Euro-Atlantic weather regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance

Climate Dynamics (2020) 54:5031–5048
https://doi.org/10.1007/s00382-020-05271-w

Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance

F. Fabiano[1] · H. M. Christensen[2] · K. Strommen[2] · P. Athanasiadis[3] · A. Baker[4] · R. Schiemann[4] · S. Corti[1]

Received: 3 December 2019 / Accepted: 24 April 2020 / Published online: 20 May 2020

Abstract
Recently, much attention has been devoted to better understand the internal modes of variability of the climate system. This is particularly important in mid-latitude regions like the North-Atlantic, which is characterized by a large natural variability and is intrinsically difficult to predict. A suitable framework for studying the modes of variability of the atmospheric circulation is to look for recurrent patterns, commonly referred to as Weather Regimes. Each regime is characterized by a specific large-scale atmospheric circulation pattern, thus influencing regional weather and extremes over Europe. The focus of the present paper is the study of the Euro-Atlantic wintertime Weather Regimes in the climate models participating to the PRIMAVERA project. We analyse here the set of coupled historical simulations (hist-1950), which have been performed both at standard and increased resolution, following the HighresMIP protocol. The models' performance in reproducing the observed Weather Regimes is assessed in terms of different metrics, focussing on systematic biases and on the impact of resolution. We also analyse the connection of the Weather Regimes with the Jet Stream latitude and blocking frequency over the North-Atlantic sector. We find that—for most models—the regime patterns are better represented in the higher resolution version, for all regimes but the NAO-. On the other side, no clear impact of resolution is seen on the regime frequency of occurrence and persistence. Also, for most models, the regimes tend to be more tightly clustered in the increased resolution simulations, more closely resembling the observed ones. However, the horizontal resolution is not the only factor determining the model performance, and we find some evidence that biases in the SSTs and mean geopotential field might also play a role.

Keywords Climate models · Atmospheric dynamics · Mid-latitudes · Weather regimes · North Atlantic · HighresMIP · Modes of variability

1 Introduction

Due to its large natural variability, the North Atlantic–European climate is one of the most difficult to predict. The peculiarity of the North Atlantic is clearly seen in climatological maps of the wintertime 500 hPa geopotential height variability, which shows two maxima in this region. The first maximum is found at high frequency—less than 5 days period—along the northeastern coast of North America, where the eddy-driven jet stream enters the Atlantic ocean. The second maximum is located in the northern part of the Atlantic Ocean, south of Iceland and northwest of the British Isles, in the center of action of the North Atlantic Oscillation (NAO) and is characterized by longer periods (5 days to a few weeks) (Blackmon et al. 1984).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00382-020-05271-w) contains supplementary material, which is available to authorized users.

✉ F. Fabiano
f.fabiano@isac.cnr.it

[1] Institute of Atmospheric Sciences and Climate (ISAC)—CNR, Bologna, Italy

[2] University of Oxford, Oxford, UK

[3] Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Bologna, Italy

[4] University of Reading, Reading, UK

Alternative views of the North-Atlantic variability. A typical approach used to describe the variability in the North Atlantic—and other mid-latitude regions—is to look for persistent and/or recurrent dynamical configurations. A great variety of dynamical variables (zonal wind, geopotential, sea level pressure, potential vorticity,...) and techniques (maxima in the pdf, clustering, ...) have been used historically to find a set of dominant "modes of variability" (Vautard 1990; Hannachi et al. 2017; Athanasiadis et al. 2010).

Recently, in particular for the North Atlantic region, most studies focus either on the geopotential height at middle/upper tropospheric levels or on the zonal wind in the lower troposphere. The first approach, which is the focus of the present paper, consists in studying the variability in terms of recurrent and persistent geopotential patterns. These patterns are commonly referred to as Weather Regimes (WRs) and can last from a few days up to three or 4 weeks (Straus et al. 2007; Dawson et al. 2012; Hannachi et al. 2017). Attempts at locating such geopotential height regimes using clustering methods have consistently identified four regimes dominating the wintertime atmospheric variability over the Euro-Atlantic region (Hannachi et al. 2017; Straus et al. 2017; Dawson et al. 2012). These regimes are:

1. The positive phase of the NAO (NAO+), characterized by a low pressure South of Iceland and a positive anomaly at lower latitudes (southern Europe, Azores);
2. The Scandinavian Blocking (SBL) pattern, showing a strong high pressure over western Scandinavia and the North Sea;
3. The Atlantic Ridge pattern (AR), characterized by a high pressure anomaly in the middle of North-Atlantic at about 55° latitude;
4. The negative phase of the NAO (NAO−), with a high pressure anomaly over Greenland and a low pressure over southern latitudes.

Weather regimes have been frequently studied in a nonlinear dynamical system perspective. The clear tendency for multimodality in the pdf of the observed field—which departs significantly from a multinormal distribution (Straus et al. 2007, 2017; Dawson et al. 2012; Corti et al. 1999)—suggests to consider the regimes as real attractors of the chaotic climate system. This hypothesis has been widely studied in simplified models (Christensen et al. 2015; Hannachi et al. 2017) and extended by analogy to complex GCMs (Palmer 1999; Corti et al. 1999).

An alternative perspective to Weather Regimes uses the latitudinal position and maximum speed of the eddy-driven Jet Stream. Woollings et al. (2010) showed that the observed distribution of the jet stream latitude is multi-modal and there are three preferred locations corresponding to different geopotential configurations. A southern shift in the jet stream is linked to a negative phase of the NAO and an anticyclonic anomaly over Greenland, while central and northern jets correspond to the positive and negative East Atlantic patterns (EA), closely related to NAO + and AR regimes (Woollings et al. 2010; Franzke et al. 2011; Madonna et al. 2017). In this framework, the SBL regime covers a wide variety of jet latitudes and is referred to as mixed jet state. As for the jet speed, systematically smaller values are found in the third sector of the NAO-EA phase space (i.e. negative NAO and negative EA), corresponding to part of the SBL and NAO-regimes. Large variability of the jet speed is found in the remaining three sectors of the NAO-EA space. An important step towards reconciling the Jet Stream latitude and Weather Regimes perspectives has been done by Madonna et al. (2017), who performed cluster analysis of the two dimensional lower tropospheric zonal wind patterns over the North Atlantic. Three out of the four "jet regimes" obtained in this way correspond to the three peaks in the one-dimensional jet latitude distribution, and the fourth (linked to the SBL regime) consists of a low or central jet in the western Atlantic that abruptly shifts northward due to the blocking high over Scandinavia. They additionally show how these "jet regimes" well match the usual weather regimes calculated from geopotential height, so building a consistent dynamical picture of the modes of variability over the Euro-Atlantic sector.

Another common and closely related analysis consists in studying the modifications of the unperturbed westerly flow in terms of blocking episodes, that is the presence of a high geopotential height anomaly that forces the jet to shift or to weaken and meander. The climatological frequency of blocked days shows two distinct maxima over northwestern Europe and Greenland, which have their counterparts in the SBL and NAO-Weather Regimes pattern (Scherrer et al. 2006; Davini et al. 2012, 2017; Madonna et al. 2017). In this sense, out of the 4 weather regimes described above, we have a single unblocked regime (NAO+) and three "blocked" regimes (SBL, AR, NAO-), although the correspondence might not be completely consistent.

Climate models. Global climate models have improved much in the representation of the large-scale circulation in the last decades, primarily as a consequence of new parameterization schemes and larger computational resources that have allowed to reach resolutions on the order of 100 km and below (Haarsma et al. 2016; Roberts et al. 2018b). However, many problems are still to be solved and the skill in reproducing some features of the observed climate will hopefully improve in future models. Focusing on the North-Atlantic, models are known to struggle to reproduce the observed jet stream variability and blocking frequency. Most CMIP5 models show no sign of multi-modality in the jet latitude pdf (Hannachi et al. 2013; Anstey et al. 2013; Iqbal et al.

2018) and have a well-known negative bias in the climatological frequency of blocking events over Europe (Anstey et al. 2013; Davini and D'Andrea 2016; Masato et al. 2013; Schiemann et al. 2017).

Also, some studies have focused on the performance of complex GCMs in correctly reproducing the observed Weather Regimes. An example in this regard is the article by Dawson et al. (2012): a low (spectral resolution T159, roughly 125 km grid spacing) and a high resolution version (T1279, roughly 16 km spacing) of the ECMWF model were compared for their skill in reproducing the regimes patterns, their frequency of occurrence, persistence and preferred transitions. The low-resolution version was found unable to correctly reproduce the SBL and AR regimes and produced too many short NAO events. The high-resolution model showed significant improvements, hinting that an increase in resolution is needed to give a reliable representation of North-Atlantic variability. Dawson and Palmer (2015) showed that some improvement is already obtained with an intermediate resolution (T511). A similar analysis has been performed by Cattiaux et al. (2013) on the CMIP3 and CMIP5 versions of the IPSL model at different resolutions. The regimes patterns were found to be better reproduced with the CMIP5 version, which included more vertical levels, and with increased resolution. In terms of frequencies, they found an overestimation of the NAO+ and SBL occurrence, which slightly improved with resolution. More recently, Strommen et al. (2019) analyzed the standard and high resolution versions of three climate models run in atmospheric-only mode (AMIP) and found no systematic improvements in the regime patterns with resolution. However, both the sharpness of the regime structure (see Sect. 3.2) and the persistence of the SBL regime improved with increased resolution.

WRs and climate change. In recent years, there has been increasing interest in studying WRs and how well climate models reproduce them, due to their importance in influencing regional weather patterns and extremes and possibly future regional changes in the climate state (Corti et al. 1999; Matsueda and Palmer 2018). In fact, much attention has been given recently not only to the future changes in the mean state of the general circulation (Hoskins and Woollings 2015), but also to changes in the modes of variability. The interest in the reproduction of WRs in GCMs and in their response to climate change is motivated by different reasons. From a dynamical systems perspective, it has been hypothesized that the first response of the system to a moderate external forcing would manifest in the frequencies of occurrence of the different modes of variability, whilst the variability patterns will remain initially unchanged (Palmer 1999; Corti et al. 1999). From a more pragmatic point of view, WRs are closely connected with regional weather types and weather extremes and constitute a clear and useful framework to study the impacts of climate change in key mid-latitude regions, like Europe and the Mediterranean (Plaut and Simonnet 2001; Yiou and Nogaj 2004; Zampieri et al. 2017; Raymond et al. 2018) or the Atlantic coast of North-America (Roller et al. 2016). Recently, the WRs framework has also been used to assess the variability of wind/solar energy production potential in Europe (Grams et al. 2017).

In this paper we report on a multi-model assessment of various diagnostics related to WRs, considering a set of coupled historical simulations using different state-of-the-art climate models at two resolutions. The main goal of the paper is to investigate the typical model biases in reproducing the observed WRs and their dynamics. Specifically, we assess whether increasing the model resolution produces a significant improvement in the diagnostics in a multi-model ensemble, following on from previous results for individual models (Dawson and Palmer 2015) and for AMIP simulations (Strommen et al. 2019). The paper is structured as follows: Sect. 2 regards the data used in the analysis and describes the models and their configurations; Sect. 3 describes the procedure used to calculate the Weather Regimes and the related diagnostics; Sect. 4 contains the main results of the paper, regarding the model performances and the improvements with increased resolution; in Sect. 5 we analyse some possible drivers of the model performance; Sect. 6 is dedicated to the final discussion and conclusions.

2 Data: models and observations

The coupled models considered in this work are those participating in the European H2020 PRIMAVERA project: CMCC-CM2 (Cherchi et al. 2019), CNRM-CM6 (Voldoire et al. 2019), EC-Earth3 (Haarsma et al. 2020), ECMWF-IFS (Roberts et al. 2018a), HadGEM3-GC31 (Williams et al. 2018), MPI-ESM1-2 (Gutjahr et al. 2019), AWI-CM-1.0 (Sein et al. 2017). All models follow the HighResMIP protocol (Haarsma et al. 2016), at nominal resolutions ranging from 250 to 25 km. For each model, two versions are available, one at standard (low-res, LR) and one at higher resolution (high-res, HR); some model produced additional intermediate resolutions as well. This is done following the philosophy of the project, which aims to assess the improvements in the representation of the observed climate due only to the increased model resolution. Therefore, the models have been tuned in their low-res version, and the high-res version is obtained by just increasing the resolution, with no additional tuning.

In Table 1, all models are listed with their basic characteristics, their effective resolutions in both the atmosphere and ocean components, and the number of ensemble

Table 1 Models used in the analysis, listed with their components (atmosphere/ocean/ice models), the atmospheric grid used for the two versions (low- and high-res), the nominal resolution and number of levels used for the atmosphere and ocean components, the number of ensemble members analysed for each resolution. Note that for HadGEM-GC31 (LL, MM, HM, HH) and ECMWF-IFS (LR, MR, HR) more than two model versions have been analyzed

Model name	CMCC-CM2	CNRM-CM6	EC-Earth3	ECMWF-IFS	MPI-ESM1	HadGEM-GC31	AWI-CM-1
Components	CAM4, NEMO, CICE	ARPEGE, NEMO, GELATO	IFS, NEMO, LIM	IFS (43r1), NEMO, LIM2	ECHAM6.3, MPIOM1.63, MPIOM1.63	UM, NEMO, CICE	ECHAM6.3, FESOM, FESIM
Atmos grid	$1^\circ \times 1^\circ$, $0.25^\circ \times 0.25^\circ$	Tl127, Tl359	Tl255, Tl511	Tco199, Tco199, Tco399	T127, T255	N96, N216, N512, N512	T63, T127
Atmos. nom. res. (km)	100, 25	250, 50	100, 50	50, 50, 25	100, 50	250, 100, 50, 50	250, 100
Atmos levels	26	91	91	91	95	85	95
Ocean nom. res. (km)	25, 25	100, 25	100, 25	100, 25, 25	40, 40	100, 25, 25, 8	50, 25
Ocean levels	50	75	75	75	40	75	47
Members	1, 1	2, 1	3, 3	8, 3, 6	1, 1	8, 1, 3, 1	1, 1

members analysed for each resolution. Note that most models increased the resolution of both the atmosphere and the ocean components, with the exception of CMCC-CM2 and MPI-ESM1, increasing the atmospheric resolution only.

Under the PRIMAVERA project, models are run both in atmosphere-only and coupled mode. Here we consider the coupled historical simulations (hist-1950), covering the range 1950–2014. For each simulation we use the geopotential height daily mean data at 500 hPa, as explained in Sect. 3.1. As a reference, we take geopotential height daily reanalysis data at 500 hPa from ERA40 (1957–1978) and ERAInterim (1979–2014), thus covering the 1957–2014 range.

3 Method and diagnostics

The following subsections describe the methodology used to calculate the Weather Regimes together with the diagnostics and metrics applied to compare observed and simulated regimes. The work-flow has been implemented in a Python software package named "WRtool", freely available at https://github.com/fedef17/WRtool.

3.1 Calculation of the weather Regimes

The calculation of Weather Regimes is performed here through clustering of the geopotential height anomalies at 500 hPa (Z500) in a reduced phase space. This approach is widely used in literature, even if many subtle variations in the actual procedure have been adopted (Hannachi et al. 2017; Straus et al. 2017). Despite these differences, the main results of the analysis are quite robust, as documented by many accurate tests in literature (Straus et al. 2007; Hannachi et al. 2013).

In this work we consider the wintertime (December–February; DJF) daily Z500 fields and proceed through the following steps.

Data pre-treatment and removal of the mean seasonal cycle. The model (and the reanalysis) data are first interpolated to a $2.5^\circ \times 2.5^\circ$ grid using bilinear interpolation. This is done mainly for practical reasons related to data-handling, i.e. to have all the data at the same spatial resolution. The interpolation does not affect the results, mainly because we are interested in large scale patterns. We then calculate the mean seasonal cycle and subtract it from the data, to obtain a timeseries of daily geopotential height anomalies. The seasonal cycle is calculated averaging the full 1957–2014 timeseries day-by-day and then performing a running mean of 20 days. Other authors used a running mean of 5 days (Dawson et al. 2012; Strommen et al. 2019) or calculated the timeseries expansion in terms of Legendre polynomials (Straus et al. 2007). We chose the 20 days running mean in order to avoid higher frequency fluctuations due to internal variability. We considered whether to detrend the seasonal cycle, in order to remove the climate change signal on the mean state: the trend in the wintertime 500 hPa geopotential height was found to be comparable to decadal variability in some regions (with a maximum of about 20 m in the south-west part of the domain) and generally smaller elsewhere (not shown). We finally decided not to apply the detrending, because considering shorter (30 years) periods introduces unwanted noise from decadal variability in the entire domain and in the WR attribution.

We select the Euro-Atlantic (EAT) domain (30°–90° N, 80° W–40° E), which is one of the most used in the literature (Dawson et al. 2012; Madonna et al. 2017; Strommen et al. 2019), although some authors considered slightly different sectors (Cassou 2008; Cattiaux et al. 2013). We checked that

the observed regimes' patterns are robust to small changes in the choice of the domain (of the order of 10°). It is worth noting that the Euro-Atlantic domain used here extends much more eastward (40°) than the usual North-Atlantic region adopted in Jet Stream studies (Woollings et al. 2010).

Calculation of the EOFs and projection onto the reduced phase space. We then calculate the Empirical Orthogonal Functions (EOFs) of the anomalies on the Euro-Atlantic domain and define a phase space of reduced dimensionality as that spanned by the first 4 EOFs, representing about 55% of the variance. The choice of the number of EOFs varies in literature from 4 (Dawson et al. 2012) up to 14 (Cassou 2008; Cattiaux et al. 2013) or more. Sensitivity tests have been performed by (Straus et al. 2007) for the Pacific-North American sector and show that the mean regime patterns are quite insensitive to this choice, but the clustering becomes less significant with a larger number of EOFs. We performed similar tests for the EAT domain, finding that regime patterns and cluster attribution are robust to this choice.

We proceed differently at this point for the models and for the reference. Repeating this step for each simulation would produce different EOFs and then different phase spaces, and would constitute an additional factor to take into account in the analysis. For example, it can also happen in some cases that the 4th model EOF is the model counterpart of the 5th EOF in the observations or so, thus heavily penalizing the model performance in some regards.

For these reasons we chose to use the same reduced phase space for all simulations, which is that spanned by the 4 leading EOFs obtained from ERA reanalysis ("phase space" in the following). The daily anomalies are then projected on this space, thus obtaining a timeseries of 4 Components (pseudo-PCs) for each simulation. Strictly speaking, these are not Principal Components, since we are not projecting on the model EOFs. Using the same space for all models allows comparing the regime patterns in a consistent way (for example, computing distances and angles). However, this choice might affect some of the metrics applied to evaluate the regime structure, so we specifically comment on this in Sect. 3.2.

Clustering and Weather Regimes attribution. We apply a K-means clustering algorithm to the models' pseudo-PCs and observation PCs, setting the number of clusters to 4. This number has been found to give the most significant clustering for the wintertime Euro-Atlantic sector (Michelangeli et al. 1995; Cassou 2008; Straus et al. 2017; Hannachi et al. 2017). Each day in the pseudo-PCs timeseries is assigned to a cluster, minimizing the distance in phase space to the cluster centroid. The mean regime patterns are calculated as composites of all the points belonging to the corresponding cluster. Once the model clusters have been calculated, their order might differ from the reference ones. Therefore a best-matching algorithm is applied, which reorders the simulated regimes so as to minimize the total RMS deviation between the observed and simulated patterns.

3.2 WR-related model diagnostics

The Weather Regimes are first calculated for the reanalysis dataset and used as a reference for the simulated WRs. The WR patterns obtained here for the ERA reanalysis are shown in Fig. 1. Using a different reanalysis like JRA55 or NCEP produces very similar results (Strommen et al. 2019; Dawson et al. 2012). In line with previous works (Dawson et al. 2012; Cattiaux et al. 2013; Strommen et al. 2019), we evaluate the model performance in reproducing WRs in terms of different metrics:

(i) Regime centroid and mean pattern. For each regime, the cluster centroid in phase space is given by the average of all pseudo-PCs assigned to that regime. This corresponds to a mean regime pattern, which is the simplest way to visualize the preferred large scale geostrophic flow configuration corresponding to that regime. A subtle difference exists between the mean pattern calculated in this way and that obtained through the composite of all daily anomalies corresponding to the cluster. The latter considers the anomalies in the full space, instead of only the 4 dimensions given by the reference EOFs. However, since higher order variability is quite uniform across the regimes, the two patterns are almost indistinguishable, and we use the reduced phase space pattern for consistency with the other metrics.

(ii) Regime "clouds" in phase space. The comparison of the simulated and observed regime patterns is good to get an idea of the model performance "at a glance". However, there is much more information stored in the statistics of the pseudo-PCs belonging to each regime. For example, the location of the centroid does not provide any information about the shape or the width of the distribution. Therefore, we apply a suitable metric to compare the model regime "clouds" in phase space. For each regime, we calculate the relative entropy of the modeled distribution with respect to the reference one. The relative entropy or Kullback–Leibler divergence (Kullback and Leibler 1951) is a measure of how much the two distributions differ from an information theory perspective: the larger the relative entropy, the further the two distributions are; a value of 0 means that the two distributions are identical. The reference and model distributions of the regime clouds are calculated through a Gaussian kernel approach for each

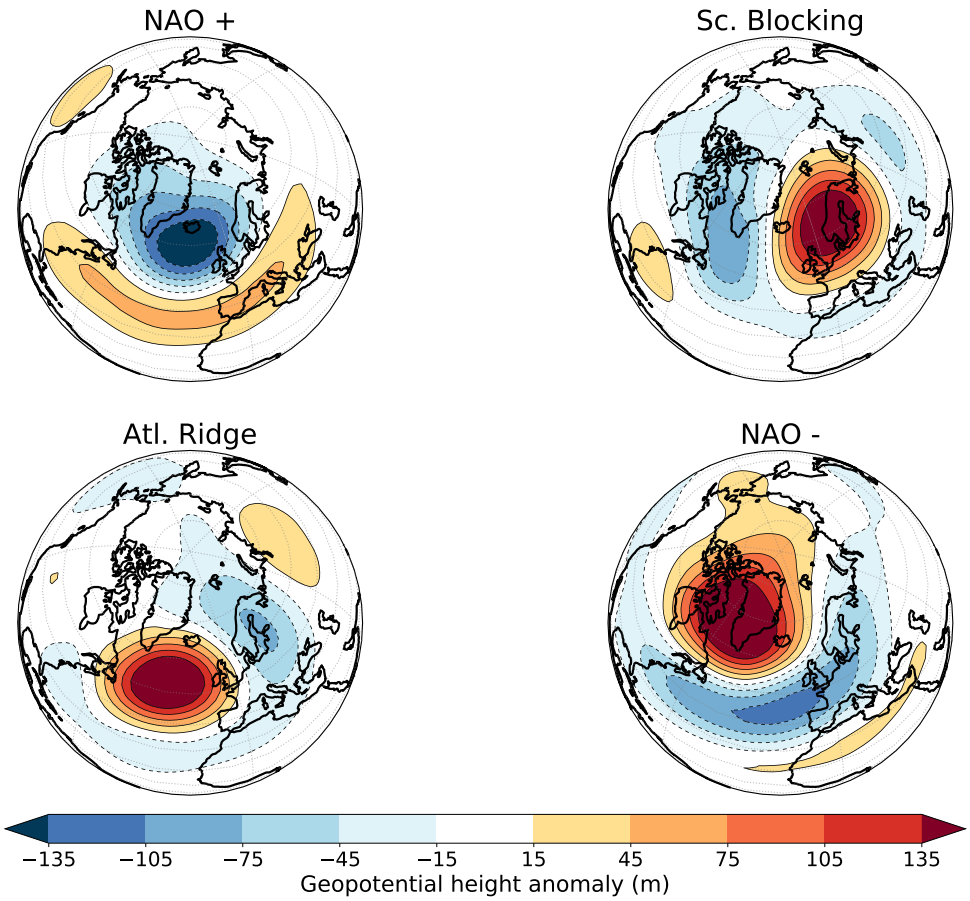


Fig. 1 Mean pattern of the weather regimes obtained from a combination of ERA40 and ERAInterim (1957–2014)

regime separately on a fixed grid, considering the first 3 pseudo-PCs only (for computational reasons). The relative entropy is given by:

$$E_{KL}(P,Q) = \sum_i P(x_i) \log\left(\frac{P(x_i)}{Q(x_i)}\right) \quad (1)$$

where P is the reference distribution obtained from ERA, Q is the simulated one and the x_i are all points in the 3-dimensional grid.

(iii) Variance ratio. Also called optimal variance ratio, this is the ratio between the mean inter-cluster squared distance and the mean intra-cluster variance. The larger the variance ratio, the more clustered are the data, giving compact clusters well apart one from the other. The variance ratios are always below unity for the data we analyze here, meaning that the clusters are somewhat overlapping. Formally, the variance ratio is obtained in the following way:

$$R_v = \frac{2\sum_{C_{ij}} \left|\mathbf{c_i} - \mathbf{c_j}\right|^2}{(n-1)\sum_i \sigma_i^2} \quad (2)$$

where the first summation is done over all the combinations (without repetition) of the clusters (C_{ij}), $\mathbf{c_i}$ and $\mathbf{c_j}$ are the respective cluster centroids, n is the number of clusters and σ_i^2 is the intra-cluster variance.

(iv) Sharpness. This measure (also called significance in previous works) is closely connected to the variance ratio. A Monte Carlo test is performed on the pseudo-PCs, to assess whether there is evidence of multi-modality and hence non-normality. Following previous works (Straus and Molteni 2004; Straus et al. 2007; Dawson et al. 2012; Dawson and Palmer 2015; Strommen et al. 2019), we first construct 1000 synthetic data series drawn from a multinormal distribution with the same length, lag-1 autocorrelation and variance as the original pseudo-PCs. The sharpness is defined as the percent of synthetic data series that have a variance ratio smaller than the original one (i.e. that are "less clustered").

(v) Regime frequencies. The frequency of occurrence of each regime is calculated simply as the ratio of days belonging to that regime to the total number of days.

(vi) Regime persistence. A regime event is defined as the set of consecutive days belonging to the same

regime. Since sometimes a single day pause breaks a long regime event, we relaxed this definition allowing for a single day belonging to a different regime in the set of a regime event. The distribution of the length of the regime events in each simulation is analysed and compared to the reference one.

(vii) Regime Jet Latitude distribution. The dynamical connection between WRs and eddy-driven Jet Stream is analyzed here calculating the distribution of the Jet Latitude index corresponding to each regime, in a similar way to what Madonna et al. (2017) did for the reanalysis. The regime-related distributions are then compared to the reference ones. The Jet latitude distributions of the hist-1950 PRIMAVERA simulations are thoroughly analyzed in Athanasiadis et al. (in prep.). Here we use their calculations to analyze the link between Euro-Atlantic WRs and Jet Stream. The calculation considers the daily mean uwind field at 850 hPa, low-pass filtered using a Lanczos filter at 10 days. The jet latitude is defined as the latitude of maximum Jet speed, after zonal averaging between 60W and 0 longitude. The index is consistent with the one defined in Woollings et al. (2010), apart from an additional filtering of grid points with orography higher than 1300 m.

(viii) Regime Blocking frequency. We also computed regime composites of the 2D blocking frequencies and compared them to the reference ones in terms of mean frequency over the EAT sector. The Blocking frequencies are calculated as in Schiemann et al. (2017) using the absolute geopotential height (AGP) blocking index (Scherrer et al. 2006). The AGP index is a two-dimensional extension of the Tibaldi and Molteni (1990) blocking index. Two conditions are needed to define a blocking event at a specific grid point: reversal of the climatological equator-pole gradient of the 500-hPa geopotential height to the south and a westerly flow to the north. There is the additional requirement that these conditions are met for at least 5 consecutive days. A thorough analysis of blocking in the hist-1950 PRIMAVERA simulations is performed in Schiemann et al. (2020). Here we are mainly interested in the connection of the blocking index with the WRs.

4 Model performance and improvements with increased resolution

A common problem in climate studies that compare coupled historical runs with the observations is that only a "single" observed history is available. If the differences between the simulated and observed climates are inside the range of the internal variability, it is very difficult to assess whether they reflect some real bias of the model or only different phases of the internal oscillations.

One way to estimate the internal variability on decadal timescales is to perform a bootstrap analysis to the available dataset. This has been done consistently for all simulations and for the observations, considering sets of 30 seasons randomly chosen (with repetitions allowed) among all available seasons between 1957 and 2014. Repeating this step 500 times gives an estimate of the distribution of the various metrics and allows to properly compare models and observations. Figures 2, 3, 4, 5, 6, 7, 8, 10 (with the exception of Fig. 9) show box plots that represent the distributions obtained through the bootstrapping. For each model, the plots show median (horizontal line), first and third quartile (boxes) and 10 and 90 percentiles (bars). For some models, more ensemble runs with the same model configuration were available (see Table 1): in those cases, the distribution median and percentiles are calculated among all members. Mean and extreme (max/min) values across the ensemble of each member's mean metrics are also shown in terms of a dot and two small triangles. At the right of the gray vertical line, three boxes are shown. The first (black box) refers to the ERA 30-year bootstraps and is obtained in the same way as for the models. The other two boxes represent average quantities among all the LR and HR models: in this case the meaning of the boxes and bars does not correspond to the true percentiles and median of the overall distribution, but are calculated as the average of the percentiles and median over all models. For the models with more than two resolutions, only the lowest and highest resolutions are taken into account for the LR and HR averages. The different number of ensemble members does not affect the multi-model average: each model weighs one regardless of the number of simulations performed.

In the following, we dedicate a paragraph to each diagnostic, aiming to rigorously assess the existence and magnitude of model biases.

4.1 Regime centroid and mean pattern

Weather Regimes' centroids (i.e. the mean anomalies) are usually poorly simulated in simplified models of the atmospheric circulation or in GCMs at low resolutions (Dawson et al. 2012), at least for some of the regimes. With state-of-the-art climate model resolutions, the patterns have greatly improved and there is a qualitatively good matching between simulated and observed regimes. More quantitatively, we calculated the distance in phase space between the simulated and reference centroids and the pattern correlation between the respective mean patterns.

Figures 2 and 3 show the performance of different models in terms of the phase space distance between simulated and

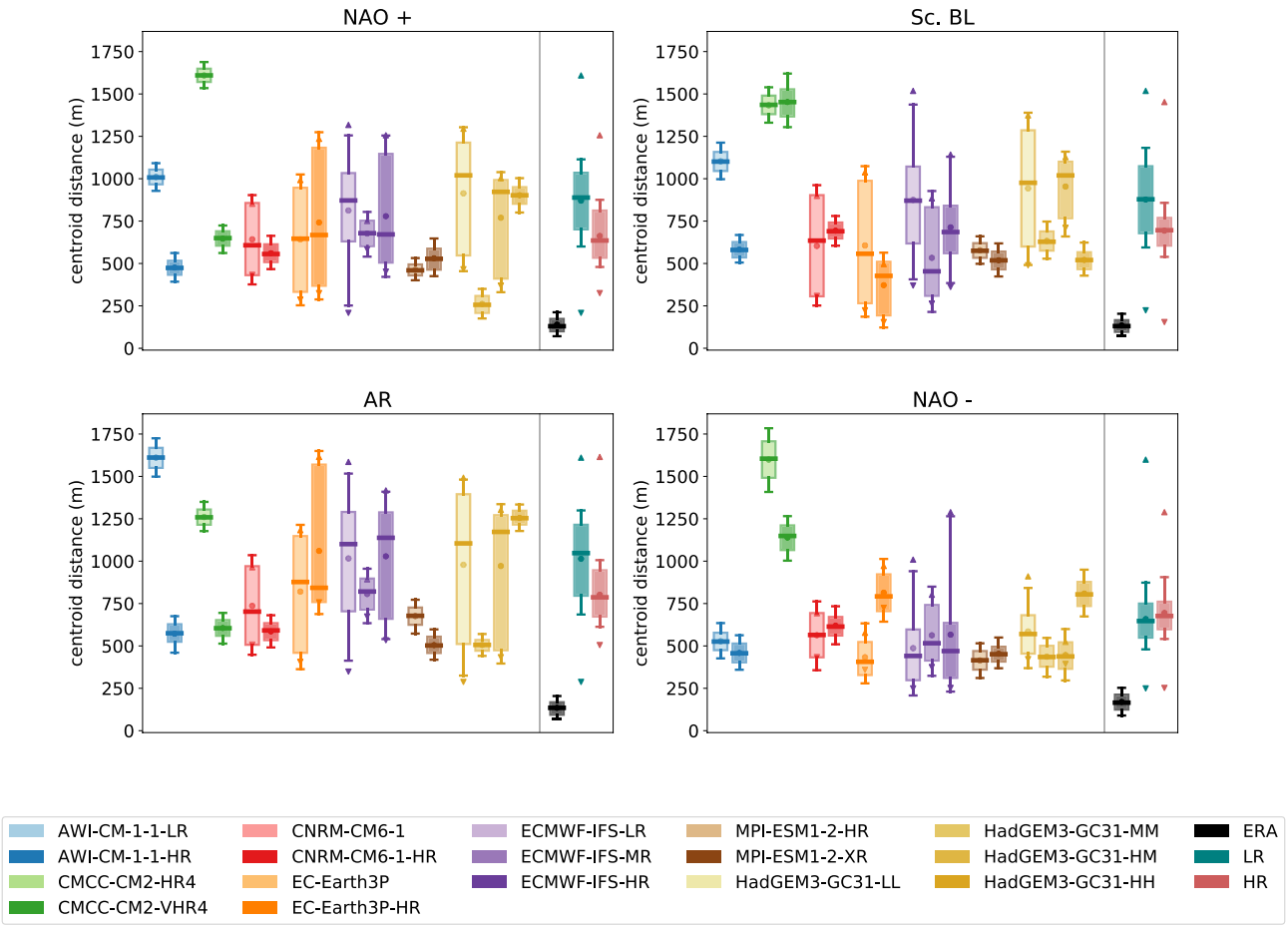


Fig. 2 Performance of the models in terms of centroid-to-centroid distance, separately for each Weather Regime. The box plots refer to the distribution of 30-yr bootstraps of each model and show mean (dot), median (horizontal line), first and third quartile (boxes) and 10 and 90 percentiles (bars). For models with more ensemble runs, mean and extreme (max/min) values across the ensemble are also shown in terms of a dot and two small triangles. At the right of the gray vertical line, three boxes are shown. The first (black box) refers to the ERA 30-year bootstraps. The other two boxes represent average quantities among all the LR and HR models and are calculated as the average of the percentiles and median over all models. More details in the text at the beginning of Sect. 4

reference centroids (Fig. 2) and in terms of the correlation between the simulated and reference mean patterns (Fig. 3). For each model, the fainter color corresponds to the lowest resolution and the stronger color to the highest resolution version. At the right end of the plot, a measure of the observed variability (black box, named "ERA") is shown along with the ensemble average of the lowest and highest resolution versions of each model.

For all regimes but the NAO-, most of the models show an improvement with increased resolution, both in terms of the centroid-to-centroid distance and of the mean pattern correlation. The improvement is not seen for NAO-, where we notice a slight degradation for most models. However, this might in part be due to the fact that the NAO-regime is already well reproduced in most models and shows in fact the smallest mean distance and the largest pattern correlation.

4.2 Regime clouds in phase space

The differences between the simulated and observed regime clouds are evaluated through the relative entropy between the two sets of distributions, as explained in Sect. 3. The results of the comparison are shown in Fig. 4. In most cases, the direction of change is consistent with the results obtained comparing the centroids and mean patterns, showing an improvement (i.e. a lower relative entropy) with high-res for most models and all regimes. It is worth recalling here that the pdfs have been corrected for the offset in the centroid position, so that the relative entropy in Fig. 4 is due only to the regime cloud shape and spread.

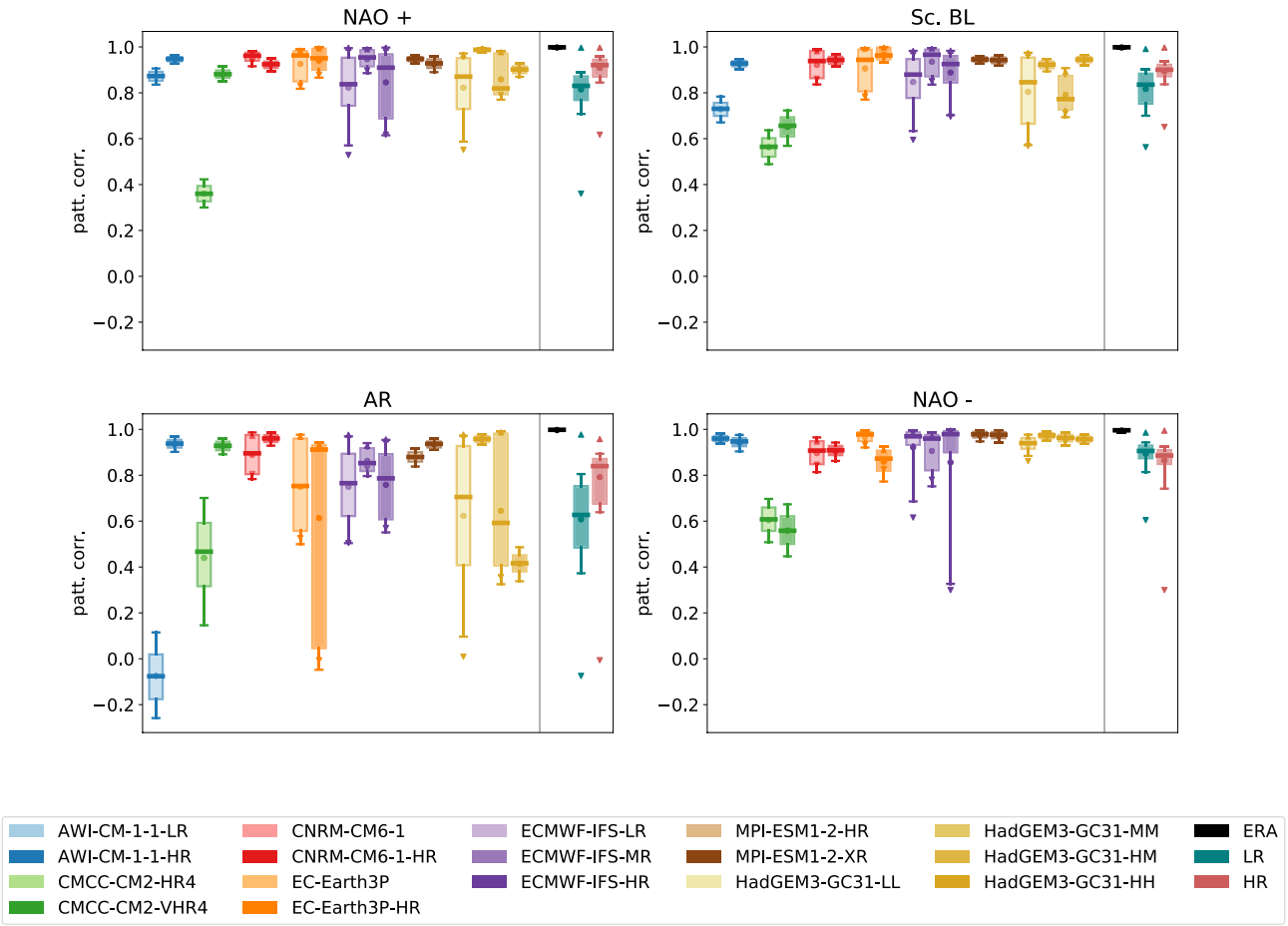


Fig. 3 Performance of the models in terms of spatial correlation between the simulated and reference mean patterns, separately for each Weather Regime. For details on the plot construction, refer to Fig. 2

4.3 Variance ratio and sharpness

The variance ratio and sharpness of the regime clusters are the metrics that most clearly show the differences between models and observations. Figures 5 and 6 show the evaluation of these two quantities for a 30-year bootstrap analysis of the models and observations.

As it is clearly seen in Fig. 5, almost all simulations have a lower variance ratio than the reanalysis, and most of them are outside the range of ERA variability. The only notable exception are the LL, MM and HM simulations of the HadGEM-GC31 model, that have values in the lower half of the observed range. This means that the models are not yet able to fully capture the regime structure of the observed geopotential field in the North Atlantic and the simulated regimes tend to be systematically less tightly clustered. Most models improve with the increased resolution, apart from MPI-ESM1-2 and HadGEM-GC31. On average, the HR model versions perform better in that the lower tail of the distribution is significantly shifted towards higher values. A small increase is also noted in the median of the distribution.

Also for the sharpness (Fig. 6) the model biases are very clear: all models have a smaller sharpness than the observation and show a much larger range of variability. Although the range of variability might be enhanced by the construction of the sharpness measure, which is highly nonlinear and threshold sensitive, the difference between models and observations remains clear and outside the range of the observed variability. In this case, the models respond in different directions to the increased resolution: some of them get better (AWI-CM-1-1, CNRM-CM6, HadGEM-GC31), other (EC-Earth3P, CMCC-CM2, ECMWF-IFS) do not show significant shifts of the distribution and MPI-ESM1 gets worse. To further complicate this picture, the change is not monotonic with increasing resolution when considering a single model: both ECMWF and HadGEM-GC31 show a different behavior between the MR and LR resolution versions and the HR and MR versions. Similarly to the variance ratio, when looking at the ensemble average of sharpness, a depopulation of the lower quartile of the distribution is seen in favour of the upper quartiles, showing

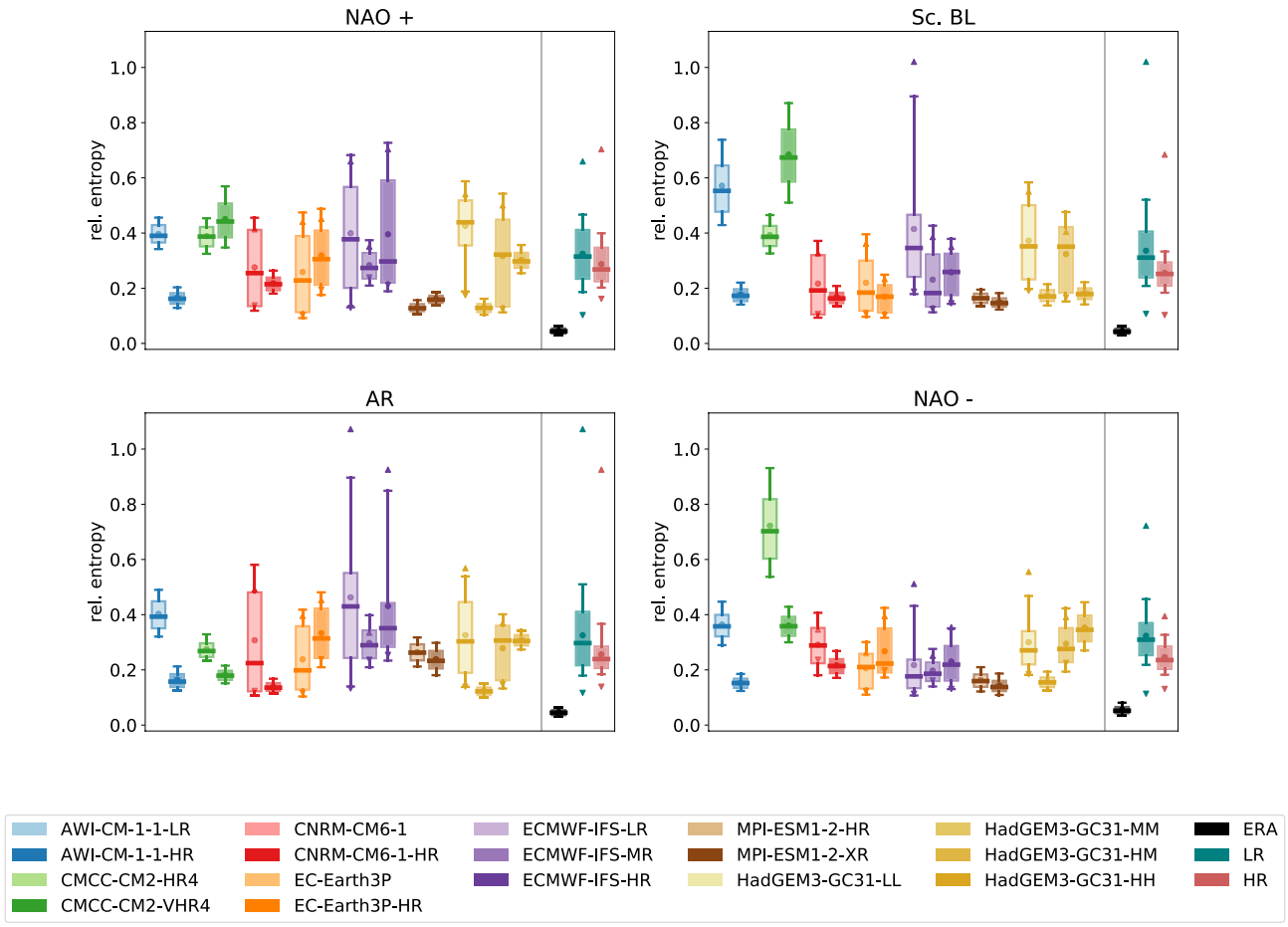
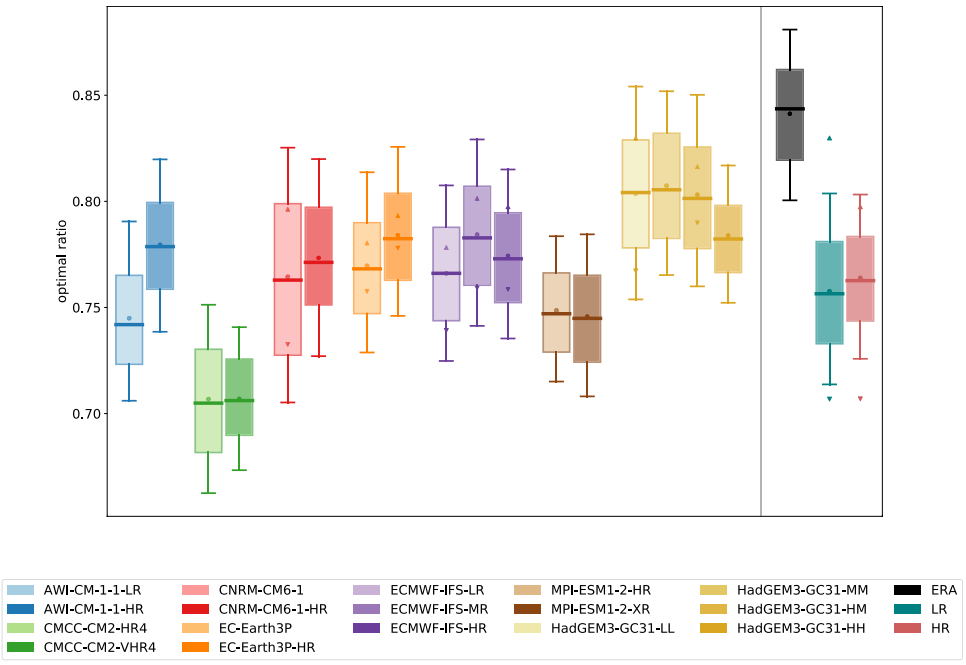


Fig. 4 Relative entropy of the regime "clouds", separately for each Weather Regime. For details on the plot construction, refer to Fig. 2

Fig. 5 Variance ratio of the WRs in the models and observations. For details on the plot construction, refer to Fig. 2



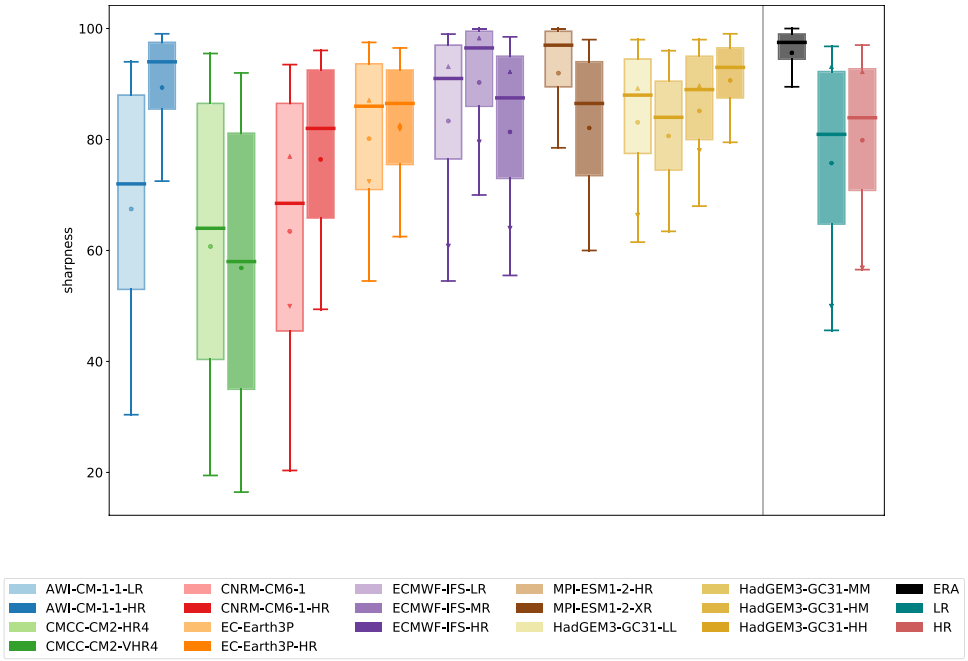


Fig. 6 As in Fig. 6, but for sharpness

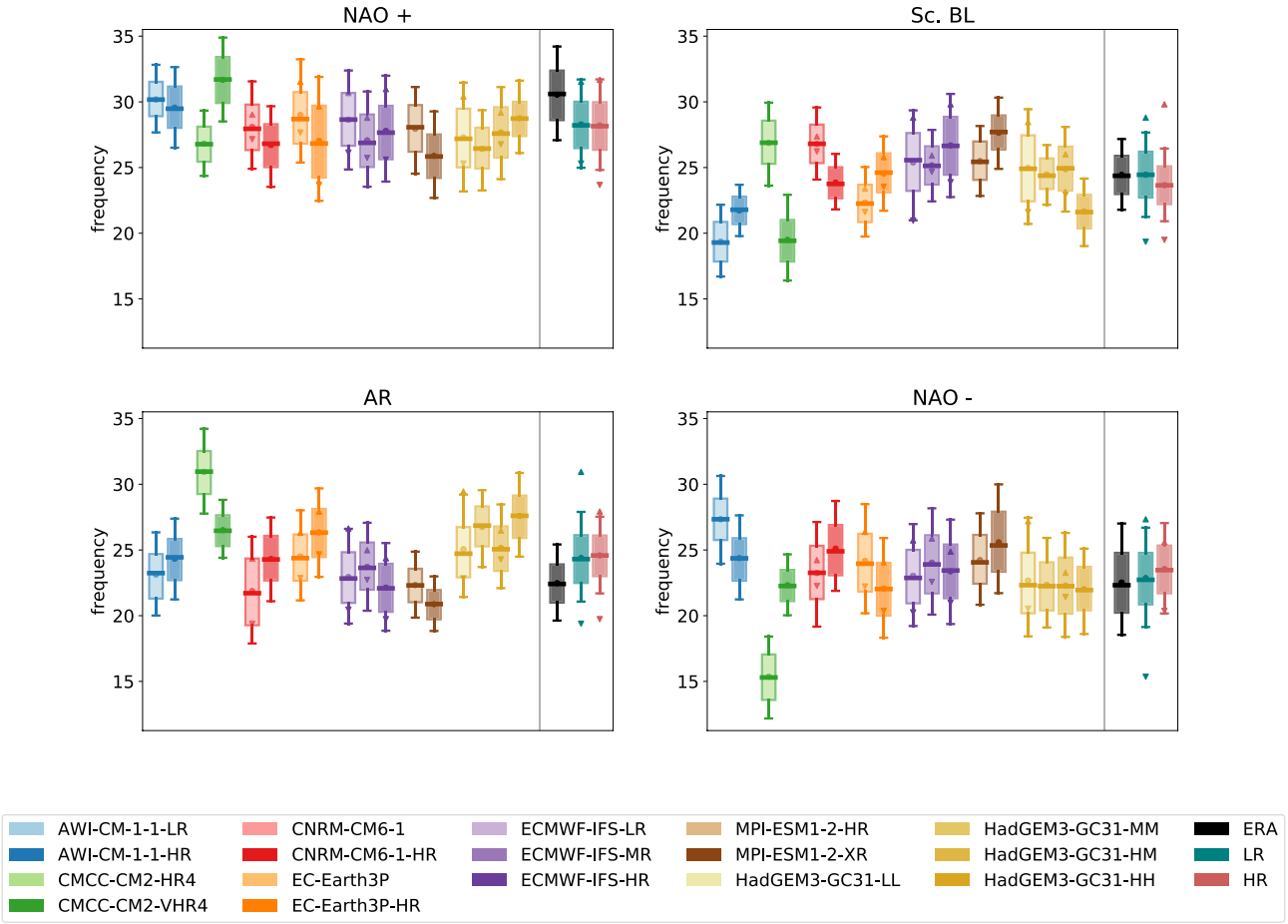


Fig. 7 Frequency of occurrence of the different regimes in models and observations. For details on the plot construction, refer to Fig. 2

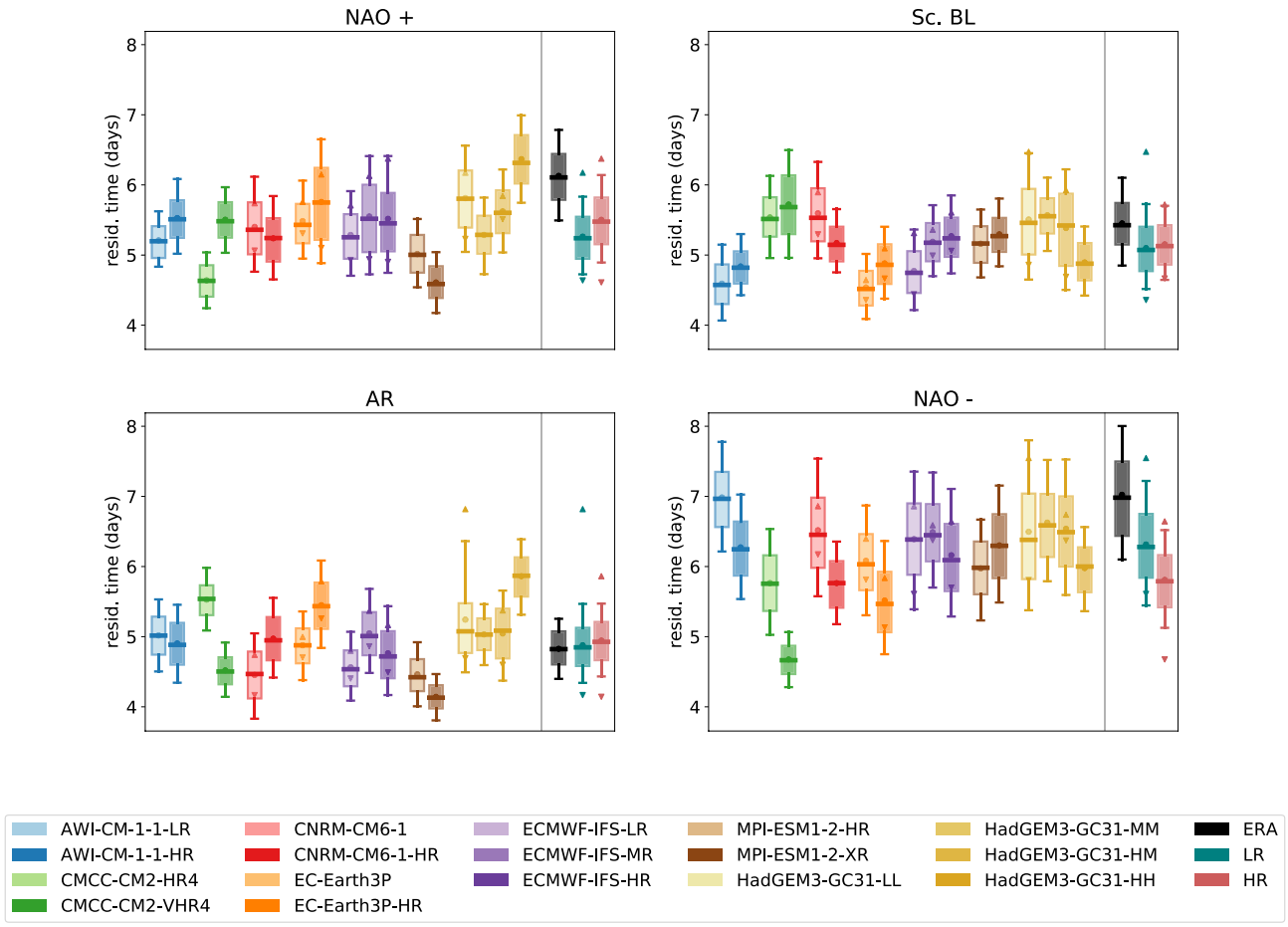


Fig. 8 Average regime duration in days. For details on the plot construction, refer to Fig. 2

an overall tendency to go towards the observed value. Strommen et al. (2019) found a consistent improvement of sharpness with increased resolution analyzing a set of AMIP simulations (3 models). Our average result across the LR and HR ensembles goes in the same direction, but the increase is not systematic, with a strong model-dependent footprint.

The projection of models' anomalies onto the reference EOFs—instead of the models' EOFs—(see Sect. 3.1 for details) might have an impact on the simulated regime structure. In particular, a positive (negative) bias can affect the variance ratio and sharpness metrics. This possibility has been tested by checking these two quantities using for each model the phase space spanned by their own EOFs. The results (see Figures S1 and S2 of the supplementary material) are consistent with those shown in Figs. 5 and 6. Although the values of both quantities are smaller when using the models' phase space and significant differences in sharpness can be seen for individual models, the general conclusions reported above remain valid.

4.4 Regime frequencies and persistence

The frequency of occurrence of Weather Regimes and their persistence are key quantities representing the dynamics of the system, and of primary importance for impact studies. Figure 7 shows the frequencies for all regimes, as obtained from the 30-yr bootstraps.

As we can see, the distributions of simulated regime frequencies overlap at least partly with the observed one, though some departures from the observations can be seen. For the NAO+ regime, a systematic underestimation of the frequency is found in almost all models, which are unable to catch frequencies in the upper half of the observed distribution. On the other side, most models overestimate the frequency of the AR regime, although the inter-model variability is much larger in this case. The frequency of NAO- and SBL regimes are better caught by the models, at least by the LR and HR averages. No systematic shift of the frequencies of the NAO+ and AR regimes is seen with increased resolution. This is clearly shown by the LR and HR averages that are remarkably

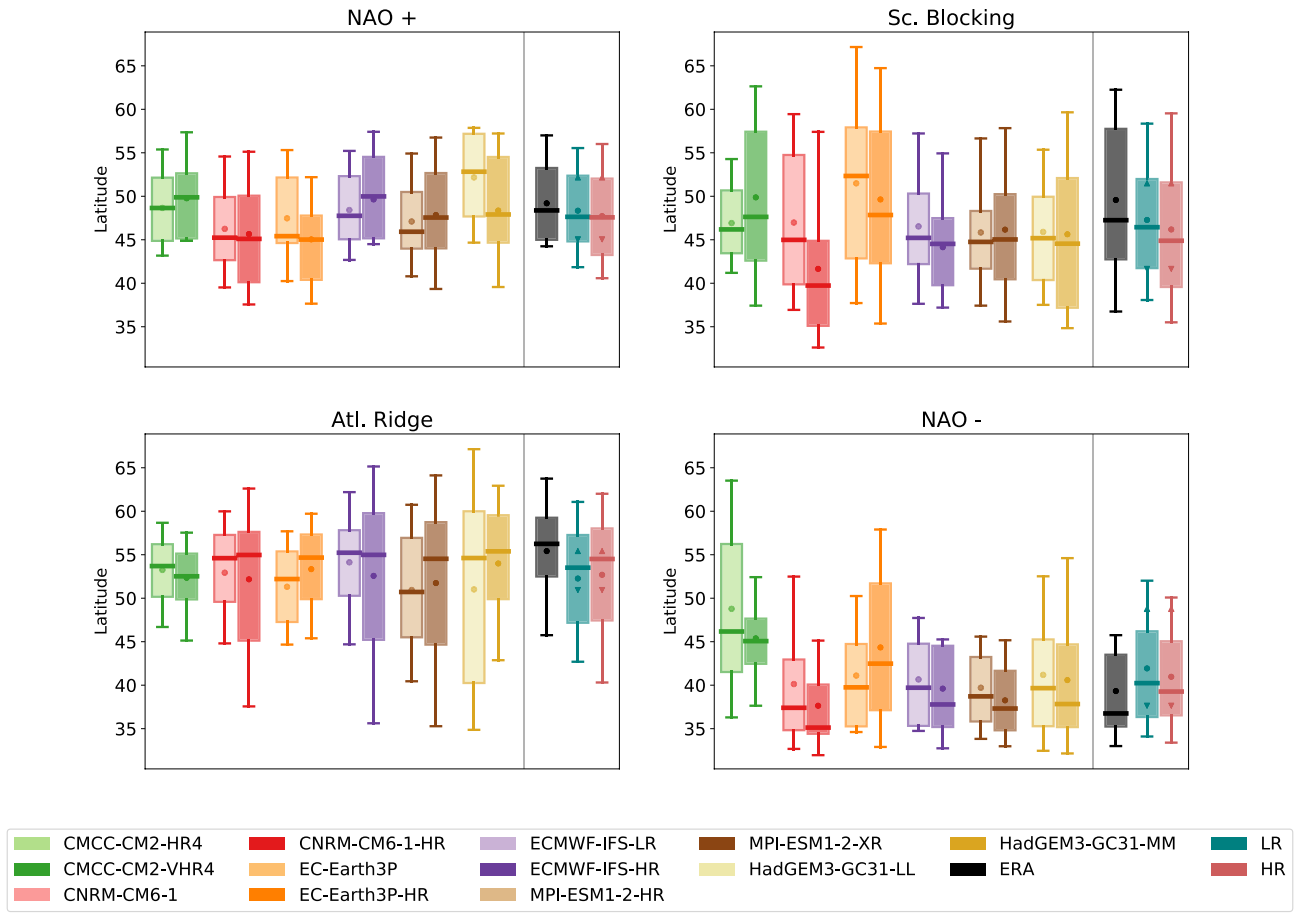


Fig. 9 Jet latitude distribution corresponding to each Weather Regime. Unlike Figs. 2, 3, 4, 5, 6, 7, 8, the distributions are simply the daily jet latitudes corresponding to each simulation, no bootstrapping is involved. The two ensemble averages are calculated as before, as the average of the median and percentiles over the ensemble

stable and have a similar range of variability. A small positive (negative) shift of the frequency distribution is seen on average for the NAO- (SBL) regime, however the different models do not behave consistently in this case.

Figure 8 shows the average regime duration in days. For both the NAO regimes, most models systematically underestimate the average regime duration. This is true on average also for the SBL regime, although some models are able to reach the observed range. The same results are obtained if the analysis is performed in terms of regime persistence probability (not shown). It is worth noting that the observed range of variability is significantly larger for the NAO-regime, meaning that there is a large interannual variability on the duration of this regime. The behaviour of the models with increased resolution gives a contrasting result: we found a higher persistence of the NAO+ regime in most models and a lower persistence—in all models but one—for the NAO-regime.

4.5 Weather Regimes and jet latitude

The link between the WRs and the peak latitude of the Jet Stream in the North-Atlantic is analyzed here in terms of the Jet latitude distributions corresponding to each regime. As assessed in Madonna et al. (2017), the observed WRs are characterized by specific shifts of the Jet Stream. NAO+, Atlantic Ridge and NAO- see prevailing central, northern and southern jets. The situation is different for Scandinavian Blocking, where the jet is usually tilted from South to North and this produces a much broader distribution of jet latitudes in the 1D framework.

Figure 9 shows the results of the same analysis for the models analyzed here. Only the models analyzed in the paper by Athanasiadis et al. (in prep.) are shown. As can be seen, the models usually do quite a good job for NAO+, where the two ensemble averages are barely distinguishable from the observed one. For AR and NAO-, the models are

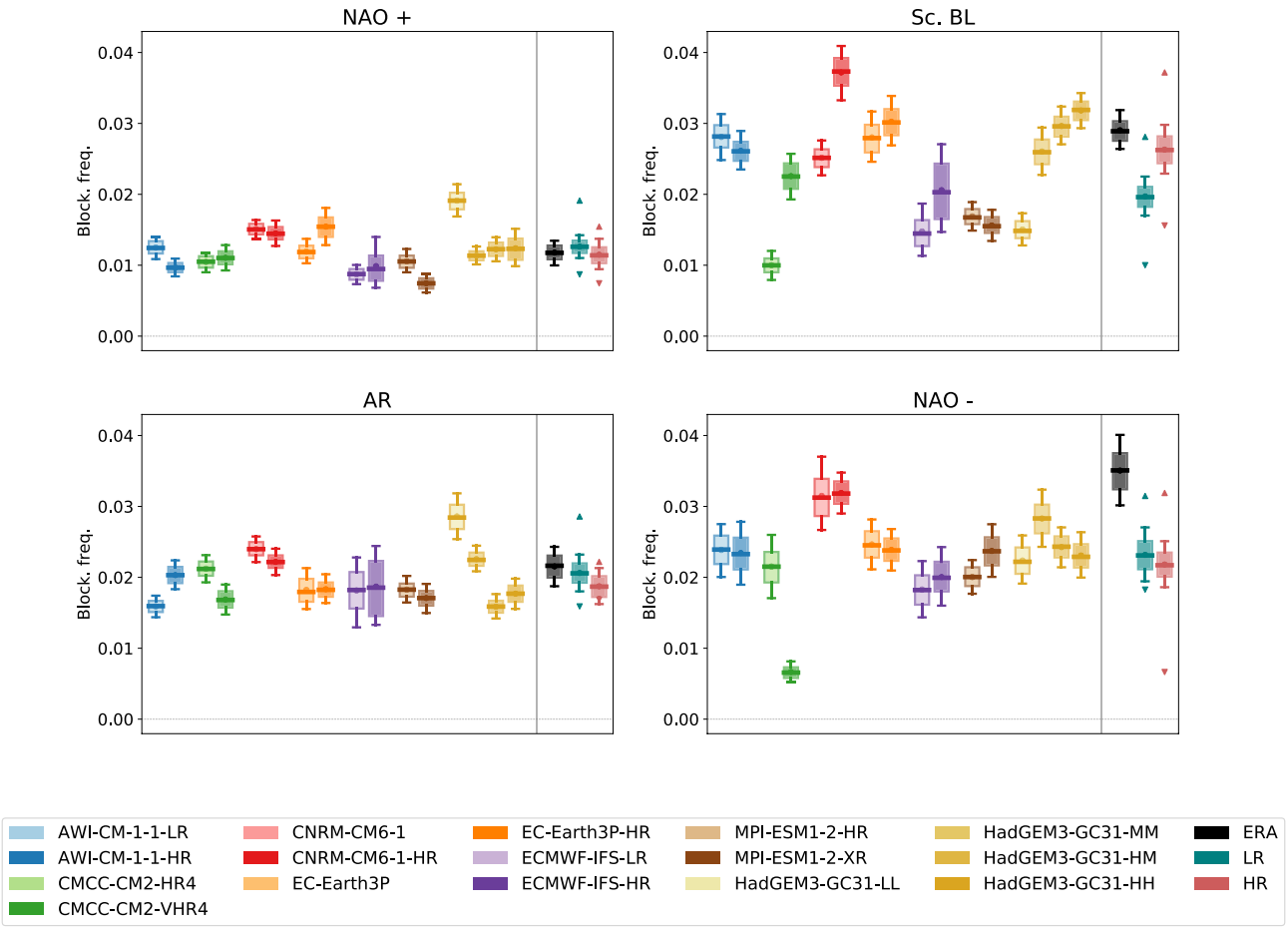


Fig. 10 The box plots show the average of the blocking frequency over the EAT sector corresponding to a certain WR (fraction of blocked days per grid point). For details on the plot construction, refer to Fig. 2

still limited in reaching the displacement seen in the real jet: for both regimes, they tend to produce too many central jets. The ensemble averages show in both cases a small improvement with increased resolution. In the case of SBL, a significant bias exists in the spread of the distributions, which tend to be confined to central latitudes for most models and systematically miss the observed northward extension.

4.6 Weather Regimes and blocking frequency

This section aims at assessing whether the models are able to reproduce the link between WRs and blocking frequency that we observe in the reanalysis. An overall assessment of the blocking performance of the PRIMAVERA models is discussed in Schiemann et al. (2020). Figure 10 shows the mean blocking frequency over the North-Atlantic sector corresponding to each regime. Numbers indicate the fraction of blocked days per grid point. For the NAO+ and AR regimes, models have a general good agreement with the observations. The model deficiencies are mostly seen for the other two regimes: both for NAO- and SBL the blocking frequency is too low in models. For the SBL, the increased resolution has a positive impact, increasing the blocking frequency in most models. On the other side, no clear effect is seen for NAO-.

5 Possible drivers of model performance

In the previous section we analyzed the results obtained for each metric both for the individual models and for the two HR and LR ensembles. An obvious question is whether it is possible to attribute the varying model performances to some simple, underlying drivers. To do this, a systematic computation of correlations between model performance and some plausible drivers has been performed. The underlying hypothesis is that the effect of these drivers might have a notable (linear) impact on the metrics, giving a signal beyond the "noise" due to other model-specific features. Concretely, we correlated all the metrics in Sect. 4 with the following quantities:

- Atmospheric/Oceanic model grid spacing;
- Model performance with respect to observations in:
  - mean SST in the North-Atlantic;
  - mean geopotential height at 500 hPa;
  - low frequency variability of the geopotential at 500 hPa (i.e. 2D standard deviation at periods longer than 5 days);
  - stationary eddies of the geopotential at 500 hPa (i.e. the time-mean departure from the zonal average);
  - blocking frequency;
- Jet latitude variability (i.e. the difference between the 90th and 10th percentiles of the jet latitude distribution).

To compute the correlations with the atmospheric/oceanic resolution we used the nominal horizontal grid spacing in km from Table 1, so that a negative correlation means that the metric increases with increasing resolution (i.e. with smaller grid spacing). For mean SST, mean geopotential, low frequency variability, stationary eddies and blocking frequency we used the pattern correlation between the observed and simulated pattern inside the EAT sector. The results are shown in Table 2, for three metrics: sharpness, variance ratio and the mean pattern correlation between the reference and simulated regimes. Bold values indicate that the p-value is lower than 0.05. For each driver/metric couple we show two values: the first is calculated for all models, the second excluding the CMCC model, which has a significantly lower number of vertical atmospheric levels (see Table 1). This was done because that model is a consistent outlier for the three metrics considered (see Figs. 3, 5 and 6) and for some of the drivers (mean geopot. field, low fr. var., stat. eddies, blocking frequency), and thus might skew correlations. In fact, computing correlations with and without CMCC produced notably different numbers in some cases. We also tried removing all outlier models separately for each metric (i.e. those lying further than 2 sigma from the multi-model average) and got very similar results.

At a first look, the sharpness is correlated with many of the proposed drivers. However, most of these correlations are spuriously produced by the outlier models and disappear or are strongly reduced when they are removed. After the filtering, there is a residual significant negative correlation of sharpness with the atmospheric grid spacing and a smaller one (non significant) with the oceanic grid spacing. This goes in the direction commented in Sect. 4, although the correlations are not very high, implying that other factors might be at play at the same time. Also, this is in line with the results by Strommen et al. (2019), which observed a systematic increase of sharpness with higher atmospheric resolution.

The agreement with the observed WR patterns shows only small correlations with the proposed drivers. The only consistent correlations before and after the filtering are found for the atmospheric grid spacing (negative, quite small). This is in line with the conclusions in Sect. 4, although the differences between the models are also playing a role here.

The variance ratio shows the most robust and interesting correlations. A small positive correlation with the mean SST pattern in the North-Atlantic becomes significant after the filtering. This suggests that models which have SSTs closer to the observed ones tend to produce more evident and well-defined regimes. Another important driver of high variance ratio is the agreement with the observed mean geopotential field and the pattern of stationary eddies. These correlations are strong both before and after the filtering. The positive correlations with the blocking frequency, low frequency variability and jet variability confirm how tightly related the different perspectives are.

6 Discussion and conclusions

Among all the metrics analyzed in Sect. 4, many show clear model biases. With "clear" we mean here that we can distinguish between models and observation just looking at some specific metrics. Some of these biases decrease in

Table 2 Pearson correlation between all possible drivers considered (rows) and three WR metrics (columns)

	Sharpness		Variance ratio		Mean WR pattern	
Atm. grid spacing	− 0.23	**− 0.54**	0.07	− 0.10	− 0.21	− 0.39
Oce. grid spacing	− 0.11	− 0.42	0.09	− 0.15	0.07	− 0.12
SST pattern	− 0.06	0.09	0.26	**0.64**	0.00	0.18
Mean geopot.	**0.63**	0.23	**0.91**	**0.81**	0.45	− 0.10
Low freq. variability	**0.66**	0.09	**0.92**	**0.77**	**0.56**	− 0.04
Stationary eddies	**0.68**	0.35	**0.83**	**0.64**	0.39	− 0.26
Blocking pattern	**0.71**	0.24	**0.93**	**0.80**	0.50	− 0.25
Jet variability	**0.61**	− 0.28	**0.89**	**0.60**	**0.72**	0.12

The value in the left column is calculated for all models, the right one excluding the CMCC model (more comments in text). Significant correlations are shown in bold font

the higher resolution versions of the models, while others seem to be insensitive to the model resolution. We will try in this Section to assess the most important model biases and which metrics most benefit from the increased horizontal resolution.

Overall, state-of-the-art coupled climate models are better at reproducing the observed North-Atlantic WRs mean patterns than their predecessors and there is some indication that this is at least partly due to increased model resolution. Dawson et al. (2012) saw a clear improvement when passing from T159 to T1279, in their single model study. Dawson and Palmer (2015) suggested that there is already a substantial improvement at T511 with respect to lower resolutions. In agreement with those studies, we also see an improvement for most models and for all regimes apart the NAO-. On the other side, Strommen et al. (2019) saw no improvement or even a slight degradation in the regimes patterns. However, these works (Dawson et al. 2012; Dawson and Palmer 2015; Strommen et al. 2019) were based on the analysis of AMIP simulations, so the results might not be directly comparable to ours. Also Cattiaux et al. (2013) saw an improvement in the WR patterns, but only between the lowest resolution (96 $\times$ 71) and all the others (from 96 $\times$ 96 to 192 $\times$ 142). For this reason, they claim there is a threshold resolution for such an improvement. The question remains of how far we want to push the goal for the simulated WR patterns. When looking at Figs. 2, 3 and 4 it is clear that the bootstrap estimate of ERA variability is much closer to zero (or 1 in the case of pattern correlation) than the models are. However, the goal for the models might be put somewhat further from the perfect match, taking into consideration that the observed internal variability might be larger than that estimated on the basis of the last 60 years. Nevertheless, some models perform better than the ensemble averages, so there seems to be still room for improvement, although this is probably not going to be obtained through increasing the horizontal resolution only.

As for the regime frequencies, we observed a systematic underestimation of the NAO+ occurrence, with a smaller overestimation of the AR regime. A systematic underestimation in the occurrence of the NAO+ was also seen in the single-model analysis by Cattiaux et al. (2013), although their model had a larger positive bias for NAO- and negative for SBL. No effect on the NAO+ frequency bias is obtained with the increased resolution, but the bias is clear and needs to be tackled in future models. We also see a systematic tendency to underestimate the NAO- and NAO+ persistence, which is in line with previous works (Strommen et al. 2019; Matsueda and Palmer 2018). We see an improvement with HR for NAO+ only and a worsening for NAO- with the increased resolution.

The quantities that most clearly distinguish between models and observation are the variance ratio and sharpness. As assessed in Sect. 4, no model is able to reach the observed value, though some of them get very close to it. For both quantities, we see a depopulation of the low-value tail of the models and a reduction in the model bias with increased resolution. However this is not systematic across the ensemble and the effect is model-dependent. A stronger signal was originally expected, in line with what Strommen et al. (2019) saw for three models. Again, the difference might be that the models are run in coupled mode here and not atmosphere-only as in Strommen et al. (2019). As analysed in Sect. 5, there seems to be some influence of the SST bias on the variance ratio, which would then tend to be higher in AMIP runs. However, we do not see an analogous correlation for sharpness, which is also related to the data autocorrelation at one day lag. Also, some role might be played by model tuning: a choice was made in the PRIMAVERA project, not to tune the increased resolution versions of the models. On one side, this assures that no other parameter changed apart from the model resolution; on the other, the increased resolution versions might be less equilibrated than the nominal ones, thus producing a contrasted output. The potentially negative effect of model tuning is expected to be stronger in the case of coupled simulations than for AMIP runs.

The correlation analysis in Sect. 5 gives some additional hints, although the signal across the ensemble may be somewhat suppressed by other model-specific features. The horizontal atmospheric resolution has a significant correlation with sharpness and a small correlation with the WR patterns, that are encouraging if summed to the results of Sect. 4. However, the most interesting metric is probably the variance ratio, which is a good candidate for a synthetic indicator of the model performance in reproducing WRs. Indeed, the variance ratio shows a robust correlation with the low frequency variability and blocking patterns, and also with the jet latitude variability, indicating the strong connection between these perspectives. On the other side, the variance ratio is determined to a certain extent by the model bias in the SSTs and the mean geopotential field over the North-Atlantic. No correlation is found between the variance ratio and the atmospheric resolution, although most models actually improve when increasing the resolution (see Sect. 4.3).

The general picture that appears is that the model performance in reproducing the wintertime Euro-Atlantic WRs tends to improve by increasing the models' horizontal resolution, regarding WR patterns, sharpness and variance ratio. No clear effect is seen on the regime frequency and persistence. However, the model horizontal resolution is clearly not the only key and the real picture is very complicated. Reducing the model biases in the mean geopotential and SST fields might have an effect on the variance ratio and hence on the regime structure. It would be very interesting in this sense to have more insight in the role of model tuning,

aside that of increased resolution, and to repeat the present analysis on a tuned high-res version of the models.

Acknowledgements The authors acknowledge support by the PRIMAVERA project of the Horizon 2020 Research Programme, funded by the European Commission under Grant Agreement 641727. The climate model simulations used in this study were performed under the PRIMAVERA project and can be accessed at the archive of the Centre for Environmental Data Analysis (CEDA). The research of HMC was supported by Natural Environment Research Council Grant number NE/P018238/1. FF wishes to thank Irene Mavilia for the development of the first version of WRtool.

References

Anstey JA, Davini P, Gray LJ, Woollings TJ, Butchart N, Cagnazzo C, Christiansen B, Hardiman SC, Osprey SM, Yang S (2013) Multi-model analysis of northern hemisphere winter blocking: model biases and the role of resolution. J Geophys Res Atmos 118(10):3956–3971

Athanasiadis P, Baker A (in prep.) Increasing model resolution, sst biases and the representation of north atlantic eddy-driven jet variability in primavera historical simulations

Athanasiadis PJ, Wallace JM, Wettstein JJ (2010) Patterns of wintertime jet stream variability and their relation to the storm tracks. J Atmos Sci 67(5):1361–1381

Blackmon ML, Lee Y, Wallace JM (1984) Horizontal structure of 500 mb height fluctuations with long, intermediate and short time scales. J Atmos Sci 41(6):961–980

Cassou C (2008) Intraseasonal interaction between the madden-julian oscillation and the north atlantic oscillation. Nature 455(7212):523

Cattiaux J, Douville H, Peings Y (2013) European temperatures in cmip5: origins of present-day biases and future uncertainties. Clim Dyn 41(11–12):2889–2907

Cherchi A, Fogli PG, Lovato T, Peano D, Iovino D, Gualdi S, Masina S, Scoccimarro E, Materia S, Bellucci A, Navarra A (2019) Global mean climate and main patterns of variability in the cmcc-cm2 coupled model. Journal of Advances in Modeling Earth Systems 11(1):185–209. https://doi.org/10.1029/2018MS001369https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001369

Christensen H, Moroz I, Palmer T (2015) Simulating weather regimes: Impact of stochastic and perturbed parameter schemes in a simple atmospheric model. Clim Dyn 44(7–8):2195–2214

Corti S, Molteni F, Palmer T (1999) Signature of recent climate change in frequencies of natural atmospheric circulation regimes. Nature 398(6730):799

Davini P, D'Andrea F (2016) Northern hemisphere atmospheric blocking representation in global climate models: twenty years of improvements? J Clim 29(24):8823–8840

Davini P, Cagnazzo C, Gualdi S, Navarra A (2012) Bidimensional diagnostics, variability, and trends of northern hemisphere blocking. J Clim 25(19):6496–6509

Davini P, Corti S, D'Andrea F, Rivière G, von Hardenberg J (2017) Improved winter european atmospheric blocking frequencies in high-resolution global climate simulations. J Adv Model Earth Syst 9(7):2615–2634

Dawson A, Palmer T (2015) Simulating weather regimes: Impact of model resolution and stochastic parameterization. Clim Dyn 44(7–8):2177–2193

Dawson A, Palmer T, Corti S (2012) Simulating regime structures in weather and climate prediction models. Geophys Res Lett 39(21):20

Franzke C, Woollings T, Martius O (2011) Persistent circulation regimes and preferred regime transitions in the north atlantic. J Atmos Sci 68(12):2809–2825

Grams CM, Beerli R, Pfenninger S, Staffell I, Wernli H (2017) Balancing europe's wind-power output through spatial deployment informed by weather regimes. Nat Clim Change 7(8):557

Gutjahr O, Putrasahan D, Lohmann K, Jungclaus JH, von Storch JS, Brüggemann N, Haak H, Stössel A (2019) Max planck institute earth system model (mpi-esm1.2) for the high-resolution model intercomparison project (highresmip). Geosci Model Dev 12(7):3241–3281, https://doi.org/10.5194/gmd-12-3241-2019, https://www.geosci-model-dev.net/12/3241/2019/

Haarsma R, Acosta M, Bakhshi R, Bretonnière PAB, Caron LP, Castrillo M, Corti S, Davini P, Exarchou E, Fabiano F, Fladrich U, Fuentes Franco R, García-Serrano J, von Hardenberg J, Koenigk T, Levine X, Meccia V, van Noije T, van den Oord G, Palmeiro FM, Rodrigo M, Ruprich-Robert Y, Le Sager P, Tourigny E, Wang S, van Weele M, Wyser K (2020) Highresmip versions of ec-earth: Ec-earth3p and ec-earth3p-hr. description, model performance, data handling and validation. Geoscientific Model Development Discussions 2020:1–37. https://doi.org/10.5194/gmd-2019-350, https://www.geosci-model-dev-discuss.net/gmd-2019-350/

Haarsma RJ, Roberts MJ, Vidale PL, Senior CA, Bellucci A, Bao Q, Chang P, Corti S, Fučkar NS, Guemas V et al (2016) High resolution model intercomparison project (highresmip v1. 0) for cmip6. Geosci Model Dev 9(11):4185–4208

Hannachi A, Barnes EA, Woollings T (2013) Behaviour of the winter north atlantic eddy-driven jet stream in the cmip3 integrations. Clim Dyn 41(3–4):995–1007

Hannachi A, Straus DM, Franzke CL, Corti S, Woollings T (2017) Low-frequency nonlinearity and regime behavior in the northern hemisphere extratropical atmosphere. Rev Geophys 55(1):199–234

Hoskins B, Woollings T (2015) Persistent extratropical regimes and climate extremes. Curr Clim Change Rep 1(3):115–124

Iqbal W, Leung WN, Hannachi A (2018) Analysis of the variability of the north atlantic eddy-driven jet stream in cmip5. Clim Dyn 51(1–2):235–247

Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22(1):79–86

Madonna E, Li C, Grams CM, Woollings T (2017) The link between eddy-driven jet variability and weather regimes in the north atlantic-european sector. Quart J R Meteorol Soc 143(708):2960–2972

Masato G, Hoskins BJ, Woollings T (2013) Winter and summer northern hemisphere blocking in cmip5 models. J Clim 26(18):7044–7059

Matsueda M, Palmer T (2018) Estimates of flow-dependent predictability of wintertime euro-atlantic weather regimes in medium-range forecasts. Quart J R Meteorol Soc 144(713):1012–1027

Michelangeli PA, Vautard R, Legras B (1995) Weather regimes: recurrence and quasi stationarity. J Atmos Sci 52(8):1237–1256

Palmer TN (1999) A nonlinear dynamical perspective on climate prediction. J Clim 12(2):575–591

Plaut G, Simonnet E (2001) Large-scale circulation classification, weather regimes, and local climate over france, the alps and western europe. Clim Res 17(3):303–324

Raymond F, Ullmann A, Camberlin P, Oueslati B, Drobinski P (2018) Atmospheric conditions and weather regimes associated with extreme winter dry spells over the mediterranean basin. Clim Dyn 50(11–12):4437–4453

Roberts CD, Senan R, Molteni F, Boussetta S, Mayer M, Keeley SPE (2018a) Climate model configurations of the ecmwf integrated forecasting system (ecmwf-ifs cycle 43r1) for highresmip. Geosci Model Dev 11(9):3681–3712. https://doi.org/10.5194/gmd-11-3681-2018, https://www.geosci-model-dev.net/11/3681/2018/

Roberts M, Vidale P, Senior C, Hewitt H, Bates C, Berthou S, Chang P, Christensen H, Danilov S, Demory ME et al (2018b) The benefits of global high resolution for climate simulation: Process understanding and the enabling of stakeholder decisions at the regional scale. Bull Am Meteorol Soc 99(11):2341–2359

Roller CD, Qian JH, Agel L, Barlow M, Moron V (2016) Winter weather regimes in the northeast united states. J Clim 29(8):2963–2980

Scherrer SC, Croci-Maspoli M, Schwierz C, Appenzeller C (2006) Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the euro-atlantic region. Int J Climatol 26(2):233–249

Schiemann R, Demory ME, Shaffrey LC, Strachan J, Vidale PL, Mizielinski MS, Roberts MJ, Matsueda M, Wehner MF, Jung T (2017) The resolution sensitivity of northern hemisphere blocking in four 25-km atmospheric global circulation models. J Clim 30(1):337–358

Schiemann R, Athanasiadis P, Barriopedro D, Doblas-Reyes F, Lohmann K, Roberts MJ, Sein D, Roberts CD, Terray L, Vidale PL (2020) The representation of Northern Hemisphere blocking in current global climate models. Weather Clim Dynam Discuss. https://doi.org/10.5194/wcd-2019-19

Sein DV, Koldunov NV, Danilov S, Wang Q, Sidorenko D, Fast I, Rackow T, Cabos W, Jung T (2017) Ocean modeling on a mesh with resolution following the local rossby radius. J Adv Model Earth Syst 9(7):2601–2614. https://doi.org/10.1002/2017MS001099, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001099

Straus D, Molteni F, Corti S (2017) Atmospheric regimes: The link between weather and the large scale circulation. Nonlinear Stochast Clim Dyn pp 105–135

Straus DM, Molteni F (2004) Circulation regimes and sst forcing: Results from large gcm ensembles. J Clim 17(8):1641–1656

Straus DM, Corti S, Molteni F (2007) Circulation regimes: Chaotic variability versus sst-forced predictability. J Clim 20(10):2251–2272

Strommen K, Mavilia I, Corti S, Matsueda M, Davini P, von Hardenberg J, Vidale PL, Mizuta R (2019) The sensitivity of euro-atlantic regimes to model horizontal resolution. Geophys Res Lett 46(13):7810–7818. https://doi.org/10.1029/2019GL082843, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL082843

Tibaldi S, Molteni F (1990) On the operational predictability of blocking. Tellus A 42(3):343–365

Vautard R (1990) Multiple weather regimes over the north atlantic: Analysis of precursors and successors. Mon Weather Rev 118(10):2056–2081

Voldoire A, Saint-Martin D, Sénési S, Decharme B, Alias A, Chevallier M, Colin J, Guérémy JF, Michou M, Moine MP, Nabat P, Roehrig R, Salas y Mélia D, Séférian R, Valcke S, Beau I, Belamari S, Berthet S, Cassou C, Cattiaux J, Deshayes J, Douville H, Ethé C, Franchistéguy L, Geoffroy O, Lévy C, Madec G, Meurdesoif Y, Msadek R, Ribes A, Sanchez-Gomez E, Terray L, Waldman R, (2019) Evaluation of cmip6 deck experiments with cnrm-cm6-1. J Adv Model Earth Syst 11(7):2177–2213. https://doi.org/10.1029/2019MS001683, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001683

Williams KD, Copsey D, Blockley EW, Bodas-Salcedo A, Calvert D, Comer R, Davis P, Graham T, Hewitt HT, Hill R, Hyder P, Ineson S, Johns TC, Keen AB, Lee RW, Megann A, Milton SF, Rae JGL, Roberts MJ, Scaife AA, Schiemann R, Storkey D, Thorpe L, Watterson IG, Walters DN, West A, Wood RA, Woollings T, Xavier PK (2018) The met office global coupled model 3.0 and 3.1 (gc3.0 and gc3.1) configurations. Journal of Advances in Modeling Earth Systems 10(2):357–380, https://doi.org/10.1002/2017MS001115, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001115

Woollings T, Hannachi A, Hoskins B (2010) Variability of the north atlantic eddy-driven jet stream. Quart J R Meteorol Soc 136(649):856–868

Yiou P, Nogaj M (2004) Extreme climatic events and weather regimes over the north atlantic: When and where? Geophys Res Lett 31(7):20

Zampieri M, Toreti A, Schindler A, Scoccimarro E, Gualdi S (2017) Atlantic multi-decadal oscillation influence on weather regimes over europe and the mediterranean in spring and summer. Global Planet Change 151:92–100