Simulating weather regimes: impact of model resolution and stochastic parameterization

Andrew Dawson · T. N. Palmer

Abstract The simulation of quasi-persistent regime structures in an atmospheric model with horizontal resolution typical of the Intergovernmental Panel on Climate Change fifth assessment report simulations, is shown to be unrealistic. A higher resolution configuration of the same model, with horizontal resolution typical of that used in operational numerical weather prediction, is able to simulate these regime structures realistically. The spatial patterns of the simulated regimes are remarkably accurate at high resolution. A model configuration at intermediate resolution shows a marked improvement over the low-resolution configuration, particularly in terms of the temporal characteristics of the regimes, but does not produce a simulation as accurate as the very-high-resolution configuration. It is demonstrated that the simulation of regimes can be significantly improved, even at low resolution, by the introduction of a stochastic physics scheme. At low resolution the stochastic physics scheme drastically improves both the spatial and temporal aspects of the regimes simulation. These results highlight the importance of small-scale processes on large-scale climate variability, and indicate that although simulating variability at small scales is a necessity, it may not be necessary to represent the small-scales accurately, or even explicitly, in order to improve the simulation of large-scale climate. It is argued that these results could have important implications for improving both global climate simulations, and the ability of high-resolution limited-area models, forced by low-resolution global models, to reliably simulate regional climate change signals.

1 Introduction

The horizontal resolutions of global atmospheric models used for operational numerical weather prediction (NWP) and of the atmospheric climate models used, for example in the fifth Coupled Model Intercomparison Project (CMIP5), are substantially different. The typical atmospheric model resolution for CMIP5 models is in the range 100–200 km (1°–2°), whereas the typical horizontal resolution used in operational NWP is much higher, in the range 16–50 km. Increased horizontal resolution in climate models has led to incremental improvements in various aspects of mean climate and variability, although many key processes appear to be fairly well represented at resolutions typical of CMIP5 models (e.g., Shaffrey et al. 2009; Dawson et al. 2011). Whilst few would doubt the desirability of being able to integrate climate models at NWP resolution, there are numerous other areas of climate model development which compete for the given computing resources: the need for ensembles of integrations, to integrate over century and longer time-scales and the need to incorporate additional Earth System complexity. It is therefore hoped that, at least in terms of large-scale climate, the difference between NWP and climate resolution should not make too quantitative a difference in the accuracy of climate simulations. This hope is supported to some extent by studies which address the impact of resolution on simulations of the time mean state of the atmosphere and variance around this state (Jung et al. 2012; Dawson et al. 2013).

On the other hand, the geometry of the climate attractor is not determined solely by its first two moments. In

A. Dawson (✉) · T. N. Palmer
Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK
e-mail: dawson@atm.ox.ac.uk

particular, considerable evidence has emerged over the last decade or more, for quasi-persistent weather regimes over different regions of the world (Straus et al. 2007; Straus 2010; Woollings et al. 2010; Pohl and Fauchereau 2012). The existence of such regimes is interesting in its own right, with implications for understanding recurring persistent regional weather patterns, but also has wider implications in the climate system. There is evidence that in a dynamical system with regime structure, the time-mean response of the system to some imposed forcing, which here could be thought of as enhanced greenhouse gas concentration, is in part determined by the change in frequency of occurrence of the naturally occurring regimes (Palmer and Weisheimer 2011; Palmer 1999, 1993; Corti et al. 1999). As such, a model which failed to simulate observed regime structures well, could qualitatively fail to simulate the correct response to this imposed forcing. The importance of regime analysis as a model diagnostic is highlighted in our companion paper (Christensen et al. 2014), which demonstrates that using regimes as a diagnostic tool provides a more insightful perspective on the suitability of different parameterization schemes than using more common techniques.

There is also increasing recognition that representation of the smaller scales, although not necessarily explicitly resolved, is beneficial for climate simulation. There is mounting evidence that stochastic parameterizations, where typically some element of randomness is introduced into physical parameterization schemes to account for model uncertainty, prove beneficial for climate simulations (e.g., Lin and Neelin 2000, 2003; Arnold et al. 2013). It has also been suggested that using stochastic computation when computing the small-scale dynamical terms in an atmospheric model could be beneficial (Palmer 2012). It was demonstrated by Düben et al. (2013) that much of the performance of a high-resolution spectral model in terms of the large-scale flow can be retained even when the precision with which the higher order wavenumbers are computed is degraded. Reduced precision computation reduces energy consumption and therefore allows higher resolution runs to be more affordable. These two approaches to introducing stochasticity into models can be thought of as equivalent in some sense, in that one is more interested in the presence of variability at small scales and the influence that this has on larger scales, rather than in simulating the small-scale processes themselves with total accuracy.

Dawson et al. (2012) demonstrated that an atmospheric model at typical CMIP5 resolution was unable to simulate realistic regimes structures in the atmosphere, and that the same model integrated at a deterministic NWP resolution provided a realistic regimes simulation. This paper builds on that work, and introduces an intermediate resolution model to determine if very high resolution is truly necessary to facilitate regime simulations. The influence of a stochastic parameterization scheme on the simulation of regimes is also investigated. This study supports the growing recognition that there is no more complex problem in computational science than that of simulating climate, and next generation climate simulators need to provide careful consideration of small-scale processes, whether these be explicitly resolved or parameterized.

This paper is organized as follows: Sect. 2 describes the models and data sets used. Section 3 describes the regimes analysis method. Section 4 describes results from analysing regimes in reanalysis. Section 5 discusses the existence of circulation regimes in the European/Atlantic flow. Section 6 then describes results from analysing regimes in model simulations of different horizontal resolutions. Section 7 then goes on to describe results from model simulations using stochastic physics. Section 8 provides a summary and conclusions from this work.

2 Datasets and models

The model used for this study is the European Centre for Medium Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) cycle 36r1. Results are presented from continuous atmosphere-only integrations of the IFS for the 45-year period 1962/63–2006/07, forced with observed sea surface temperatures and sea ice fields. The IFS is integrated at different horizontal resolutions: a high-resolution T1279 configuration, a medium-resolution T511 configuration, and a low-resolution T159 configuration. These triangular truncations correspond to approximate grid spacings of 16 km, 40 km and 125 km respectively. A horizontal resolution of T159 is typical of the resolution of the atmospheric component in the CMIP5 climate models (and is admittedly not considered low resolution in the context of CMIP5), whereas T1279 is ECMWF's operational deterministic forecast resolution, and is more typical of short–medium range NWP forecast model resolution. The T511 resolution configuration is used to 'fill the gap' between these two extremes. The vertical discretization and most of physical parameterizations are the same for all three resolution configurations (Jung et al. 2012). These model configurations therefore represent a relatively clean comparison between different horizontal resolutions.

For each of the lower resolution T159 and T511 configurations we have an additional configuration which is the identical to the base configuration except with the stochastic physics parameterization scheme turned on. These configurations are designed to investigate the effect of representing model uncertainty in the indirect representation of the unresolved smaller scales in the lower resolution models. The stochastic physics scheme used is that described by Buizza et al. (1999). The choice of this scheme, rather

than the more advanced stochastically perturbed parametrization tendencies (SPPT) scheme (Palmer et al. 2009), was necessitated by the desire to use the same model cycle as the existing T1279 integration produced by the Athena project (Jung et al. 2012; Kinter et al. 2013). In the studies performed here, the T511 deterministic and stochastic physics integration and both the T159 deterministic and stochastic integrations were performed with the model cycle used in the Athena integrations, but using the ECMWF supercomputer. For this reason the T159 deterministic integration is statistically equivalent to, but not bit reproducible with, the T159 integration discussed in Dawson et al. (2012).

The models are compared to reanalysis data for the same 45-year time period. The primary reanalysis data set used is composed of the ECMWF 40-year reanalysis (ERA-40; 1962–1988) and the ECMWF Interim reanalysis (ERA-Interim; 1989–2007). The same SST data set was used to force the models as was used in reanalysis. This data set is referred to as ERA throughout the study. For reference comparison we also use data from the National Centers for Environmental Prediction (NCEP) twentieth century Reanalysis data set (Compo et al. 2011). This data set is referred to as NCEP throughout the study.

3 The regimes diagnostic

The regime analysis method used in this study is identical to that presented in Dawson et al. (2012). The analysis is based on daily fields of wintertime (December–March; DJFM) geopotential height on the 500 hPa pressure surface. A seasonal cycle is computed by averaging the seasonal time series at each grid point over all years, and is then smoothed by the application of a 5-day running mean before being subtracted from the daily time series to produce a daily anomaly time series at each grid point.

Our aim is to cluster maps of daily geopotential height anomalies into groups of similar anomaly patterns. In order to do this effectively it is necessary to reduce the dimensionality of the problem. Without such reduction it would be extremely difficult, if not impossible, to determine any significant clustering (see Sect. 5). To reduce the dimensionality empirical orthogonal function (EOF) analysis is applied to the region of interest, which in this study is the European/Atlantic domain defined by the sector $30° - 90°N$, $80°W - 40°E$, and only a limited number of EOFs are retained. Prior to the EOF analysis all data sets are interpolated a $2.5° \times 2.5°$ grid to ensure that variability is considered on the same range of horizontal scales for all data sets. The interpolation is done by truncating the spectral coefficients of geopotential height to 63 wavenumbers then evaluating the gridded data from these T63 coefficients. The EOFs are computed using standard area weighting. The principal components (PCs), the time series of the retained EOF patterns, form the coordinates of a reduced phase space. Each of the PCs has variance equal to the eigenvalue of the mode it corresponds to. This weighting is used in preference to standardized (unit variance) PCs because it prevents unfair weight being given to modes that represent only a small amount of variance. Each point in this reduced phase space represents an atmospheric anomaly state expressed as a linear combination of the retained EOF patterns. In this study we choose to retain the first 4 EOFs, as this number is sufficient to account for at least 50 % of the total variance in reanalysis and models. We note that the same analysis has been conducted retaining the first 11 EOFS, which account for over 80 % of the variance in all cases, and the results are not changed. This reduction of dimensionality can be thought of as imposing an implicit filter on the anomaly time series, since by retaining only a limited number of EOFs the set of possible states are restricted to fairly large-scale patterns, which in turn are associated with variability on longer time scales than typical transient disturbances.

Clusters in the reduced phase space are identified by applying the k-means clustering algorithm (e.g., Michelangeli et al. 1995; Straus et al. 2007). This clustering procedure aims to locate preferred regions of the reduced phase space which can be interpreted in the framework of regimes. The method seeks to partition the phase space into k clusters. This partition is constructed so that the ratio of variance between cluster centroids to the average intra-cluster variance is maximized. This condition corresponds to desiring cluster centroids to be far apart and for the points within each cluster to be close together. Since the standard algorithm for solving the k-means problem is a heuristic algorithm, and is not guaranteed to solve for the globally optimal partition, it is necessary to apply the clustering multiple times with different initial centroids, and use the partition that maximizes the previously mentioned variance ratio.
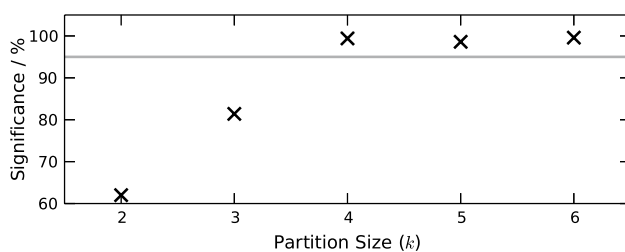
3.1 Significance of cluster partitions

The null hypothesis when applying cluster analysis is that there are no regimes. This implies that the probability density function (pdf) of the underlying phase space follows a multi-normal distribution. In order to assess if this null hypothesis can be rejected for a given cluster partition, Monte Carlo simulations using a large number of synthetic data sets are applied, as in Dawson et al. (2012), Straus et al. (2007). The cluster analysis is applied to 500 synthetic data sets, each one modelled on the PCs of the data set being tested. The synthetic PCs are independent Markov processes with the same lag-1 autocorrelation, mean and standard deviation as the corresponding PC in the real data set. The

significance of the clustering is reported as the percentage of times that the optimal variance ratio found by clustering the real data set exceeds the optimal variance ratio found by clustering the synthetic data sets. Large values of this significance indicate that the optimal variance ratio for the given partition is unlikely to have been found by chance in a data set whose pdf follows a multi-normal distribution.

3.2 Choosing the number of clusters

The choice of k is arbitrary in k-means cluster analysis, and is an *a priori* assumption. In this study we have chosen to use $k = 4$, meaning we seek to partition the phase space into 4 clusters. This choice is based upon computing different sized cluster partitions for ERA and assessing their significance. Figure 1 shows the significance of cluster partitions for $k = 2, 3, \ldots, 6$. It is clear that partitioning into 2 or 3 clusters does not yield significant results, but for $k = 4, 5, 6$ the results are significant. The $k = 4$ partition is chosen as it is the smallest partition size that is significant. There is no evidence of extra significance with 5 or more clusters, hence there is no evidence that more than 4 clusters should be considered. This is a reasonable choice since in general the addition of extra clusters makes it easier for the algorithm to isolate pockets in the data, whilst not necessarily confining the clusters to potentially physically meaningful areas of phase space.

Christiansen (2007) has raised questions about the ability of k-means clustering to determine the correct number of clusters in a data set. The object of this study is to use the regimes framework as a model diagnostic tool. The aim is to use cluster analysis to diagnose non-gaussian structures in a way that can be applied to both observations and models. In this context the specific concerns over the ability of the k-means algorithm to detect the "correct" number of clusters become less relevant, as one would expect a good model to be able to reproduce significant clusters from observations, even if the number of clusters is not "correct". Other authors have raised objections over actually defining unambiguous circulation regimes in the European/

Atlantic flow (Fereday et al. 2008). This is a more complex discussion, which is covered thoroughly in Sect. 5.

3.3 Comparing model clusters to reanalysis

In order to compare model clusters to reanalysis it is necessary to be able to measure how different the spatial pattern of any particular modelled cluster is from its reanalysis counterpart. To do this an error metric is defined as the length of the vector between projections of the model cluster and the corresponding reanalysis cluster in a common phase space. The common phase space is produced by projecting the spatial pattern of both the modelled and reanalysis clusters onto the EOFs obtained from reanalysis. Denoting a cluster centroid map from reanalysis as the row vector $\mathbf{c}_0$, and the corresponding cluster centroid map from a model as $\mathbf{c}_m$, these projections are:

$$\mathbf{p}_0 = \mathbf{c}_0 \mathbf{E}, \qquad \mathbf{p}_m = \mathbf{c}_m \mathbf{E},$$

where $\mathbf{E}$ is a matrix containing the n retained EOFs in its columns. Each projection is simply a point in n-dimensional space, and the error metric is then the length of the vector between these two points:

$$\text{error} = \|\mathbf{p}_m - \mathbf{p}_0\|.$$

This error metric is analogous to a root-mean-squared error (RMSE) of the spatial pattern with respect to reanalysis, and will be referred to as such. The mean RMSE over all clusters in a partition will also be considered, which provides a measure of how well the partition as a whole compares to reanalysis.

4 Observed European/Atlantic circulation regimes

Figure 2 shows composites of 500 hPa geopotential height anomaly associated with each of the four clusters for the European/Atlantic sector. The clusters are presented in order of their climatological frequency of occurrence. The significance of this cluster partition is 99.8 % (as indicated in Fig. 1) as determined by the test described in Sect. 3. Therefore these clusters represent regimes. The high significance indicates that the variability of large-scale circulation in the European/Atlantic sector has non-Gaussian structure.

Clusters 1 and 4 (Fig. 2a, d) are referred to as NAO+ and NAO− regimes respectively, as they are consistent with the spatial patterns of the positive and negative phases of the North Atlantic Oscillation (e.g., Woollings et al. 2010; Hurrell and Deser 2010). The second most frequent cluster (Fig. 2b) has a positive geopotential height anomaly centered over Scandinavia with negative anomalies to the east and west. This pattern is referred to as the blocking

Fig. 1 Significance of clustering for the ERA data set as a function of cluster partition size k. The 95 % significance level is indicated by the *grey horizontal line*

Fig. 2 Cluster centroid maps of 500 hPa geopotential height for ERA. The maps are composites of daily geopotential height anomalies for each day the given cluster is active. The climatological frequency of occurrence for each cluster is given in next to the name. **a** NAO+ cluster, **b** BL cluster, **c** AR cluster and **d** NAO− cluster
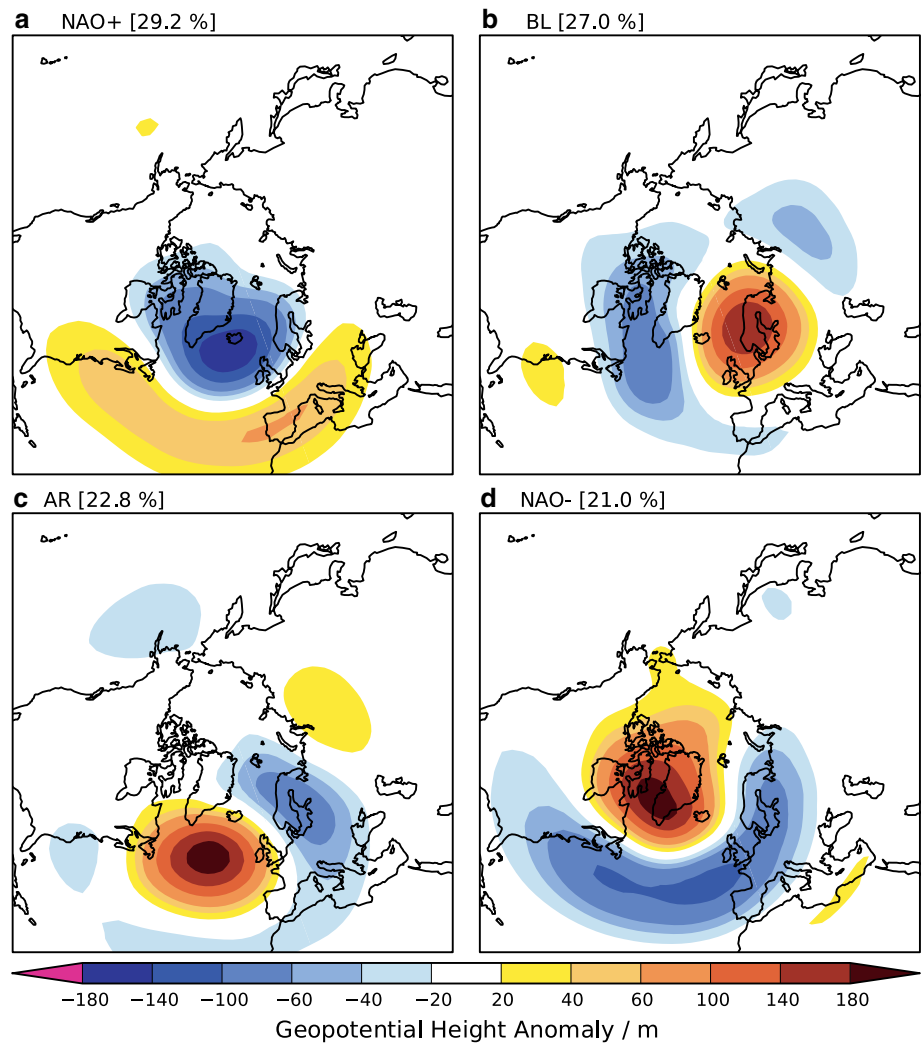
(BL) regime as it resembles closely the anomalous flow during European blocking episodes. The third most frequent cluster (Fig. 2c) consists of a positive geopotential height anomaly over the North Atlantic and a negative anomaly over Scandinavia and eastern Europe. This cluster is referred to as the Atlantic Ridge (AR) regime, although we note the pattern has much in common with the East Atlantic pattern (e.g., Wallace and Gutzler 1981). Each of these regimes has associated patterns of surface temperature and precipitation anomalies which are consistent with the physical interpretation of the regimes.

To understand what kind of differences we might expect between reanalysis and simulated clusters, we apply the same methodology to the NCEP data set. This gives us an idea of the differences one can expect in spatial patterns and frequencies of occurrence of each cluster, and the significance of the clustering itself, due to using a different reanalysis system. The $k = 4$ cluster partition for the NCEP data set has a significance value of 99.4 %, which is highly significant. The clusters for this data set are shown as the

coloured contours in Fig. 3, with the overlaid grey contours showing the equivalent pattern from ERA (Fig. 2). Simply by comparing the two sets of contours it is clear that the spatial patterns of each NCEP cluster are very similar to the corresponding ERA cluster. We also note that the climatological frequency of occurrence of each NCEP cluster is within 1 % of the corresponding ERA cluster. Given these details it would be hard to argue that the clusters found here are an artefact of the particular reanalysis system used.

The fact that the same regimes can be identified in two independent reanalyses demonstrates the robustness of the European/Atlantic circulation regimes. For example, applying the same analysis in the Pacific region yields less consistent results. The same clustering procedure was applied to the domain $140°E − 80°W$, $30° − 87.5°N$, which covers the whole Pacific Ocean and has the same latitudinal extent as the European/Atlantic sector. An additional filtering step, a 10-day Lanczos filter (Duchon 1979) applied to the anomaly time series as recommended by Straus et al. (2007), was required in order to find significance in the

Fig. 3 As Fig. 2 but for NCEP twentieth century Reanalysis. The *grey contours* are the corresponding contours from the ERA data set for comparison. The RMS error relative to the corresponding ERA cluster is given at the *top right* of each *panel*. **a** NAO+ cluster, **b** BL cluster, **c** AR cluster and **d** NAO− cluster
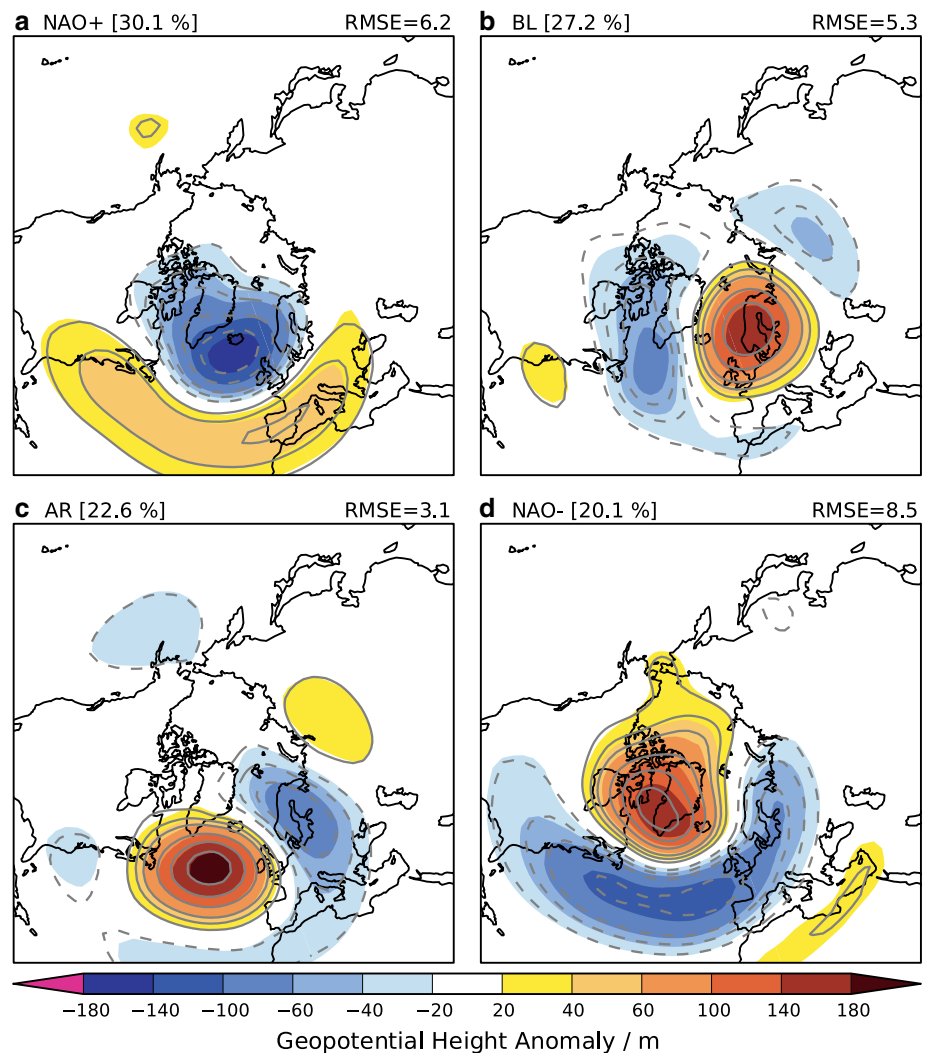

clustering in the Pacific. The results of this analysis are shown in the top row of Fig. 4, and in the bottom row of Fig. 4 for NCEP. The clusters from ERA have significance of 93.6 %, and those from NCEP are 95.0 % significant. The two reanalyses find clusters which are broadly physically similar in terms of the spatial patterns, but there are significant deviations in the locations and specific shapes of the patterns. This demonstrates that the similarity of the European/Atlantic regimes in the two reanalyses is not simply inevitable, but is a genuine indicator of robustness.

5 Existence of regimes in the European/Atlantic circulation

The four regimes in the European/Atlantic domain are qualitatively similar to those found in other studies (Michelangeli et al. 1995; Cassou 2008; Barrier et al. 2013), and appear to be robust between reanalysis products. However, some authors have questioned whether it is even possible to define circulation regimes in the European/Atlantic region, particularly using cluster analysis. For example, Fereday et al. (2008) concluded that it is not possible to determine an unambiguous clustering for European/Atlantic circulation, and therefore not possible to define regimes. This apparent contradiction to the current work is now briefly discussed, and a demonstration of why Fereday et al. (2008) cannot detect circulation regimes is provided.

Fereday et al. (2008) attempt to find clusters in 2-month samples of mean sea level pressure (MSLP). They use k-means cluster analysis, but do not use any kind of spatial or temporal filtering, or reduction of dimensionality prior to clustering. The key differences in between their analysis and ours are the length of the temporal samples, the dimensionality of the space clustering is done in and the variable used for the analysis. We have re-applied our methodology, using combinations of different sample (season) lengths, phase space dimension and analysis variable. For each methodology we compute the significance of the $k = 4$ cluster partition. The results of this are shown in Table 1.
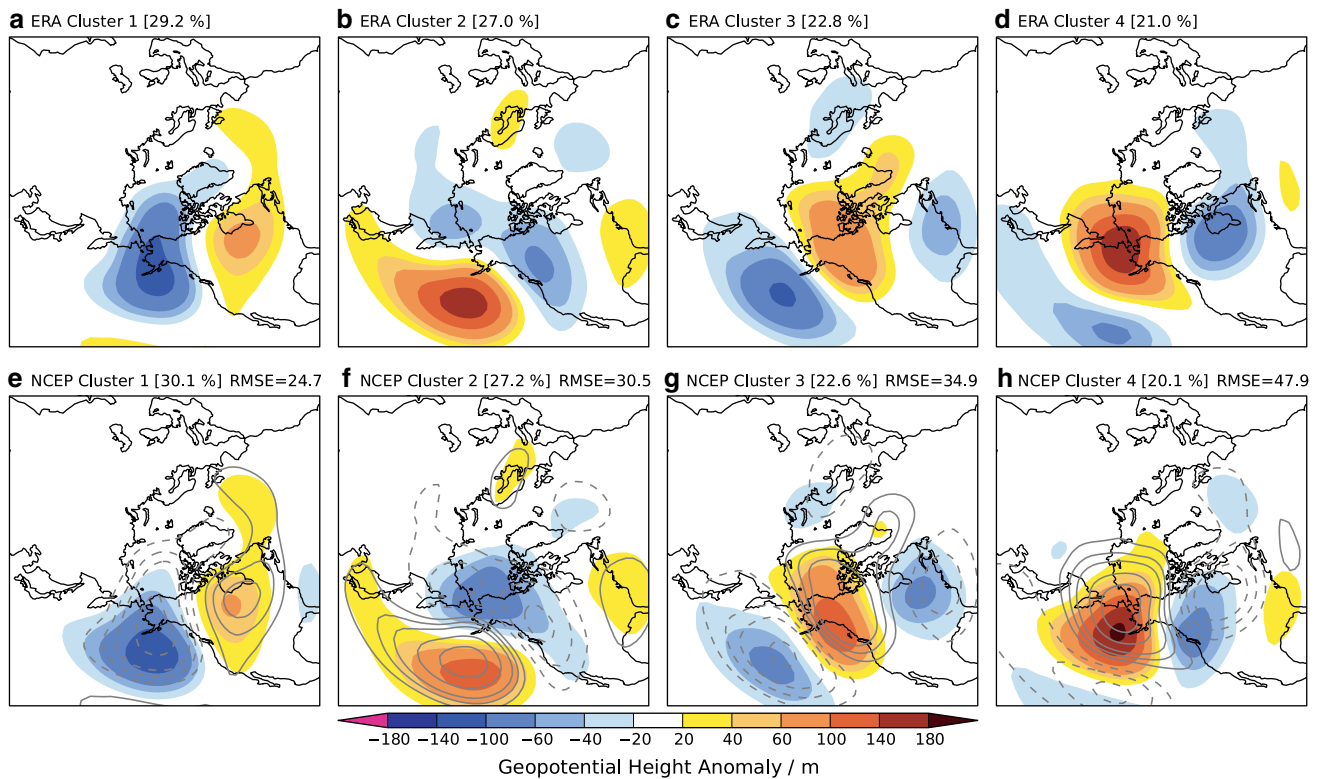
Fig. 4 Cluster centroid maps of 500 hPa geopotential height for the Pacific region for ERA and NCEP twentieth Century Reanalysis. The *upper panels* (**a–d**) are for ERA and the *lower panels* (**e–h**) are for NCEP. The *grey contours* in the *lower panels* are the corresponding contours from ERA for comparison. The RMS error relative to the corresponding ERA cluster is given at the *top right* of each of the *lower panels*

Firstly we will consider analysis of geopotential height on the 500 hPa pressure level (Z_{500}; first row in Table 1). For a low-dimensional (4-d) clustering space it is possible to find significance using either a 4-month sample length or a 2-month sample length, although the significance value is somewhat lower for the 2-month sample length case. For a higher-dimensional (50-d) space it is still possible to find significant clusters when using a 4-month sample length, but a 2-month sample length is not sufficient to adequately sample such a large space. This suggests that the combination of a large (250-d) clustering space and a short sample length used by Fereday et al. (2008) is severely limiting their chances of finding significant clusters.

These tests were also done using mean sea level pressure (MSLP; second row in Table 1). These show that even for a low-dimensional (4-d) space, a 2-month sample length is not adequate when using MSLP as the analysis variable. For the larger 50-d space neither the 4 or 2-month sample lengths are sufficient. This is likely to be because MSLP is a noisier field, containing much of the smaller scale variability from within the planetary boundary layer and is therefore less directly linked to the large-scale balanced dynamics which determines regime structure. This is in contrast to geopotential height on the 500 hPa pressure

Table 1 Significance of the 4-cluster partition for different combinations of phase space size, sample length and analysis variable

Variable	4-d phase space		50-d phase space	
	4-month sample (%)	2-month sample (%)	4-month sample (%)	2-month sample (%)
Z500	99.8	95.2	95.0	86.2
MSLP	98.8	90.4	92.2	80.2

surface, which is a comparatively smooth field typically with variations on large (synoptic and greater) scales. In summary it is likely that Fereday et al. (2008) fail to find significant regimes in the European/Atlantic circulation because they used a clustering space that is too large and a sample size that is too small, as well as their choice of variable aggravating the situation.

## 6 Modelled European/Atlantic circulation regimes

Having addressed the representation of regimes in observations (reanalysis) we now turn attention to the simulation of circulation regimes in models. Figure 5 shows the

$k = 4$ cluster partition for the high-resolution T1279 model configuration, again the clusters are presented in order of climatological frequency of occurrence. These clusters are as presented in Dawson et al. (2012). They are highly significant (98.6 %) and can be interpreted as regimes in the European/Atlantic circulation. One can easily make a one-to-one correspondence between the modelled and reanalysis clusters, and note that the patterns of each modelled cluster are very similar to the corresponding pattern in reanalysis.

The equivalent clusters for the low-resolution T159 model configuration are shown in Fig. 6. For reasons discussed in Sect. 2 these clusters are not identical to those shown in Dawson et al. (2012) and show a reasonably high (97.2 %) level of significance. However, as will now be described, significance of the clusters does not necessarily imply a realistic representation of regime behaviour. The order of the clusters in terms of their climatological

frequency of occurrence is different to reanalysis or the high-resolution configuration. Although the NAO+ regime is most frequent, the T159 configuration shows NAO− to be the second most frequent cluster in contrast to reanalysis where this is the least frequent regime. This is consistent with the T159 configuration results shown in Dawson et al. (2012). The clusters identified by Dawson et al. (2012) were not consistent with the BL or AR regimes, but in this integration it is possible to identify these regimes. However, the BL regime is the least frequently occurring regime in this case, which is in stark contrast to observations. The spatial patterns of the individual clusters are generally not well represented in the T159 configuration, with the possible exception of the NAO+ regime. The NAO− regime is too zonally constrained, and the BL and AR regimes are both shifted away from the locations they occur in reanalysis. The higher significance and one-to-one matching of regime patterns with reanalysis suggest that this T159

Fig. 5 As Fig. 2 but for the T1279 model configuration. The *grey contours* are the corresponding contours from the ERA data set for comparison. The RMS error relative to the corresponding ERA cluster is given at the *top right* of each panel. **a** NAO+ cluster, **b** BL cluster, **c** AR cluster and **d** NAO− cluster
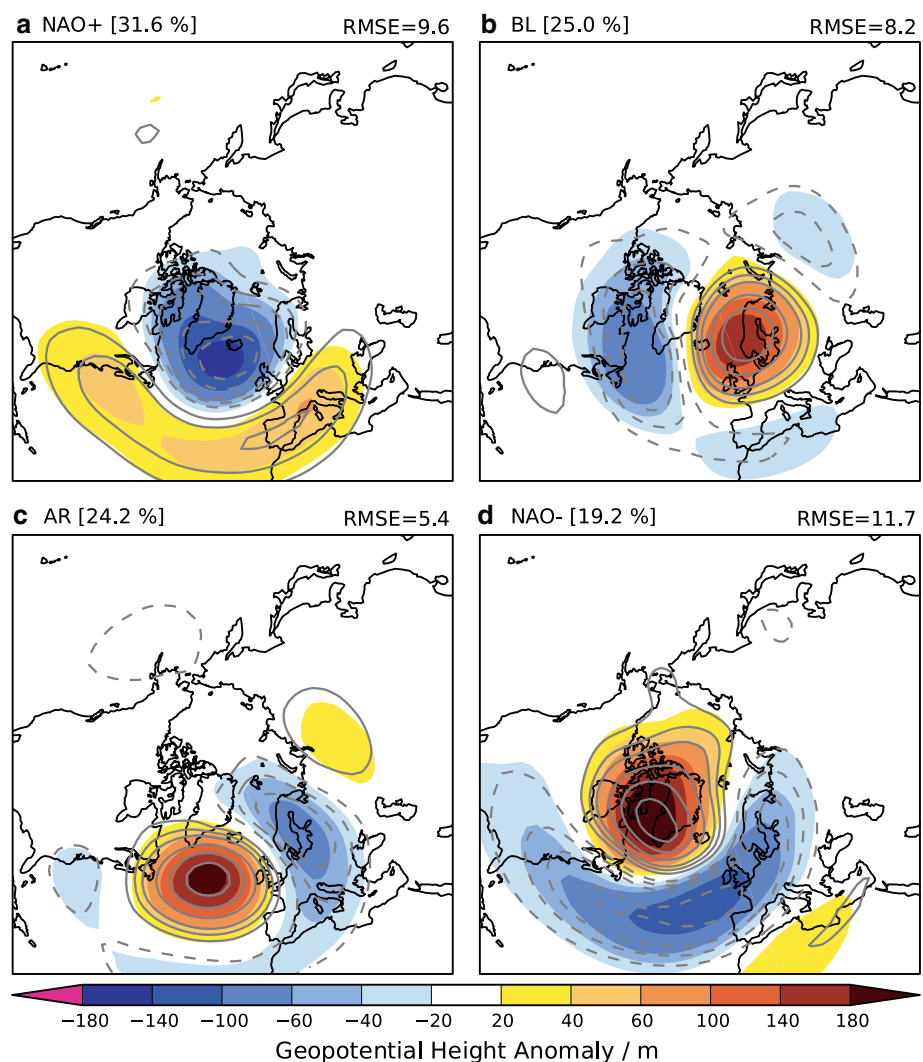
Fig. 6 As Fig. 2 but for the T159 model configuration. The *grey contours* are the corresponding contours from the ERA data set for comparison. The RMS error relative to the corresponding ERA cluster is given at the *top right* of each panel. **a** NAO+ cluster, **b** NAO− cluster, **c** AR cluster and **d** BL cluster
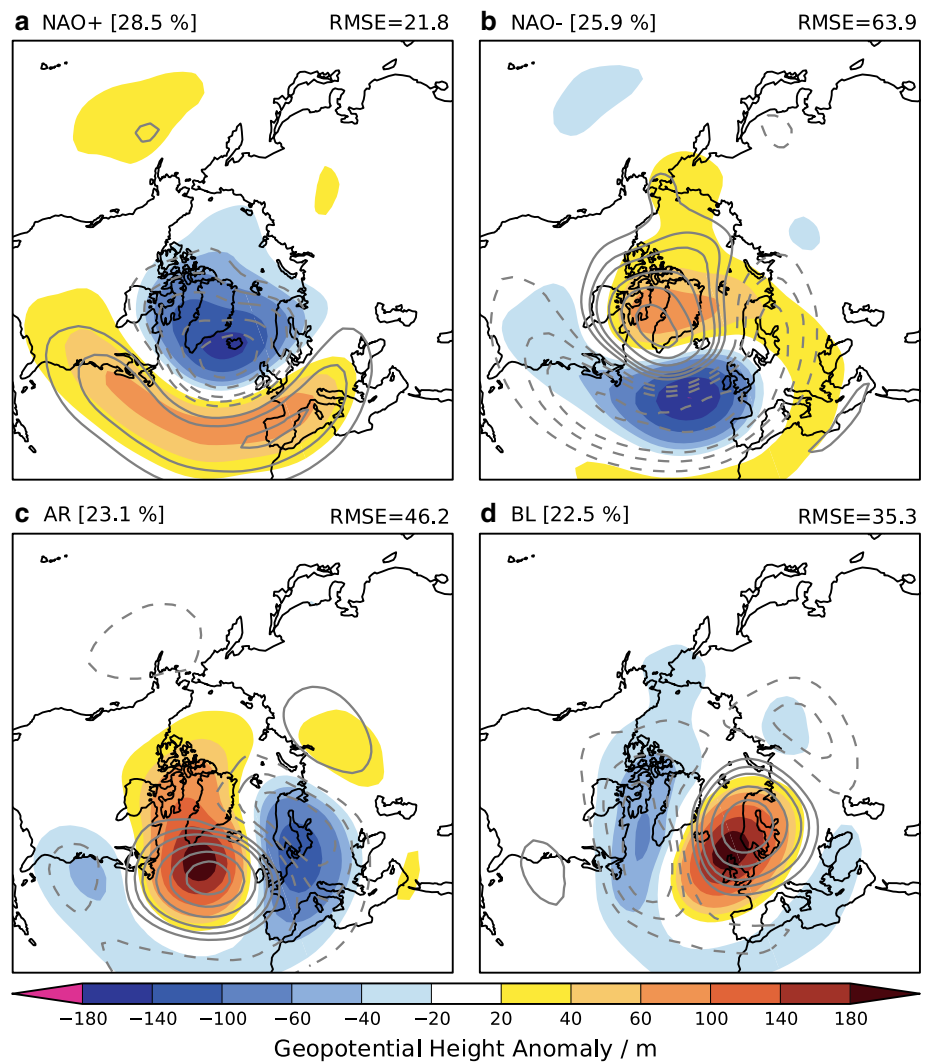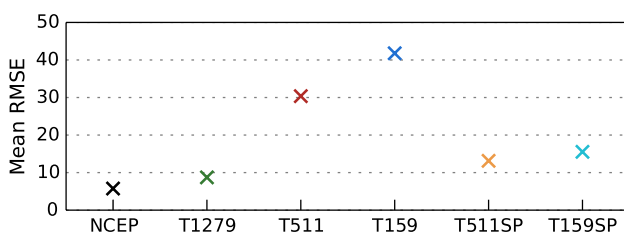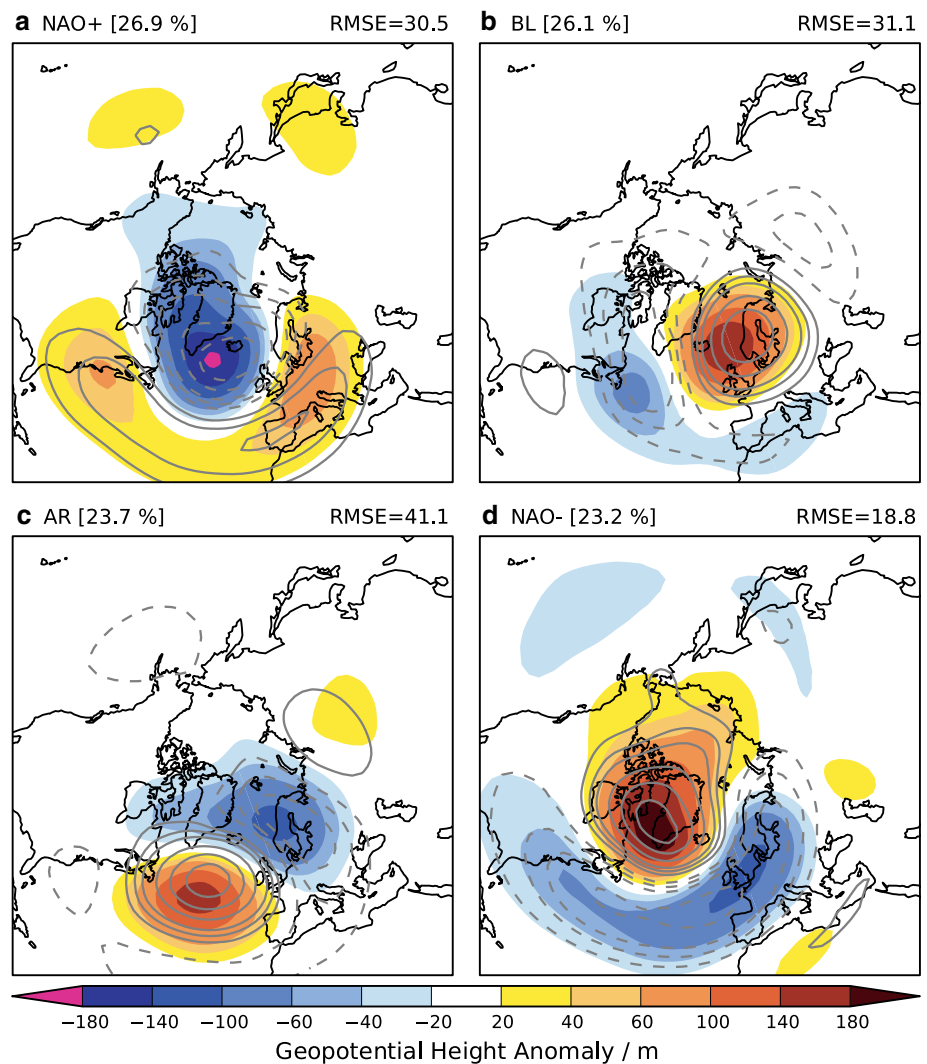

configuration is performing somewhat better in this integration than in the integration discussed in Dawson et al. (2012). However, the inaccurate representation of cluster frequencies and large differences in the spatial patterns compared to reanalysis are erroneous features that are common across both studies. These errors mean that despite the apparent slight improvement we see in the T159 configuration over the previous study, our conclusion must remain that the T159 model has a generally poor representation of regimes.

Having determined that at T1279 we get a realistic simulation of circulation regimes, and at T159 resolution regimes are not simulated realistically, the obvious question is: would an intermediate-resolution configuration be sufficient to accurately simulate these regimes, or is T1279 resolution really necessary to capture the important processes? T1279 is far too high a resolution to be practical for use in climate simulations, so it would be ideal if a lower resolution could be used, with much of the same benefits

in terms of the regime simulations. Figure 7 shows clusters from the intermediate T511 resolution model configuration. These clusters are highly significant (97.2 %), allowing their interpretation as regimes, and can be matched one-to-one with the reanalysis clusters. The quality of the spatial patterns is less good than with the T1279 configuration, but they are somewhat better than those in the T159 configuration. The major improvement the T511 configuration makes over the T159 configuration is in the regime frequencies. The T511 configuration produces the clusters in the same order as ERA, with the notable divide between the two more frequent NAO+ and BL regimes and the two less frequent AR and NAO− regimes. The intermediate-resolution configuration therefore represents a significant improvement upon the low-resolution T159 model, but has not been able to match the performance of the very-high-resolution T1279 configuration.

An overall comparison of the errors relative to ERA in the spatial patterns of the clusters for each model

a NAO+ [26.9 %] RMSE=30.5

b BL [26.1 %] RMSE=31.1

c AR [23.7 %] RMSE=41.1

d NAO- [23.2 %] RMSE=18.8

−180 −140 −100 −60 −40 −20 20 40 60 100 140 180

Geopotential Height Anomaly / m

Fig. 8 Root-mean-square errors in the spatial patterns of the cluster centroids relative to ERA averaged over all four clusters

configuration and NCEP is given in Fig. 8. The figure shows the mean RMS error over all four clusters. From this it is clear that the error in T1279 is actually comparable to the error between two reanalysis systems, and that the error in T159 is considerably larger. It also shows that the T511 configuration does make some improvement over T159 with respect to reproducing the actual spatial patterns of the regimes.

Another important aspect of a regime simulation is the temporal persistence of the regimes. Figure 9 shows the distribution of persistence (in days) of each regime. Persistence is defined as the number times the given cluster persists for the given number of days. This is done by breaking the cluster time series into non-overlapping sections containing data points in the same cluster cluster, then counting the number of sections of each length. A persistence of 1 day refers to a section of length 1 day. The shading around the ERA persistence distribution is an indicator of uncertainty, obtained through a bootstrapping method. A large number of randomized realizations of the ERA cluster time series, the output of the clustering which tells us which cluster the atmospheric state is assigned to for each day in the analysis, were generated using sampling with replacement. Each realization has the same length as the original time series (45 years) and keeps whole seasons together, since the individual days within a season are not independent, but seasons as a whole are independent of

Fig. 9 Distributions of the persistence of each cluster for ERA and the T159, T511 and T1279 model configurations. The last bin includes all events with persistence of more than 20 days. The *shaded region* around the ERA persistence distribution is the ±1 standard deviation interval obtained from a bootstrap re-sampling method. **a** NAO+ cluster, **b** BL cluster, **c** AR cluster and **d** NAO− cluster
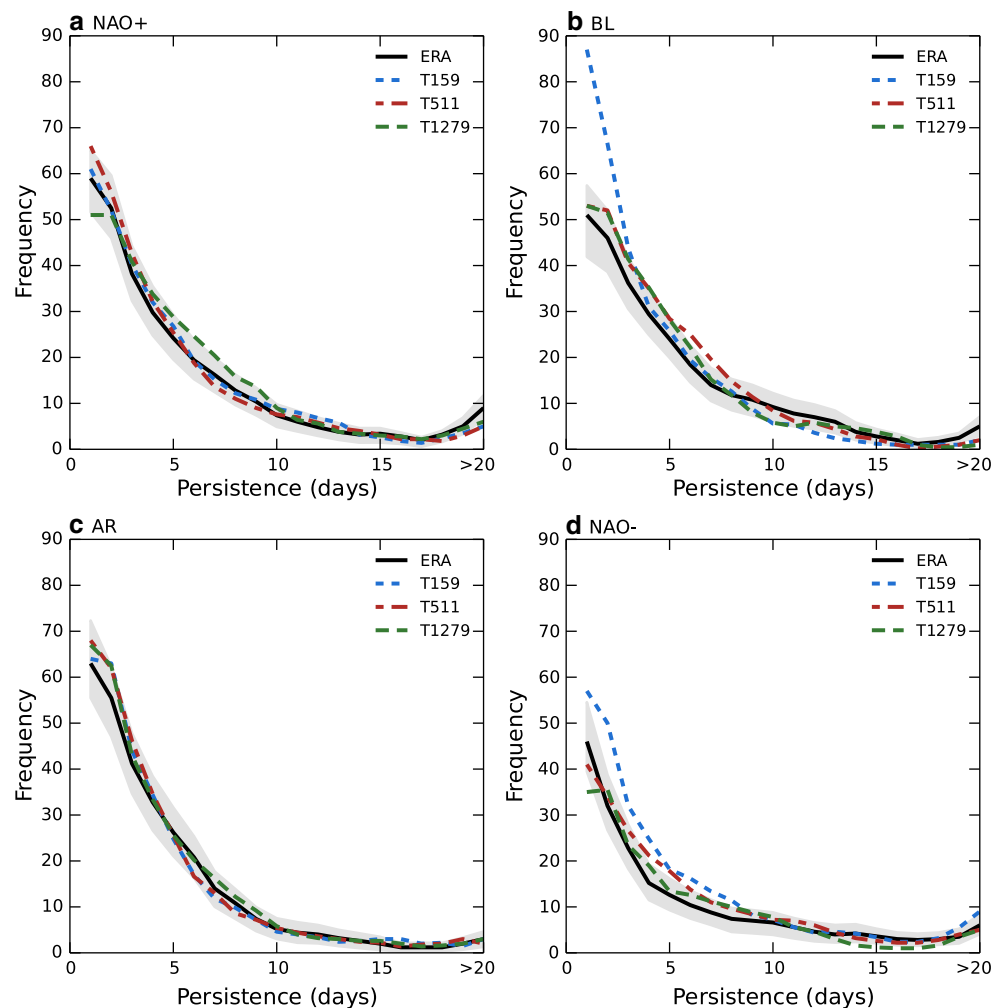

each other. For each realization the distribution of persistence was calculated for each cluster. The mean and standard deviation of these histograms was then computed. The shading in Fig. 9 represents the ±1 standard deviation interval around the mean for ERA. This provides a simple indicator of whether or not the persistence of the model cluster should be considered as the same as the persistence of the cluster in ERA. However, it does not necessarily imply that model persistence outside this shading is significantly different from the persistence in ERA. All three configurations have realistic persistence for the NAO+ and AR regimes, although the T1279 configuration perhaps has too much preference for 5–10 day NAO+ events. The persistence of the BL regime is simulated well at high resolution, but there is a clear over representation of short 1–3 day events in the T159 configuration, and the T511 configuration is somewhat overly persistent on the 5–7 day time-scale. The T1279 and T511 configurations have a realistic representation of NAO− persistence also. The T159 configuration has an erroneous preference for shorter lived (1–8 day) visits to the NAO− cluster.

7 Improvements through stochastic parameterization

Our current assessment is that a low-resolution T159 model is not sufficient to simulate regime behaviour, and even though it is possible to do better at a higher T511 resolution, it is still necessary to have extremely high T1279 resolution to simulate regimes realistically. This is a problematic conclusion for climate modelling. T1279 resolution is out of the question for use in coupled climate models due to its immense computational expense. Although T511 resolution may perhaps be a viable option in the near future, simply waiting for the computing power to match our requirements is not. It would therefore be ideal if it were possible to get some of the benefits demonstrated by these high-resolution models through better representation of the unresolved small scales in lower resolution models, without actually having to explicitly represent the smaller scales.

We now discuss the results of applying the regime diagnostic to configurations of the same model used previously, but with the stochastic physics parameterization scheme turned on. The cluster centroids calculated from a T511

Fig. 10 As Fig. 2 but for the T511 with stochastic physics model configuration. The *grey contours* are the corresponding contours from the ERA data set for comparison. The RMS error relative to the corresponding ERA cluster is given at the *top right* of each panel. **a** NAO+ cluster, **b** BL cluster, **c** AR cluster and **d** NAO− cluster
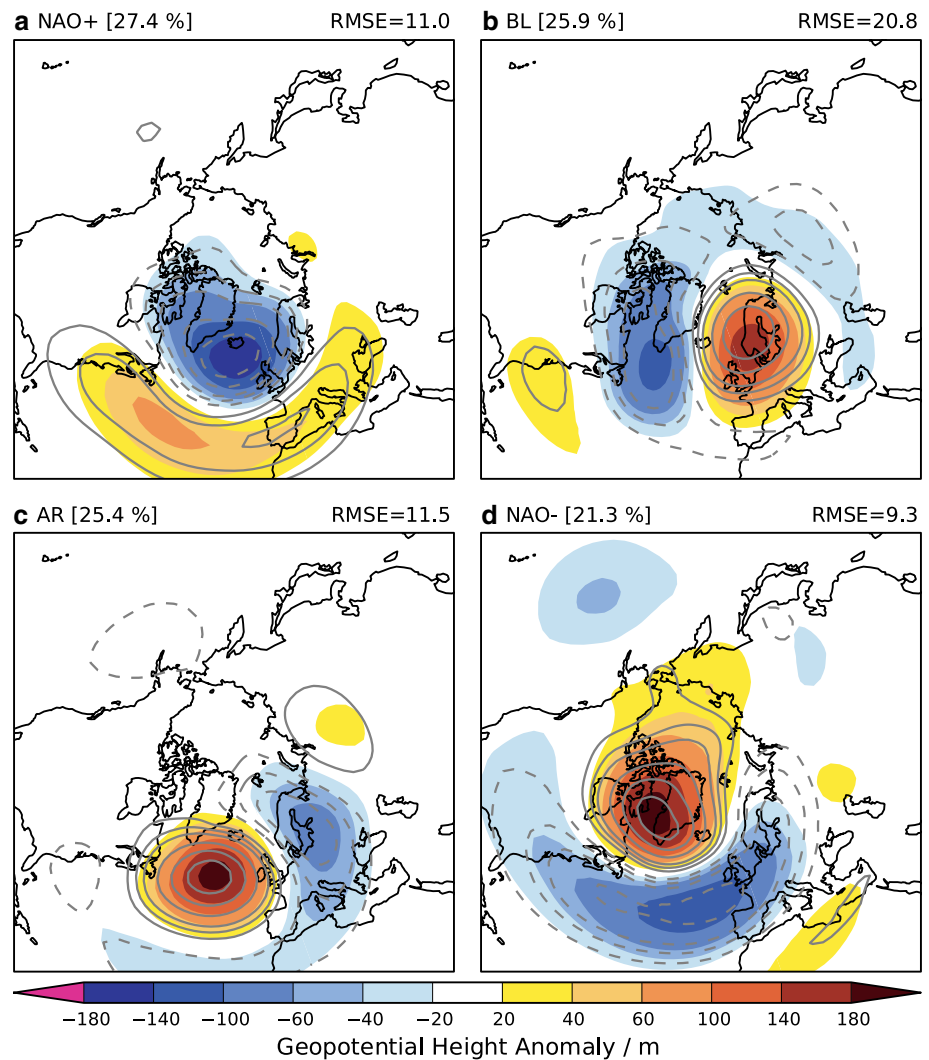

model configuration including stochastic physics (T511SP) are shown in Fig. 10. These clusters have a significance of 97.8 % and are therefore interpreted as regimes. The spatial patterns of each regime are improved upon significantly from the T511 configuration without stochastic physics, with lower errors across all four clusters. The AR regime is much improved with the ridge itself being shifted northeast into almost the same location as in ERA, and the corresponding low anomaly is shifted into a more realistic position over the Baltic Sea rather than being predominantly over the North Sea and the north coast of Norway. The centers of the BL regime are also shifted making their locations closer to ERA's BL regime.

The cluster centroids calculated from the T159 model configuration including stochastic physics (T159SP) are shown in Fig. 11. Again, these clusters have high significance (98.8 %) and are therefore interpreted as regimes. The addition of a stochastic physics scheme represents a considerable improvement over the standard T159

configuration. Most obviously the order of the regimes in terms of climatological frequency of occurrence is now the same as observations. Secondly, the representation of the spatial patterns of the regimes is greatly improved. The patterns are visually more realistic compared with the overlaid contours of the ERA regimes, and the RMS errors are greatly reduced for all four clusters.

It appears that the introduction of a stochastic physics scheme is enormously beneficial for the simulation of the spatial patterns of European/Atlantic circulation regimes. This is evident from Fig. 8 where both T511SP and T159SP have mean errors more comparable to the T1279 configuration than the T511 or T159 configurations without stochastic physics. Christensen et al. (2014) showed that in a highly simplified model of the atmosphere the introduction of a stochastic parameterization scheme has significant benefits for the simulation of the system's naturally occurring regimes, compared with a deterministic parameterization or a perturbed parameter ensemble. Christensen et al.

Fig. 11 As Fig. 2 but for the T159 with stochastic physics model configuration. The *grey contours* are the corresponding contours from the ERA data set for comparison. The RMS error relative to the corresponding ERA cluster is given at the *top right* of each panel. **a** NAO+ cluster, **b** BL cluster, **c** AR cluster and **d** NAO− cluster
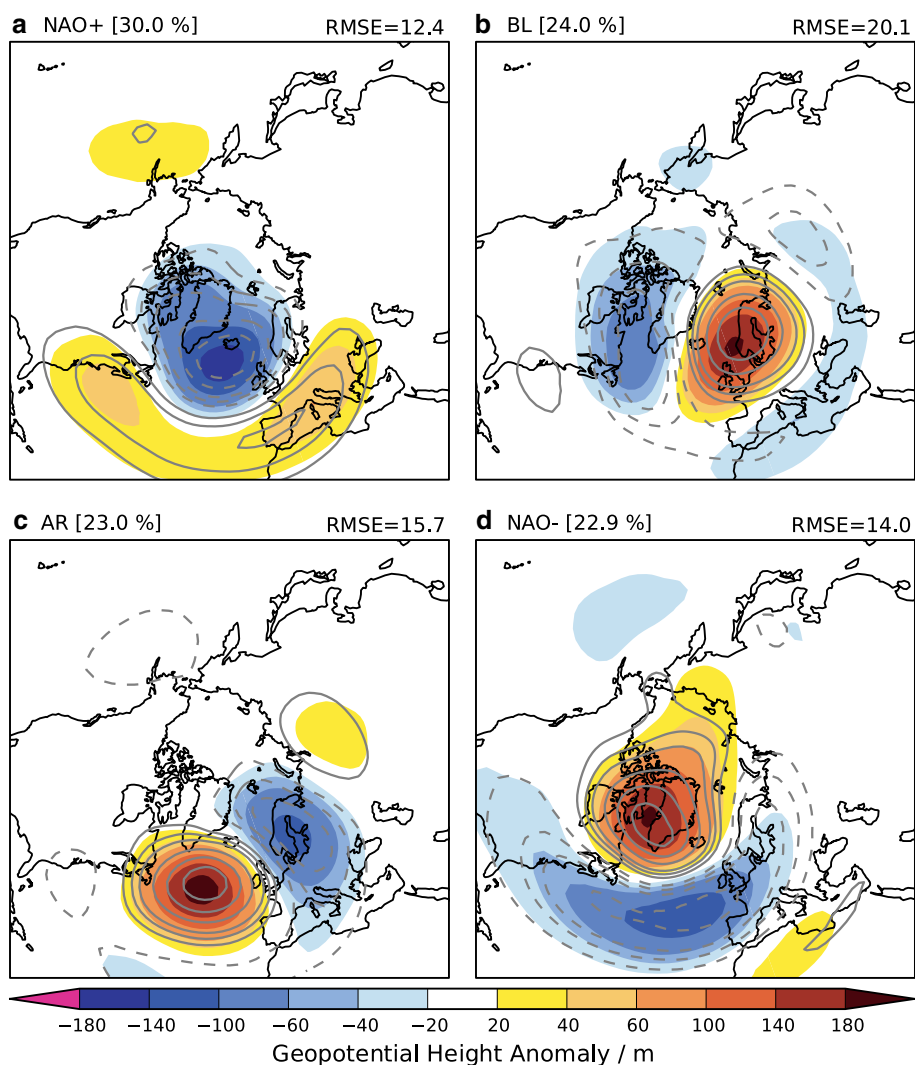
(2014) determined that the simplified model with a stochastic parameterization was more able to able to visit multiple distinct regions of phase space and hence produce realistic regimes, whereas the model had a poorer representation of regimes when deterministic parameterizations were used. The results presented here suggest that this finding could generalise to a full complexity atmospheric model, and is likely not an artefact of using the simplified model.

Whilst it is obvious that the climatological frequencies of the individual clusters in the T159SP configuration are improved over the standard T159 configuration, it is more difficult to assess any improvement in the T511SP frequencies. However, we can determine easily that they are not made any worse. Figure 12 shows the mean absolute error in climatological frequency relative to ERA over all four clusters. This gives an idea of how well the models represent the frequency of occurrence across all four clusters. The T1279 and T511 models have an appreciably larger error than the NCEP, but are of the same order as
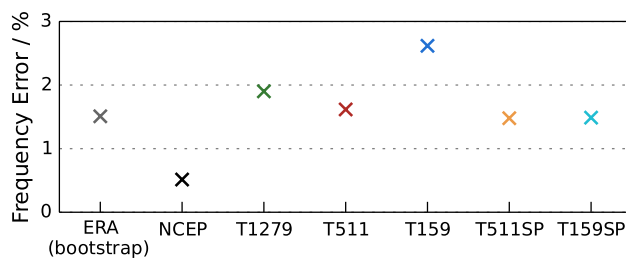
Fig. 12 Absolute error in climatological frequency of occurrence (%) relative to ERA averaged over all four clusters. The error for ERA is the bootstrapped error estimate which is detailed in Sect. 7

one another. The errors in T159 are much higher, as one would expect given the different ordering of the regimes in the T159 configuration. It is clear that the errors for T511SP and T159SP are of a similar magnitude to the standard T1279 and T511 configurations, suggesting that the introduction of stochastic physics does not negatively
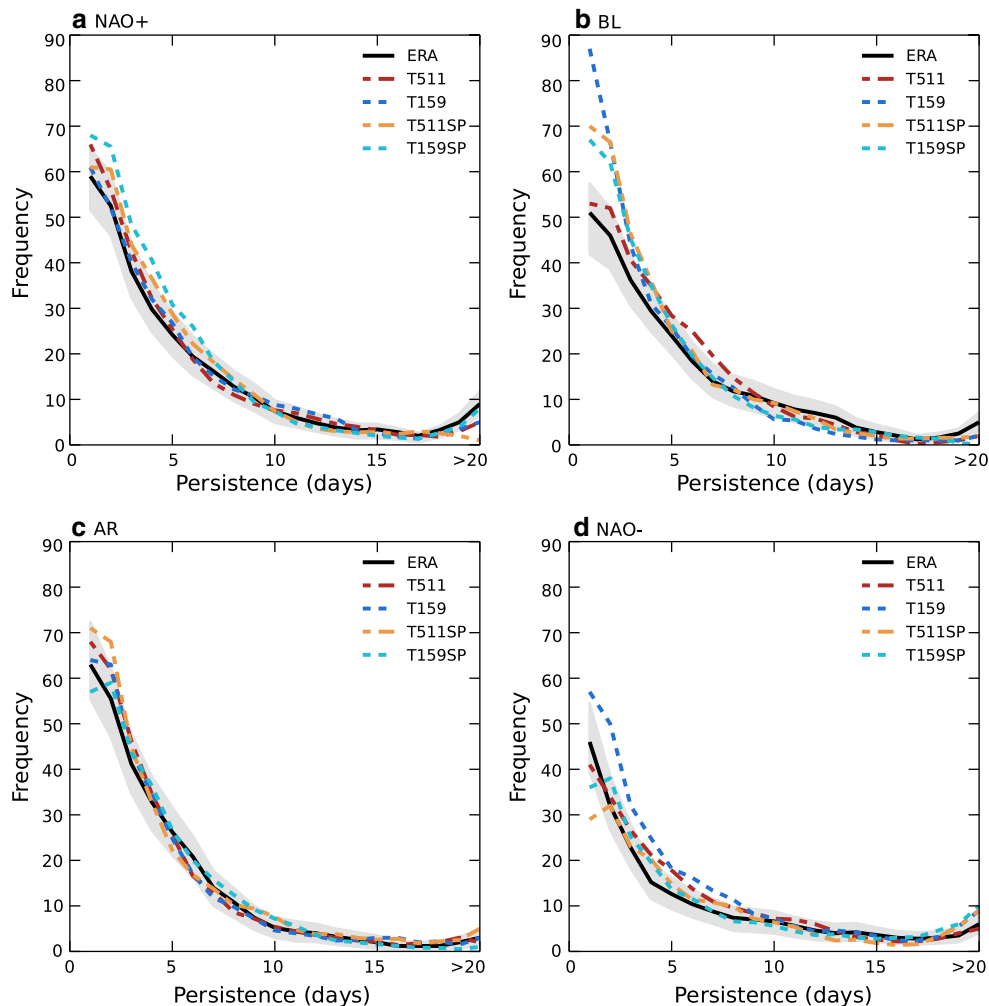
Fig. 13 Distributions of the persistence of each cluster for ERA and the T159 and T511 model configurations with and without stochastic physics. The last bin includes all events with persistence of more than 20 days. The *shaded region* around the ERA persistence distri-bution is the ±1 standard deviation interval obtained from a bootstrap re-sampling method. **a** NAO+ cluster, **b** BL cluster, **c** AR cluster and **d** NAO− cluster

impact the model representation of cluster frequencies. The error between NCEP and ERA provides some expectation of the error in climatological frequencies, but since the differences in climatological frequency are quite subtle we also provide a second bootstrapped estimate. This estimate uses the same method as was used for persistence, whereby a large number of randomized realizations of the ERA cluster time series were generated, and the absolute error relative to ERA was computed for for each cluster in each realization. The mean of these errors over realizations was then calculated which yields a bootstrapped value of the mean absolute error due to sampling variability. This error is of the same magnitude as the errors in the T511, T1279, T511SP and T159SP model configurations, suggesting that although their errors are larger than that between ERA and NCEP, they are of a similar size to what we might expect given sampling uncertainty.

We also examine the influence of stochastic physics on the persistence of the regimes. Figure 13 shows the distribution of persistence (in days) of each regime for T159SP and T511SP, with the standard configurations for reference. The introduction of stochastic physics into the T511 resolution configuration makes little difference in terms of persistence for the NAO+, AR and NAO− regimes. However, for the BL regime the introduction of stochastic physics makes the persistence worse for up to 4 day events, but makes some improvement in the 5–10 day range. At T159 resolution the introduction of stochastic physics makes the persistence of the NAO+ regime less realistic, with shorter events occurring too often. However, there are significant improvements to the persistence for the BL and NAO− regimes. It appears that the introduction of stochastic physics has a smaller impact on the persistence of the regimes than it does on the spatial patterns of the regimes themselves, although we note

that the stochastic physics does not make the persistence significantly worse in any case.

8 Conclusions

Using a regime diagnostic four clusters are identified in the wintertime European/Atlantic geopotential height field in ERA. These regimes are highly significant and are interpreted as circulation regimes. The same four regimes are identified in an independent reanalysis data set, suggesting that the regimes are a robust feature of the European/Atlantic wintertime flow. The same regime diagnostic is applied to an atmospheric model at three different horizontal resolutions. A configuration with very high horizontal resolution has a realistic simulation of regimes, in terms of both the spatial patterns of the flow regimes and their temporal characteristics. A configuration with low resolution, typical of the atmospheric resolution of modern coupled climate models, has a comparatively poor simulation of regimes. The spatial patterns of the regimes are dissimilar to reanalysis and the frequency with which each regime occurs is erroneous. Some of the regimes are also less persistent than the observed regimes. This result demonstrates the importance of representing small-scale process for the simulation of large-scale climate. It is likely that both more realistic orography (Jung et al. 2012) and more realistic representation of Rossby wave breaking processes, which are known to be important in maintaining persistent anomalies (Woollings et al. 2008; Masato et al. 2012), are responsible for the better performance of the very-high-resolution model. A configuration with intermediate horizontal resolution is able to partially bridge the gap between the very high and the low-resolution configurations, with significant improvements to the climatological frequency of occurrence and persistence of each regime, and smaller improvements to the spatial patterns of the regimes. However, this intermediate resolution is not sufficient to produce a simulation as realistic as the very-high-resolution configuration.

The introduction of a stochastic physics scheme into the low and intermediate-resolution configurations provided significant benefits. The intermediate-resolution configuration with stochastic physics had more realistic spatial patterns, with simulated regimes comparable to the very-high-resolution configuration. The stochastic physics scheme did not alter the frequency of occurrence or persistence of the regimes in a detrimental way. In the low-resolution configuration the stochastic physics scheme had a bigger impact, not only improving the spatial patterns of the regimes, but also having significant positive impact on the climatological frequency of occurrence and persistence of the regimes. As with the intermediate-resolution configuration, the stochastic physics scheme shifts the performance of the

regime simulation towards that of the very-high-resolution configuration. These results are consistent with results from simplified models of the atmosphere, where the introduction of stochastic parameterization provides a better representation of small-scale transience than deterministic parameterizations, and as a consequence allows for more realistic regime structures to be simulated (Christensen et al. 2014). These results are also particularly important, as they suggest that although small-scale processes appear to be important for the simulation of large-scale circulation regimes, it may not be necessary for these small scales to be explicitly represented in numerical models. Given that performing climate integrations at very high resolution is essentially infeasible, stochastic parameterizations could represent a computationally efficient way forward for improving the representation of the small-scale processes that are important to large-scale climate by allowing us to use coarser resolutions without losing the demonstrated benefits of very high resolution.

The results presented here also support a new more computationally efficient approach to integrating the dynamical core of weather and climate models (Palmer 2012). In this approach, not all scales are represented and integrated with the same level of precision and accuracy (Düben et al. 2013). In particular, the use of energy-efficient stochastic hardware to integrate the small-scale components of the dynamical core can lead to substantial reductions in power consumption and increases in computational speed. The results here suggest that a T1279 dynamical core, where scales between wave numbers 511 and 1279 are represented e.g. by single precision numbers and integrated on approximate chips, could show the same levels of performance as the T1279 model itself. These ideas will be tested in future studies.

Studies of low-dimensional dynamical systems have provided evidence that suggests the response of a system with naturally occurring regimes under external forcing is manifested primarily in terms of the frequency of occurrence of these regimes. If the atmosphere exhibits regime behaviour, then this implies that it is necessary for climate models to represent such regime behaviour in the atmosphere. We have demonstrated that the atmospheric resolution typical of CMIP5 climate models is likely not sufficient to represent realistic regimes, and that much higher resolution is needed. This suggests that current predictions of regional climate change may be questionable. It is also particularly relevant to regional climate modelling, where a low-resolution global model is often used to provide boundary conditions for a higher resolution limited area model. If these global driving models fail to simulate regimes realistically due to lack of horizontal resolution then the boundary conditions supplied to the embedded limited area model may be erroneous, and cause the regional model to generate an

unrealistic realization of regional climate and variability. However, this study has also demonstrated that the effect of a stochastic physics parameterization scheme is to improve the regime simulation in coarser resolution models. This is a much more promising conclusion than was presented by Dawson et al. (2012), since stochastic parameterization is very cheap compared to model resolution it potentially presents an affordable way to gain significant improvements from our climate models.

This study has focussed on atmosphere-only model integrations, with prescribed SST boundary conditions. Whilst this provides a clean way to compare different atmospheric model configurations it is not necessarily representative of how the atmosphere may behave in a fully coupled climate model. Straus et al. (2007) demonstrated the important role of SST forcing on regime structures. It is possible that the consistency of the modelled and observed regimes is in part due to the realistic prescribed SST boundary condition. If the SST surface boundary condition provided by a coupled ocean model represented large scale variability poorly then it may result in a less realistic atmospheric regime simulation, even at high resolution. Due to additional errors and biases in the oceanic component of coupled climate models, and the almost inevitable parameter tuning that is so often required for the coupling process itself, it is unlikely that one would see such a large improvement moving from the low to very-high-resolution model configurations, and even the very positive impact of the stochastic physics scheme may be lessened in the more complex coupled system. However, it has been demonstrated that even modest improvements to oceanic resolution in a coupled climate model can significantly reduce biases and allow the atmospheric model component to produce more realistic climate and variability (Scaife et al. 2011). This suggests that the improvements to atmospheric regime behaviour noted in this study may still be of relevance in a coupled model scenario.

References

Arnold HM, Moroz IM, Palmer TN (2013) Stochastic parametrizations and model uncertainty in the Lorenz '96 system. Philos Trans R Soc A 371: doi:10.1098/rsta.2011.0479

Barrier N, Treguier A-M, Cassou C, Deshayes J (2013) Impact of the winter North-Atlantic weather regimes on subtropical sea-surface height variability. Clim Dynam 41:1159–1171. doi:10.1007/s00382-012-1578-7

Buizza R, Miller M, Palmer TN (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Q J R Meteor Soc 125:2887–2908. doi:10.1002/qj.49712556006

Cassou C (2008) Intraseasonal interaction between the Madden-Julian oscillation and the North Atlantic oscillation. Nature 455:523–527. doi:10.1038/nature07286

Christensen HM, Moroz IM, Palmer TN (2014) Simulating weather regimes: impact of stochastic and perturbed parameter schemes in a simple atmospheric model. Clim Dynam. doi:10.1007/s00382-014-2239-9

Christiansen B (2007) Atmospheric circulation regimes: Can cluster analysis provide the number? J Clim 20:2229–2250. doi:10.1175/JCLI4107.1

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RR, Grant AN, Groisman PY, Jones PD, Kruk M, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ (2011) The twentieth century reanalysis project. Q J R Meteor Soc 137:1–37. doi:10.1002/qj.776

Corti S, Molteni F, Palmer TN (1999) Signature of recent climate change in frequencies of natural atmospheric circulation regimes. Nature 398:799–802. doi:10.1038/19745

Dawson A, Matthews AJ, Stevens DP (2011) Rossby wave dynamics of the extra-tropical response to El Niño: importance of the basic state in coupled GCMs. Clim Dynam 37:391–405. doi:10.1007/s00382-010-0854-7

Dawson A, Palmer TN, Corti S (2012) Simulating regime structures in weather and climate prediction models. Geophys Res Let 39:L21805. doi:10.1029/2012GL053284

Dawson A, Matthews AJ, Stevens DP, Roberts MJ, Vidale PL (2013) Importance of oceanic resolution and mean state on the extra-tropical response to El Niño in a matrix of coupled models. Clim Dynam 41:1439–1452. doi:10.1007/s00382-012-1518-6

Düben PD, McNamara H, Palmer TN (2013) The use of imprecise processing to improve accuracy in weather & climate prediction. J Comput Phys 271:2–18. doi:10.1016/j.jcp.2013.10.042

Duchon CE (1979) Lanczos filtering in one and two dimensions. J Appl Meteorol 18:1016–1022. doi:10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2

Fereday DR, Knight JR, Scaife AA, Folland CK, Philipp A (2008) Cluster analysis of North Atlantic-European circulation types and links with tropical pacific sea surface temperatures. J Clim 21:3687–3703. doi:10.1175/2007JCLI1875.1

Hurrell JW, Deser C (2010) North Atlantic climate variability: the role of the North Atlantic oscillation. J Mar Syst 79:231–244. doi:10.1016/j.jmarsys.2009.11.002

Jung T, Miller MJ, Palmer TN, Towers P, Wedi N, Achuthavarier D, Adams JM, Altshuler EL, Cash BA, Kinter JL III, Marx L, Stan C, Hodges KI (2012) High-resolution global climate simulations with the ECMWF model in Project Athena: experimental design, model climate, and seasonal forecast skill. J Clim 25:3155–3172. doi:10.1175/JCLI-D-11-00265.1

Kinter JL III, Cash B, Achuthavarier D, Adams J, Altshuler E, Dirmeyer P, Doty B, Huang B, Marx L, Manganello J, Stan C, Wakefield T, Jin E, Palmer T, Hamrud M, Jung T, Miller M,

Towers P, Wedi N, Satoh M, Tomita H, Kodama C, Nasuno T, Oouchi K, Yamada Y, Taniguchi H, Andrews P, Baer T, Ezell M, Halloy C, John D, Loftis B, Mohr R, Wong K (2013) Revolutionizing climate modelling—project Athena: a multi-institutional, international collaboration. Bull Am Meteorol Soc 94:231–245. doi:10.1175/BAMS-D-11-00043.1

Lin JW-B, Neelin JD (2000) Influence of a stochastic moist convective parameterization on tropical climate variability. Geophys Res Let 27:3691–3694. doi:10.1029/2000GL011964

Lin JW-B, Neelin JD (2003) Toward stochastic deep convective parameterization in general circulation models. Geophys Res Let 30:1163. doi:10.1029/2002GL016203

Masato G, Hoskins BJ, Woollings TJ (2012) Wave-breaking characteristics of midlatitude blocking. Q J R Meteor Soc 138:1285–1296. doi:10.1002/qj.990

Michelangeli P-A, Vautard R, Legras B (1995) Weather regimes: recurrence and quasi stationarity. J Atmos Sci 52:1237–1256. doi:10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2

Palmer TN (1993) Extended-range atmospheric prediction and the Lorenz model. Bull Am Meteorol Soc 74:49–65. doi:10.1175/1520-0477(1993)074<0049:ERAPAT>2.0.CO;2

Palmer TN (1999) A nonlinear dynamical perspective on climate prediction. J Clim 12:575–591. doi:10.1175/1520-0442(1999)012<0575:ANDPOC>2.0.CO;2

Palmer TN (2012) Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. Q J R Meteor Soc 138:841–861. doi:10.1002/qj.1923

Palmer TN, Weisheimer A (2011) Diagnosing the causes of bias in climate models—why is it so hard? Geophys Astro Fluid 105:351–365. doi:10.1080/03091929.2010.547194

Palmer TN, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts GJ, Steinheimer M, Weisheimer A (2009) Stochastic parametrization and model uncertainty, Tech. Rep. 598, European Centre for Medium-Range Weather Forecasts, Reading, UK

Pohl B, Fauchereau N (2012) The southern annular mode seen through weather regimes. J Clim 25:3336–3354. doi:10.1175/JCLI-D-11-00160.1

Scaife AA, Copsey D, Gordon C, Harris C, Hinton T, Keeley S, O'Neill A, Roberts M, Williams K (2011) Improved Atlantic winter blocking in a climate model. Geophys Res Let 38:L23703. doi:10.1029/2011GL049573

Shaffrey LC, Stevens I, Norton WA, Roberts MJ, Vidale PL, Harle JD, Jrrar A, Stevens DP, Woodage MJ, Demory M-E, Donners J, Clark DB, Clayton A, Cole JW, Wilson SS, Connolley WM, Davies TM, Iwi AM, Johns TC, King JC, New AL, Slingo JM, Slingo A, Steenman-Clark L, Martin GM (2009) UK-HiGEM: the new UK high resolution global environment model. Model description and basic evaluation. J Clim 22:1861–1896. doi:10.1175/2008JCLI2508.1

Straus DM (2010) Synoptic-eddy feedbacks and circulation regime analysis. Mon Weather Rev 138:4026–4034. doi:10.1175/2010MWR3333.1

Straus DM, Corti S, Molteni F (2007) Circulation regimes: chaotic variability versus SST-forced predictability. J Clim 20:2251–2272. doi:10.1175/JCLI4070.1

Wallace JM, Gutzler DS (1981) Teleconnections in the geopotential height field during the Northern Hemisphere winter. Mon Weather Rev 109:784–812. doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2

Woollings T, Hoskins B, Blackburn M, Berrisford P (2008) A new Rossby wave-breaking interpretation of the North Atlantic Oscillation. J Atmos Sci 65:609–626. doi:10.1175/2007JAS2347.1

Woollings T, Hannachi A, Hoskins B (2010a) Variability of the North Atlantic Eddy-driven jet stream. Q J R Meteor Soc 136:856–868. doi:10.1002/qj.625

Woollings T, Hannachi A, Hoskins B, Turner A (2010b) A regime view of the North Atlantic Oscillation and its response to anthropogenic forcing. J Clim 23:1291–1307. doi:10.1175/2009JCLI3087.1