Detection of interannual ensemble forecast signals over the North Atlantic and Europe using atmospheric circulation regimes

Received: 21 May 2021 | Revised: 5 November 2021 | Accepted: 8 November 2021

DOI: 10.1002/qj.4213

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Detection of interannual ensemble forecast signals over the North Atlantic and Europe using atmospheric circulation regimes

Swinda K.J. Falkena[1] | **Jana de Wiljes**[1,2] | **Antje Weisheimer**[3,4] | **Theodore G. Shepherd**[5]

[1]Department of Mathematics and Statistics, University of Reading, Reading, UK

[2]Institute for Mathematics, Universität Potsdam, Potsdam, Germany

[3]NCAS, Department of Physics, Atmospheric, Oceanic and Planetary Physics (AOPP), University of Oxford, Oxford, UK

[4]ECMWF, Reading, UK

[5]Department of Meteorology, University of Reading, Reading, UK

Correspondence
S.K.J. Falkena, Department of Mathematics and Statistics, University of Reading, Reading RG6 6AX, UK
Email: s.k.j.falkena@pgr.reading.ac.uk

Funding information
Centre for Doctoral Training in Mathematics of Planet Earth, with funding from the UK Engineering and Physical Sciences Research Council (EPSRC), Grant/Award Numbers: EP/L016613/1; Deutsche Forschungsgemeinschaft (DFG), Grant/Award Numbers: SFB1294/1, 318763901

Abstract

To study the forced variability of atmospheric circulation regimes, the use of model ensembles is often necessary for identifying statistically significant signals as the observed data constitute a small sample and are thus strongly affected by the noise associated with sampling uncertainty. However, the regime representation is itself affected by noise within the atmosphere, which can make it difficult to detect robust signals. To this end we employ a regularised *k*-means clustering algorithm to better identify the signal in a model ensemble. The approach allows for the identification of six regimes for the wintertime Euro-Atlantic sector and leads to more pronounced regime dynamics, compared to results without regularisation, both overall and on sub-seasonal and interannual time-scales. We find that sub-seasonal variability in the regime occurrence rates is mainly explained by changes in the seasonal cycle of the mean climatology. On interannual time-scales relations between the occurrence rates of the regimes and the El Niño Southern Oscillation (ENSO) are identified. The use of six regimes captures a more detailed response of the circulation to ENSO compared to the common use of four regimes. Predictable signals in occurrence rate on interannual time-scales are found for the two zonal flow regimes, namely a regime consisting of a negative geopotential height anomaly over the Norwegian Sea and Scandinavia, and the positive phase of the NAO. The signal strength for these regimes is comparable between observations and model, in contrast to that of the NAO-index where the signal strength in the observations is underestimated by a factor of 2 in the model. Our regime analysis suggests that this signal-to-noise problem for the NAO-index is primarily related to those atmospheric flow patterns associated with the negative NAO-index as we find poor predictability for the corresponding NAO− regime.

KEYWORDS

atmospheric circulation regimes, ensemble data, interannual variability, regularised k-means clustering, signal-to-noise

1 | INTRODUCTION

Atmospheric circulation regimes, or weather regimes, provide a way to study the low-frequency variability in the atmosphere (Hannachi *et al.*, 2017). Many studies have looked into their identification (e.g., Michelangeli *et al.*, 1995; Straus and Molteni, 2004) and numerous research efforts have focused on the relation between regimes and local weather (e.g., Cassou *et al.*, 2005; Ortizbeviá *et al.*, 2011). It is of interest to know whether the occurrence of the different regimes varies on both (sub-)seasonal and interannual time-scales in a predictable way. Such forced (i.e., non-stationary) variability can be caused by links between circulation regimes and other patterns of related climate variability, for example sudden stratospheric warmings on sub-seasonal time-scales (e.g., Charlton-Perez *et al.*, 2018; Domeisen *et al.*, 2020) or the El Niño Southern Oscillation (ENSO) on interannual time-scales (e.g., Drouard and Cassou, 2019; Lee *et al.*, 2019). A better understanding of the processes guiding the non-stationary regime dynamics can help improve predictions of the regimes themselves, as well as the consequences for local and regional weather. To this end it is important to robustly identify predictable regime variations, given the inevitable presence of noise, which can conceal these possibly weak signals within the regime dynamics.

The common approach for identifying circulation regimes is to apply a *k*-means clustering algorithm to the 500 hPa geopotential height (e.g., Michelangeli *et al.*, 1995; Cassou *et al.*, 2005; Straus *et al.*, 2007). Applying this method to reanalysis data has shown consistent results between studies for, for example, the Northern Hemisphere or the Euro-Atlantic sector. However, reanalysis data only provide one realisation of reality and thus are sensitive to sampling uncertainty in detecting the non-stationary signal. Reanalysis data mostly cover only the last 40 years (e.g., ERA-Interim), which is too short to reliably identify any non-stationary regime behaviour, especially on interannual time-scales, but also on sub-seasonal time-scales. For example, when one is interested in the effect of ENSO on the occurrence rate of the circulation regimes in winter, one only has a few years of data consisting of roughly 120 days each when using ERA-Interim. For six regimes, on average occurring 20 days each year, a few days more being assigned to one regime can significantly affect the regime frequencies. This makes it difficult to distinguish any signal from the noise.

One way to increase the sample size, and thus identify a more robust signal of the predictable component of the variability, is to use the UNSEEN method, in which model ensemble members with different lead times are pooled to create a very large ensemble (e.g., Thompson *et al.*, 2017; Kelder *et al.*, 2020). Here, the lead times are beyond the deterministic predictability limit, and there is no skill for predicting individual weather events. The ensemble members can thus be treated as plausible alternative realisations of reality, which are all equally affected by the sources of predictable variability. In line with this approach, we use hindcast ensemble data of the ECMWF seasonal forecast system to study sub-seasonal and interannual regime dynamics. As the model has high levels of interannual ENSO forecast skill (Johnson *et al.*, 2019), using seasonal forecasts allows for a more precise study of the interannual dynamics and effects of, for example, ENSO on the regimes. Here, we are not primarily concerned with the initial condition problem of weather forecasting, but rather require a high-resolution model with a small bias on the slightly longer time-scales of interest for this study.

A difficulty here is that models are imperfect and exhibit a wide spread in their regime frequencies compared to reanalysis data (e.g., Fabiano *et al.*, 2020). This behaviour may reflect the "signal-to-noise paradox" whereby models are more noisy than the real world (see further discussion below). It is exemplified by the domain dependence of the regimes, particularly the negative counterpart of the Atlantic Ridge (AR−), when considering the six regimes identified using the ECMWF SEAS5 hindcast ensemble in Figure 1. Here, six regimes are considered instead of the commonly used four as this was identified to be the optimal number when using gridpoint data and allows for a more detailed description of the variability in the atmospheric circulation (Falkena *et al.*, 2020). For ERA-Interim data, no such domain dependence of the regimes is found, despite the smaller sample size. This domain dependence of the regimes within the model is undesirable from a physical and useability perspective.

When identifying the circulation regimes, the presence of noise can hide possible regime variability signals. Specifically, small deviations in the distance of data to the regimes can result in them being assigned to a different regime, obscuring the "true" signal (see further discussion in Section 2). To avoid the misinterpretation of the regime signal, we implement a regularisation within the clustering method to strengthen the non-stationary signal by penalising noise. Specifically, we add information from the model ensemble to obtain a better informed regime identification method. Similar forms of regularisation, designed to increase persistence in time, have been successfully employed to detect robust and meaningful regimes in the climate context (e.g., Horenko, 2010; Falkena *et al.*, 2020). Since the regularisation is empirical, we monitor its effect by quantifying the trade-off between accuracy and information, and assessing whether what it does is physically sensible.

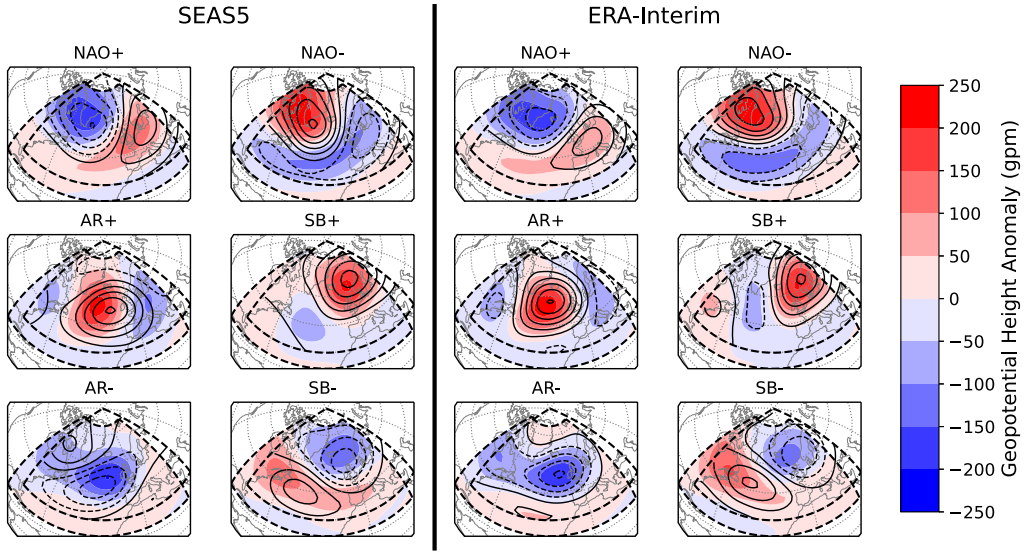

FIGURE 1 The regimes identified for the ECMWF SEAS5 hindcast ensemble members (left) and ERA-Interim (right) using standard *k*-means clustering for two slightly different domains (indicated by the dashed boxes). They are the positive and negative phases of the North Atlantic Oscillation (NAO), the Atlantic Ridge (AR+) and its negative counterpart (AR−), and the Scandinavian Blocking (SB+) and its negative counterpart (SB−). Regimes for domain A (20–80°N, 90°W–30°E) are indicated by the colours and those for domain B (30–90°N, 80°W–40°E) by the contours, following the same 50 gpm difference between contour levels [Colour figure can be viewed at wileyonlinelibrary.com]

Our focus is on identifying non-stationary regime dynamics in the Euro-Atlantic sector during winter. In this region the North Atlantic Oscillation (NAO) is the dominant mode of variability and most models show moderate skill for its prediction during winter (Scaife *et al.*, 2014; Baker *et al.*, 2018; Weisheimer *et al.*, 2019). It is expected that this skill can extend, at least in part, to the regime dynamics, where the regularisation can help to better identify the signal. In this way the regime dynamics as obtained using the regularised clustering method can possibly help understand the so-called signal-to-noise paradox observed for the NAO. This paradox relates to the observation that most models are better at predicting the real world NAO than the NAO of their own ensemble members (Eade *et al.*, 2014; Scaife and Smith, 2018), while the variance of single ensemble members and of the observations is comparable. Studying the predictable signal of regime occurrence on interannual time-scales, using the regularised clustering results, can firstly indicate whether the signal-to-noise paradox extends to the regime dynamics, and secondly point to the dynamical flow patterns which are poorly represented within the model and thus warrant special focus when trying to resolve the paradox.

In the next section we start by discussing the problem setting of using clustering methods to identify circulation regimes in model ensembles exhibiting a wide spread in their regime representation, and show a motivational example for the regularisation method proposed to handle this spread and identify a more robust signal. This method, and the data used, are then discussed in some detail in Section 3, followed by a discussion on the choice of the regularisation constraint in Section 4 using several criteria. The results for the regime dynamics are presented in Section 5, looking into the effect of the regularisation and into the non-stationary signal on both sub-seasonal and interannual time-scales, as well as discussing the signal-to-noise problem. We end with a summary and discussion of the results in Section 6.

2 | MOTIVATION

The method of *k*-means clustering, which is the standard approach for identifying circulation regimes, deals poorly with data containing a lot of noise. The reason for this is that *k*-means clustering assigns every data point to the cluster centre, or regime, that it is closest to, even if only by a tiny margin. This makes it overly sensitive to noise, especially when the signal-to-noise ratio is small. A consequence is that the identification procedure lacks robustness and the informational gain is small. The following example visualises this issue by means of one possible scenario.

Consider Figure 2a which shows the distribution of ensemble members over three different clusters, and Figure 2b showing the (theoretical) distributions of data over two regimes (green, left and orange, right) when they are equally likely (top) and when the orange regime is

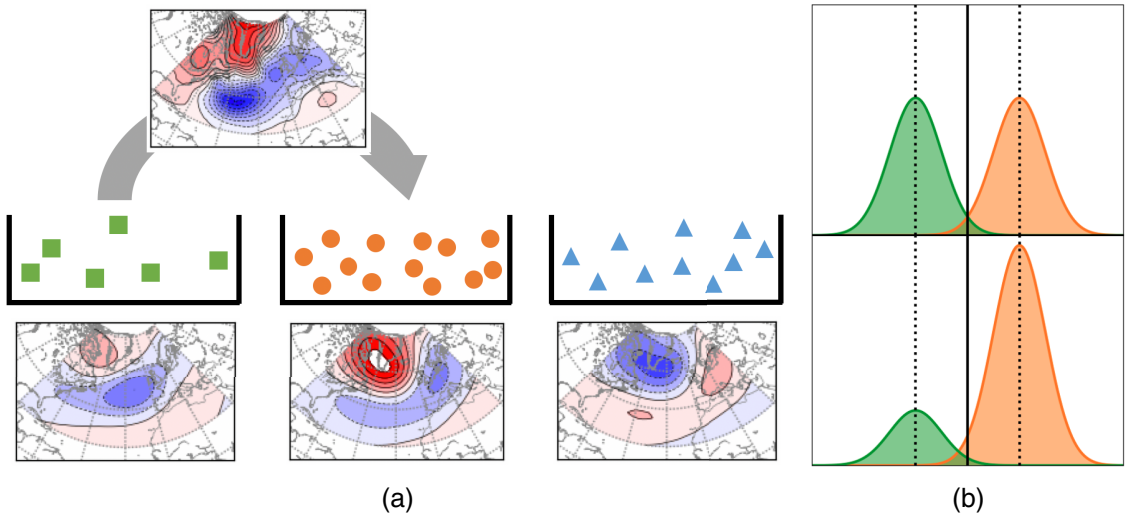

FIGURE 2 (a) An example of a possible reassignment of an ensemble member. The green squares, orange circles and blue triangles give the distribution of the ensemble members over the different regimes, that is, each marker indicates to which regime the corresponding ensemble member is assigned, but its location within the bin does not provide any further information. The arrow shows the desired reassignment of a data-point, which might plausibly be associated with either the green square or the orange circle regime. To better understand this reassignment, in (b) the probability distribution of data around the cluster centres (dotted lines) is shown for equally likely regimes (top), and a situation where the orange (right) regime is more likely, that is, of higher amplitude (bottom). When the two regimes are equally likely, a point in the middle (solid line) has an equal probability of belonging to either of the regimes. However, when the orange (right) regime is more likely than the green one (left), data that lie half-way between have a larger probability of belonging to the orange regime than to the green one, and thus might better be reassigned to the orange regime. Such reassignments can help to reduce the effect of noise and identify a more robust signal [Colour figure can be viewed at wileyonlinelibrary.com]

more likely (bottom; note this is an exaggerated visualisation for illustration purposes). At a fixed time t it is possible that a data point, that is, an ensemble member, falls in between two (or more) regimes (e.g., the clusters associated with the green squares and orange circles). Here a standard k-means clustering assigns it to the regime it is closest to in distance, which is valid if the regimes are equally likely. However, due to the effect on the regimes of external forcing, such as ENSO, this is not always the case. If one regime is known to be more likely than the neighbouring one at that point in time, then it would be prudent to assign the ensemble member to the more likely regime. In this way the regime assignment of an ensemble member is not solely determined by its distance to the cluster centres, but also by a prior likelihood set by the distribution of the ensemble members over the regimes, which is picking up a non-stationary signal. Effectively, noise is being penalised. For example, in our visualisation the cluster comprised of the ensemble members indicated by the orange circles is more likely, that is, occurs a lot more over all ensemble members, at a given time t. The shown ensemble member in distance falls between the green square and orange circle regimes, that is, only slightly left of the solid line in Figure 2b. To assign it to the green square regime based on this small difference in distance places more weight on the noise than on the signal. For that reason it can be better to assign it to the orange circle regime which has a higher probability as shown in Figure 2b.

The aim is then to design a clustering method that penalises noise, to mitigate incorrect assignments of data points as exemplified above. One possible way to achieve this is to regularise the k-means clustering method by implementing a constraint enforcing a level of similarity between the ensemble members at each moment in time (Bishop, 2006 gives a discussion of different types of regularisation). This design has a physically meaningful basis as the preferred regimes should, on average, be represented by the overall ensemble, and if one regime is more populated than usual at a particular time, it makes it more likely that borderline cases belong to that regime (Figure 2). By introducing a constraint that prioritises similarities over small deviations it is possible to distinguish more pronounced regime behaviour, that is, to discriminate better between the regimes. The underlying assumption here is that the distribution of the ensemble members over the regimes changes in time due to external factors such as ENSO and that the regularisation helps to better identify such weak non-stationary signals.

Of course, such an algorithm could be over-confident. A curb on over-confidence is provided by accuracy. The overall goal is to keep the accuracy at a reasonably high

level while significantly increasing the information gain (in the entropy sense) of the derived regime model. Specifically, we favour regimes with more informative dynamics over those that fit the data slightly better, since this can be advantageous in identifying weak signals. In the next Section we describe how to quantify this trade-off between accuracy and informativeness.

3 | DATA AND METHODS

When it comes to regime analysis of model ensemble data there are two approaches one can take. The first is to assign the model data to the regimes obtained from reanalysis data (e.g., Ferranti *et al.*, 2015; Grams *et al.*, 2018). The second approach is to compute the regimes from the model data itself (e.g., Dawson and Palmer, 2015; Matsueda and Palmer, 2018). This latter approach means that the regimes identified can differ from those of reanalysis data. On the other hand, it includes a natural bias correction of the model data in the regime representation. Here we choose the latter approach.

Before discussing how to implement the constraint discussed in the previous section in the k-means clustering algorithm for model ensemble data, we first need to detail the data used for the identification of the circulation regimes and the distribution of the data over them. Then we describe the regularised clustering method and lastly we discuss the use of regression analysis for the identification of a predictable non-stationary signal and link this approach to more commonly used methods.

3.1 | Data

Daily 500 hPa geopotential height from the ECMWF hindcast ensemble of SEAS5 (Johnson *et al.*, 2019) is used for the identification of the circulation regimes. The ensemble has 51 members for a 01 November start date and is available from 1981 to 2016. We use the 500 hPa geopotential height on a 2.5° by 2.5° grid for a domain covering the Euro-Atlantic sector. To analyse the robustness of the obtained regimes, we consider two slightly different domains; 20° to 80°N, 90°W to 30°E (domain A) and 30° to 90°N, 80°W to 40°E (domain B) (both are commonly used in literature, e.g., Cassou *et al.*, 2005; Dawson *et al.*, 2012). The months December to March are considered using daily data (00:00 UTC), meaning forecast lead times of over a month are used, for which there is no longer any discernible effect of the atmospheric initial conditions. Note that this loss of memory of the initial conditions does not imply there is no predictable variability, as other processes such as the month within the season or the phase of ENSO affect the circulation. We compute anomalies with respect to a DJFM average climatology in order not to make any assumptions on the (sub-)seasonal variability in the background climatology (Section 5.2 and Falkena *et al.*, 2020 have further reasoning on this point).

To reduce the effect of weather noise, preprocessing methods are often used to focus on the larger-scale, predictable variability. In Fabiano *et al.* (2020) an Empirical Orthogonal Function (EOF) analysis was used to reduce the dimensionality of the data in a model ensemble, but a large spread in the centroid distance and spatial correlation of the regimes was still found. This indicates that this way of preprocessing is not sufficient to reduce the effect of noise on the identified regimes. Other methods of preprocessing the data, such as using a low-pass filter, could filter out some of the noise within the model as well. However, these methods can also lead to biases in the resulting regimes. For example, in Falkena *et al.* (2020) it was found that the use of a low-pass filter affects the regime frequencies. Therefore, we focus on adapting the clustering method instead of preprocessing the data, to identify a more robust regime signal. This way we do not lose any information present in the data and avoid possibly introducing a bias by preprocessing. The regimes obtained for the SEAS5 hindcast data using the clustering method described in the following section are identified with the corresponding regimes in the ERA-Interim reanalysis (Dee *et al.*, 2011) as obtained in Falkena *et al.* (2020). Thus the regimes for SEAS5 and for ERA-Interim are slightly different (Figure 1), which allows for bias within the model. For consistency, the same period of 36 years for which the SEAS5 data are available is considered for ERA-Interim.

3.2 | Regularised clustering method

Let the considered ensemble data be of the form $x_{t,n} \in \mathbb{R}^{T\times N\times D}$, where T is the length of the time series, N the number of ensemble members and D the spatial dimension of the data (here latitude×longitude). The aim of clustering the data is to find k cluster centres $\Theta = (\theta_1, ..., \theta_k) \in \mathbb{R}^{k\times D}$ (regimes) that best represent the data. The assignment of individual data points to the different clusters is given by the weights, or affiliation vector, $\Gamma = \{\gamma_1(t,n), ..., \gamma_k(t,n)\} \in \mathbb{R}^{k\times T\times N}$. This can be understood as the probability of a data-point belonging to each of the different regimes.

Identifying the circulation regimes means that we need to find the optimal parameters for the cluster centres Θ and the data-affiliations Γ. To achieve this, a cost function, also referred to as the averaged clustering functional (in its

discrete form), is minimized (Franzke *et al.*, 2009):

$$\mathbf{L}(\Theta,\Gamma)=\sum_{t=0}^{T}\sum_{n=1}^{N}\sum_{i=1}^{k}\gamma_i(t,n)g(x_{t,n},\theta_i). \tag{1}$$

Here $g(x_{t,n},\theta_i)$ is the distance between the cluster centre and a data point, for which the L_2-norm $\left(\left\{\sum_{lat,lon}\left[x_{t,n}(lat,lon)-\theta_i(lat,lon)\right]^2\right\}^{1/2}\right)$ weighted by the cosine of latitude is used. The affiliations $\gamma_i(t,n)\geq 0$ are normalised following

$$\sum_{i=1}^{k}\gamma_i(t,n)=1, \qquad \forall t\in[0,T], \quad \forall n\in[1,N]. \tag{2}$$

In practice, the $\gamma_i(t,n)$ values obtained via the optimisation are mostly equal to zero or one. In that case the data-points are unambiguously assigned to one of the regimes. In traditional k-means clustering, the assignment of Γ often does not exhibit persistence in time or with respect to the different ensemble members. This can be a sign of misinterpreting noise to be the signal. The aim is to mitigate this effect in order to identify a robust signal.

Previous studies have introduced a constraint within the clustering method to increase the temporal persistence of the regimes (Horenko, 2010; de Wiljes *et al.*, 2014; Falkena *et al.*, 2020). Here we expand on that idea by implementing a constraint on the similarity between the ensemble members at every time-step, with the aim of identifying a more robust regime signal, as discussed in Section 2. This constraint takes the form

$$\frac{1}{2}\sum_{i=1}^{k}\sum_{n_1,n_2=1}^{N}|\gamma_i(t,n_1)-\gamma_i(t,n_2)|\leq\phi\cdot C_{\mathrm{eq}}, \qquad \forall t\in[0,T], \tag{3}$$

where the sum over n_1,n_2 is taken over all combinations of two ensemble members, that is $\sum_{n_1=1}^{N-1}\sum_{n_2=n_1+1}^{N}$. The division by two ensures that differences are not counted twice. C_{eq} is the maximum value that can be attained by the sum on the left-hand side and is given by

$$C_{\mathrm{eq}}=\frac{N}{2}\left(N-\frac{N}{k}\right). \tag{4}$$

The maximum of C_{eq} is reached if the ensemble members are equally distributed among the k regimes. Thus ϕ represents the strength of the constraint relative to the maximum value C_{eq}. One can think of ϕ as the proportion of the ensemble members that is not affected by the constraint. Note that, by expressing the constraint in this way, its strength, as given by ϕ, is independent of the ensemble size or the number of regimes. Instead, the ensemble size and number of regimes enter into C_{eq}. Since we implement the constraint separately for every time step we do not make any assumptions on the form of the non-stationarity, but only ensure the algorithm can better discriminate between the regimes at a given time.

The regimes are obtained by minimising the clustering functional $\mathbf{L}$ in Equation (1) subject to the constraints in Equations (2) and (3). This minimisation is done in two steps:

1. For fixed Γ, minimise $\mathbf{L}$ over all possible values of Θ.
2. For fixed Θ, minimise $\mathbf{L}$ over all possible values of Γ.

The first step is realised via standard k-means clustering, while for the second we employ linear programming, that is, optimisation of a linear function subject to constraints, using the Gurobi package for Python (Gurobi Optimization, 2019). When the difference between subsequent $\mathbf{L}$ values comes below a set threshold, the computation is terminated. Analogously to standard k-means clustering, this presented algorithm only finds local minima. Therefore we run it at least 50 times starting from different initial seeds in an attempt to heuristically infer a global minimum (this approach is referred to as simulated annealing in the literature). The run with the lowest $\mathbf{L}$ value is then selected as the final result.

Falkena *et al.* (2020) identified $k=6$ to be the optimal number of regimes for representing the wintertime circulation over the Euro-Atlantic region for ERA-Interim using an information criterion (Bayesian), and we therefore choose 6 as the number of regimes considered for this study. While $\gamma_i(t,n)\in\{0,1\}$ within the standard k-means clustering, this is no longer the case for all time-steps or ensemble members when incorporating the constraint. Specifically $\gamma_i(t,n)\notin\{0,1\}$ when it is not possible to numerically obtain a solution on the bounds of the admissible set of the optimisation problem; in that case $\gamma_i(t,n)$ is between 0 and 1. We use this as an indication that those data-points cannot be unambiguously assigned to one of the regimes and employ it to define a 'no-regime' category. Note that this means that, even if $\gamma_i(t,n)$ for some t, n is very close to 1 for a regime, it is still assigned to be no-regime. Using this definition of a no-regime category, the number of data-points assigned to it increases approximately linearly with ϕ (not shown), that is, the stronger the constraint (lower ϕ), the more data are assigned to the no-regime category.

3.3 | Estimation of signal strength

When discussing the obtained regimes and their non-stationary dynamics, we use a regression analysis to identify the strength of the signal in comparison with that in observations. Assume there is a true signal given by $c(t)$.

For an observational time series $y(t)$ we can then write

$$y(t) = ac(t) + e_y(t), \tag{5}$$

where $e_y(t)$ represents noise. Note that we explicitly allow for the possibility that the index (e.g., the NAO) we consider is only a projection of the "true" signal, which probably is not perfect, hence $a \leq 1$. In a similar way, we have a statistical model for the time series of an ensemble member $x_i(t)$ given by

$$x_i(t) = bc(t) + e_{x_i}(t), \tag{6}$$

now with a coefficient b as the model likely is imperfect. For the ensemble mean $\bar{x}(t)$ we then obtain

$$\bar{x}(t) = bc(t) + e_{\bar{x}}(t). \tag{7}$$

We regard the ensemble mean as the best estimate of the signal, and ask how well it can predict the observations. Thus we regress $y(t)$ onto $\bar{x}(t)$, that is, estimating $y(t) = A\bar{x}(t) + E_y(t)$, which yields $A = a/b$ as the regression coefficient. This is the ratio of signal strengths with the model prediction being well calibrated if $a = b$.

The regression coefficient thus provides information on the signal strength without having to explicitly address the noise of the observations, nor of the model. Since estimates of the noise in the observations (i.e., e_y) are especially uncertain, it is beneficial to avoid having to quantify them when estimating the signal strength. The regression coefficient a/b can be linked to more conventional measures of signal strength such as the Anomaly Correlation Coefficient (ACC) or the Ratio of Predictable Components (RPC) (Eade *et al.*, 2014). One can derive that

$$\mathrm{ACC} = \frac{a}{b}\frac{\sigma_{\bar{x}}}{\sigma_y}, \qquad \mathrm{RPC} = \frac{a}{b}\frac{\sigma_{x_i}}{\sigma_y}, \tag{8}$$

where $\sigma_{y,\bar{x},x_i}$ are the standard deviations of the residuals for the respective variables. So long as $\sigma_{x_i} \approx \sigma_y$, the RPC provides the same information as the regression coefficient. However the regression coefficient is a more robust estimate, as it does not require estimating the noise.

4 | SELECTION OF REGULARISED REGIME MODEL

Using a constraint to regularise the outcome of the clustering algorithm requires choosing a suitable constraint value ϕ. This value determines the regime model that is used for the subsequent analysis of occurrence rates and non-stationarity. An appropriate value is highly dependent on the considered application and needs to be determined accordingly. Yet it is possible to employ several different selection criteria to aid the decision process, independently of the underlying physical processes. In the next section we introduce two main criteria that can be used for the constraint selection, as well as a check on the domain dependence of the regimes. We then evaluate these measures for the considered problem and discuss the arguments to arrive at our final choice of ϕ.

4.1 | Selection criteria

Numerous methods exist for deciding on the best model for the data at hand, that is, to find an optimal value of the constraint. For example, for the choice of the optimal number of circulation regimes, researchers have used verification by synthetic datasets (e.g., Straus *et al.*, 2007), a classifiability index (e.g., Michelangeli *et al.*, 1995), an information criterion (e.g., O'Kane *et al.*, 2013) or cross-validation (Quinn *et al.*, 2020). These methods can be used not only to determine the number of circulation regimes, but also the values of other hyper-parameters (e.g., Falkena *et al.*, 2020; Quinn *et al.*, 2020). In this section we introduce three criteria, namely the Bayesian Information Criterion, Shannon entropy and a domain robustness measure, which can all be used to inform the choice of a suitable constraint value ϕ.

4.1.1 | Bayesian information criterion

Information criteria are a popular tool for model selection. The aim is to find a balance between the accuracy and complexity of the model (in the spirit of Occam's razor), that is, between how well the model fits the data and the number of parameters required (Burnham and Anderson, 2004). There are many different information criteria variants, where versions of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are the most widely used. In the context of our application, the BIC is preferable to the AIC as the considered dataset is high-dimensional (by using grid-point data) and in contrast to the AIC the BIC factors the dimension of the dataset into the penalty term (Falkena *et al.*, 2020 discuss this further). The BIC is given by

$$\begin{aligned}\mathrm{BIC} &= -2\log\{\mathcal{L}(\widehat{\theta}|\mathrm{data})\} + K\log(m)\\ &= m\log(\widehat{\sigma}^2) + K\log(m),\end{aligned} \tag{9}$$

where m is the dimension of the data and K the number of parameters needed to describe the clusters. $\mathcal{L}(\widehat{\theta}|\mathrm{data})$

is the likelihood of the model parameters given the data, which can be expressed using the error variance $\widehat{\sigma}^2$. Note that the error variance, being a measure of the accuracy of the regimes, is estimated using the clustering functional $\mathbf{L}$ in Equation (1). When computing the BIC, we need to determine m and K. Each ensemble member has dimension $D \times T$, and the main question is in what way the number of ensemble members is incorporated. There are two choices that can be made for this. Firstly, one can simply use the number of ensemble members N. However, the regularisation constrains the number of combinations between any two ensemble members $\binom{N}{2}$ instead of the assignment of each of the N ensemble members individually. For this reason, we decide to use $\binom{N}{2}$ as the dimension of the ensemble, which yields $m = D \times \binom{N}{2} \times T$ and $K = k \times D + T \times (k-1) \times \phi \cdot \binom{N}{2}$.

4.1.2 | Shannon entropy

The second method considered to identify a suitable constraint value ϕ is to calculate the entropy, or information content. Entropy has already been used occasionally in evaluating model performance for circulation regimes (Fabiano *et al.*, 2020), and some studies use it as a way of correcting information criteria for the distribution of the model residuals (Murari *et al.*, 2019; Rossi *et al.*, 2020). The goal of regularising the clustering algorithm is to identify a stronger non-stationary signal, that is, to increase the amount of information in the resulting regime dynamics. Therefore, the use of an information measure, such as entropy, follows naturally from the aim of implementing the constraint. Here we use the Shannon entropy (Shannon, 1948; Shannon and Weaver, 1949), which is given by

$$H(p_i) = -\sum_{i=0}^{k} p_i \log_2 p_i, \tag{10}$$

where p_i, $i \neq 0$, is the occurrence rate of regime i and p_0 is the occurrence rate of no-regime. The Shannon entropy is low for an equal distribution of the data over the regimes. On the other hand it is larger for a more unequal distribution in which there is a stronger signal. Note that the information gain by enforcing a less equal distribution of the data over the regimes comes at the cost of reduced accuracy.

4.1.3 | Domain robustness

Lastly, we discuss the domain robustness of the regimes to see whether a choice of constraint is suitable. By domain robustness we refer to the domain dependence of the regimes as obtained by the algorithm. Using standard k-means clustering, the regimes of the SEAS5 hindcast data are found to be domain dependent, as discussed in the Section 1 (Figure 1). From a physical perspective this is undesirable, as the conclusions on the regime dynamics can then depend on the chosen domain. To transform this domain dependence into a verification criterion for the constraint value, we consider the clustering results for both domains A and B and compute the pattern correlation between the two sets of regimes over the overlapping section of these domains. The average pattern correlation is then computed as a measure of how well the regimes for the two domains match. Alternatively, one could take the lowest pattern correlation to indicate the quality of the match, that is, what is the worst matching regime. This yields very similar results.

4.2 | Selecting the constraint value

In this section the three discussed criteria are evaluated for the considered data. In Figure 3 these criteria are shown for a range of ϕ, where the BIC and Shannon entropy are shown for the two domains considered. For the BIC, which strikes a balance between accuracy and complexity, the optimal value is located at its minimum. In contrast, for the entropy, which compares information content and complexity, a higher value indicates a better result. The pattern correlation between the two domains ideally is as high as possible, indicating robustness of the regimes with respect to the choice of domain.

The BIC attains its minimum at $\phi = 0.96$ for both domains, indicating that for that constraint value the accuracy of the regime-representation of the data is still high. When looking at the results, we find that these regimes and the assignment of the data to them are very similar to those without the constraint, indicating that the constraint is too weak to have a strong impact. Furthermore, we find that this minimum depends strongly on the number of ensemble members considered, for example, using 26 members the minimum of the BIC is found for $\phi = 0.92$. While the BIC is generally a good method to select certain hyper-parameters, the goal of the regularisation is not just to identify the best statistical model and attain the highest accuracy, but also to obtain a more pronounced regime signal. To this end it can be desirable to lose some accuracy, by using a stronger constraint value, for gaining information.

The Shannon entropy indicates an optimal constraint value around $\phi = 0.92 - 0.94$, which is slightly stronger, that is, a lower ϕ, than the optimum indicated by the BIC. This slightly lower ϕ is where most information, or signal strength, is gained by constraining the data. It shows that

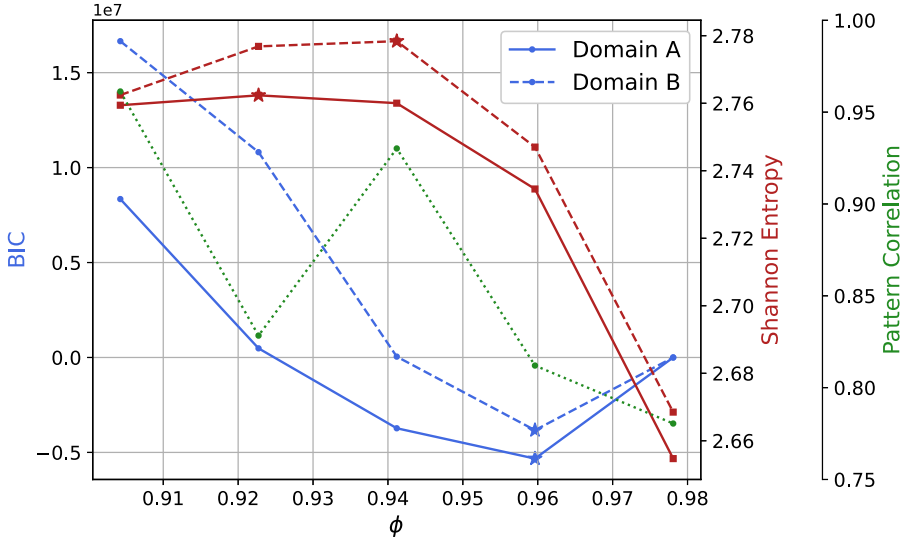

FIGURE 3 The BIC (blue, circles) and Shannon entropy (red, squares) for the two domains considered. The average pattern correlation between the regimes for the two domains (green, dotted) is shown as well. Stars indicate the lowest (blue) or highest (red) value, suggesting a suitable value for the constraint ϕ [Colour figure can be viewed at wileyonlinelibrary.com]

to identify a stronger signal, that is, higher entropy, one loses some accuracy, that is, higher BIC. Since the aim of implementing the constraint on the ensemble similarity is to identify a stronger signal, we decide to use the entropy results as the main guidance, and select the high end of the range where the entropy is maximized, that is $\phi = 0.94$, to not lose too much accuracy. Also, for this constraint value the regimes are barely domain dependent, as indicated by the high average pattern correlation.

5 | REGIME DYNAMICS

For this suitable value of $\phi = 0.94$ we study the resulting regime dynamics. We start by discussing the regimes themselves and their overall occurrence rates, after which we turn to the non-stationary signals that can be identified. When discussing the non-stationarity we also consider the results obtained using standard k-means clustering, to further look into the effect of the regularisation. We look at variability on both the sub-seasonal and interannual time-scales. For the interannual variability we discuss whether there is any predictable signal for the regime occurrence rates and compare this to the signal identified for an NAO-index.

5.1 | Effect of the constraint

The constraint affects the assignment of the data to the regimes, and thus their occurrence rates. In Figure 4 the average occurrence rates of the regimes are shown for standard and constrained k-means clustering. For the unconstrained result, as well as ERA-Interim, the occurrence rates are close to an equal distribution of the data over all six regimes (dotted line in Figure 4). In contrast, the occurrence rates differ significantly from an equal distribution when $\phi = 0.94$ is used as a constraint. Despite the relatively weak constraint, several regimes, such as the NAO+, have occurrence rates whose range barely overlaps with that of a uniform distribution (when corrected for the no-regime occurrence rate). This shows that the constraint helps to discriminate better between the different regimes, identifying a stronger regime signal within the SEAS5 data. It follows that the uniformity of the occurrence rates found for the ERA-Interim regimes is potentially due to a lack of discrimination between the regimes, as there are not enough data available to control the noise.

The geographical regime structures are mostly unaffected by the constraint of $\phi = 0.94$, as can be seen in Figure 5. Only AR− changes noticeably, now having a weak positive Z500 anomaly over Greenland. Overall these regimes correspond well to the regimes obtained for ERA-Interim, as can be seen in Table 1. The exception is AR−. However, the sample size for the regime identification in ERA-Interim is limited. Therefore, the poorer regime correspondence between ERA-Interim and the constrained results for AR− does not necessarily mean that the constrained SEAS5 AR− regime is incorrect. It may instead indicate that assigning SEAS5 data to the ERA-Interim regimes might not be the best approach for identifying a robust and statistically significant regime signal.

Overall the regularisation ensures that NAO+ occurs more often, while NAO−, AR+ and SB− occur less often, than the unconstrained results. To study in detail how the assignment of the data changes between the unconstrained and constrained results, we look at the contingency table given in Table 2. For the cases where significant amounts of data are reassigned to a different regime by using the constraint (over 5,000, in bold) we compute composites (Figure 6) to look into the

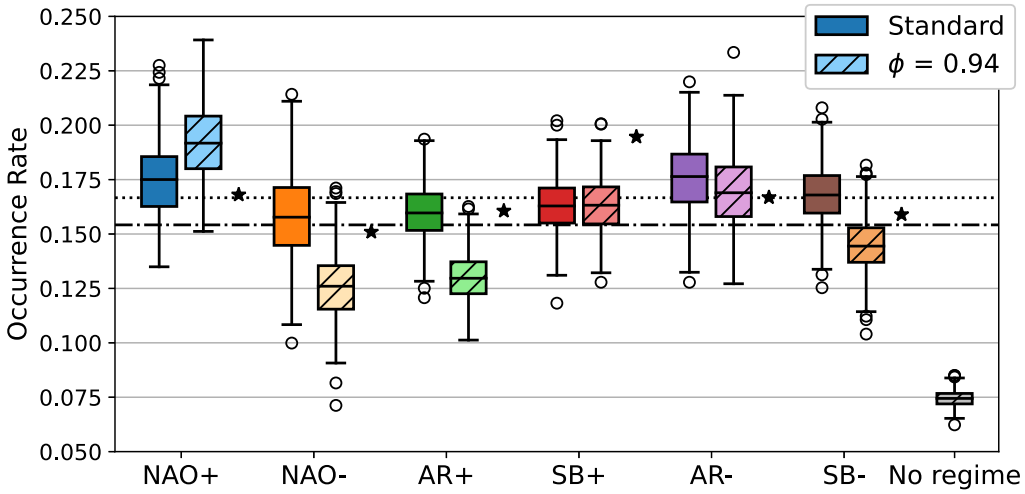

FIGURE 4 The overall occurrence rates of the different regimes for the results with and without a constraint. The boxes show the interquartile range (IQR) for bootstrapping with one (random) ensemble member per year, the whiskers extend 1.5 times the IQR on top of this (99.3% of the data fall within this range) and the circles are outlier points. Stars indicate the ERA-Interim values. The dotted line gives the 1/6 line indicating an equal distribution of the data over the regimes, while the dash-dotted line indicates an equal distribution after correcting for the no-regime rate. Note that there are no ERA-Interim data assigned to the no-regime category by the way this category is defined using the outcome of the regularised algorithm [Colour figure can be viewed at wileyonlinelibrary.com]

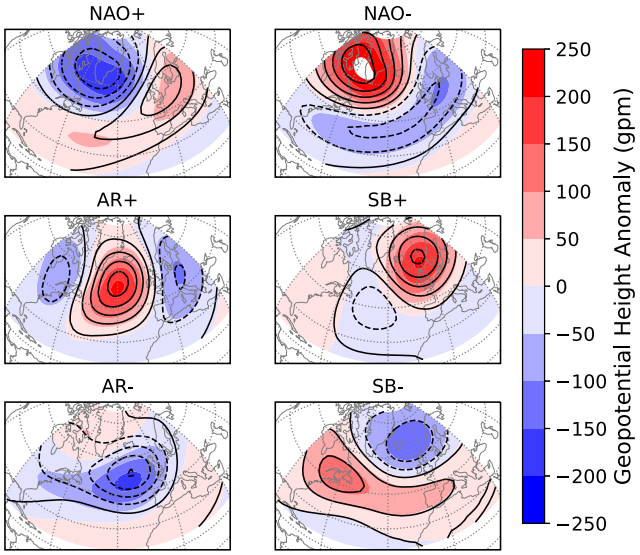

FIGURE 5 The regimes for domain A using a constraint value of $\phi = 0.94$ (colour shading) and without constraint (contours, with the same 50 gpm difference between contour levels) [Colour figure can be viewed at wileyonlinelibrary.com]

Z500 anomaly structure of this data and interpret the changes.

A large proportion of the data that without constraint was assigned to AR− is assigned to NAO+, explaining the latter's increase in occurrence rate. In turn, AR− now contains a substantial part of the data that without constraint was assigned to NAO−. This reflects the change in the AR− regime with a higher positive anomaly in the north for the regularised results, which is exemplified by the composites shown in Figure 6. It also can be linked to the slight strengthening of the positive Z500 anomaly for NAO−, as data with a relatively weak anomaly move to the AR− regime. Interestingly NAO+ loses some of its data points to SB+ when the constraint is used. This concerns data with a northwestern negative and northeastern positive Z500 anomaly (Figure 6) where the balance of regime assignment is shifted by the regularisation. The unconstrained NAO+ has a relatively high positive Z500 anomaly with its centre over the North Sea, which is lower when the constraint is used, corresponding to the positive Z500 anomaly of SB+ being slightly weaker and located further south. The decrease in occurrence rate of AR+ is due to data having slightly off-centre positive anomaly areas now being assigned to NAO− or SB+. Data assigned to SB− when no constraint is used form the largest part of the no-regime set of the data, accounting for the majority of the SB− decrease in occurrence rate.

Changes in transition probabilities between the regimes as a consequence of the implementation of the constraint are roughly in line with changes in the occurrence rates (not shown). That is, regimes that occur more often become more persistent and are less likely to transition to another regime, and the other way around for regimes that occur less often. One notable change is the increase in the number of transitions from NAO− to AR−, which is due to the change in the AR− regime ensuring both regimes have an area of positive Z500 anomalies over Greenland (Figure 5). This change is one-way, as there is

TABLE 1 The pattern correlation of the SEAS5 regimes (standard and regularised with constraint value $\phi = 0.94$) with the ERA-Interim regimes for domain A

	NAO+	NAO−	AR+	SB+	AR−	SB−
Standard	0.92	0.98	0.97	0.91	0.95	0.96
Regularised	0.96	0.96	0.95	0.95	0.71	0.96

TABLE 2 A contingency table for the assignment of the SEAS5 data to the regimes for the results without constraint and for $\phi = 0.94$

	$\phi = 0.94$							
Unconstrained	**NAO+**	**NAO−**	**AR+**	**SB+**	**AR−**	**SB−**	**No-regime**	**Total**
NAO+	*27,994*	0	1,443	**5,367**	54	1,297	2,934	39,089
NAO−	0	*23,301*	87	37	**9,432**	310	2,069	36,227
AR+	14	3,409	*25,867*	2,349	480	426	2,981	35,526
SB+	412	1,132	208	*28,005*	3,881	50	2,670	36,358
AR−	**12,512**	0	640	71	*22,190*	795	2,757	38,965
SB−	1,946	104	722	685	1,474	*29,265*	3,254	37,450
Total	42,878	27,946	28,967	36,514	37,511	32,143	16,656	222,615

Note: Each column indicates the regime assignment following standard *k*-means clustering of the constrained regimes. Values over 5,000 data points are in *italic* for the same regime and **bold** for a different regime.

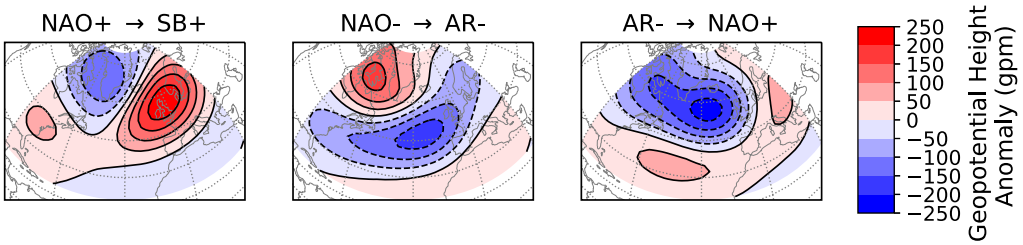

FIGURE 6 Composites of the most frequent regime reassignments between the results without constraint and for $\phi = 0.94$ (as shown in bold in Table 2). Shown are composites of data which without constraint are assigned to NAO+, NAO− and AR−, but which with constraint are assigned to SB+, AR− and NAO+, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

no increase in the transition probability from AR− into NAO−.

The above discussion of the effect of regularising the clustering algorithm shows that the constraint works as expected. That is, the occurrence rates of the regimes become more distinct, indicating that a more pronounced regime signal is identified. The changes in the regimes are in line with changes in the assignment of the data to them and no unexpected changes in transition probabilities are found. In the next sections we turn to discussing the non-stationary behaviour of the regimes. We start with a brief discussion of the sub-seasonal signal, followed by a more detailed study of the interannual signal.

5.2 | Sub-seasonal variability

Since we consider anomaly data with respect to a constant background climatological state, it is expected that there is a seasonal signal in the occurrence rates of the regimes. For SEAS5 this is shown in Figure 7 by the dash-dotted black line for the regularised results. The sub-seasonal variability obtained using standard *k*-means clustering is shown as well (dotted black line), exhibiting similar behaviour in time as found for the regularised results. The sub-seasonal variability for ERA-Interim falls within the ensemble spread of the SEAS5 results. This variability is not shown since the sample size is too small to draw reliable conclusions.

NAO+ exhibits the largest variability throughout the season with a maximum occurrence rate close to 0.3 in mid-January and a minimum below 0.1 in March, with the identified variability being amplified by the regularisation. All other regimes exhibit a seasonal cycle as well, where the amplitude of the variations differs between the regimes. We see that AR+ and AR− have a peak in occurrence rate in February, whereas NAO−, SB+ and SB− have a minimum in January and/or February. Most of this

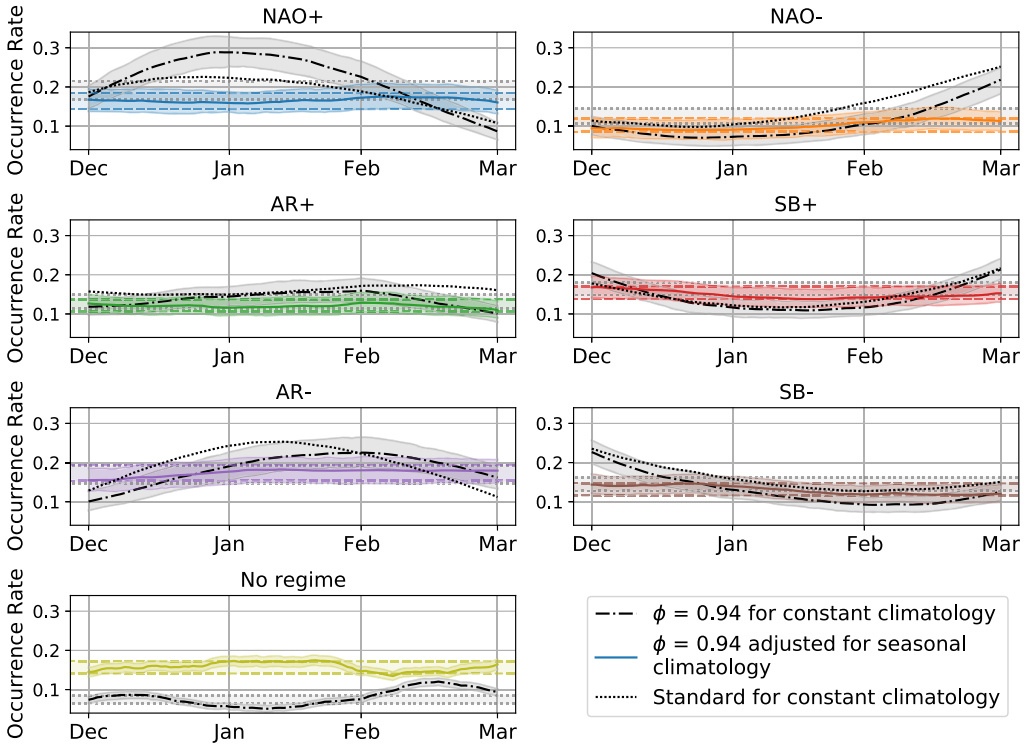

FIGURE 7 The sub-seasonal variation in occurrence rates (30-day running mean) of the different regimes for the constrained SEAS5 results with respect to a constant climatology (black, dash-dotted line) and corrected for a seasonally varying background state (colour, solid line), with the shaded area indicating the two-standard deviation range. The areas bounded by the horizontal dotted and dashed lines give the 2-standard deviation noise level for the constant and seasonal climatology results, respectively. Error bounds are determined using bootstrapping with three members for every 30-day period in each year. In addition the sub-seasonal variation obtained using standard *k*-means clustering for SEAS5 is shown by the black dotted line [Colour figure can be viewed at wileyonlinelibrary.com]

variability exceeds the sampling uncertainty as shown by the shaded area bounded by the grey dotted lines. Looking at the variability on sub-seasonal time-scales found in other studies, comparison is difficult because the number of regimes considered is different (e.g., Cortesi *et al.*, 2021). The seven year-round regimes of Grams *et al.* (2017) come closest and show comparable dynamics for AR+, AR− (Atlantic trough) and SB− (Scandinavian trough), while for NAO+ (zonal regime) an opposite signal emerges with lowest occurrence rates in January (when looking at DJFM). This may be linked to the different way in which a no-regime state is determined. A similar story holds for NAO− (Greenland blocking) although the difference is less robust.

To study whether this sub-seasonal variability is solely due to the changing background state within winter, we correct for this effect. This is done by computing an anomaly dataset with respect to a sub-seasonal climatology, instead of a fixed one, and assigning the obtained Z500 fields to the closest of the regimes shown in Figure 5. The sub-seasonal climatology is computed by fitting a fourth-order polynomial to the daily averaged fields. The assignment of the sub-seasonal anomaly data is done by first computing the distance of the data to the regimes and then minimising the clustering functional $\mathbf{L}$ over all values of Γ subject to the constraint (for $\phi = 0.94$), that is, we apply the second step of the constrained clustering algorithm (Section 3.2). This ensures that the corrected occurrence rates are comparable with the standard constrained results. Note that this will result in more data being assigned to no-regime, as the minima of $\mathbf{L}$ for the sub-seasonal anomaly data are likely to differ slightly from those of the anomalies with respect to a constant climatology. An alternative approach would be to cluster the data again after having removed the sub-seasonal climatology, however this would require rerunning the clustering for each choice of the sub-seasonal climatology one wishes to consider, whereas the approach taken here does not require re-running the clustering.

The sub-seasonal occurrence rates corrected for the seasonal cycle of the mean climatology are shown in colour (solid line) in Figure 7. As expected the occurrence rate of no-regime has increased, with an approximate doubling of the number of data points that are difficult to assign. This leads to an overall decrease in the occurrence rate of

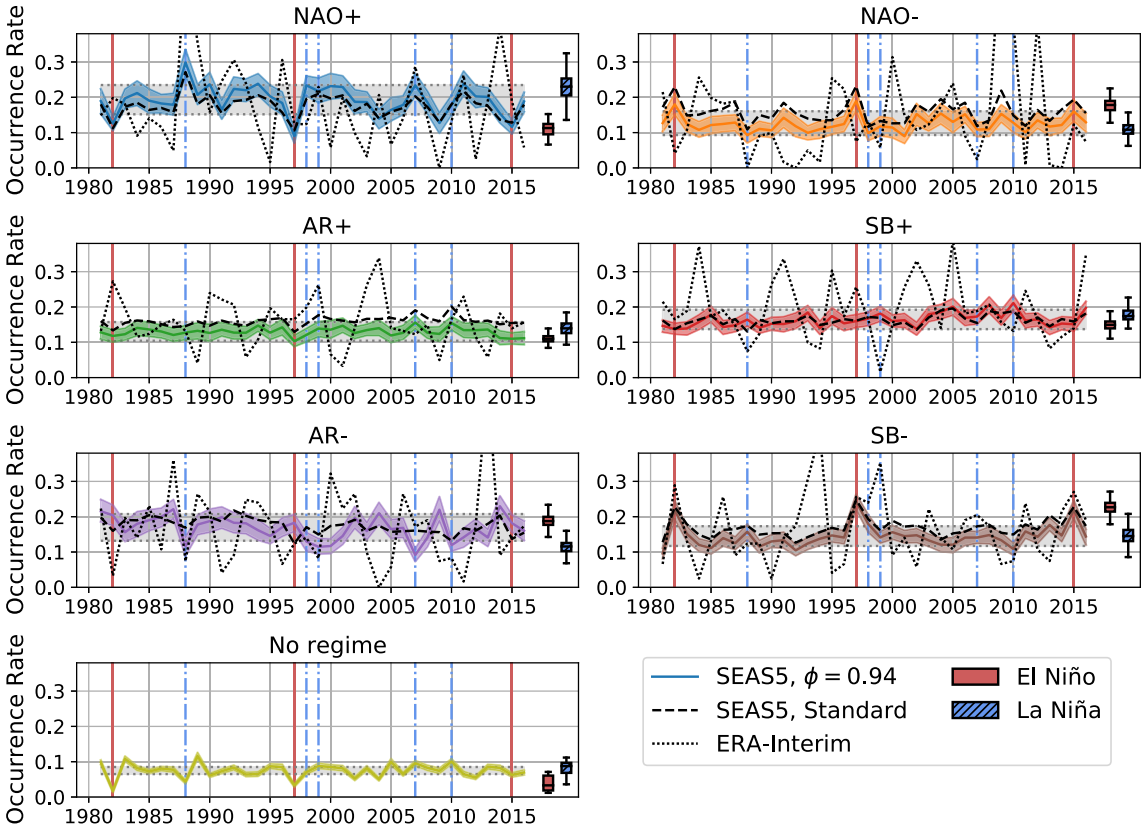

FIGURE 8 The yearly winter occurrence rates of the different regimes for the constrained results (colour), where the year indicated on the axis corresponds to December of that winter. The grey bands give a noise level, the dashed black line shows the unconstrained SEAS5 results and the dotted black line the ERA-Interim occurrence rates. The bandwidth is given by the 2-standard deviation range when using bootstrapping with 25 ensemble members, where for the noise level ensemble members for different (random) years are considered. On the right the average occurrence rates of very strong El Niño years (indicated by the solid vertical red lines) and strong La Niña years (vertical dash-dotted blue lines) are shown [Colour figure can be viewed at wileyonlinelibrary.com]

the regimes, which is largest for NAO+. The sub-seasonal variability is significantly reduced after correcting for the background climatology, and for all regimes the average falls within the sampling uncertainty. There still is some variability, for example, SB− appears to be more likely in early winter, but this is not statistically significant. Thus, we do not find any significant sub-seasonal variability in the regime occurrence rates when using a sub-seasonal climatology. We conclude that the seasonal cycle in the occurrence rates primarily reflects the seasonal cycle in the mean climatology, rather than any seasonal cycle in the variability itself.

Note that any attempt to correct for a varying climatology will be dependent on the choice of sub-seasonal climatology. For example, when one uses a 90-day running mean as reference climatology (e.g., Grams *et al.*, 2017) it is possible that part of the sub-seasonal signal in occurrence rates seen in Figure 7 remains. Furthermore, often meteorological data are grouped according to the season (e.g., DJF) and sub-seasonal variations are not considered to first order. Thus, we deem it better to use a fixed climatology before clustering and identify the sub-seasonal signal afterwards. This way there is no assumption on the form of the sub-seasonal climatology, which could affect the regimes themselves and the attribution of the data to them.

5.3 | Interannual variability

In Figure 8 the wintertime interannual variability in occurrence rates is plotted for the constrained results (colour), as well as for standard k-means clustering for comparison (black dashed). These results are with respect to a constant climatology; the interannual variability when correcting for a seasonal background climatology is comparable (not shown). The signal identified in the interannual variability, as indicated by the SEAS5 ensemble mean, is slightly stronger for the regularised results than that obtained without constraint. The average (over all regimes) mean standard deviation of the interannual occurrence rate for the bootstrapped results (using 25 members) is 0.029 with a 95% confidence interval of [0.027, 0.031] for the regularised

results, while it is 0.026 [0.024, 0.028] without the constraint. Thus the variability on interannual time-scales is slightly amplified by using the regularisation. In this and the next paragraph the discussion is focused on the regularised results. Note that the standard deviation of single ensemble members is of the same order as that of ERA-Interim, albeit slightly smaller on average (0.087 versus 0.104). For NAO+, AR− and SB− we find strong interannual variability in the ensemble mean occurrence rates, whereas for AR+ and SB+ no significant signal is found. Interestingly, the interannual variability in the occurrence rate of NAO− is weaker than that of NAO+, suggesting a smaller predictable signal.

The majority of the signal coincides with El Niño or La Niña years, as shown in the boxplots on the right-hand side of Figure 8. Here we refrain from separating early (November and December) and late (January and February) winter as is sometimes done (e.g., Moron and Plaut, 2003; Ayarzaguena *et al.*, 2018), because we would not want to include November (due to the initialisation on 01 November). During winters in which there was a very strong El Niño, indicated by the solid red lines, we find an increase in occurrence of SB− and NAO−, and a decrease of NAO+. In those years there also is less data that cannot be attributed to one of the regimes, indicating that the ensemble members are more similar in their dynamics. On the other hand, we see an increase of NAO+ and decrease of AR− occurrence during strong La Niña winters, indicated by the dash-dotted blue lines, with the highest NAO+ value in 1988–1989 for both SEAS5 and ERA-Interim. This is in line with previous studies looking into links between ENSO and the NAO (Toniazzo and Scaife, 2006; Li and Lau, 2012; Ayarzaguena *et al.*, 2018; also Figure 10 below), although here we capture this relation in more detail by using the regime variability. Note that data assigned to either NAO+ or SB− would tend to be assigned to the positive phase of the NAO when considering only four regimes. However, their response to a strong El Niño is opposite in sign, that is, SB− becomes more frequent while NAO+ occurs less often. The distinction between these two regimes thus allows better understanding of the details of the response of the circulation to ENSO.

To see whether the SEAS5 ensemble provides a predictive signal for the ERA-Interim occurrence rates, we regress the ERA-Interim annual occurrence rates against those of SEAS5 (Section 3.3). The scatter plots for this regression are shown in Figure 9. The slopes and p-values are given in Table 3 for each of the regimes for both the constrained and unconstrained results. In addition, the Bayes factor is given. The Bayes factor is the ratio of the probability of the data given a hypothesis for two different hypotheses H_1 and H_2, that is, $P(D|H_1)/P(D|H_2)$ (Kass and Raftery, 1995) and has been recently used in climate studies (Kretschmer *et al.*, 2020). Here, the first hypothesis H_1 is the linear regression model and the second hypothesis H_2 is a constant occurrence rate following the overall value. A value above 1 indicates that H_1 is more likely than H_2, while the converse is true for a value below 1. To have strong evidence towards the hypothesis of linear regression, the Bayes factor would have to be much larger than 1. The Bayes factor allows for the comparison of different hypotheses, whereas the p-value only indicates whether the null hypothesis can be rejected without providing an alternative (Wagenmakers, 2007; Shepherd, 2021).

This linear regression analysis indicates that there is a predictive signal for NAO+ and SB− with p-values below 0.05 using $\phi = 0.94$, while without the constraint only the NAO+ signal is found to be significant at the 95% level (Table 3). The Bayes factor for both constrained regimes is substantially larger than 1, albeit not very large. This constitutes positive, but not yet particularly strong, evidence that the signal seen in the model is reflected in the observations (Kass and Raftery, 1995, in which values of 3–20 are said to constitute positive evidence, while values over 20 yield strong evidence). Note that these two regimes are characterised by a zonal flow pattern. Comparing the regularised result with the standard approach, the constraint adds a significant predictable signal for SB− which would not otherwise have been found. The regression coefficient is around 1 for both the NAO+ and SB− regimes, indicating just as strong a signal in SEAS5 as in ERA-Interim, as discussed in Section 3.3. For NAO− the regression coefficient is around 1 as well, but this is not significant as indicated by the high p-value and Bayes factor close to 1. No predictable signal is found for the other three regimes either.

The absence of a significant signal for NAO− is intriguing, as we do obtain a signal for NAO+. Interestingly, we obtain a strong predictable signal for the NAO− regime by applying multiple linear regression using the NAO+ and SB− occurrence rates (i.e., regressing the observed NAO− onto the SEAS5 NAO+ and SB−; last column of Table 3). For these regimes the response of the occurrence rate of SB− to a strong El Niño is similar to that of NAO−, but that of NAO+ is opposite (Figure 8). The Bayes factor here is very large and constitutes strong evidence towards this being a real signal. Hence the NAO− regime is predictable from SEAS5, just not from the SEAS5 NAO− regime signal itself. Again regularisation significantly improves the predictability, as that of the standard k-means results has a considerably larger p-value and smaller Bayes factor. Both the NAO+ and SB− regime patterns project well onto the positive phase of the NAO-index, which could in part explain the strong signal obtained from these two regimes for the predictability of the NAO− regime, which projects

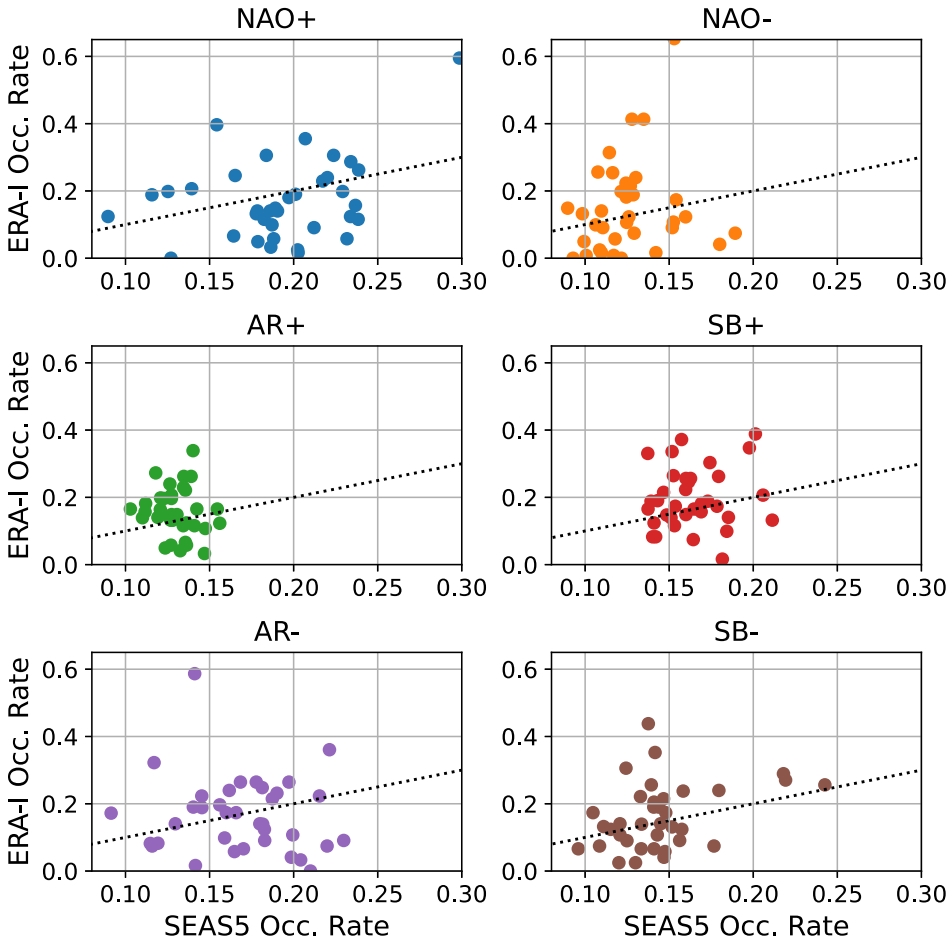

FIGURE 9 Scatter plots of the annual winter occurrence rates of ERA-Interim against those of the SEAS5 ensemble mean for each of the six regimes. The dotted lines show a one-to-one relation [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 The results for linear regression of the regime occurrence rates, linear regression of the NAO-index and multiple linear regression (MLR) of the ERA-Interim NAO− against the SEAS5 ensemble mean NAO+ and SB− occurrence rates, for both the regularised ($\phi = 0.94$) and standard results

	Regime	NAO+	NAO−	AR+	SB+	AR−	SB−	NAO-index	MLR	NAO−
$\phi = 0.94$	Regression coeff.	0.99	0.98	−0.69	0.53	−0.32	1.14	1.87	NAO+	−1.35
									SB−	−1.76
	p-value	0.04	0.35	0.50	0.49	0.58	0.03	0.01		0.03
	Bayes factor	8.81	1.60	1.28	1.29	1.18	13.20	53.82		54.10
Standard	Regression coeff.	1.23	1.11	−0.22	0.65	1.29	1.02		NAO+	−1.44
									SB−	−1.72
	p-value	0.05	0.15	0.81	0.51	0.14	0.10			0.06
	Bayes factor	7.39	3.04	1.03	1.26	3.15	4.42			16.10

on its negative phase (in line with the negative regression coefficients).

The discussion of the signal-to-noise problem of the North Atlantic is often focused on the NAO-index. Here, we compare the regression results of the regimes with those of the NAO-index. We compute the NAO-index as the first principal component of the daily 500 hPa geopotential height fields for December to March (Weisheimer *et al.*, 2017). The yearly NAO-index is then computed as the average index over all days in that winter and shown as the dashed black line in Figure 10 for SEAS5. The regression for this NAO-index is shown in Figure 11 with the coefficient and statistics given in Table 3. The signal for this NAO-index is strong with a Bayes factor of over 50

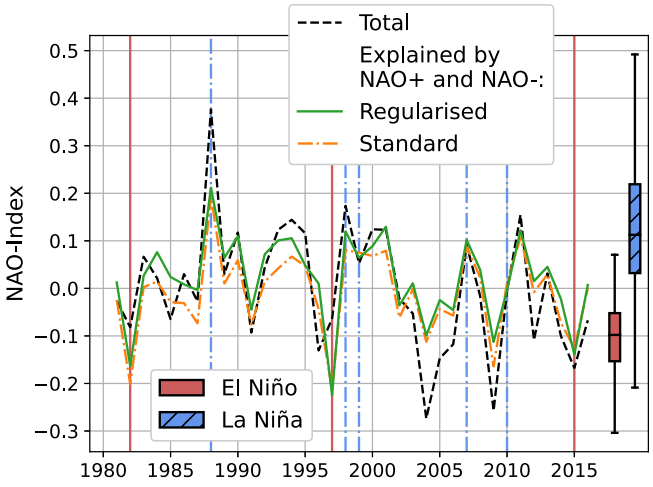

FIGURE 10 The SEAS5 NAO-index (black, dashed) and an approximation using the NAO+ and NAO− occurrence rates and projected NAO-indices for the regularised (green, solid) and standard (orange, dash-dotted) approaches. On the right the average NAO-indices of very strong El Niño years (indicated by the solid vertical red lines) and strong La Niña years (vertical dash-dotted blue lines) are shown [Colour figure can be viewed at wileyonlinelibrary.com]

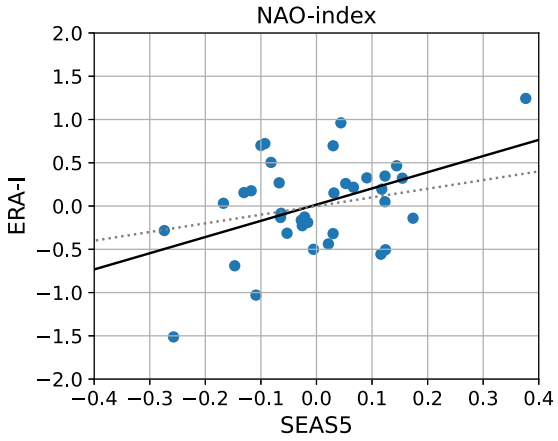

FIGURE 11 Linear regression of the NAO-index (black solid line), being the winter average of the first principal component of the DJFM daily Z500. The dotted line shows a one-to-one relation [Colour figure can be viewed at wileyonlinelibrary.com]

(comparable to that attained for the NAO− regime using NAO+ and SB− as predictors). The regression coefficient is roughly 2, indicating that the SEAS5 model is underpredicting the signal in the observed NAO-index by about a factor of 2. Assuming that the variance of the error is comparable between observations and model, as discussed in Section 3.3, this result is in line with previous studies on the signal-to-noise paradox for the NAO where RPC ≈ 2 has been found as a lower bound (Eade *et al.*, 2014; Scaife and Smith, 2018). (We also computed the RPC directly and found a value of around 2.)

Thus, there is a significant difference in the model representation of the regimes compared to that of the NAO-index. While for the NAO-index we see an underestimation of the signal in the model compared to observations, in line with the signal-to-noise paradox, this is not the case for the signal in the occurrence rates of the two zonal regimes NAO+ and SB−, which are the regimes with interannual predictability. In order to analyse whether this discrepancy is due to only considering the *occurrence rates* of the regimes, we need to address whether there is a possible signal-to-noise problem in the *amplitude*, that is, strength, of the regimes. To this end we compute the average NAO-index for each regime, that is, averaging the NAO-index over all days assigned to a regime, for both SEAS5 and ERA-Interim. The results are shown in Table 4. As expected, the NAO+ and NAO− regimes contribute most to the respective phases of the NAO-index. Approximating the NAO-index in SEAS5 using the annual occurrence rates and average NAO-indices for these two regimes, that is, multiplying the NAO− and NAO+ regime NAO-indices from Table 4 by their annual occurrence as shown in Figure 8 and adding the two, provides a good estimate of the NAO-index variability for both regularised and standard results, as can be seen in Figure 10. In addition we compute the average annual winter NAO-indices for each regime in SEAS5, which are found to be uncorrelated with their respective regime frequencies (not shown). This indicates that the regime occurrence and the regime strength (in terms of its projection on the NAO-index) are independent. Hence we do not find evidence of a signal-to-noise problem in relation to the regime strengths, for example, a regime is not weak when its occurrence rate is high.

This leaves us with the discrepancy between the signal strength for the regime frequencies and for the NAO-index. Using the regularisation, we found that SEAS5 has a predictable signal for the two zonal regimes with a regression coefficient around 1. On the other hand, no signal was found for the non-zonal regimes and the NAO− signal was not manifest directly, though it could be detected from NAO+ and SB−. Thus, the signal-to-noise paradox for the NAO-index might be linked to certain regimes being poorly represented within the model, that is, the NAO-index cannot provide all the relevant information of the atmospheric flow structure for predictability in the Euro-Atlantic sector. It is not necessarily the case that the amplitude of the predictable signal in response to remote forcings such as ENSO is too weak (Scaife and Smith, 2018), but rather that the signal is only present in part of the dynamics, while other aspects are incompletely represented. The first regime to consider in this regard is NAO−, which represents a blocking over Greenland, as it is unsuccessfully predicted from the SEAS5 NAO− regime,

TABLE 4 The average NAO-index for each of the six regimes in both SEAS5 (regularised and standard) and ERA-Interim

Regime	NAO+	NAO−	AR+	SB+	AR−	SB−
SEAS5 Regularised	1.26	−1.79	−0.31	−0.16	−0.33	0.69
SEAS5 Standard	1.33	−1.53	−0.53	−0.47	0.30	0.69
ERA-Interim	1.38	−1.69	−0.31	0.06	0.06	0.31

even though a strong signal has been identified using the NAO+ and SB− regimes. This also points to the negative phase of the NAO being at the heart of the signal-to-noise problem.

A potential reason for the lack of a predictable signal in NAO− could lie in the role of stratospheric sudden warmings (SSWs), which are known to induce negative NAO states (Baldwin and Dunkerton, 2001; Hitchcock and Simpson, 2014; Domeisen *et al.*, 2020). Portal *et al.* (2021) have shown that seasonal forecast models, including the SEAS5 model studied here, tend to overpredict the SSW response to ENSO. Consistent with that, there is a strong NAO− response to ENSO (Figure 8), which is not seen in observations. However, if SSWs are playing an important role in seasonal predictability, then this role will be difficult to assess from the limited sample size provided by the reanalysis record.

6 | CONCLUSION AND DISCUSSION

To identify a regime variability signal in model hindcast ensemble data, a constraint on the similarity between ensemble members has been implemented. In this way a stronger and more informative regime signal is identified by considering the trade-off between accuracy and entropy. Different criteria are used to identify the optimal settings for the regularisation method, yielding an optimal constraint value. This optimal value is sufficiently strong to increase the information gain (as indicated by the entropy), but not so strong to lose a lot of accuracy (as indicated by the BIC). The constraint helps better discriminate between the different regimes, which is reflected in the overall occurrence rates of the regimes being more distinct. The regime patterns themselves are not strongly affected, increasing confidence in this approach.

When considering the non-stationary regime dynamics, we find that the average sub-seasonal variability is primarily determined by variability in the average background climatology. When one looks at a seasonal climatology, such as the DJF average, a large part of the found variability will remain. A question when removing a background climatology based on daily averages is whether one is removing part of the signal, for the differences in regime occurrence throughout the season do reflect the changes in the background climatology. Therefore, we regard it as cleaner to consider a constant climatology within the season and account for the background variability in the interpretation.

On interannual time-scales the NAO+, NAO−, AR− and SB− regimes show significant variability, which is enhanced by the regularisation compared to standard *k*-means clustering results. In large part this is related to ENSO, with El Niño leading to SB− and NAO− being more frequent, while La Niña corresponds to increased frequencies of NAO+ and decreased frequencies of AR−. In this respect, it would be interesting to look at the response for early and late winter separately, as those responses might differ (e.g., Moron and Plaut, 2003; Ayarzaguena *et al.*, 2018). When considering only four regimes, most data here assigned to either NAO+ or SB− would be allocated to the NAO+ regime. This separation between NAO+ and SB− could help better understand the relation between ENSO, the Madden–Julian Oscillation (MJO) and the North Atlantic circulation where previous studies found a more frequent NAO+ after MJO phases 1–5 in El Niño years using four regimes (Lee *et al.*, 2019), whereas the overall signal is a decrease of the NAO-index during a positive ENSO (Figure 10). El Niño was found to enhance the teleconnections between the MJO and the regimes, which is also apparent in the increase of NAO− after MJO phases 7 and 8. On the other hand La Niña on average inhibits such teleconnections between the MJO and the four regimes studied. It would be interesting to examine the teleconnection signal for the additional regimes (AR− and SB−) discussed here to see whether these relations can be captured in more detail.

We have used linear regression to identify the signal on interannual time-scales, as it allows a direct estimation of the ratio of signal strengths between observations and model, without requiring estimation of the noise levels. The SEAS5 ensemble has a predictable signal for the occurrence rates of the two zonal regimes (NAO+ and SB−), but not for the other regimes. The Bayes factors show a substantial improvement of predictability, especially for SB−, with the constraint compared to the standard results. Interestingly, a strong predictive signal for NAO− is obtained by considering multiple linear regression using the model NAO+ and SB−, whereas no such

signal is found using the model NAO− frequencies. The regression coefficients that come out of the linear regression are around 1 for both NAO+ and SB−, indicating that the SEAS5 signal is of the same magnitude as that found in ERA-Interim. This implies there is no signal-to-noise paradox for these two flow regimes. Note that also for NAO− the regression coefficient is around 1, but is not statistically significant.

In contrast we find that for an NAO-index the regression analysis results in a regression coefficient of 2, in line with previous studies on the signal-to-noise paradox for the North Atlantic sector that show that the model underpredicts the observed NAO by a similar factor (e.g., Eade *et al.*, 2014). Our regime analysis suggests that the NAO signal-to-noise paradox largely manifests itself in the non-zonal phase of the NAO (and related regimes), that is, in its negative phase rather than its positive phase. Improving the regime representation of NAO− (and AR−) within the SEAS5 model could not only improve the regime dynamics, but could also help shed light on the signal-to-noise paradox over the Euro-Atlantic domain.

The proposed regularisation approach allows for identifying a more informative regime signal and its application is not limited to variability on sub-seasonal and interannual time-scales, as examined in this study. It could be employed in the context of other regime behaviour studies that rely on ensemble data and might lead, as shown here, to a substantial information gain when studying model ensemble data. There is potential for applications in forecasting of regimes where, especially in the initial stages, the ensemble members are expected to exhibit even more similar dynamics than discussed here. Furthermore, this approach could help shed light on biases present in models when it comes to their regime representation and dynamics.

ACKNOWLEDGEMENTS

SKJF was supported by the Centre for Doctoral Training in Mathematics of Planet Earth, with funding from the UK Engineering and Physical Sciences Research Council (EPSRC) (grant EP/L016613/1). The research of JdW has been partially funded by Deutsche Forschungsgemeinschaft (DFG) – SFB1294/1, 318763901.

AUTHOR CONTRIBUTIONS

Swinda K.J. Falkena: conceptualization; formal analysis; investigation; methodology; project administration; validation; visualization; writing – original draft. **Jana de Wiljes:** conceptualization; methodology; supervision; writing – review and editing. **Antje Weisheimer:** conceptualization; supervision; writing – review and editing. **Theodore G. Shepherd:** conceptualization; supervision; writing – review and editing.

ORCID

Swinda K.J. Falkena https://orcid.org/0000-0002-2317-9213
Antje Weisheimer https://orcid.org/0000-0002-7231-6974

REFERENCES

Ayarzaguena, B., Ineson, S., Dunstone, N.J., Baldwin, M.P. and Scaife, A.A. (2018) Intraseasonal effects of El Niño-Southern Oscillation on North Atlantic climate. *Journal of Climate*, 31(21), 8861–8873.

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(15), 7808–7817.

Baldwin, M.P. and Dunkerton, T.J. (2001) Stratospheric harbingers of anomalous weather regimes. *Science*, 294(5542), 581–584.

Bishop, C.M. (2006) *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Burnham, K.P. and Anderson, D.R. (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2), 261–304.

Cassou, C., Laurent, T. and Phillips, A.S. (2005) Tropical Atlantic influence on European heat waves. *Journal of Climate*, 18(15), 2805–2811.

Charlton-Perez, A.J., Ferranti, L. and Lee, R.W. (2018) The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 144, 1140–1151.

Cortesi, N., Torralba, V., Lledó, L., Manrique, A., Nube, S. and Reviriego, G. (2021) Yearly evolution of Euro-Atlantic weather regimes and of their sub-seasonal predictability. *Climate Dynamics*, 56, 3933–3964.

Dawson, A. and Palmer, T.N. (2015) Simulating weather regimes: impact of model resolution and stochastic parameterization. *Climate Dynamics*, 44, 2177–2193.

Dawson, A., Palmer, T.N. and Corti, S. (2012) Simulating regime structures in weather and climate prediction models. *Geophysical Research Letters*, 39(21), 1–6.

de Wiljes, J., Putzig, L. and Horenko, I. (2014) Discrete nonhomogeneous and nonstationary logistic and Markov regression models for spatiotemporal data with unresolved external influences. *Communications in Applied Mathematics and Computational Science*, 9(1), 1–46. https://doi.org/10.2140/camcos.2014.9.1.

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., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, 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.

Domeisen, D.I.V., Grams, C.M. and Papritz, L. (2020) The role of North Atlantic-European weather regimes in the surface impact of sudden stratospheric warming events. *Weather and Climate Dynamics*, 1, 373–388.

Drouard, M. and Cassou, C. (2019) A modeling- and process-oriented study to investigate the projected change of ENSO-forced wintertime teleconnectivity in a warmer world. *Journal of Climate*, 32(23), 8047–8068.

Eade, R., Smith, D., Scaife, A.A., Wallace, E., Dunstone, N., Hermanson, L. and Robinson, N. (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?. *Geophysical Research Letters*, 41(15), 5620–5628.

Fabiano, F., Christensen, H.M., Strommen, K., Athanasiadis, P., Baker, A., Schiemann, R. and Corti, S. (2020) Euro-Atlantic weather regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance. *Climate Dynamics*, 54, 5031–5048.

Falkena, S.K.J., de Wiljes, J., Weisheimer, A. and Shepherd, T.G. (2020) Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, 146, 2801–2814.

Ferranti, L., Corti, S. and Janousek, M. (2015) Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, 141, 916–924.

Franzke, C., Horenko, I., Majda, A.J. and Klein, R. (2009) Systematic metastable atmospheric regime identification in an AGCM. *Journal of the Atmospheric Sciences*, 66(7), 1997–2012.

Grams, C.M., Beerli, R., Pfenninger, S., Staffell, I. and Wernli, H. (2017) Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7(8), 557–562.

Grams, C.M., Magnusson, L. and Madonna, E. (2018) An atmospheric dynamics perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. *Quarterly Journal of the Royal Meteorological Society*, 144, 2577–2591.

Gurobi Optimization (2019) Gurobi Optimizer Reference Manual. Beaverton, OR. Available at: http://www.gurobi.com.

Hannachi, A., Straus, D.M., Franzke, C.L.E., Corti, S. and Woollings, T. (2017) Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere. *Reviews of Geophysics*, 55(1), 199–234.

Hitchcock, P. and Simpson, I.R. (2014) The downward influence of stratospheric sudden warmings. *Journal of the Atmospheric Sciences*, 71(10), 3856–3876.

Horenko, I. (2010) Finite element approach to clustering of multidimensional time series. *SIAM Journal on Scientific Computing*, 30(3), 62–83.

Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S.P., Mogensen, K., Zuo, H. and Monge-Sanz, B.M. (2019) SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3), 1087–1117.

Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.

Kelder, T., Müller, M., Slater, L.J., Marjoribanks, T.I., Wilby, R.L., Prudhomme, C., Bohlinger, P., Ferranti, L. and Nipen, T. (2020) Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes. *npj Climate and Atmospheric Science*, 3(47), 1–13.

Kretschmer, M., Zappa, G. and Shepherd, T.G. (2020) The role of Barents–Kara sea ice loss in projected polar vortex changes. *Weather and Climate Dynamics*, 1, 715–730.

Lee, R.W., Woolnough, S.J., Charlton-Perez, A.J. and Vitart, F. (2019) ENSO modulation of MJO teleconnections to the North Atlantic and Europe. *Geophysical Research Letters*, 46, 13,535–13,545.

Li, Y. and Lau, N.-C. (2012) Impact of ENSO on the atmospheric variability over the North Atlantic in late winter – role of transient eddies. *Journal of Climate*, 25, 320–342.

Matsueda, M. and Palmer, T.N. (2018) Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, 144, 1012–1027.

Michelangeli, P.-A., Vautard, R. and Legras, B. (1995) Weather regimes: recurrence and quasi-stationarity. *Journal of Atmospheric Sciences*, 52(8), 1237–1256.

Moron, V. and Plaut, G. (2003) The impact of El Nino-Southern Oscillation upon weather regimes over Europe and the North Atlantic during boreal winter. *International Journal of Climatology*, 23, 363–379.

Murari, A., Peluso, E., Cianfrani, F., Gaudio, P. and Lungaroni, M. (2019) On the use of entropy to improve model selection criteria. *Entropy*, 21(4), 1–12.

O'Kane, T.J., Risbey, J.S., Franzke, C., Horenko, I. and Monselesan, D.P. (2013) Changes in the metastability of the midlatitude Southern Hemisphere circulation and the utility of nonstationary cluster analysis and split-flow blocking indices as diagnostic tools. *Journal of the Atmospheric Sciences*, 70(3), 824–842.

Ortizbeviá, M.J., SánchezGómez, E. and Alvarez-García, F.J. (2011) North Atlantic atmospheric regimes and winter extremes in the Iberian peninsula. *Natural Hazards and Earth System Science*, 11(3), 971–980.

Portal, A., Ruggieri, P., Palmeiro, F.M., García-Serrano, J., Domeisen, D.I.V. and Gualdi, S. (2021) Seasonal prediction of the boreal winter stratosphere. *Climate Dynamics*. https://doi.org/10.1007/s00382-021-05787-9.

Quinn, C., Harries, D. and Kane, T.J.O. (2020) Dynamical analysis of a reduced model for the North Atlantic Oscillation. *Journal of the Atmospheric Sciences*, 78(5), 1647–1671.

Rossi, R., Murari, A., Gaudio, P. and Gelfusa, M. (2020) Upgrading model selection criteria with goodness of fit tests for practical applications. *Entropy*, 22(4), 1–13.

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

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(9), 3307–3314.

Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

Shannon, C. and Weaver, W. (1949) *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Shepherd, T.G. (2021) Bringing physical reasoning into statistical practice in climate-change science. *Climatic Change*, 169(2). https://doi.org/10.1007/s10584-021-03226-6.

Straus, D.M., Corti, S. and Molteni, F. (2007) Circulation regimes: chaotic variability versus SST-forced predictability. *Journal of Climate*, 20(10), 2251–2272.

Straus, D.M. and Molteni, F. (2004) Circulation regimes and SST forcing: results from large GCM ensembles. *Journal of Climate*, 17(8), 1641–1656.

Thompson, V., Dunstone, N.J., Scaife, A.A., Smith, D.M., Slingo, J.M., Brown, S. and Belcher, S.E. (2017) High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, 8(107), 1–6.

Toniazzo, T. and Scaife, A.A. (2006) The influence of ENSO on winter North Atlantic climate. *Geophysical Research Letters*, 33(24). https://doi.org/10.1029/2006GL027881.

Wagenmakers, E.-J. (2007) A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review*, 14(5), 779–804.

Weisheimer, A., Decremer, D., MacLeod, D., O'Reilly, C., Stockdale, T.N., Johnson, S. and Palmer, T.N. (2019) How confident are predictability estimates of the winter North Atlantic Oscillation?. *Quarterly Journal of the Royal Meteorological Society*, 145(S1), 140–159.

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D.A. and Palmer, T.N. (2017) Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quarterly Journal of the Royal Meteorological Society*, 143, 917–926.

How to cite this article: Falkena, S.K.J., de Wiljes, J., Weisheimer, A. & Shepherd, T.G. (2021) Detection of interannual ensemble forecast signals over the North Atlantic and Europe using atmospheric circulation regimes. *Quarterly Journal of the Royal Meteorological Society*, 1–20. Available from: https://doi.org/10.1002/qj.4213