemble for each ensemble size up to J (~100 times for each ensemble subsample) and recompute γ. Figure 8 shows this variation for the two leading modes in DJ and JF. We then fit a function to extract an estimate of the value of γ in the limit $J \to \infty$. For more details, see Appendix A.1.

3 | RESULTS

We now present the analysis of the GloSea5 winter hindcasts. We first consider the DJ and JF subseasons, before proceeding to examine the full winter season (DJF). We perform the optimal detection already described herein, on monthly mean MSLP data for each season/subseason, over the North Atlantic region (102° W-20° E, 0°–80° N). We apply the noise and signal/noise truncations described in Section 2.4 and then examine the two leading modes produced by the final steps of the optimal detection analysis described earlier herein. In each case, North's rule (North *et al.*, 1982) shows that these modes are distinct – the values of $(\lambda_1 - \lambda_2)/\lambda_1(2/N)^{1/2}$ (i.e., the ratio of the difference between the first and second eigenvalues, $\lambda_1 - \lambda_2$, to the estimate of the spread, $\lambda_1(2/N)^{1/2}$) are all >1: DJ, 1.02; JF, 1.30; DJF, 1.31. We note that these ratios are still <2, and so, as (North *et al.*, 1982) suggests, there may still be a degree of mixing between these modes (especially in DJ).

3.1 | DJ subseason

Figure 2 shows the two leading predictable mode estimates over the North Atlantic region, both time series (PCs: Figure 2a,b) and spatial patterns (EOFs: Figure 2c,d) from the optimal detection analysis of DJ. We have retained 22 noise EOFs ($K = 22$) and five optimal filters ($R = 5$) (Figure 1). The first leading predictable mode in DJ MSLP explains 35% of the variance in the ensemble mean and has an NAO-like dipole structure (Figure 2c). This mode has a time series (PC) that is significantly correlated with the observed NAO (0.45, $P < 0.05$), but not the observed Niño 3.4 index for the preceding November (Figure 2a). The global extent of this mode is revealed by regressing the ensemble-mean MSLP onto this PC (Figure 2e). Correlating this PC with model SSTs from the preceding November (to avoid capturing ocean responses to the atmosphere mode) shows significant but weak correlations of this mode with the SSTs over the Indian Ocean and west northern Atlantic Ocean (Figure 2g). Similar correlation patterns are produced when November SSTs are correlated with a station-based NAO index (constructed as the observed NAO index: Lisbon–Reykyavik) computed from the ensemble-mean MSLP (Supporting Information Figure S4).

The second predictable mode in DJ explains 19% of the variance in the ensemble mean. The time series (PC) is significantly correlated with an observed Niño 3.4 index (0.80, $P < 0.05$), but not with the NAO index. Regressing the ensemble-mean MSLP onto this PC reveals a structure similar to the ENSO-related Pacific North America (PNA) wave-train pattern (Wallace and Gutzler, 1981), with a notable low pressure centred over the UK (Figure 2d,f). This PC is strongly correlated with tropical Pacific SSTs from the preceding November in a characteristic El Niño structure (Figure 2h) .

Interestingly, these leading predictable modes are very similar, spatially and in time, to the two leading modes from a standard EOF analysis of the ensemble mean (Supporting Information Figure S1), suggesting that, for DJ, the forced responses in these predictable modes may be

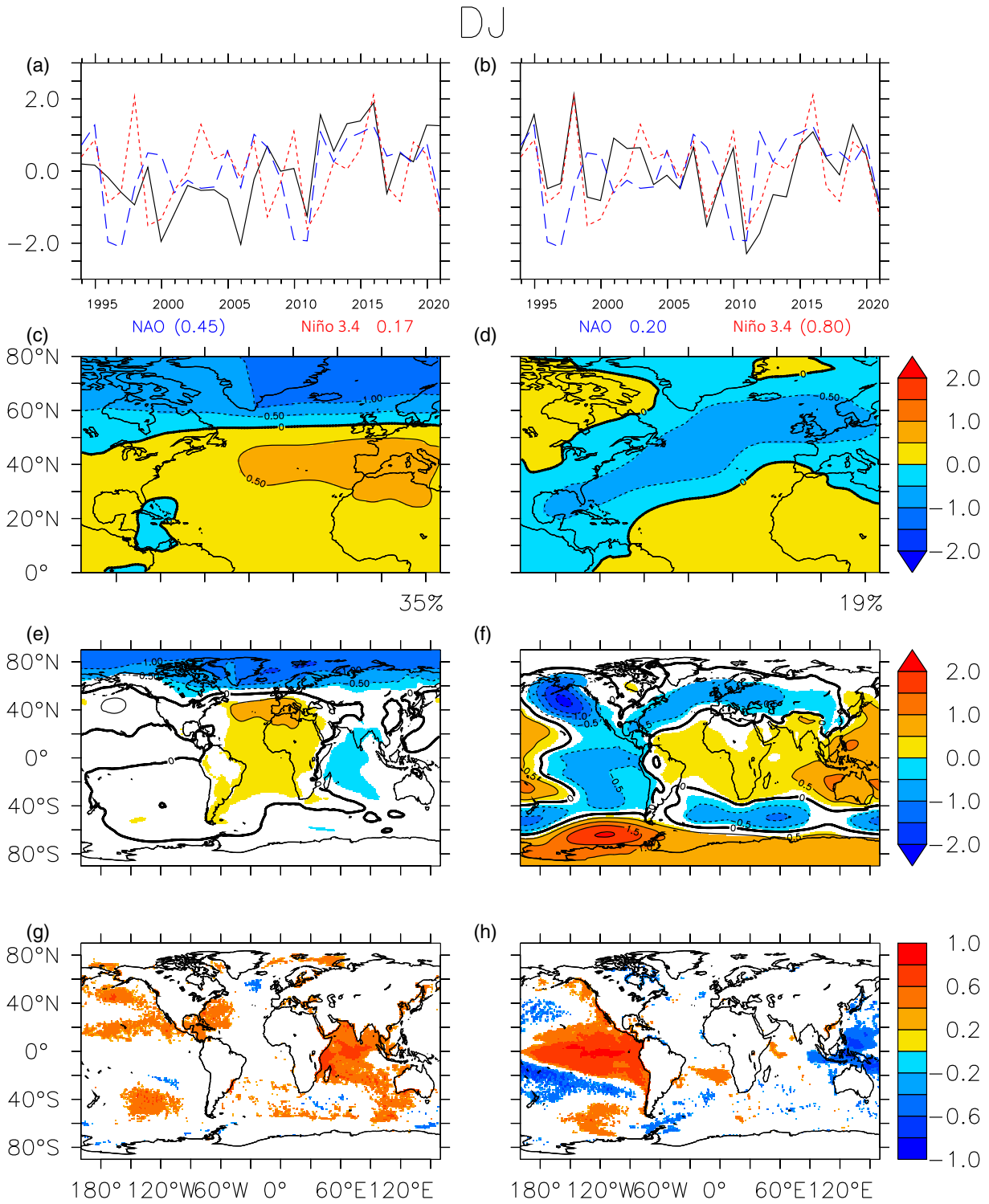

FIGURE 2 The first two predictable modes of December–January (DJ) mean-sea-level pressure (MSLP; retaining 22 noise empirical orthogonal functions and the first five signal-to-noise-maximizing empirical orthogonal functions). (a) Black-solid: time series (principal component [PC]) of the first predictable mode. Blue-dashed: observed DJ North Atlantic oscillation index (Climatic Research Unit). Red-dotted: Niño 3.4 index (hindcast ensemble mean sea-surface temperature from preceding November averaged over 120°–170° E, 5° S–5° N). Correlations between the PC and each index are shown below the panel. Correlations in parentheses are significant ($P < 0.05$). Time axis shows year that hindcast was started (e.g., October–November). (c) First predictable mode of DJ MSLP. Percentage below panel is fraction of the variance this mode explains of the ensemble mean. (e) Hindcast ensemble-mean MSLP regressed onto the first (standardized) PC. Units: hPa/standard deviation of PC. Contour interval 0.5 hPa. (g) Ensemble-mean hindcast sea-surface temperature from the preceding November correlated with PC. In (e) and (g), regions are only where the correlation is significant ($P < 0.05$). (b, d, f, h) as (a, c, e, g) but for second predictable mode. [Colour figure can be viewed at wileyonlinelibrary.com]

sufficiently strong that the 28-member ensemble is large enough for a standard EOF analysis to give a good estimate of the predictable modes. However, it would be difficult to know this was the case without the comparison with the optimal detection results.

3.2 | JF subseason

Figure 3 shows the two leading predictable mode estimates over the North Atlantic region in JF. We have retained 11 noise EOFs ($K = 11$) and three optimal filters ($R = 3$) (Figure 1). The first predictable mode (Figure 3c) explains 30% of the ensemble mean variance and is again a north–south dipole pattern, but with pronounced tilt. The PC (Figure 3a) is significantly correlated with the observed Niño 3.4 index from the preceding November (0.77, $P < 0.05$), but not the JF NAO. The global regression (Figure 3e) reveals a PNA-like pattern again, with significant correlations with tropical Pacific SSTs from the preceding November (Figure 3g).

The second mode (Figure 3d) explains 13% of the ensemble mean variance and features a band of anomalous high pressure between the Labrador Sea and the Iberian Peninsula. The PC (Figure 3b) is not significantly correlated with either the Niño 3.4 or the NAO index, and the regression (Figure 3f) weakly projects onto a PNA-like pattern, consistent with the weak tropical SST correlations (Figure 3h).

For JF, unlike DJ, the leading predictable modes are less similar to those from a standard EOF analysis applied to the ensemble mean (Supporting Information Figure S2); the optimal detection captures more of the El Niño variability in the leading mode, suggesting that the standard EOF analysis results are more contaminated with internal variability, perhaps due to the longer time since initialization, compared with DJ.

3.3 | DJF season

Over the full winter season (DJF), we retain 15 noise EOFs ($K = 15$) and four optimal filters ($R = 4$) (Figure 1). The first predictable mode explains 38% of the ensemble mean variance, and is again a north–south dipole pattern (Figure 4c). The PC is weakly, but significantly, correlated with the observed November Niño 3.4 index (0.39, $P < 0.05$) but not the DJF NAO. The global regression (Figure 4e) reveals a PNA-like pattern again, with weak but significant correlations with tropical Pacific SSTs from the preceding November (Figure 4g).

The second mode (Figure 4d) explains 16% of the ensemble mean variance, and the PC is strongly correlated (0.89, $P < 0.05$) with the November Niño 3.4 index (Figure 4b). The regression (Figure 4f) is again a PNA-like pattern, consistent with the strong El Niño-like tropical SST correlations (Figure 4h).

Again, these leading predictable modes over the North Atlantic region for DJF are very similar to those from a standard EOF analysis applied to the ensemble mean (Supporting Information Figure S3), suggesting that, with 28 ensemble members, averaging over the extended season may be enough to extract the predictable modes. Although, again, we can only know this by comparison with the optimal detection results.

In summary, the two leading predictable modes over the North Atlantic region during winter are an NAO-like mode and a PNA-like mode. The NAO-like mode is more prominent in early winter (DJ) and is significantly correlated with the observed NAO, whereas the PNA-like mode is more prominent in late winter (JF) but maintains a significant correlation with a November El Niño index throughout winter.

3.4 | Drivers

In JF, the time evolution of the leading mode over the North Atlantic (Figure 3a) is significantly correlated with a Niño 3.4 index of the preceding November hindcast SSTs. This relationship is clearer in Figure 3g, where correlations show a clear El Niño-like pattern across the Pacific. We can be confident, therefore, that the JF leading predictable mode is primarily a PNA-like mode driven by El Niño SST anomalies.

Whereas the JF PNA-like mode is associated with an El Niño SST, the NAO-like DJ mode is associated with a strong preceding November SST signal in the Indian Ocean in DJ (Figure 2g). Further examination reveals that the majority of this Indian Ocean signal is due to the warming trend in the region projecting onto the trend in the NAO-like mode PC (Figure 2a). Removing these trends in the DJ PC1 and November SST greatly reduces this signal (Figure 5a) but does not remove it entirely. Comparing a region box-average of November detrended Indian Ocean SSTs with a detrended DJ PC1 (Figure 5b) reveals a significant remaining correlation (0.56) – a similar, but slightly weaker, correlation (0.47), is found using the ensemble mean hindcast November SSTs. This detrended DJ PC1 also remains significantly correlated with the observed DJ NAO (0.39). A similar picture emerges if we use a station-based (Lisbon–Reykyavik) NAO index computed from the ensemble mean (Supporting Information Figure S4).

This correlation suggests that variability in the Indian Ocean is potentially responsible for some of variability

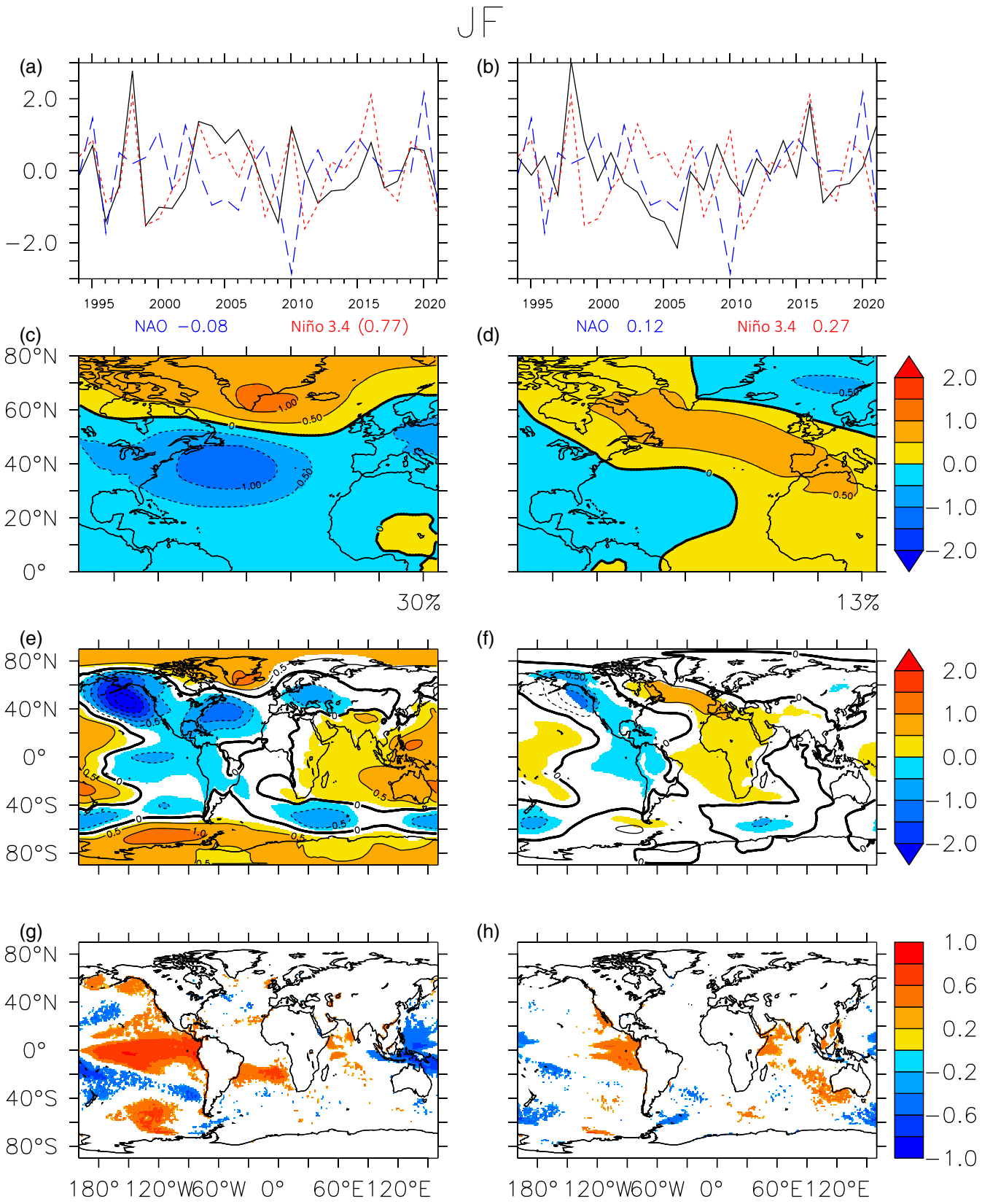

FIGURE 3 As Figure 2, but for January–February (JF), retaining 11 noise empirical orthogonal functions and three signal-to-noise-maximizing empirical orthogonal functions. Again, the time axis shows year that hindcast was started (e.g., October–November); hence, 1996 would mean January 1997, as forecast was initialized in October–November 1995. [Colour figure can be viewed at wileyonlinelibrary.com]

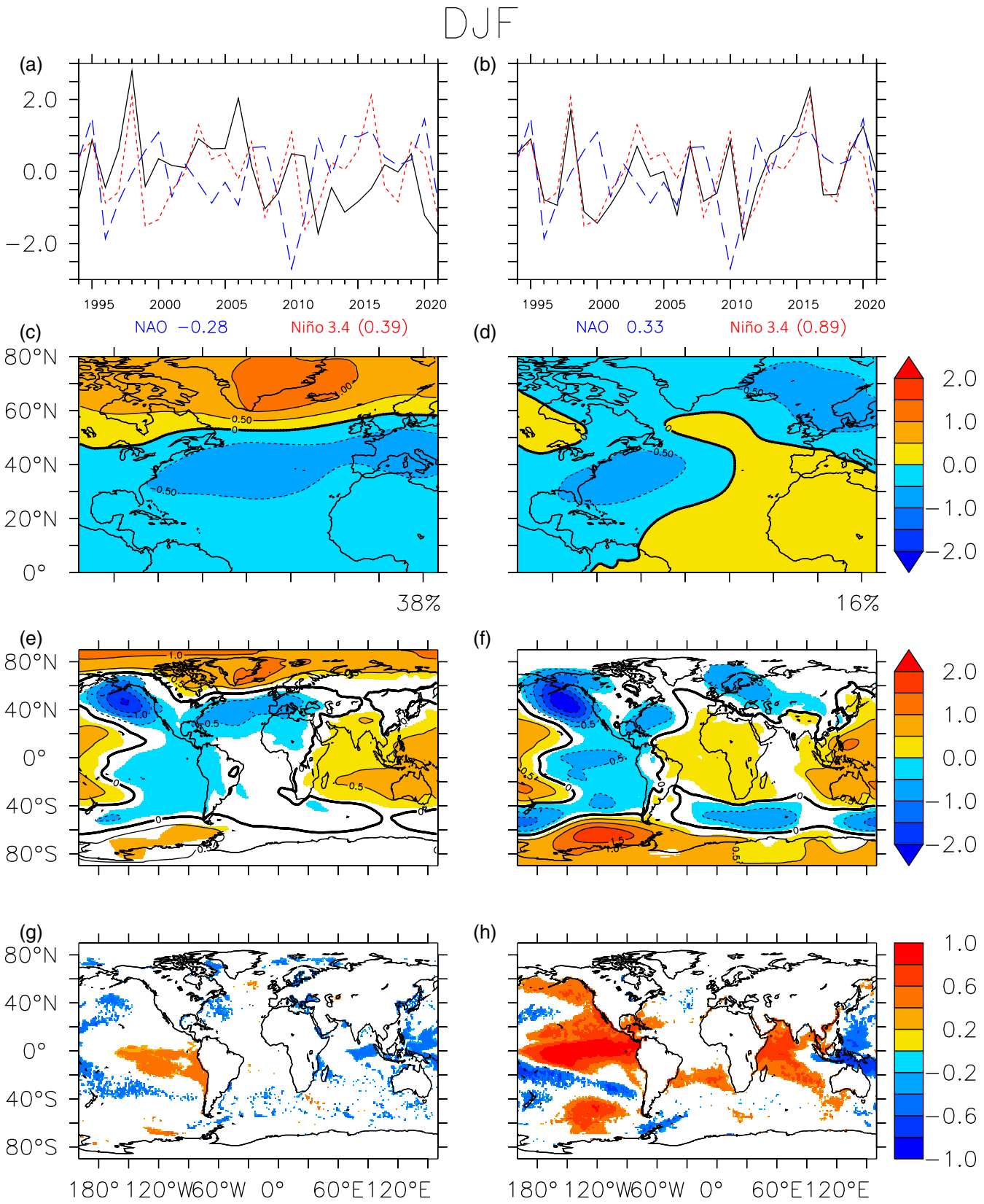

FIGURE 4 As Figure 2, but for December–February (DJF), retaining 15 noise EOFs and four signal-to-noise-maximizing EOFs (T15 SN0-3). [Colour figure can be viewed at wileyonlinelibrary.com]

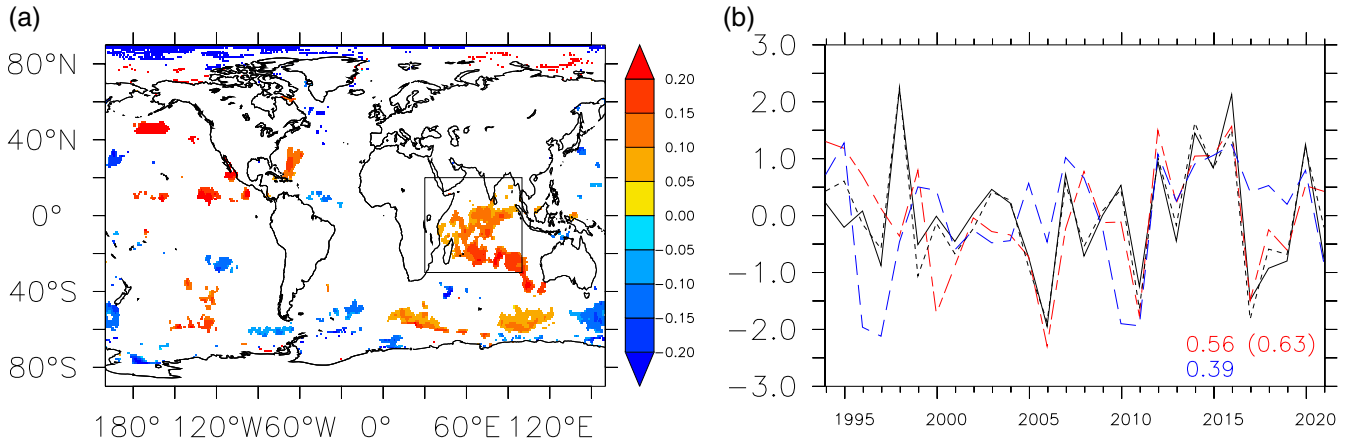

FIGURE 5 (a) Regression of detrended November ensemble-mean hindcast sea-surface temperature (SST) onto detrended December–January (DJ) first principal component (PC1; Figure 2a) – as Figure 2g. Units: °C/standard deviation. Shaded regions are where the correlation between SST and the PC is significant ($P < 0.05$). (b) Black-solid: index of detrended November ensemble-mean hindcast SST averaged over Indian Ocean box in (a). Black-dotted: as black-solid index, but only averaging over shaded (significant, $P < 0.05$) region in box in (a). Red dash-dotted: detrended DJ PC1 (Figure 2a). Blue dashed: North Atlantic oscillation index (Climatic Research Unit; as Figure 2). Red figure in lower right: correlation between Indian Ocean index (black) and detrended DJ PC1 (red). Correlation between dashed line and detrended DJ PC1 is given in parentheses. Blue figure in lower right: correlation between detrended DJ PC1 and observed DJ North Atlantic oscillation index (Climatic Research Unit; as Figure 6c). [Colour figure can be viewed at wileyonlinelibrary.com]

in early winter DJ NAO-like predictable mode. Such a link is consistent with previous studies examining the impact of tropical SSTs on the global circulation (Molteni *et al.* (2015); Scaife *et al.* (2017); Abid *et al.* (2021)). Rodwell and Folland (2002) demonstrated that the observed NAO can be partly reproduced in an atmospheric general circulation model forced by observed SSTs, with observed and model NAO having a correlation of 0.41. Further studies suggest that this NAO variability is driven from the Tropics (Hoerling *et al.*, 2001) but that the amplitude of the NAO signal is smaller than observed (Hurrell *et al.*, 2004). An earlier study by Farrara *et al.* (2000) had suggested a link between the NAO and the Indian Ocean. Later studies demonstrated the Indian Ocean warming trend could drive a trend in the model NAO (Hoerling *et al.* (2004); Bader and Latif (2005); SanchezGomez *et al.* (2008)). Furthermore, Cassou (2008) showed that the phase of the Madden–Julian oscillation in the Indian Ocean influences the NAO. A further recent study by Hardiman *et al.* (2020) demonstrates a clear link between the Indian Ocean dipole and the predictability of the 2019–2020 NAO, detailing a teleconnection pathway that extends over the Atlantic and into the polar stratosphere, leading to an enhanced positive NAO. Examination of Figure 5 suggests that a significant fraction of the variability in this NAO-like predictable mode is either driven by Indian Ocean SST variability (both the trend and the variability about that trend) or both are responding to a common external forcing. These previous studies suggest that it is the Indian Ocean SST variability that is driver in this case and that models significantly underestimate the observed magnitude of this response. It is also important to acknowledge that some of the predictability of the NAO-like mode may arise from the atmospheric initial conditions (Stockdale *et al.*, 2015).

3.5 | Comparison with observations

How does the spatial pattern of the NAO-like leading predictable mode compare with the corresponding mode in observations? We can estimate this by repeating the regression of MSLP onto the first PC (Figure 2e) using observed MSLP rather than the hindcast ensemble-mean MSLP.

The agreement between the spatial patterns for observations (Figure 6a) and the model (Figure 6b) is strikingly clear across the whole Atlantic. However, the magnitude of the regression pattern for the observations is approximately three times larger than for the hindcast. This scaling is further illustrated in Figure 6c, which compares the NAO-like time series produced by projecting MSLP of the hindcast ensemble members and observations onto the DJ NAO-like mode (Figure 2c). The observed NAO-like indices usually fall within the hindcast ensemble range, but the variance of the ensemble mean hindcast NAO-like index is much lower than that of the observations in spite of the correlation between the two. This is the signal-to-noise problem described by Eade *et al.* (2014), Dunstone *et al.* (2016), and Scaife and Smith (2018).

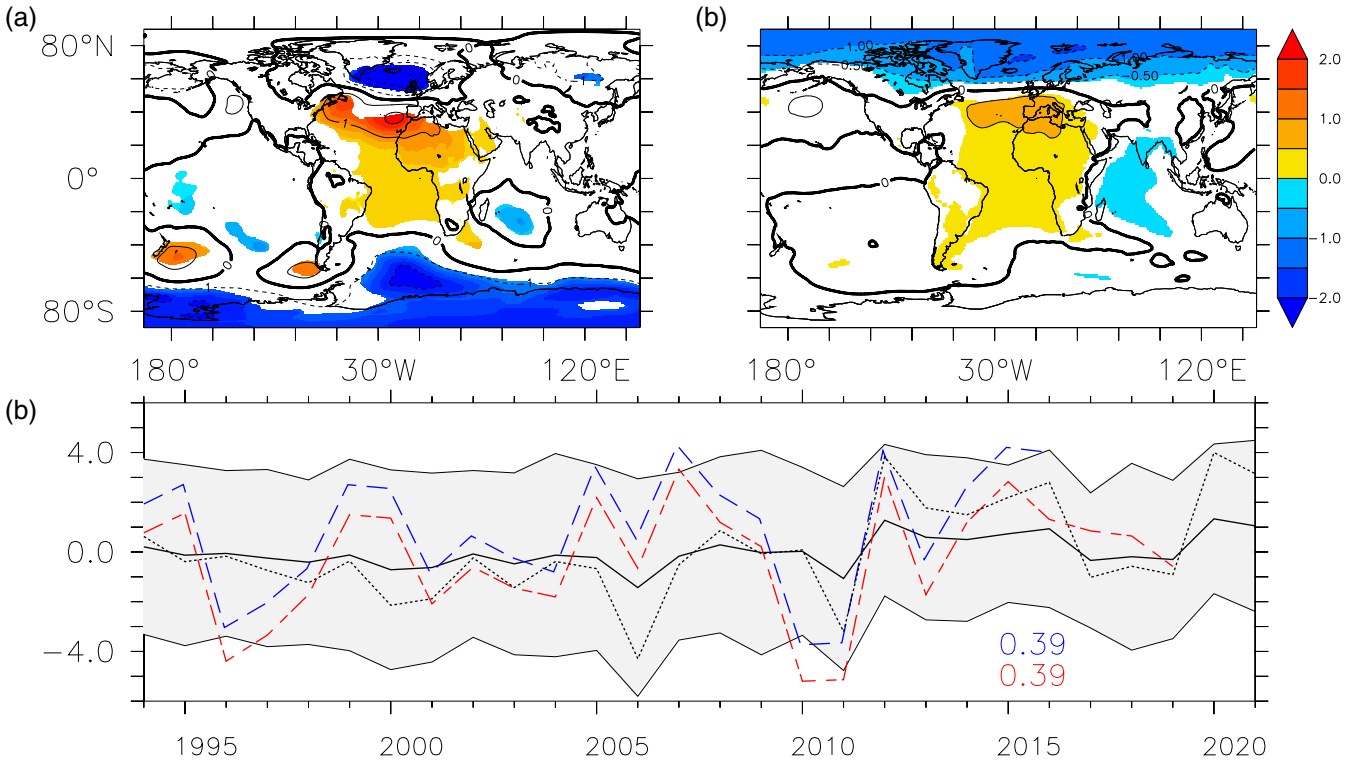

FIGURE 6 Comparison of December–January (DJ) North Atlantic oscillation (NAO)-like predictable mode in model and observations. (a) Observed (European Centre for Medium-Range Weather Forecasts Reanalysis–Interim) DJ mean-sea-level pressure (MSLP) regressed onto standardized first DJ principal component (PC; DJ PC1 Figure 2a). Units: hPa/standard deviation. (b) DJ model ensemble-mean MSLP regressed onto standardized DJ PC1. Shaded regions are where the correlation between MSLP and the PC is significant ($P < 0.05$). (c) Model ensemble-mean NAO index (black solid) and multiplied by 3 (black dotted). The 2σ (standard deviation) ensemble spread is shaded in grey. Red dash-dotted: NAO index in European Centre for Medium-Range Weather Forecasts Reanalysis–Interim observations (computed by projecting observations onto the first DJ empirical orthogonal function, Figure 2a). Blue dashed: as red dash-dotted, but for National Oceanic and Atmospheric Administration 20C RA. [Colour figure can be viewed at wileyonlinelibrary.com]

Repeating the same analysis for the JF leading predictable (PNA-like) mode reveals a similar, more global, consistency between the observations and the hindcast (Figure 7), and there is also closer agreement on the magnitude of the observed and hindcast regression patterns.

To further investigate the signal-to-noise problem in this relationship between the amplitude of the predictable signals in the model and in observations, we use the approach outlined in Section 2.5 and described in detail in the Appendix. Eade *et al.* (2014) compute an RPC to quantify the scaling between the predicted and observed responses. Here, we make no assumptions about how to construct the leading predictable modes (i.e., by taking differences between observed MSLP at two points to construct an NAO index), which may be biased spatially toward the observations, but simply use the leading modes that emerge from the optimal detection analysis. This allows us to use all the spatial information contained in the regression patterns (Figure 6a,b) to compute the best scaling, γ, between these patterns in the presence of noise. This provides a complementary approach to Eade *et al.* (2014) to quantifying the signal-to-noise problem. The relationship between these two approaches (γ and RPC) is discussed in the Appendix. The estimate of this scaling ratio γ (between the amplitude of the predictable mode in observations and its amplitude in the hindcasts; e.g., Obs = γModel)) is shown in Figure 8. This analysis suggests that the NAO-like mode in DJ in observations has an amplitude that is about three times larger than in the hindcasts, consistent with previous estimates of scale of the signal-to-noise problem (Eade *et al.* (2014); Dunstone *et al.* (2016); Scaife and Smith (2018)). If we repeat this analysis for the second (PNA-like) mode in DJ (Figure 2d), the amplitude is also larger in observations than in the hindcasts, but the discrepancy in amplitude is much less, about 1.4 times larger.

Repeating this analysis for both JF predictable modes reveals a similar picture: the leading PNA-like predictable mode has again a scaling of about 1.4 (as in DJ), and the second JF mode has a slightly higher scaling of 2. These results suggests that the signal-to-noise problem does not affect *all* predictable modes equally.

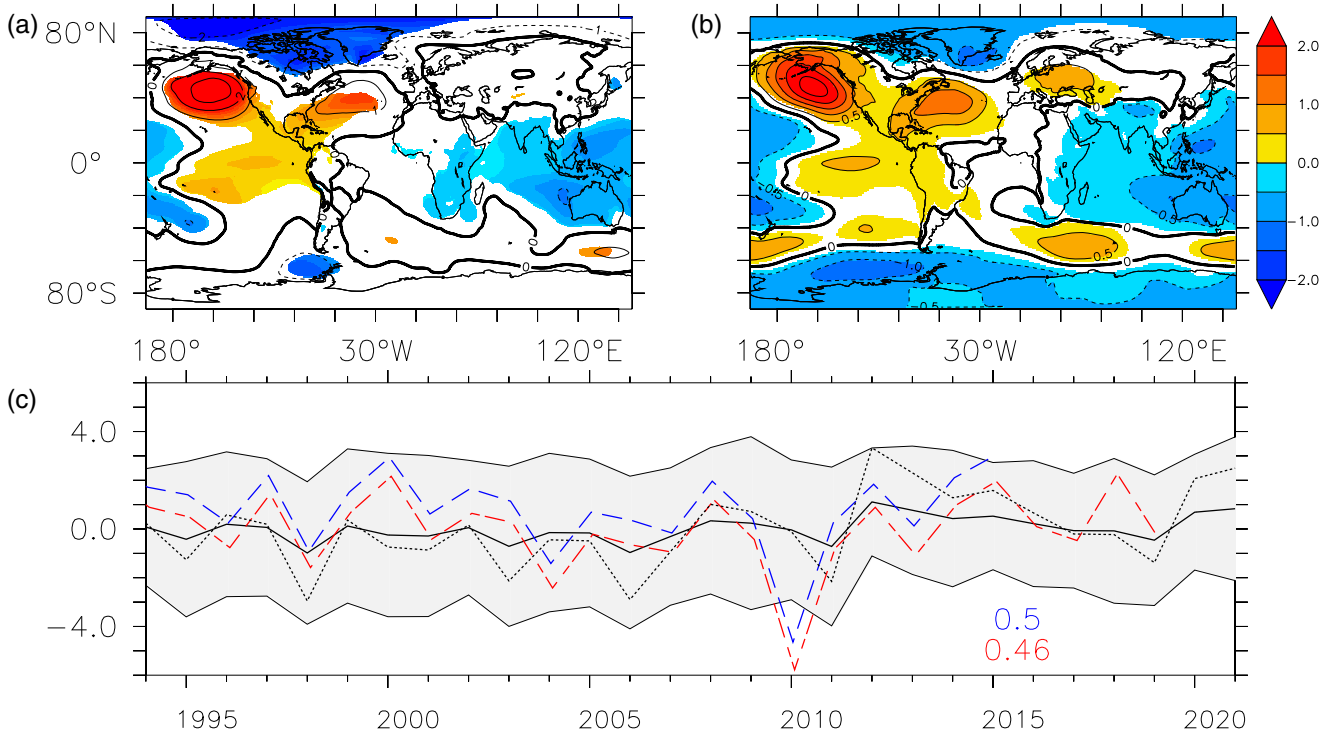

FIGURE 7 As Figure 6, but for January–February leading mode. [Colour figure can be viewed at wileyonlinelibrary.com]

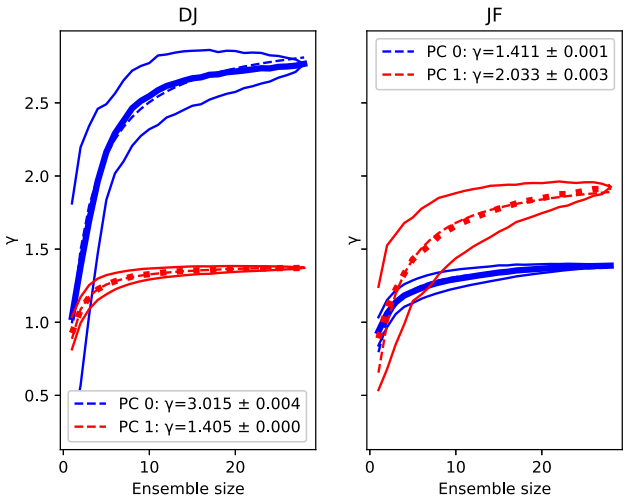

FIGURE 8 The variation of ratio (γ) of the amplitude of the observed to model predictable mode for the first two leading modes in December–January (DJ) and January–February (JF), with size of the model ensemble. First principal component (PC1; blue solid) is the North Atlantic oscillation-like mode in Figure 2e; PC2 (red dotted) is the Pacific North America-like forced mode in Figure 2f. Dashed lines are a fit using the function in Equation A.10. Thin lines show ±1 standard deviation across the subsampling ensemble. [Colour figure can be viewed at wileyonlinelibrary.com]

Overall, our results suggest that the space–time structure of the predictable NAO-like mode is well captured in the GloSea5 hindcasts, but crucially its amplitude is significantly underestimated, and this muted amplitude may point to errors in specific model processes involved in the representation of the NAO in forecast models, such as eddy feedback (Scaife *et al.*, 2019) or ocean–atmosphere coupling (Ossó *et al.*, 2019), and affect its response to Indian Ocean SSTs.

4 | CONCLUSIONS

In this study we have investigated the predictable "modes" (space–time patterns common to all ensemble members) of North Atlantic atmospheric circulation variability in a set of 28 winter (DJF) seasonal hindcasts spanning the period 1993–2021. A key motivation was the puzzling evidence of a "signal-to-noise paradox" (Eade *et al.* (2014); Scaife *et al.* (2014); Scaife and Smith (2018)) in seasonal forecast systems that exhibit skill in predicting the NAO, including GloSea5 (Baker *et al.* (2018)), wherein ensemble-mean forecasts exhibit higher correlation skill when evaluated against observations than when evaluated against a single ensemble member from the same prediction system. A novel aspect of our analysis was the use of a powerful method (Venzke *et al.* (1999)) to identify the predictable modes in the presence of contaminating noise. This method enabled us to study the evolution of the predictable signals through the winter basis (DJF). We also develop and use a method of quantifying the

scaling between the predictable modes in the observations and the hindcast ensemble, which complements the approach of Eade *et al.* (2014). The key findings are as follows:

- We identified two predictable modes over the North Atlantic region: a "PNA-like mode", a response to tropical SST variability, and an "NAO-like mode". The PNA-like mode is the leading predictable mode over the North Atlantic during late winter (JF) but not in early winter (DJ), when it is the second mode. The NAO-like mode is the leading predictable mode over the North Atlantic in DJ. Its spatial pattern in DJ closely resembles the canonical NAO pattern and it is significantly correlated with the observed NAO index ($r \sim 0.45$). Though these modes are separated according to North's rule (North *et al.*, 1982), this separation is not very large, and so there may be a degree of mixing between them.
- Whereas the PNA-like forced mode has a strong link to the tropical Pacific SSTs, the NAO-like mode appears linked to Indian Ocean SSTs in the preceding November. Part of this link is due to mutual trends in Indian Ocean SSTs and the NAO-like mode time series during 1993–2021, but the correlations persist even when these trends are removed. This is consistent with many previous studies and suggests that Indian Ocean SST variability drives at least part of the predictability of the NAO-like mode in GloSea5.
- The amplitude of the NAO-like mode in DJ is around three times larger in observations than is found in the model ensemble mean. This is evidence of the signal-to-noise problem discussed by Scaife *et al.* (2014) and Eade *et al.* (2014). Our results suggest that, for the GloSea5 winter forecasts, this signal-to-noise problem is particularly associated with the NAO-like mode. Our estimate of the scaling is larger than that of Eade *et al.* (2014), potentially due to the separation of the predictable modes in our approach. The PNA-like mode also appears to be weaker in the hindcasts than in observations, but the discrepancy in amplitude is much less (factor ~1.4) than for the NAO-like mode. This suggests that the weaker response of the forecast system compared with observations may not be global, but may be more confined to specific processes or teleconnection pathways. This could suggest that specific processes or teleconnection pathways may be muting the forcing signal from the Indian Ocean. It is also possible that processes governing the influence of atmospheric initial conditions on the predictability of the NAO-like mode are poorly represented in the model (Stockdale *et al.* (2015); O'Reilly *et al.* (2018); Nie *et al.* (2019)). However, the evidence for this problem arising from errors in the initial conditions is weakened by recent studies showing that uninitialized models also exhibit a signal-to-noise problem (Klavans *et al.*, 2019).

Our results, consistent with Hardiman *et al.* (2020), suggest that an important source of predictability in the NAO may be the Indian Ocean SSTs, but that key mechanisms, processes, or teleconnection pathways that communicate tropical variability to the NAO may be muted or damped in models in comparison with the real world. This suggests the need for focused attention on teleconnections to the NAO to resolve the signal-to-noise problem.

The predictable modes we have examined in this study are specific to the GloSea5 seasonal forecast model, so our conclusion herein may not extend to other seasonal forecast models. However, GloSea5 is not the only system to exhibit the signal-to-noise problem (Athanasiadis *et al.* (2017); Baker *et al.* (2018); Klavans *et al.* (2019)), and so it would clearly be valuable to apply the same analysis used here to other forecast systems to see whether similar results are obtained. Furthermore, there is strong evidence that the signal-to-noise problem is not limited to seasonal predictions of the NAO (Athanasiadis *et al.* (2017); Klavans *et al.* (2019)). Thus, the methods used in this study could usefully be applied to other predictions.

We note that in DJ and DJF, for the 28-member ensemble, a standard EOF analysis of the ensemble mean would have been sufficient to extract the predictable modes. However, we only know this by comparison with the optimal detection results. Is there a simpler method of judging whether an ensemble is large enough to confidently recover the predictable modes from a standard EOF analysis? This would certainly be a valuable area of future work.

A limitation of our method is it assumes that predictable signals can be separated into linearly independent components. In reality, nonlinear interactions may well exist; for example, the impact of the quasi-biennial oscillation may be sensitive to the phase of ENSO or vice versa (Garfinkel and Hartmann (2008); Hansen *et al.* (2016)). Investigating the possibility of such nonlinear interactions would be another valuable area for future work.

AUTHOR CONTRIBUTIONS

Daniel L.R. Hodson: conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; writing – original draft; writing – review and editing. **Rowan T. Sutton:** conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing – review and editing. **Adam A. Scaife:** data

curation; funding acquisition; resources; writing – review and editing.

ACKNOWLEDGEMENTS

Daniel L.R. Hodson and Rowan T. Sutton acknowledge support from the Natural Environment Research Council (NERC) through funding of the National Centre for Atmospheric Science (NCAS) and in particular the North Atlantic Climate System Integrated Study (ACSIS) programme. Adam A. Scaife was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. We would like to thank Stephanie Johnson for assistance with the hindcast data transfer and Len Shaffrey for helpful discussions. We wish to acknowledge use of the Ferret program for analysis and graphics in this paper. Ferret is a product of the National Oceanic and Atmospheric Administration's Pacific Marine Environmental Laboratory (information is available at http://ferret.pmel.noaa.gov/Ferret/). We would also like to thank the anonymous reviewers whose comments led to significant improvements of this article.

DATA AVAILABILITY STATEMENT

The datasets used in this paper are all publicaly available and referenced in the Data and Methods section.

ORCID

Daniel L.R. Hodson https://orcid.org/0000-0001-7159-6700
Rowan T. Sutton https://orcid.org/0000-0001-8345-8583
Adam A. Scaife https://orcid.org/0000-0002-5189-7538

REFERENCES

Abid, M.A., Kucharski, F., Molteni, F., Kang, I.-S., Tompkins, A.M. and Almazroui, M. (2021) Separating the Indian and Pacific Ocean impacts on the euro-Atlantic response to ENSO and its transition from early to late winter. *Journal of Climate*, 34, 1531–1548.

Allen, M.R. and Smith, L.A. (1997) Optimal filtering in singular spectrum analysis. *Physics Letters A*, 234, 419–428. https://doi.org/10.1016/s0375-9601(97)00559-8.

Athanasiadis, P.J., Bellucci, A., Scaife, A.A., Hermanson, L., Materia, S., Sanna, A., Borrelli, A., MacLachlan, C. and Gualdi, S. (2017) A multisystem view of wintertime NAO seasonal predictions. *Journal of Climate*, 30, 1461–1475.

Bader, J. and Latif, M. (2005) North Atlantic oscillation response to anomalous Indian Ocean SST in a coupled GCM. *Journal of Climate*, 18, 5382–5389.

Baker, L.H., Shaffrey, L.C., Sutton, R.T., Weisheimer, A. and Scaife, A.A. (2018) An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic oscillation in multimodel seasonal forecasts. *Geophysical Research Letters*, 45, 7808–7817. https://doi.org/10.1029/2018gl078838.

Bell, V.A., Davies, H.N., Kay, A.L., Brookshaw, A. and Scaife, A.A. (2017) A national-scale seasonal hydrological forecast system: development and evaluation over Britain. *Hydrology and Earth System Sciences*, 21, 4681–4691. https://doi.org/10.5194/hess-21-4681-2017.

Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Ménard, C.B., Edwards, J.M., Hendry, M.A., Porson, A., Gedney, N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M., Grimmond, C.S.B. and Harding, R.J. (2011) The joint Uk land environment simulator (jules), model description–Part 1: energy and water fluxes. *Geoscientific Model Development*, 4, 677–699 https://gmd.copernicus.org/articles/4/677/2011/.

Blockley, E.W., Martin, M.J., McLaren, A.J., Ryan, A.G., Waters, J., Lea, D.J., Mirouze, I., Peterson, K.A., Sellar, A. and Storkey, D. (2014) Recent development of the met office operational ocean forecasting system: an overview and assessment of the new global foam forecasts. *Geoscientific Model Development*, 7, 2613–2638 https://gmd.copernicus.org/articles/7/2613/2014/.

Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J. and Shelly, A. (2012) Unified modeling and prediction of weather and climate: a 25-year journey. *Bulletin of the American Meteorological Society*, 93, 1865–1877 https://journals.ametsoc.org/view/journals/bams/93/12/bams-d-12-00018.1.xml.

Cassou, C. (2008) Intraseasonal interaction between the madden–Julian oscillation and the North Atlantic oscillation. *Nature*, 455, 523–527.

Chen, D., Cane, M.A., Kaplan, A., Zebiak, S.E. and Huang, D. (2004) Predictability of el niño over the past 148 years. *Nature*, 428, 733–736. https://doi.org/10.1038/nature02439.

Christiansen, B. (2018) Ensemble averaging and the curse of dimensionality. *Journal of Climate*, 31, 1587–1596.

Clark, R.T., Bett, P.E., Thornton, H.E. and Scaife, A.A. (2017) Skilful seasonal predictions for the european energy industry. *Environmental Research Letters*, 12, 024002. https://doi.org/10.1088/1748-9326/aa57ab.

Compo, G.P., Whitaker, J.S. and Sardeshmukh, P.D. (2006) Feasibility of a 100-year reanalysis using only surface pressure data. *Bulletin of the American Meteorological Society*, 87, 175–190. https://doi.org/10.1175/bams-87-2-175.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N. and Vitart, F. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597 https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.828.

Dobrynin, M., Domeisen, D.I.V., Müller, W.A., Bell, L., Brune, S., Bunzel, F., Düsterhus, A., Fröhlich, K., Pohlmann, H. and Baehr, J. (2018) Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, 45, 3605–3614. https://doi.org/10.1002/2018gl077209.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M. and Knight, J. (2016) Skilful predictions of the winter North Atlantic oscillation one year ahead. *Nature Geoscience*, 9, 809–814. https://doi.org/10.1038/ngeo2824.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L. and Robinson, N. (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Research Letters*, 41, 2014GL061146. https://doi.org/10.1002/2014gl061146.

Ely, C.R., Brayshaw, D.J., Methven, J., Cox, J. and Pearce, O. (2013) Implications of the North Atlantic oscillation for a UK–Norway renewable power system. *Energy Policy*, 62, 1420–1427. https://doi.org/10.1016/j.enpol.2013.06.037.

Farrara, J.D., Mechoso, C.R. and Robertson, A.W. (2000) Ensembles of AGCM two-tier predictions and simulations of the circulation anomalies during winter 1997–98. *Monthly Weather Review*, 128, 3589–3604.

Garfinkel, C.I. and Hartmann, D.L. (2008) Different ENSO teleconnections and their effects on the stratospheric polar vortex. *Journal of Geophysical Research*, 113, D18114. https://doi.org/10.1029/2008jd009920.

Hansen, F., Matthes, K. and Wahl, S. (2016) Tropospheric QBO–ENSO interactions and differences between the Atlantic and pacific. *Journal of Climate*, 29, 1353–1368. https://doi.org/10.1175/jcli-d-15-0164.1.

Hardiman, S.C., Dunstone, N.J., Scaife, A.A., Smith, D.M., Knight, J.R., Davies, P., Claus, M. and Greatbatch, R.J. (2020) Predictability of European winter 2019/20: Indian Ocean dipole impacts on the NAO. *Atmospheric Science Letters*, 21, e1005.

Hoerling, M.P., Hurrell, J.W. and Xu, T. (2001) Tropical origins for recent North Atlantic climate change. *Science*, 292, 90–92.

Hoerling, M.P., Hurrell, J.W., Xu, T., Bates, G.T. and Phillips, A.S. (2004) Twentieth century North Atlantic climate change. Part II: understanding the effect of Indian Ocean warming. *Climate Dynamics*, 23, 391–405.

Hunke, C.H., Lipscomb, W.H., Turner, A.K., Jeffery, N. and Elliot, S. (2015) Cice: the los alamos sea ice model documentation and software user's manual version 5.1 la-cc-06-012. *Technical Report*., Los Alamos National Laboratory. https://svn-ccsm-models.cgd.ucar.edu/cesm1/alphas/branches/cesm1_5_alpha04c_timers/components/cice/src/doc/cicedoc.pdf.

Hurrell, J.W., Hoerling, M.P., Phillips, A.S. and Xu, T. (2004) Twentieth century North Atlantic climate change. Part I: Assessing determinism. *Climate Dynamics*, 23, 371–389.

Jin, E.K., Kinter, J.L., Wang, B., Park, C.K., Kang, I.S., Kirtman, B.P., Kug, J.S., Kumar, A., Luo, J.J., Schemm, J., Shukla, J. and Yamagata, T. (2008) Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Climate Dynamics*, 31, 647–664. https://doi.org/10.1007/s00382-008-0397-3.

Jones, P.D., Jonsson, T. and Wheeler, D. (1997) Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-West Iceland. *International Journal of Climatology*, 17, 1433–1450. https://doi.org/10.1002/(sici)1097-0088(19971115)17:13%3C1433::aid-joc203%3E3.0.co;2-p.

Klavans, J.M., Clement, A.C. and Cane, M.A. (2019) Variable external forcing obscures the weak relationship between the NAO and North Atlantic multidecadal SST variability. *Journal of Climate*, 32, 3847–3864.

MacLachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereday, D., Scaife, A.A., Gordon, M., Vellinga, M., Williams, A., Comer, R.E., Camp, J., Xavier, P. and Madec, G. (2015) Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141, 1072–1084. https://doi.org/10.1002/qj.2396.

Madec, G. (2016) Nemo Ocean engine. Technical Report. 27, l'Institut Pierre-Simon Laplace. https://www.nemo-ocean.eu/wp-content/uploads/NEMO_book.pdf.

Molteni, F., Stockdale, T.N. and Vitart, F. (2015) Understanding and modelling extra-tropical teleconnections with the indo-pacific region during the northern winter. *Climate Dynamics*, 10, 3119–3140.

Nie, Y., Scaife, A., Ren, H.-L., Comer, R.E., Andrews, M., Davis, P. and Martin, N. (2019) Stratospheric initial conditions provide seasonal predictability of the North Atlantic and Arctic oscillations. *Environmental Research Letters*, 14, 034006.

North, G.R., Bell, T.L., Cahalan, R.F. and Moeng, F.J. (1982) Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, 110, 699–706.

O'Reilly, C.H., Weisheimer, A., Woollings, T., Gray, L. and MacLeod, D. (2018) The importance of stratospheric initial conditions for winter North Atlantic oscillation predictability and implications for the signal-to-noise paradox. *Quarterly Journal of the Royal Meteorological Society*, 145, 131–146.

Ossó, A., Shaffrey, L., Dong, B. and Sutton, R. (2019) Impact of air– sea coupling on northern hemisphere summer climate and the monsoon–desert teleconnection. *Climate Dynamics*, 53, 5063–5078.

Palin, E.J., Scaife, A.A., Wallace, E., Pope, E.C.D., Arribas, A. and Brookshaw, A. (2016) Skillful seasonal forecasts of winter disruption to the U.K. transport system. *Journal of Applied Meteorology and Climatology*, 55, 325–344. https://doi.org/10.1175/jamc-d-15-0102.1.

Rodwell, M.J. and Folland, C.K. (2002) Atlantic air–sea interaction and seasonal predictability. *Quarterly Journal of the Royal Meteorological Society*, 128, 1413–1443.

SanchezGomez, E., Cassou, C., Hodson, D.L.R., Keenlyside, N., Okumura, Y. and Zhou, T. (2008) North Atlantic weather regimes response to Indian-western Pacific Ocean warming: a multi-model study. *Geophysical Research Letters*, 35, L15706 .

Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D., Folland, C.K., Gordon, M., Hermanson, L., Knight, J.R., Lea, D.J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A.K., Smith, D., Vellinga, M., Wallace, E., Waters, J. and Williams, A. (2014) Skillful long-range prediction of European and north American winters. *Geophysical Research Letters*, 41, 2014GL059637. https://doi.org/10.1002/2014gl059637.

Scaife, A.A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M., MacLachlan, C., Martin, N., Nie, Y., Ren, H.-L., Roberts, M., Robinson, W., Smith, D. and Vidale, P.L. (2019) Does increased atmospheric resolution improve seasonal climate predictions? *Atmospheric Science Letters*, 20, e922.

Scaife, A.A., Comer, R.E., Dunstone, N.J., Knight, J.R., Smith, D.M., MacLachlan, C., Martin, N., Peterson, K.A., Rowlands, D., Carroll, E.B., Belcher, S. and Slingo, J. (2017) Tropical rainfall, rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, 143, 1–11. https://doi.org/10.1002/qj.2910.

Scaife, A.A. and Smith, D. (2018) A signal-to-noise paradox in climate science. *Npj Climate and Atmospheric Science*, 1, 28. https://doi.org/10.1038/s41612-018-0038-4.

Siegert, S., Stephenson, D.B., Sansom, P.G., Scaife, A.A., Eade, R. and Arribas, A. (2016) A bayesian framework for verification and recalibration of ensemble forecasts: how uncertain is nao predictability? *Journal of Climate*, 29, 995–1012 https://journals.ametsoc.org/view/journals/clim/29/3/jcli-d-15-0196.1.xml.

Smith, D.M., Scaife, A.A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L.F., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F.J., Dunstone, N.J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., Kimoto, M., Merryfield, W.J., Mignot, J., Mochizuki, T., Modali, K., Monerie, P.-A., Müller, W.A., Nicolí, D., Ortega, P., Pankatz, K., Pohlmann, H., Robson, J., Ruggieri, P., Sospedra-Alfonso, R., Swingedouw, D., Wang, Y., Wild, S., Yeager, S., Yang, X. and Zhang, L. (2020) North Atlantic climate far more predictable than models imply. *Nature*, 583, 796–800.

Stockdale, T.N., Molteni, F. and Ferranti, L. (2015) Atmospheric initial conditions and the predictability of the Arctic oscillation. *Geophysical Research Letters*, 42, 1173–1179.

Sutton, R.T. and Hodson, D.L.R. (2003) Influence of the ocean on North Atlantic climate variability 1871–1999. *Journal of Climate*, 16, 3296–3313. https://doi.org/10.1175/1520-0442(2003)016%3C3296:iotoon%3E2.0.co;2.

Trenberth, K.E. and Caron, J.M. (2000) The southern oscillation revisited: sea level pressures, surface temperatures, and precipitation. *Journal of Climate*, 13, 4358–4365. https://doi.org/10.1175/1520-0442(2000)013%3C4358:tsorsl%3E2.0.co;2.

Venzke, S., Allen, M.R., Sutton, R.T. and Rowell, D.P. (1999) The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *Journal of Climate*, 12, 2562–2584. https://doi.org/10.1175/1520-0442(1999)012%3C2562:tarotn%3E2.0.co;2.

von Storch, H. and Zwiers, F.W. (1999) *Statistical analysis in climate research*. Cambridge: Cambridge University Press.

Walker, G.T. and Bliss, E.W. (1932) World weather v. *Memoirs of the Royal Meteorological Society*, 4, 53–84 https://www.rmets.org/sites/default/files/ww5.pdf.

Wallace, J.M. and Gutzler, D.S. (1981) Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, 109, 784–812 https://journals.ametsoc.org/view/journals/mwre/109/4/1520-0493_1981_109_0784_titghf_2_0_co_2.xml.

Walters, D.N., Best, M.J., Bushell, A.C., Copsey, D., Edwards, J.M., Falloon, P.D., Harris, C.M., Lock, A.P., Manners, J.C., Morcrette, C.J., Roberts, M.J., Stratton, R.A., Webster, S., Wilkinson, J.M., Willett, M.R., Boutle, I.A., Earnshaw, P.D., Hill, P.G., MacLachlan, C., Martin, G.M., Moufouma-Okia, W., Palmer, M.D., Petch, J.C., Rooney, G.G., Scaife, A.A. and Williams, K.D. (2011) The met office unified model global atmosphere 3.0/3.1 and jules global land 3.0/3.1 configurations. *Geoscientific Model Development*, 4, 919–941 https://gmd.copernicus.org/articles/4/919/2011/.

Wilks, D. (1995) *Statistical Methods in the Atmospheric Sciences*, Vol. 373–398. Elsevier, UK: Academic Press.

Zwiers, F.W. (1990) The effect of serial correlation on statistical inferences made with resampling procedures. *Journal of Climate*, 3, 1452–1461 http://www.jstor.org/stable/26196360.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hodson, D.L., Sutton, R.T. & Scaife, A.A. (2023) Signal-to-noise and predictable modes of variability in winter seasonal forecasts. *Quarterly Journal of the Royal Meteorological Society*, 149(755), 2598–2616. Available from: https://doi.org/10.1002/qj.4522

APPENDIX A

A.1 Estimating the predictable response ratio γ

In Section 2.5 we discuss evidence that the amplitudes of observed predictable modes (or forced modes) have a greater amplitude than predictable modes in the GloSea5 hindcasts. In this section we derive an estimate of the ratio of these amplitudes, γ.

Consider a common predictable mode, $\mathbf{X}^{\mathrm{F}}$ that is the (outer) product of a spatial pattern $\underline{a}$ and the time variation of that pattern $\underline{x}^{\mathrm{F}}$ (a time series):

$$\mathbf{X}^{\mathrm{F}} = \underline{a}\,\underline{x}^{\mathrm{F}}, \tag{A.1}$$

where $\underline{x}^{\mathrm{F}}$ has unit length:

$$\underline{x}^{\mathrm{F}}\underline{x}^{\mathrm{F}^{\mathrm{T}}} = 1. \tag{A.2}$$

Here, bold symbols are matrices (dimensions of space × time), and underscored symbols (e.g., $\underline{a}$) are vectors (dimension of either space or time).

A hindcast model ensemble member

$$\mathbf{X}_j = \mathbf{X}^{\mathrm{F}} + \boldsymbol{\mu}_j \tag{A.3}$$

is composed of the predictable mode (or forced response) $\mathbf{X}^{\mathrm{F}}$ and a noise term (with a time mean of zero) $\boldsymbol{\mu}_j$ ($N(0, \sigma_\mu^2)$), where j is an index for each ensemble member. Here, $\boldsymbol{\mu}_j$ represents all the atmospheric variability that is not due to the external forcing. We assume $\mathbf{X}^{\mathrm{F}}$ has a zero time mean and a variance σ_{F}^2 ($\mathbf{X}^{\mathrm{F}} \sim N(0, \sigma_{\mathrm{F}}^2)$). The ensemble mean is

$$\overline{\mathbf{X}} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{X}_j. \tag{A.4}$$

We define the true (observed) amplitude of this mode as a factor of γ larger; hence, our single real-world

ensemble member is given by $\mathbf{X}^{\mathrm{o}}$:

$$\mathbf{X}^{\mathrm{o}} = \gamma \mathbf{X}^{\mathrm{F}} + \epsilon, \tag{A.5}$$

where ϵ is a noise term ($N(0, \sigma_\epsilon^2)$). If we regress each "spatial point" of $\mathbf{X}^{\mathrm{o}}$ onto the time evolution of $\mathbf{X}^{\mathrm{F}}$ ($\underset{-}{x}^{\mathrm{F}}$), the regression slope estimate $\underset{-\mathrm{o}}{\hat{\beta}}$ is

$$\underset{-\mathrm{o}}{\hat{\beta}} = \frac{\mathbf{X}^{\mathrm{o}} \underset{-}{x}^{\mathrm{F}\mathrm{T}}}{\underset{-}{x}^{\mathrm{F}} \underset{-}{x}^{\mathrm{F}\mathrm{T}}} = \mathbf{X}^{\mathrm{o}} \underset{-}{x}^{\mathrm{F}\mathrm{T}}, \tag{A.6}$$

which follows from Equation (A.2).

Similarly, for the ensemble mean $\overline{\mathbf{X}}$:

$$\underset{-\mathrm{m}}{\hat{\beta}} = \frac{\overline{\mathbf{X}} \underset{-}{x}^{\mathrm{F}\mathrm{T}}}{\underset{-}{x}^{\mathrm{F}} \underset{-}{x}^{\mathrm{F}\mathrm{T}}} = \overline{\mathbf{X}} \underset{-}{x}^{\mathrm{F}\mathrm{T}}. \tag{A.7}$$

$\underset{-\mathrm{o}}{\hat{\beta}}$ and $\underset{-\mathrm{m}}{\hat{\beta}}$ are vectors with a "spatial" dimension (i.e., labelled by the model spatial grid points). We can now regress these vectors together to give an estimate of the scaling $\hat{\gamma}$:

$$\hat{\gamma} = \frac{\underset{-\mathrm{o}-\mathrm{m}}{\hat{\beta}}\, \underset{-\mathrm{o}-\mathrm{m}}{\hat{\beta}^{\mathrm{T}}}}{\underset{-\mathrm{m}-\mathrm{m}}{\hat{\beta}}\, \underset{-\mathrm{m}-\mathrm{m}}{\hat{\beta}^{\mathrm{T}}}} = \frac{(\mathbf{X}^{\mathrm{o}} \underset{-}{x}^{\mathrm{F}\mathrm{T}})(\overline{\mathbf{X}} \underset{-}{x}^{\mathrm{F}\mathrm{T}})^{\mathrm{T}}}{(\overline{\mathbf{X}} \underset{-}{x}^{\mathrm{F}\mathrm{T}})(\overline{\mathbf{X}} \underset{-}{x}^{\mathrm{F}\mathrm{T}})^{\mathrm{T}}}. \tag{A.8}$$

Using Equations A.4,A.3, and A.5, and that the expectation value of the product of two uncorrelated noise terms is zero, we arrive at

$$\hat{\gamma} = \frac{\gamma a^2}{a^2 + \frac{1}{J}\sigma_\mu^2} \tag{A.9}$$

$$= \frac{\gamma}{1 + \frac{\alpha}{J}}, \tag{A.10}$$

where $\alpha = \sigma_\mu^2 / a^2$.

We now have a simple form for $\hat{\gamma}$, our biased estimate of γ, in terms of ensemble size J and two fitting parameters, α and the true value of γ. We can now use this expression with multiple subsamples (of varying sizes, $1 \ldots J$) of the full ensemble and fit, Equation (A.10), to obtain our unbiased estimate of γ (Figure 8).

A.2 Relationship of γ to RPC

The scaling factor γ is directly related to the RPC discussed by Eade *et al.* (2014) in their analysis of the GloSea5 NAO hindcasts. We examine this relationship in this section.

If we consider two variables y and x, the least-squares regression slope estimate $\hat{\beta}(y, x)$ between these two variables is related to their correlation $r(y, x)$ by

$$\hat{\beta}(y, x) = r(y, x)\left(\frac{\sigma_y}{\sigma_x}\right), \tag{A.11}$$

where σ_y^2 is the variance of y, and similarly for x.

This is because

$$\hat{\beta}(y, x) = \frac{y x^{\mathrm{T}}}{x x^{\mathrm{T}}}$$

$$r(y, x) = \frac{y x^{\mathrm{T}}}{(y y^{\mathrm{T}} x x^{\mathrm{T}})^{1/2}}.$$

The RPC is defined as the ratio of the magnitude of the signal in the observation to that in the model hindcast. Eade *et al.* (2014) define the RPC as

$$\mathrm{RPC}(x^{\mathrm{o}}, \overline{x}) = \frac{r(x^{\mathrm{o}}, \overline{x})}{\sqrt{\sigma_{\mathrm{sig}}^2 / \sigma_{\mathrm{tot}}^2}},$$

where x^{o} are the observations and $\overline{x}$ is the model ensemble mean. σ_{sig}^2 is the variance of the ensemble mean ($\sigma_{\overline{x}}^2$), and σ_{tot}^2 is the mean of the (time) variance of each ensemble member. Examining this relation, we can see that if $\sigma_{\mathrm{tot}}^2 = \sigma_{\mathrm{o}}^2$ (the mean variance over time of the ensemble was equal to the variance over time of the observations), then we would have Equation (A.11). This implies that, if the model and observations are drawn from the same distribution (and hence have the same variance), the RPC is simply the regression slope between the two.

The RPC is computed between the observations and the model ensemble mean at every grid point s, and so varies across the spatial domain. As discussed in Appendix A.1, γ is a scaling factor that applies to a spatial predictable mode, and so γ is constant across the spatial domain.

If, for the moment, we assume that both the model and observations can be explained by a single predictable mode, as in Equations A.1,A.3, and A.5, then the RPC at a grid point s can be expressed as

$$\mathrm{RPC}(X_s^{\mathrm{o}}, \overline{X_s}) = \frac{\boldsymbol{X}_s^{\mathrm{o}} \overline{\boldsymbol{X}_s}^{\mathrm{T}}}{(\boldsymbol{X}_s^{\mathrm{o}} \boldsymbol{X}_s^{\mathrm{o}\mathrm{T}} \overline{\boldsymbol{X}_s \boldsymbol{X}_s}^{\mathrm{T}})^{1/2}} \left(\frac{\sigma_{\mathrm{tot}}}{\sigma_{\mathrm{sig}}}\right)$$

$$= \frac{\boldsymbol{X}_s^{\mathrm{o}} \overline{\boldsymbol{X}_s}^{\mathrm{T}}}{\overline{\boldsymbol{X}_s \boldsymbol{X}_s}^{\mathrm{T}}} \left(\frac{\sigma_{\mathrm{tot}}}{\sigma_{\mathrm{o}}}\right).$$

Substituting in Equations A.5 and A.4:

$$\begin{aligned}\mathrm{RPC}_s &= \frac{(\gamma X_s^{\mathrm{F}} + \epsilon)(X_s^{\mathrm{F}} + \frac{1}{J}\sum_j \mu_j)^T}{(X_s^{\mathrm{F}} + \frac{1}{J}\sum_j \mu_j)(X_s^{\mathrm{F}} + \frac{1}{J}\sum_j \mu_j)^T}\left(\frac{\sigma_{\mathrm{tot}}}{\sigma_{\mathrm{o}}}\right)\\ &= \frac{\gamma\frac{a_s^2}{T-1}}{(\frac{a_s^2}{T-1} + \frac{1}{J}\sigma_\mu^2)}\left(\frac{\sigma_{\mathrm{tot}}}{\sigma_{\mathrm{o}}}\right)\\ &= \frac{\gamma}{1 + \frac{(T-1)\sigma_\mu^2}{Ja_s^2}}\left(\frac{\sigma_{\mathrm{tot}}}{\sigma_{\mathrm{o}}}\right),\end{aligned}$$

where a_s is a from Equation (A.1), at a particular grid point s, and T are the number of time points.

This shows that the RPC at grid point s is proportional to γ weighted by the magnitude of the predictable mode pattern at that grid point. We can make a further simplification if we assume that the variance of the observations σ_{o} is the same as the variance of each ensemble member σ_{tot} (which is likely, as models tend to be tuned so that this is the case), in which case the RPC at s becomes

$$\mathrm{RPC}_s = \frac{\gamma}{1 + \frac{(T-1)\sigma_\mu^2}{Ja_s^2}}.$$

In practice, the model and observations will be the sum over M predictable modes X_m^{F}; hence, we can rewrite Equations A.3 and A.5 as

$$\begin{aligned}\boldsymbol{X}_j &= \sum_m \boldsymbol{X}_m^{\mathrm{F}} + \boldsymbol{\mu}_j\\ \boldsymbol{X}^{\mathrm{o}} &= \sum_m \gamma_m \boldsymbol{X}_m^{\mathrm{F}} + \boldsymbol{\epsilon}.\end{aligned}$$

Alternatively, we can write this in a matrix format (a standard EOF or singular value decomposition description):

$$\begin{aligned}\boldsymbol{X}_j &= \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{\mu}_j\\ \boldsymbol{X}^{\mathrm{o}} &= \boldsymbol{E}\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{\epsilon},\end{aligned}$$

where $\boldsymbol{E}$ is a space $\times M$ matrix, Λ an $M \times M$ matrix, $\boldsymbol{P}^{\mathrm{T}}$ an $M \times T$ matrix, such that

$$\begin{aligned}\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}} &= \boldsymbol{I}\\ \boldsymbol{P}\boldsymbol{P}^{\mathrm{T}} &= \boldsymbol{I},\end{aligned}$$

and $\boldsymbol{\Gamma}$ is an $M \times M$ diagonal matrix containing the γ_m along the diagonals.

Using this more general approach, RPC_s becomes

$$\mathrm{RPC}_s = \frac{\boldsymbol{E}_s\boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{E}_s^{\mathrm{T}}}{\boldsymbol{E}_s\boldsymbol{\Lambda}^2\boldsymbol{E}_s^{\mathrm{T}} + \frac{1}{J}\sigma_\mu^2}$$

(if we again, assume that $\sigma_{\mathrm{o}} = \sigma_{\mathrm{tot}}$).

This means that the RPC is the sum of the γ_m for all modes, weighted by the variance ($\boldsymbol{\Lambda}^2$) of each mode, and the value of the predictable mode at that spatial point, all normalized by the equivalent in the model ensemble mean. In other words, whereas the γ_m reflect the scalings of each mode, the RPC is a weighted sum of these scalings. In the case where the $\gamma_m = \gamma$ (i.e., are identical – all predictable modes are scaled by the same factor), $\boldsymbol{\Gamma} = \gamma\boldsymbol{I}$ and in the limit of large J (ensemble size), RPC and γ would be identical. We believe however, from the evidence presented in this article, that the scaling for predictable modes of climate may not be identical and that this may warrant examination of the different physical processes associated with the different predictable modes.