How well does CMIP6 capture the dynamics of Euro-Atlantic weather regimes, and why?

Joshua Dorrington[1,*], Kristian Strommen[1,*], and Federico Fabiano[2]

[1]Department of Atmospheric, Oceanic, and Planetary Physics, University of Oxford, UK
[2]Institute of Atmospheric Sciences and Climate (ISAC-CNR), Bologna, Italy
[*]These authors contributed equally to this work

Correspondence: Joshua Dorrington (joshua.dorrington@physics.ox.ac.uk)

Abstract. Even the most advanced climate models struggle to reproduce the observed wintertime circulation of the atmosphere over the North Atlantic and Western Europe. During winter, this particularly challenging region is dominated by eddy-driven and highly non-linear flows, which are often studied from the perspective of regimes - a small number of qualitatively distinct atmospheric states. Poor representation of regimes associated with persistent atmospheric blocking events, or variations in jet

5   latitude, degrade the ability of models to correctly simulate extreme events. In this paper we leverage a recently developed hybrid approach - which combines both jet and geopotential height data - to assess the representation of regimes in 8,400 years of historical climate simulations drawn from CMIP6, CMIP5 and HighResMip. We show that these geopotential-jet regimes are particularly suited to the analysis of climate data, with considerable reductions in sampling variability compared to classical regime approaches. We find that CMIP6 has a considerably improved spatial regime structure, and a more trimodal eddy-driven

10   jet, relative to CMIP5, but still struggles with underpersistent regimes, and too little European blocking, when compared to reanalysis. Reduced regime persistence can be understood, at least in part, as a result of jets that are too fast and eddy feedbacks on the jet stream that are too weak - structural errors that do not noticeably improve in higher resolution models.

1   Introduction

The wintertime atmospheric circulation in the Euro-Atlantic region is highly complex, featuring non-linear Rossby wave break-

15   ing (Woollings et al., 2008), complex orography (White et al., 2019), variability in the position, speed and angle of the low-level jet, and external forcings from ocean heat fluxes (Delworth and Zeng, 2016; Delworth et al., 2017), Arctic sea ice (Barnes and Screen, 2015), stratospheric signals (Domeisen et al., 2020) and tropical teleconnections (Rodríguez-Fonseca et al., 2016; Jiménez-Esteve and Domeisen, 2018). In the absence of a strong tropical jet, as in the North Pacific, the importance of eddy variability is amplified (Eichelberger and Hartmann, 2007; Lee and Kim, 2003), and the meridional tilt of the storm track,

20   induced by the precise details of the regional stationary wave pattern, adds further complexity (Brayshaw et al., 2009; Gerber and Vallis, 2009).

A powerful approach that is well-suited for understanding and characterising the fundamentally non-linear behaviour of this region is to think in terms of regime dynamics - a discretisation of the continuous atmospheric state into a small number of qualitatively distinct flow patterns. There are two main and complementary perspectives on Euro-Atlantic regimes in the

25   literature: one based on variations in the latitude of the eddy-driven jet stream (Woollings et al., 2010; Franzke et al., 2011; Madonna et al., 2017) and another aiming to capture the circulation structure of persistent anticyclonic blocks in the field of geopotential height anomalies (Michelangeli et al., 1995; Fabiano et al., 2020). Climate models have historically struggled to represent these non-linear variability modes (Anstey et al., 2013; Davini and D'Andrea, 2016), and these biases unfortunately persist in the latest generation of climate models, as captured by the 6th round of the Climate Model Intercomparison Project

30   (CMIP6). CMIP6 struggles to produce sufficiently long-lived blocking events (Davini and D'Andrea, 2020), as also seen from a regime perspective in Fabiano et al. (2020) where NAO- events are insufficiently persistent, and where a model bias towards too frequent Atlantic ridging is seen. There is also a continued bias towards an overly zonal jet stream (Oudar et al., 2020) in CMIP6, although with notable improvements in jet structure and storm track representation relative to earlier models (Harvey et al., 2020). Understanding how the changes in jet and blocking regime structure relate to each other is an important aspect

35   of correctly resolving the North Atlantic circulation, yet little work has been done in a multimodel context to understand how model errors in these different components of the circulation relate to one another.

However, as pointed out in Dorrington and Strommen (2020) (hereafter DS20), the identification of robust circulation regimes via K-means clustering of, e.g., 500 hPa geopotential height anomalies, is fraught with complications, due to low statistical significance of multimodality, substantial sampling variability in the exact regime patterns obtained, and difficulties

40   in choosing the precise number of regimes to use. In particular, it is not possible to have confidence that the identified regime patterns remain invariant across different time periods. Since the time-invariance of regime patterns is assumed - often implicitly - by many applications of regimes to both predictability (Strommen et al., 2019a) and, crucially, climate change (Corti et al., 1999), this has seriously complicated the practical use of regimes to date.

This is in stark contrast to the regime behaviour of the eddy-driven North-Atlantic jet stream, which features visually obvious

45   trimodality in its daily latitude distribution, with peaks robustly centred at particular latitudes. DS20 posited that the reason the strong trimodality of the jet latitude perspective doesn't result in an unambiguous multimodality in the Z500 phase space is partly due to the confounding influence of the *longitudinal* variability of the jet, as measured by its speed. Indeed, the jet speed is distributed unimodally, experiences considerable interdecadal variability (Woollings et al., 2014), and is largely decoupled from the jet latitude, with the two quantities being uncorrelated on daily time scales (cf. Supplementary Figure 1).

50   Several studies have highlighted the different nature of the latitudinal and longitudinal variability of the jet. Besides its large decadal variability, the jet speed appears to be mostly unpredictable on seasonal time scales, unlike the jet latitude (Parker et al., 2019; Strommen, 2020). The two quantities also respond differently to thermal forcing (Baker et al., 2017), implying their response to climate change may be very different. As such, it is natural to want to treat these two quantities separately, and indeed this approach is often adopted (Woollings and Blackburn, 2012; Barnes and Polvani, 2013b). DS20 essentially

55   argued that this separation should also be used in the context of regime analysis, and that not doing so is the source of the aforementioned complications (lack of time invariant patterns, low statistical significance, and so on).

Concretely, by regressing out the influence of jet speed from the principal components of Z500 prior to K-means clustering, DS20 found three highly stable regimes in the ERA20C reanalysis associated with anticyclonic blocking states and that approximately corresponded to the preferred jet latitudes. This approach can be understood as decomposing the circulation (as
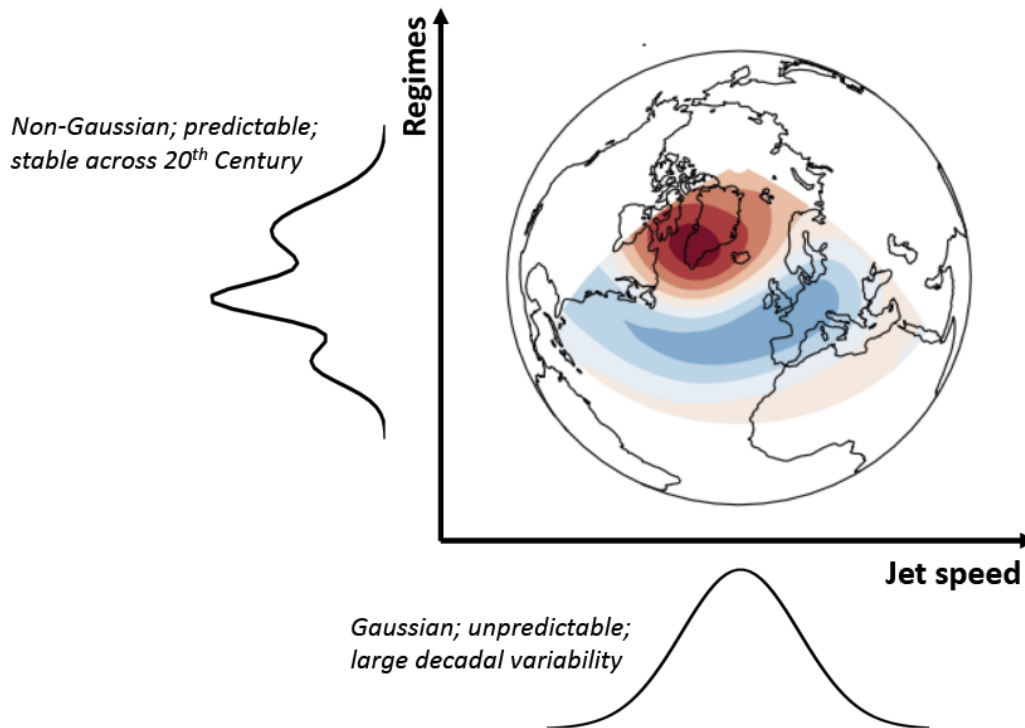
Figure 1. A schematic of the conceptual framework used in this paper. The Euro-Atlantic circulation, visualised here using the North Atlantic Oscillation (filled contours), is decomposed into orthogonal (i.e., uncorrelated) modes of jet variability: the longitudinal variability ('pulsing'), as measured by the Gaussian jet speed, and the latitudinal variability ('wobbling' or 'meandering'), as measured using non-Gaussian regimes. These two modes are then studied independently: the jet speed with linear methods (linear regression), and regimes with non-linear methods (changes to persistence and occurrence).

60   measured by Z500) into components influenced by the longitudinal and latitudinal variability of the jet, with the former being measured by the Gaussian jet speed and the latter by non-Gaussian regimes. A schematic of this framework is shown in Figure 1. Identifying a stable, reproducible regime framework, in which the sampling variability in regime patterns is negligible, allows the regime approach to be applied to the study of interdecadal variability, model comparison and climate change in a simplified way, with increased confidence that results obtained genuinely derive from features of the underlying circulation,

65   rather than instabilities in the clustering methodology.

    In this paper, we apply this new framework – which we now term as *geopotential-jet regimes* due to their hybrid nature - to analyse the regime representation in the historical ensemble of CMIP6 climate models, and characterise improvements over the earlier generation of CMIP5 models. We will demonstrate that geopotential-jet regimes are easily identified and very well represented in climate models, indicating their suitability for such analysis. In particular, they provide a comprehensive

70   assessment of historical regime variability, model bias, and the physical predictors of realistic regime representation within

Reanalysis name	Generating Centre	Year of production	Time period used	Observations included	Assimilating Model	Resolution (xy,z,t)
20CRv2	NOAA-CIRES	2011	1900-2010	surface observations	GFS (2008 experimental)	(210km, 28 levels, 6 hrs)
ERA20C	ECMWF	2016	1900-2010	surface pressure and marine wind observations	IFS Cy38r1	(125km, 91 levels, 3 hrs)
CERA20C	ECMWF	2018	1901-2010	surface, wind and ocean observations	IFS Cy41r2	(125km, 91 levels, 3 hrs)
ERA5	ECMWF	2020	1950-2010	surface, satellite and upper air observations	IFS Cy41r2	(31km, 137 levels, 1hr)
20CRv3	NOAA-CIRES-DOE	2021	1900-1980	surface observations	GFS 14.01	(60km, 64 levels, 6 hrs)

Table 1. A summary of the 5 reanalysis products analysed in this paper. For references to full descriptions of the products, see main text.

current state-of-the-art climate models. We will also show that they add further clarity and insight into earlier studies, such as how the latitudinal variability of the jet relates to its speed (Woollings et al., 2018).

In section 2 we describe the model data and variables we analyse, provide a full description of the metrics and regime identification methodology we use, and describe the physical metrics we have considered as possible explanatory variables for model regime representation. Section 3 analyses how realistically and reliably climate models are able to reproduce regime patterns found in reanalysis, and shows that the three geopotential-jet regimes found in DS20 are also optimal for the analysis of CMIP data. We also consider the relation between geopotential-jet regime and jet latitude regime representation. Section 4 looks at the occurrence and persistence of those regimes in models, and explores both mean-state biases and historical variability. Section 5 analyses the covariability in models' mean-state with regime statistics in order to gain insight into the causes of regime structure and model biases. Finally we summarise our key findings in section 6.

2 Data and methodology

2.1 Data

In order to assess the uncertainty in the historical record of regimes, we make use of 5 different reanalysis products. In the case of reanalyses with multiple ensemble members, we always use the first member. For each reanalysis we make use of boreal winter (DJF) geopotential height at 500 hPa (Z500) and zonal wind speed data at 850 hPa (U850), at a daily temporal resolution and on a 1 degree grid, from ERA20C (Poli et al., 2016), the extended ERA5 (Hersbach et al., 2020), CERA20C (Laloyaux et al., 2018), 20CRv2 (Compo et al., 2011), and 20CRv3 (Slivinski et al., 2021). All available data covering the time period 1900-2010, available at the time of writing, was used. A summary of each product is given in Table 1.

We analyse equivalent model geopotential height and wind speed data drawn from the 5th (CMIP5) and 6th (CMIP6) phases of the coupled model inter-comparison project: multi-centre ensembles of earth-system models representing the state-of-the-

art in global climate modelling in 2011 and 2020 respectively. We analyse the historical experiments for 31 single-member CMIP6 models (Eyring et al., 2016) detailed in SI table 1, and a total of 71 ensemble members from 28 distinct CMIP5 models (Taylor et al., 2012) detailed in SI table 3. These historical experiments consist of coupled uninitialised climate runs forced with historical greenhouse gas and aerosol forcings over the 20th century, after a spin-up from a free-running pre-industrial

95 control run.

In section 5 we also make use of model data produced as part of the PRIMAVERA project. These coupled simulations all follow the HighResMIP protocol (Haarsma et al., 2016), and are therefore initialised in 1950, following a short 50-year spin-up. The simulations span the 65 years between 1950 and 2015, and use the same historical forcings as CMIP6.

In total, we consider more than 8400 DJF seasons of climate model data, deriving from 128 ensemble integrations of 76

100 different climate models, originating from 18 independent modelling centres. We are therefore able to provide a comprehensive analysis of the current state-of-the-art in climate model regime representation.

2.2 Jet stream metrics

The structure of the low-level eddy-driven Atlantic jet can be summarised through the jet latitude and jet speed indices introduced in Woollings et al. (2010), which we here compute with the simplified methodology of Parker et al. (2019). In brief, the

105 jet speed index is defined as the maximum (oriented Eastward) of latitudinally averaged 850 hPa zonal wind speed, smoothed to remove synoptic fluctuations, over the Atlantic domain [0–60W,15-75N]. The jet latitude index is complementarily defined as the latitude at which the jet speed is maximum on a given day. The smoothing timescale varies in the literature, but here we apply a 5-day lowpass filter to the winds. Note that this methodology defines a unique latitude and speed for every day, and so cannot account for split jets or insignificant jet activity, and also does not contain any information on the meridional tilt of the

110 jet.

As the probability distribution of the jet latitude in reanalysis is trimodal, it is difficult to summarise deficiencies in model jet latitude distribution through simple summary statistics such as the mean or variance. In order to holistically quantify the error in the jet latitude and jet speed distributions in models, we make use of the Wasserstein distance. Also known as the "Earth-mover distance", this metric provides a natural way to compare two probability distributions, and can be understood

115 informally as a measure of how much of the probability density of one distribution must be shifted to transport it into another. Formally, the Wasserstein distance (as used here) is defined as the integrated difference between two distributions' cumulative density functions. This distance was introduced to the atmospheric literature in Ghil (2015) and has since been applied to the analysis of both simple climate attractors (Robin et al., 2017) and fully coupled climate models (Vissio et al., 2020), the latter of which uses it to evaluate model error, just as we do here. Like any distance measure, values close to zero indicate smaller

120 errors.

Finally, a simpler measure of how well a model captures the jet latitude distribution can be given by identifying how many peaks a given model distribution has (i.e., whether it is trimodal, bimodal or unimodal) and the location of these peaks. To do this objectively, we used the python algorithm scipy.signal.find_peaks to locate, for a given distribution, all peaks with a height of at least 0.01 which are separated from each other by at least 4 degrees. The numbers were chosen based on inspecting the

125 distribution for ERA5 and the CMIP6 multimodel mean. Note that this methodology excludes 'shoulders' from being classified as peaks.

2.3 Regime computation and metrics

Following well established methodology (Michelangeli et al., 1995; Cassou, 2008; Dawson et al., 2012), we identify atmospheric regimes by partitioning a low-dimensional truncated phase space using the K-means clustering algorithm. For all
130 datasets analysed, we take this space as the 4 leading principal components of the daily mean 500hPa geopotential height field (Z500) restricted to Boreal winter (DJF) and the Euro-Atlantic region [30-90N,80W-40E]. The K-means algorithm is non-deterministic as it requires a random-seed at initialisation, which can cause convergence to different local minima. In order to assure repeatability we perform each K-means clustering 100 times with different seeds, and use the result that maximises the inter-to-intracluster variance ratio (as defined in Straus and Molteni (2004)). The standard approach, is to apply K-means
135 clustering directly to this space and we refer to the regimes this approach produces as *Classical Circulation Regimes*.

As discussed in the introduction, DS20 argued that robust identification of non-linear regime structure in this Z500 phase space is confounded by the noisy, predominately linear, variability of the Euro-Atlantic jet speed. They showed that regressing out the influence of jet speed allowed more stable clusters to be identified. We refer to regimes produced by applying K-means to the residual principal components of Z500 - once the jet speed has been regressed out - as *Geopotential-Jet Regimes*. In all
140 cases we use the principal components of each dataset and the jet speed regression coefficients of each dataset to perform this analysis. While regimes are identified in the residual phase space, when analysing spatial circulation patterns we return to the full Z500 space over the Euro-Atlantic domain, by compositing the fields for all days assigned to a given regime.

At many points in this paper we wish to compare the regime composites identified in different datasets, or in different subsamples of a single dataset. This requires a mapping between the two sets of K clusters, which we obtain by maximising
145 the average area-weighted pattern correlation between matched pairs, using linear sum assignment (Crouse, 2016). In all cases where pattern correlation is referred to, we mean this area-weighted pattern correlation of matched regime composites.

We are interested in identifying stable atmospheric regime patterns that can be found with consistency in different datasets covering different time periods - i.e. where sampling variability is minimised. To aid in this analysis we make use of the stationarity metric introduced in DS20, which we here term the *regime stability*. To calculate regime stability for a given
150 dataset, we cluster the full dataset, as well as a number of subsamples of the dataset using the same method. We then compute the pattern correlation between the full dataset's regimes - termed the reference clusters- and each of the subsamples' regimes. The regime stability is the average of these subsample pattern correlations, and may be computed for a specific regime pattern or averaged again over all patterns. For perfectly stable regimes we would therefore obtain regime stability equal to 1. We choose to use 30-year contiguous slices of each dataset to generate subsamples, so as to properly account for inter-annual correlations
155 that would be lost in a bootstrap sampling approach. An illustrative example of this process for the ERA20C dataset is shown in Figure 2 for four classical circulation regimes. For all reanalyses and models we use a ten-year spacing between subsamples i.e. (1910-1940, 1920-1950,.. etc.). For models with multiple ensemble members, each member is subsampled and treated separately.
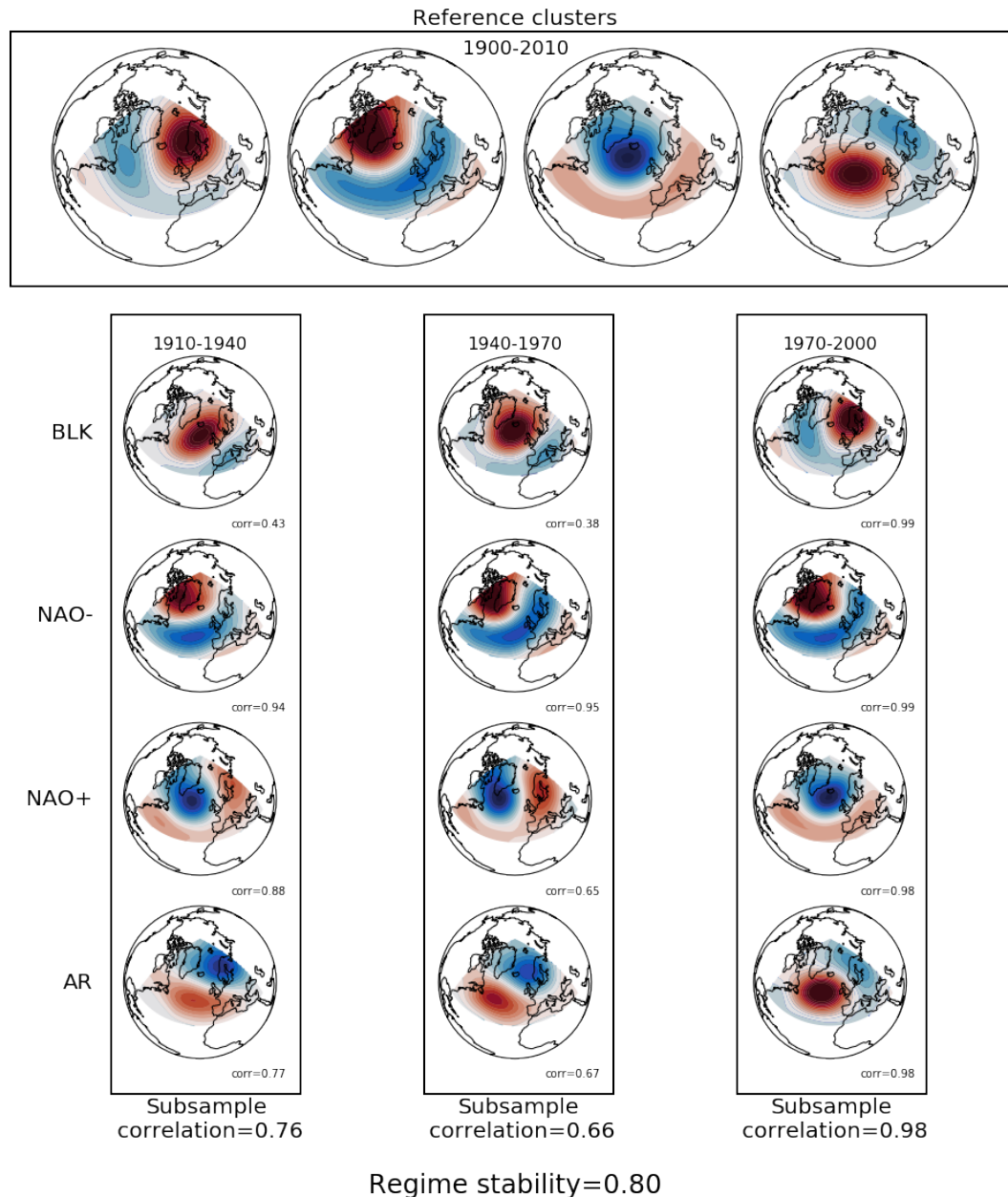
Figure 2. A schematic highlighting how regime stability is calculated, shown for 4 classical circulation regimes found in ERA20C. Reference regimes are obtained from the full length of the dataset, in this case 110 DJF seasons. Regimes are also calculated for a number of subsamples of the dataset covering different time periods, and the regime patterns are mapped to the reference regime with which they have the highest pattern correlation. These correlations are then averaged across regimes and subsamples to give an average measure of regime stability. Note that while only 3 independent subsamples are shown here in this illustrative schematic, overlapping time periods are used in practice to increase the number of samples.

As well as estimating how stable regimes in a dataset are, for model data we may wish to know how realistic those regime
160   patterns are: to what extent do they reflect regimes seen in reanalysis? To answer this question we also introduce the *regime
fidelity* metric, which is computed in the same way as the regime stability, except that the reference clusters are taken from
reanalysis, while the subsample clusters are taken from model data. Fidelity calculated with respect to different reanalysis
products will give slightly differing values, however the results are qualitatively equivalent, and so for convenience, all regime
fidelities reported in this paper are calculated with respect to ERA20C's reference regimes. Again, a fidelity of 1 in a model
165   would indicate regime patterns exactly identical to those of ERA20C.

We will characterise the temporal variability of regimes through the commonly-used *regime occurrence* and *regime persistence* metrics. Regime occurrence is defined as the fraction of days in the time-period considered assigned to a particular
regime, while regime persistence is defined as the probability that a regime event will persist from one day to the next. Persistence is calculated by fitting a first-order Markov chain to the data. The Markovian assumption is a good fit for the regime
170   lifetime distribution, with only very slight deviations in the fractions of events lasting < 3 days (see Strommen et al. (2019a)
and Supplementary Figure 2).

When considering temporal variability, it is useful to separate out days which do not strongly resemble any of the regime
patterns, which we refer to as Neutral days. This is done by calculating the pattern correlation of each day's Z500 anomaly field
with the full-data regime composites, and considering only those days with a pattern correlation ≥ 0.4 to represent active regime
175   events. This threshold value was chosen to be as low as possible while still ensuring every regime was associated with above
average levels of persistent blocking. This was done by computing the frequency of Atlantic persistent blocking events in each
regime for different neutral thresholds in ERA20C, as detailed and shown in Supplementary Figure 3. For the chosen threshold,
the AR, NAO- and BLK regimes feature above average levels of Atlantic ridging, Greenland blocking and European blocking
respectively, concretely demonstrating a connection between the regimes and long-lived anticyclonic blocking anomalies. The
180   spatial regime composites restricted to only non-Neutral days are essentially the same, featuring only slightly higher-amplitude
anomalies but no change in spatial structure. The Neutral days themselves have vanishing composites. Therefore, for simplicity
of presentation we apply the neutral filtering only when considering temporal variability.

2.4   Description of predictive model parameters

Several properties of the model mean-state are examined for their potential to explain inter-model spread in regime behaviour.
185   These quantities, which we will refer to as 'predictors', were chosen based on their prominence in the literature, and are now
described and motivated. Abbreviations used in discussion and plots are also indicated. Note that data is always interpolated
onto a common 1 degree regular grid before computation (except for Gulf stream sea surface temperature gradients, see below),
and that for quantities computed between 1979 and 2015, all available data in this period is used (i.e., for CMIP5 models this
means 1979 to 2005).

190   **Jet speed**   The mean daily, DJF jet speed across every available year. While our framework views the jet speed variability
        as being orthogonal to regime variability, this does not exclude possible relationships between the mean jet speed and

mean regime structure. For example, it is known that the mean strength of the jet is related to its latitudinal variability (Woollings et al., 2018), and this relationship may also be expected to manifest in a regime context. Abbreviation: JetSpeed.

Arctic sea ice The mean Arctic (40-90N) sea ice concentration in November, between 1979 and 2015. November sea ice has been extensively discussed as potentially exerting an influence on the jet and NAO on both seasonal (Deser et al. (2007); Strong and Magnusdottir (2011); Dunstone et al. (2016); Wang et al. (2017) to name a few) and climate time scales (see Barnes and Screen (2015) for an overview), by acting as a source of stationary Rossby waves. In Strommen (2020) it was shown that these links are also visible in a regime context. We only consider the November means here, as these may be expected to act causally on regime variability (which is always computed using DJF data) without being immediately confounded by the fact that the winter jet also influences sea ice. While restricting to November avoids such coupling issues within a given season, we cannot rule out that the climatic mean November sea ice state is intimately coupled to the mean DJF regime variability. Two further provisos should be noted. Firstly, many studies concerned with seasonal teleconnections focus on the impact of just the Barents-Kara region: we choose here to focus on the entire Arctic, as this is common in studies focusing on longer climate time scales. Secondly, a small number of models were found to be essentially ice free in November: the PRIMAVERA models MPI-ESM1-2-HR, MPI-ESM1-2-XR, and the CMIP5 version of EC-Earth (all ensemble members). To avoid these extreme outliers obscuring the analysis, these simulations were not included when computing correlations or ridge regressions involving sea ice metrics: they are still included for all other metrics. Abbreviation: Arctic Nov-mean.

Eddy forcing strength A key regime metric considered is the persistence of regime events, which is closely related to the persistence of the eddy-driven jet. Many studies have demonstrated that jet anomalies are reinforced through transient eddies, either via direct forcing or eddy-meanflow feedbacks (e.g. Robinson (1996); Lorenz and Hartmann (2003)). This suggests that eddy forcing is important for regime persistence, a hypothesis which was confirmed in Strommen (2020). Inspired by the metric used in ibid, as well as the NAO-focused eddy metric of Scaife et al. (2019), we introduce in this paper a new, simple metric for measuring the strength of the eddy forcing on the jet. We compute the daily eddy momentum flux convergence of 250hPa winds:

$$E_{250} := \frac{\partial(-u''_{250} v''_{250})}{\partial y}$$

where u''_{250} (resp. v''_{250}) is the zonal (resp. meridional), 2-6 day bandpass filtered wind field at 250hPa. Positive values of this quantity correspond to regions where the transient eddies are accelerating the westerlies (Hoskins et al., 1983). By taking DJF means of E_{250}, one obtains a field with a leading EOF resembling a northward, zonal shift of the eddies (cf. Supplementary Figure 4). Large values of the corresponding principal component PC_{250} would therefore be expected to be associated with an eddy-induced northward shift of the jet. The strength of the eddy forcing is then estimated as the regression coefficient of the normalised PC_{250} index against the (non-normalised) DJF mean jet latitude. This number can plausibly be interpreted as measuring how many degrees north the jet shifts in response to a unit measure

225   of anomalous, northwards eddy activity. The PC_{250} index was found to correlate strongly, but not perfectly, with the metric used in Strommen (2020). Unlike the metric in ibid, however, the regression coefficient here does not implicitly assume the existence of a multimodal jet, which is important given that not all models considered have one. Note that a consistent choice of sign for the EOF across data sets is maintained by asserting that each EOF computed should enjoy a positive pattern correlation with the 1st EOF of ERA5. Abbreviation: EF-JetLat-Regr.

230   **Stratospheric variability** To measure stratospheric variability, we computed the standard deviation of monthly zonal mean zonal winds at 10hPa at the equator (5S-5N). The region and pressure level were chosen so as to capture the quasi-biennial oscillation (QBO) in a simple manner. The QBO is now well resolved in many CMIP6 and PRIMAVERA models (Richter et al., 2020), unlike in CMIP5 (Butchart et al., 2018), but several models still show almost no meaningful stratospheric variability. We found that the standard deviation computed here was effective at discriminating between

235   models: models with an unresolved QBO show almost zero monthly variability, while models with a more realistic QBO show variability comparable to that of ERA5. The potential influence of the stratosphere on Euro-Atlantic regimes has been noted in previous studies (Charlton-Perez et al., 2018; Strommen, 2020; Fabiano et al., 2021). Its importance for modulating jet stream variability more broadly is well known (Kidston et al., 2015; Domeisen et al., 2020). Abbreviation: QBO Std10.

240   **Gulf Stream SST gradient** The Gulf stream region exhibits a strong sea surface temperature (SST) gradient which has been suggested as contributing to regime variability by acting as a source of heat flux anomalies (O'Reilly et al., 2016). We measured this gradient by computing longitudinal means of monthly mean SSTs in the region 30S-65N, 75W-30W and computing the regression coefficient of these against latitude. Months were restricted to November through February, and latitudes are counted from south to north: the gradient thus computed is therefore always negative. The SST data is

245   interpolated to a regular 0.5 degree grid prior to regression (with land-points masked out), in order to allow for a cleaner separation between reanalysis/high-resolution models (which have sharp gradients) and low-resolution models (which have weak gradients). Broadly speaking, this gradient may be viewed as giving a measure of the resolution of the Gulf stream, and is therefore closely linked to the strength of atmosphere-ocean coupling, with sharper gradients indicative of potentially stronger coupling. The importance of realistic coupling for Euro-Atlantic atmospheric variability has been

250   emphasised in several studies (cf. Small et al. (2019); Athanasiadis et al. (2020); Bellucci et al. (2021b); Zhang et al. (2021) and references therein). Abbreviation: NA SST-grad.

    Atmospheric horizontal resolution The hypothesis that relatively high horizontal resolution may be crucial for the realistic simulation of Euro-Atlantic regimes was first raised in Dawson et al. (2012) using a single model with one ensemble member, and has since been examined in increasingly larger ensembles with multiple models using both forced SST

255   (Strommen et al., 2019b) and coupled simulations (Fabiano et al., 2020). The results of these studies are somewhat inconsistent, with each suggesting different impacts of increased resolution: the results of this paper will shed further light on this mismatch. The horizontal resolutions are measured as the approximate grid spacing at the equator (in

kilometres), and were compiled by referring to tables in the CMIP5, CMIP6 and PRIMAVERA implementation papers. The resolution of reanalysis data is set, by convention, as 1km. Abbreviation: Atm Res.

260   In order to explore possible links between these quantities and the regimes' metrics, which may give a hint of their influence on the regime structure, we perform a multi-linear regression analysis over the whole dataset in Section 5. All predictors and metrics are standardized to zero mean and unit variance before performing the regression. Individual missing values in the predictors (at most 5 over 70 values) have been filled with the dataset mean. For models with more than one member, the multi-member mean is considered for both the predictors and the regime metrics. Cross-correlations between the predictors

265   are generally low (below 0.3 for most combinations), with the exception of QBO-std10 and jet speed (-0.37), QBO-std10 and atm-res (-0.32), and jet speed and EF-JetLat-regr (-0.30) (cf. Supplementary Figure 5). To assure that the partial colinearity is not influencing the regression, we adopted a ridge regression strategy, that adds a penalty for large regression coefficients (Hoerl and Kennard, 1970). A-posteriori check showed that the regression coefficients calculated in this way differ by less that 1% from those of the standard multi-linear regression. We compute the optimal set of n predictors (for n between 2 and

270   6), maximising the R^2 score (i.e. the explained variance) of the regression model, and choose n with consideration of the adjusted R^2 score which, as shown in Supplementary Figure 6, indicates that using all 6 predictors substantially improves the explanatory power of the model.

   Importantly, we note that we initially considered a wider set of predictors, including Arctic sea ice and polar vortex variability, as well as the mean North Atlantic SSTs. These were then pruned down to those listed by allowing the ridge regression

275   algorithm to select an optimal set of predictors, as described.

3   Climatological representation of regimes

3.1   Geopotential-jet regimes in reanalysis

In DS20, three geopotential-jet regimes, capturing a negative North Atlantic Oscillation (NAO-), Atlantic Ridge (AR) and European Blocking pattern (BLK) were identified as being particularly reproducible in different periods of the 20th century.

280   We briefly review the features of those regimes here. Figure 3 shows composites of DJF Z500, U850, jet latitude and jet speed for these three geopotential-jet regimes as computed with the ERA20C reanalysis over the period 1900-2010 (regime composites computed using other reanalyses are qualitatively identical). The AR regime features a low-latitude ridge, with an anomalous low to the North-East, and tends to feature a Northerly, zonally-oriented jet latitude on the poleward flank of the ridge, with reduced wind speed on the equatorward flank. The NAO-, also referred to in the literature as a Greenland blocking

285   pattern, features a strong anticyclone over Greenland, and low pressure in the mid-Atlantic, and a correspondingly strong reduction in wind speeds equatorward of the block. The eddy-driven jet is displaced southward, and will tend to merge with the high-level subtropical jet (Li and Wettstein, 2012). While almost all Southerly jet events are assigned to the NAO- regime, the correspondence is not one-to-one, and there are many days where a central jet is coincident with the NAO- geopotential-jet regime. The BLK pattern features a meridional dipole with a high over Scandinavia, suggesting a wavy jet structure. The BLK
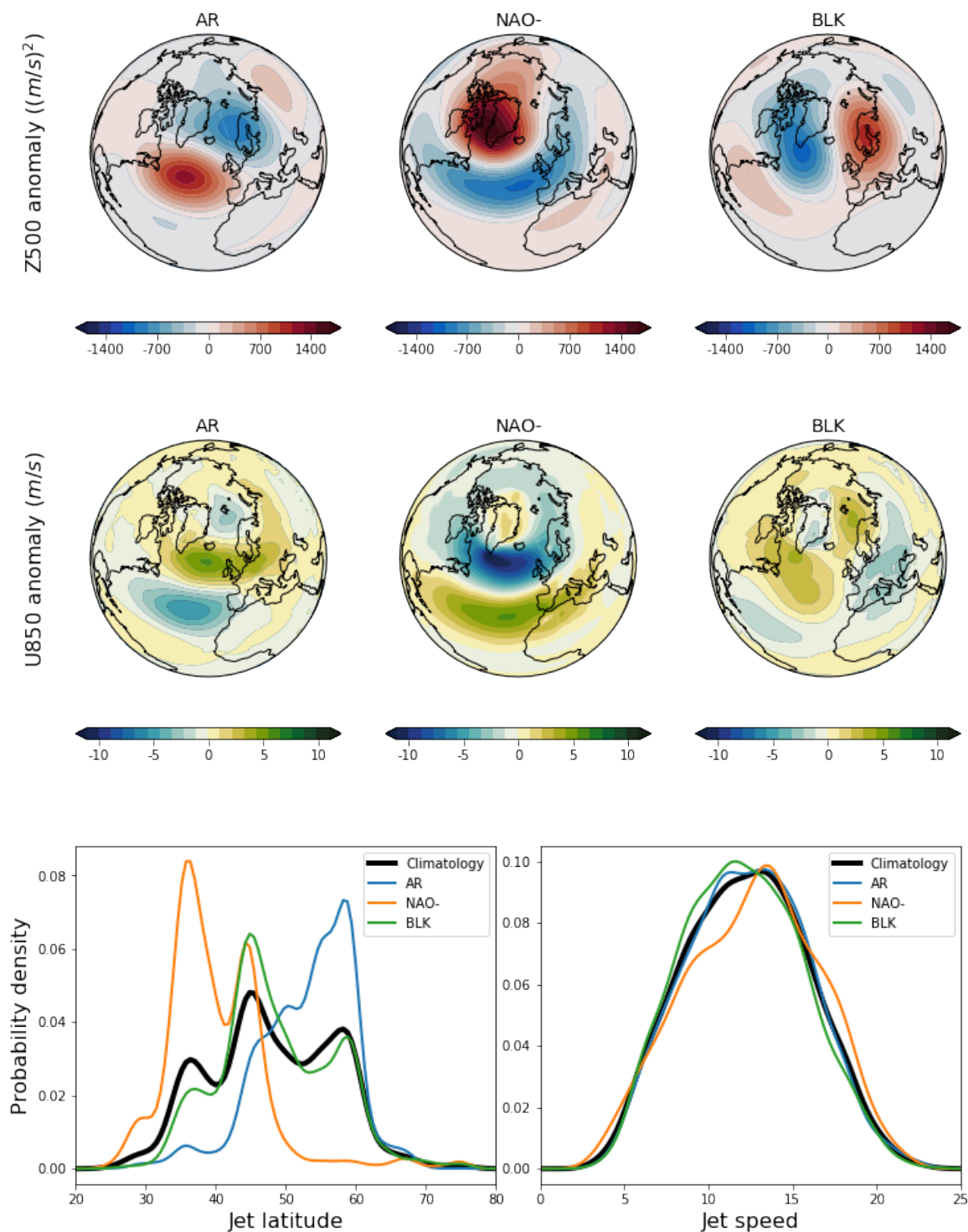
Figure 3. Regime composites for 3 geopotential-jet regimes in ERA20C. Top: 500 hPa geopotential height anomalies. Middle: 850hPa zonal wind speed anomalies. Bottom left: distribution of jet latitude. Bottom right: distribution of jet speed.

290   regime is the most weakly coupled to the jet position, featuring a trimodal latitude distribution, but it does feature a preference for central jet activity, and with the same pattern of positive poleward wind anomalies and negative equatorward anomalies as in the AR regime, the latter of which produce the low wind speeds over continental Europe which are characteristic of a European blocking.

     By regressing out jet speed from the principal components prior to clustering, our regime identification prioritises features of

295   the Z500 field that link to the spatial structure of the jet stream, rather than variations in the strength of the jet. Nevertheless, the NAO- regime does feature a small but statistically significant coincidence with faster jet speeds, indicating a weak nonlinear component to the jet–blocking relationship.

3.2   Geopotential-jet regimes in CMIP5 and CMIP6

For all climate model data sets, subsamples were constructed by taking 30-year windows of DJF data at 10-year intervals,

300   such that 4 subsamples would be computed for a 60-year dataset, and 9 subsamples for a 110-year dataset. This was done to provide a good number of subsamples for each model or reanalysis product while also maintaining a reasonable degree of independence between each sample. Contiguous windows were used rather than bootstraps in order to correctly sample any possible interdecadal variability in regime patterns.

     The representation of regime structure in reanalyses, and in the CMIP5 and CMIP6 model ensembles, is shown in Figure 4,

305   as summarised in the regime stability and fidelity metrics defined in section 2. Metrics were computed using both the classical circulation regime and geopotential-jet regime frameworks, and for cluster numbers varying from 2 to 10.

     Just as was found using only ERA20C data in DS20, the regime stability in the reanalysis ensemble is maximal for 3 geopotential-jet regimes, and in most cases the geopotential-jet regimes are more stable than their classical circulation counterparts, with the only exception being the two-regime case, where the classical circulation regimes capture the two phases of

310   the NAO with high robustness. Quite strikingly, we see that both the CMIP5 and CMIP6 ensembles show the same stability maxima for 3 geopotential-jet regimes as reanalysis, identifying them with even more confidence than the NAO dipole in the classical circulation regime framework. That we see high stability for these three regimes, even in uninitialised climate models, indicates that this is not simply a chance feature of 20th century weather variability, but instead a property inherent to the structure of the Euro-Atlantic circulation. This validates the hypothesis and conclusions of DS20 regarding the three-regime

315   approach. We note however that the extended set of 5 regimes also considered in DS20 does not appear robust when considering multiple reanalyses and model data. Instead, a secondary stability peak is seen in reanalysis when using 7 regimes, both for classical and geopotential-jet regimes. This more detailed discretisation of the circulation may be useful for impact-focused work, but is not as clearly captured by models as the coarser three-regime picture. We therefore leave the analysis of the 7 regime picture to future work.

320   There are no statistically significant differences in the regime stability of the CMIP5 and CMIP6 ensembles, but the CMIP6 models have robustly higher regime fidelity than CMIP5, showing clear progress towards a more accurately resolved Euro-Atlantic circulation. The increase in regime fidelity between CMIP5 and CMIP6 is largest for those cluster numbers previously suggested to represent meaningful regimes: 2 and 4 clusters for classical circulation regimes, and 3 clusters for geopotential-jet
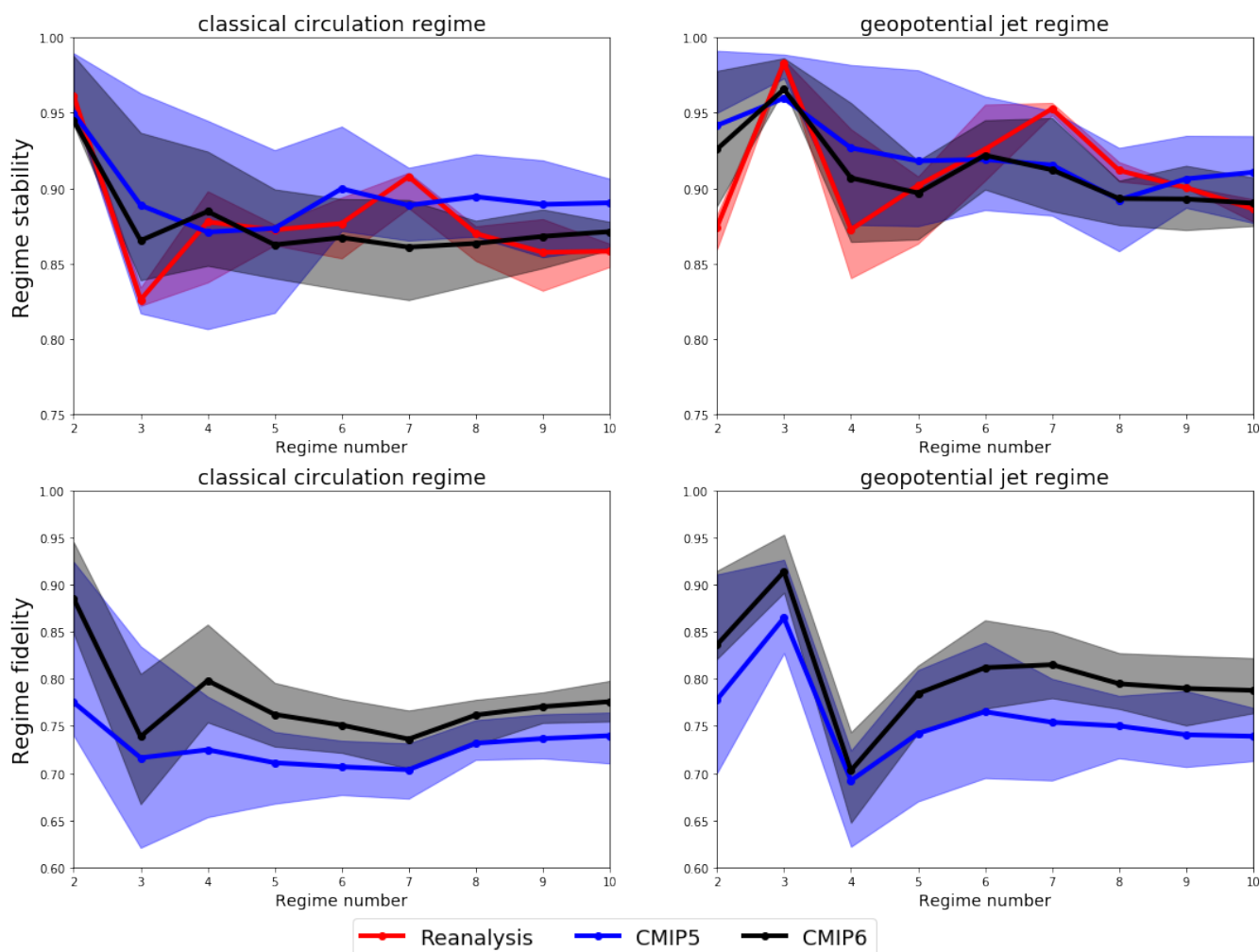
Figure 4. Regime stability and fidelity, calculated with 30-year subsamples, as a function of regime number, for both the classical circulation and geopotential-jet regime frameworks. Lines show the mean value for each data set, while shading shows the ensemble interquartile range.
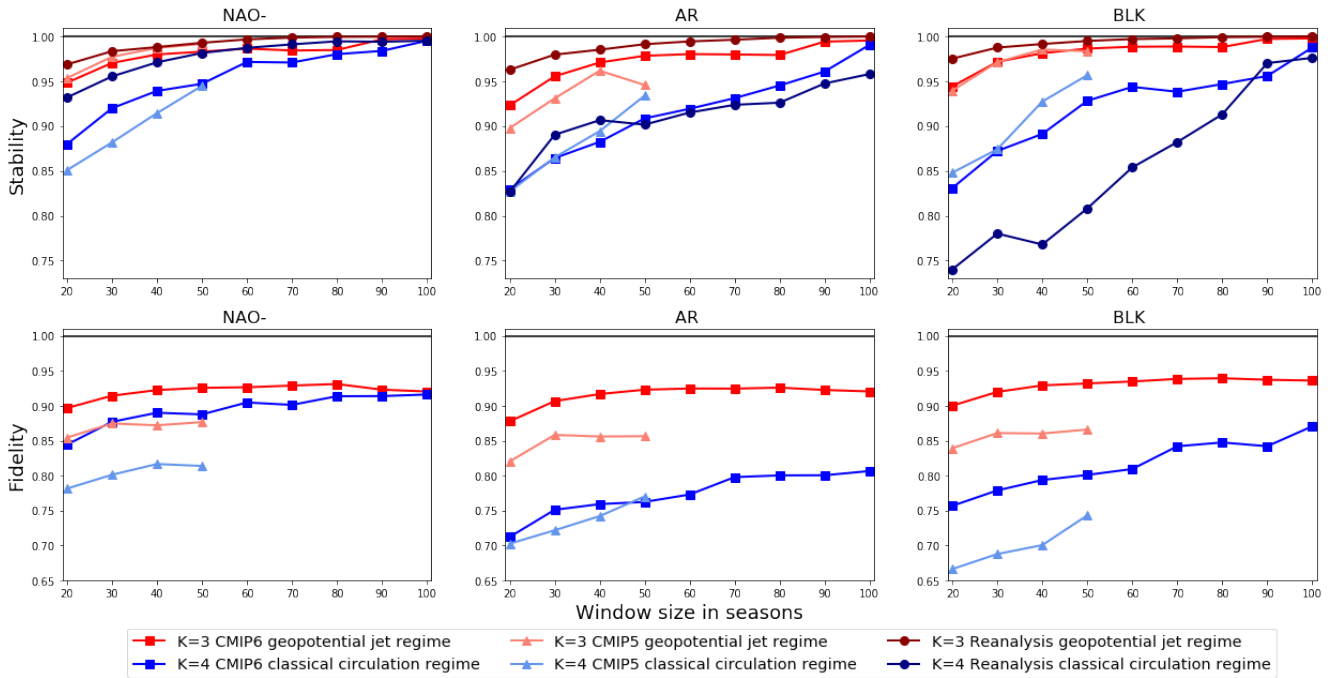
Figure 5. Regime stability and fidelity as a function of subsample window size. Ensemble mean values are shown for each atmospheric regime pattern as identified using both 4 classical circulation regimes and 3 geopotential-jet regimes.

regimes. This is what we would expect if these regimes really do have a physical basis - clusters found for a non-physical cluster
number will be to some extent arbitrary, and so we would not expect a model to reproduce the reanalysis clusters particularly
well. In contrast, for physically motivated regimes, a model with improved physics and resolution should be expected to more
accurately reproduce the real-world regime structure. This therefore gives us further reason to consider these particular regime
frameworks as meaningful partitions of the phase space. For both stability and fidelity, the spread of the CMIP6 models is
reduced compared to CMIP5, indicating a general convergence in ensemble behaviour, and CMIP6 more closely follows the
variations in stability seen for reanalysis, with a sharper maxima for 3 geopotential-jet regimes and minima for 4 classical
circulation regimes than is visible in CMIP5.

The fidelity and stability of the three geopotential-jet regimes - AR, NAO- and BLK - considered separately are contrasted
with their equivalents in the 4 classical circulation regime framework in Figure 5. Here we also assess the sensitivity of our
results to the subsample window size used to calculate stability/fidelity, which is varied from 20 to 100 years while maintaining
the same 10-year sampling interval. The geopotential-jet regimes are robustly more stable and have higher fidelity in all datasets
for all sampling windows. This is particularly pronounced in regime fidelity for the AR regime, and in both metrics for the
BLK regime, while the classical circulation regime framework already captures the NAO- regime relatively well. For the
geopotential-jet regimes, both CMIP ensembles have slightly lower stability than seen in reanalysis, most notable in the 20-

year subsamples, and the fidelity of CMIP6 is uniformly ∼ 5% higher than in CMIP5, regardless of regime or subsample
340   window size considered.

3.3   Jet latitude regimes in CMIP5 and CMIP6

There are also substantial improvements in the representation of the jet latitude regimes themselves in CMIP6. Figure 6 shows
the fraction of models in both CMIP ensembles with one, two, or three jet latitude peaks, as well as the average central latitude at
which those peaks occur. While a larger fraction of CMIP6 models properly resolve the trimodal jet latitude distribution than in
345   CMIP5, the main improvement comes from a large reduction in the number of models with unimodal jet latitude distributions.
The positioning of the Southern and Central peaks is improved, and the number of peaks occurring at latitudes avoided in
reanalysis is reduced. There are still issues in correctly resolving the Northern jet latitude peak, however (associated with the
AR regime), and many of the models with bimodal jet latitude distributions are missing this peak. However, even in CMIP5,
half the models considered produce a trimodal jet distribution, giving a less pessimistic assessment of model performance than
350   presented in Anstey et al. (2013), which found only a few models could capture jet trimodality. This disagreement likely results
from the much larger number of models included in this paper.

The full distributions of jet speed and jet latitude in CMIP5, CMIP6 and reanalysis are shown in Supplementary Figure 7.
In order to holistically quantify the model error in these distributions we make use of the Wasserstein distance introduced in
section 2. Like any distance measure, values close to zero indicate smaller errors. The cumulative density function of the jet
355   latitude and jet speed Wasserstein distances between each model and the multi-reanalysis mean jet distributions are shown in
Figure 7. The average Wasserstein distance is 33% and 38% lower in CMIP6 than in CMIP5 for the jet latitude and jet speed
distributions respectively. For the jet latitude this is particularly a result of the reduction in the number of models with large
errors in CMIP6, as visible in the diverging of the cumulative density functions for distances exceeding 0.002.

Of course biases in jet speed and latitude are not independent of each other, and models that do better at representing one
360   aspect of the Euro-Atlantic circulation generally do better at representing other aspects. Figure 8a) shows a positive correla-
tion between the jet speed and jet latitude Wasserstein distances, significant at the 1% level–a relationship that is seemingly
stronger in the CMIP6 ensemble. Panel b) shows that a smaller jet latitude Wasserstein distance is in turn predictive of a model
having high fidelity in the 3 geopotential-jet regime patterns, with strong negative correlations between the two measures again
significant at the 1% level. Jet speed errors are weakly negatively correlated to regime fidelity (corr=-0.25 – -0.4), but this can
365   be explained entirely by the covariability with jet latitude errors; when we regress out the Wasserstein distance in jet latitude
from the Wasserstein distance in jet speed and correlate the residual with regime fidelity as in panel c), we see there is no
remaining correlation. The reverse approach - computing the residual of jet latitude distance after regressing out jet speed -
produces a residual which is still strongly correlated with regime fidelity (not shown). This suggests errors in jet latitude are
related to errors in jet speed and in spatial regime structure, but that biases in the model jet speed do not independently relate
370   to the quality of the anticyclonic regime structure, although we cannot rigorously assert causal relationships from a post-hoc
analysis. Results are very similar when correlating jet errors with the fidelity of the 4 classical circulation regimes, except the
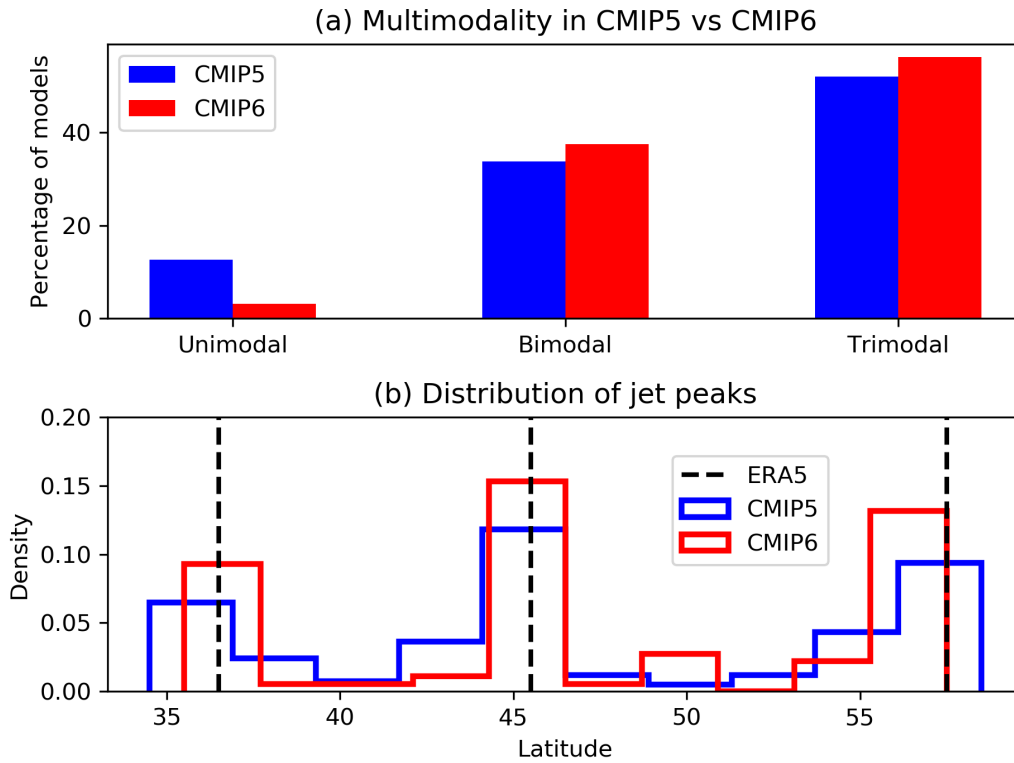negative correlation between jet latitude errors and fidelity is weaker.

16

Figure 6. a) The percentage of climate models identified as having one, two, or three peaks in their DJF jet latitude distributions. b) Histograms of the latitude at which climate models produce jet latitude peaks, with vertical lines showing the true location of peaks found in ERA5.
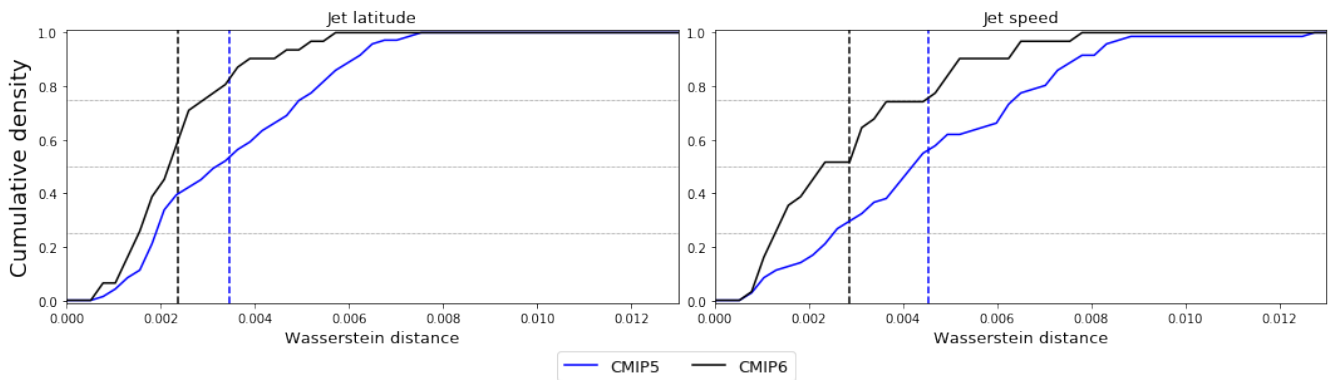

Figure 7. Cumulative density functions over model ensembles of the Wasserstein distance between model jet latitude and jet speed distributions and the multi-reanalysis mean distributions. Dashed vertical lines indicate the ensemble mean distances.
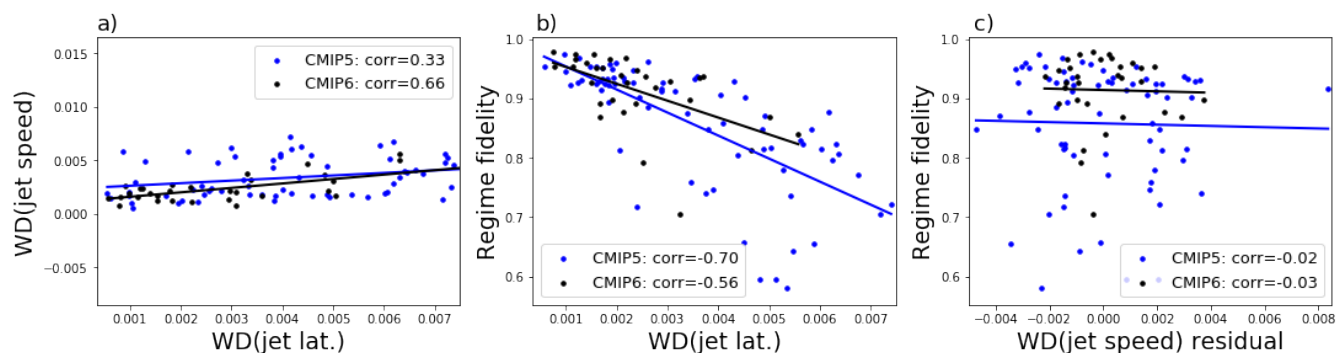
Figure 8. a) Correlations between the Wasserstein distances in jet latitude and jet speed in each climate model simulation. b) The same but showing correlation between Wasserstein distance in jet latitude and the fidelity of the three geopotential-jet regimes. c) the correlation between regime fidelity and the residual of the jet speed distance after covariability with the jet latitude distance has been regressed out. Correlations in panels a) and b) are significant at the 1% level, while no significant correlations are present in panel c).

4 Historical regime variability

In the previous section we have established that CMIP6 models, and even many CMIP5 models, have an adequate represen-
tation of geopotential-jet regime patterns - the positions of the highs and lows associated with blocking anomalies. We now
turn our attention to the representation of the temporal characteristics of those regimes – their lifetime and probabilities of
occurrence – as well as their historical variability.

Regime occurrence and persistence, as defined in section 2 have been computed in 30-year rolling windows for all reanalyses
and all CMIP models, in order to highlight the interdecadal climatic variations in the regimes. The ensemble-mean 30-year
regime occurrence for each dataset is shown in Figure 9, where all models/reanalyses that have data for each full 30-year period
are included in the average. The full ensemble spread is shaded, and for the CMIP6 and reanalysis datasets, long-term mean
values are also indicated.

Observational variability in regime occurrence is quite substantial, and is markedly higher for the Neutral and NAO- regime
than for the AR and BLK regimes, which show less pronounced variation. The previously documented mid-century peak in
NAO- occurrence that has been related to reductions in forecast skill in Weisheimer et al. (2017) is visible here, and we now also
document a clear decrease in occurrence of the Neutral regime over 1970-1990. Between the reanalyses there is a fair amount
of spread in the exact occurrence statistics, which is notable given the almost perfect equivalence between regime patterns
found in the reanalyses. This is most pronounced in the early century Neutral and AR regimes where reanalyses disagree by as
much as 3-4%, but can also be seen right into the modern era, such as in the case of the BLK regime occurrence.

The biases in CMIP6 are similar to those in CMIP5: Neutral flow states are too common and the BLK regime occurs too
infrequently, both by ~2%. There is no noticeable bias in the NAO- or AR occurrence, which represents an improvement in the
NAO- relative to CMIP5. These coupled models do not reproduce the actual historical variability seen in reanalysis, and the
ensemble mean occurrence is almost flat across the century, consistent with the perspective that slow SST variability – which
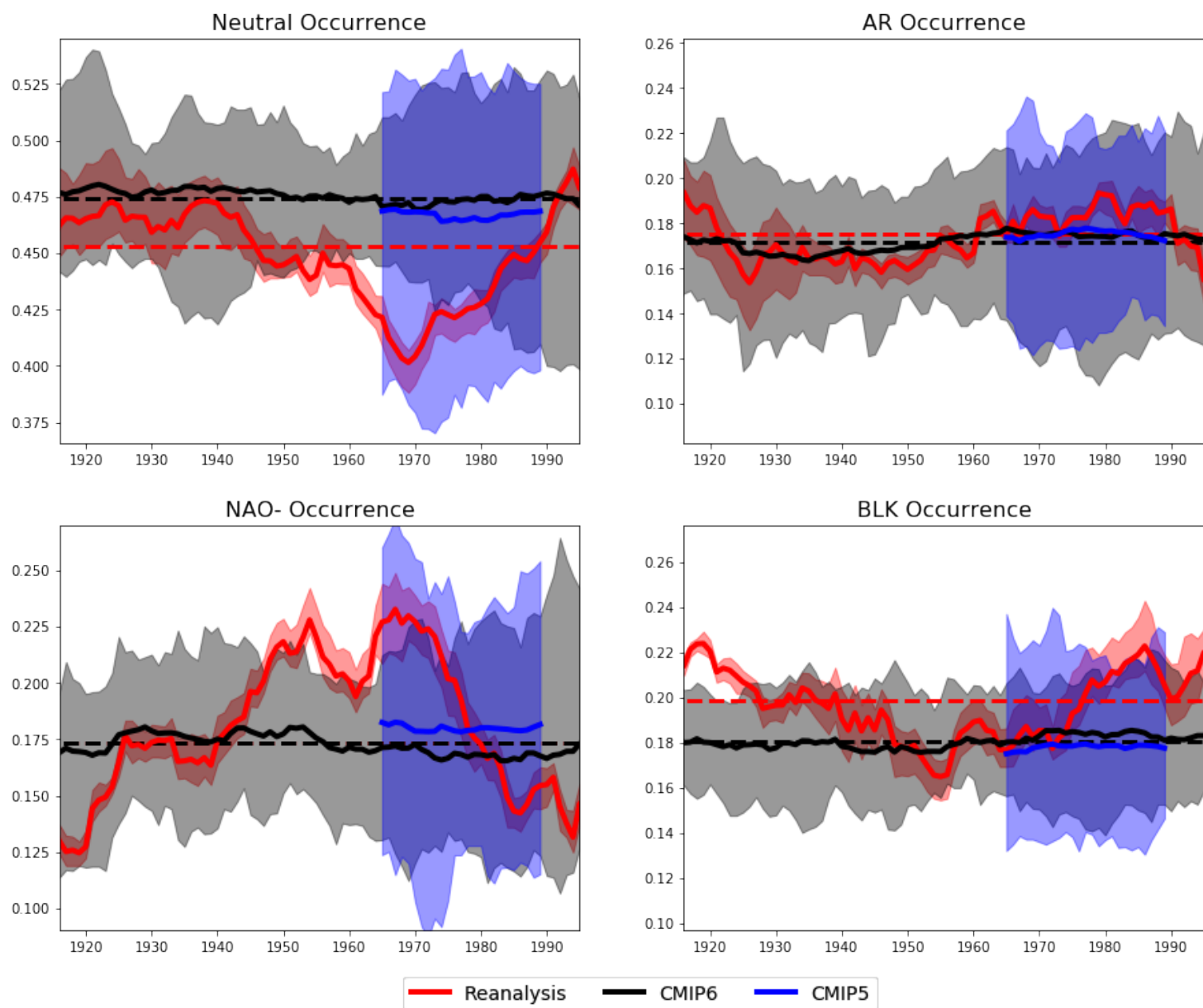
Figure 9. 30-year rolling windows of regime occurrence for the multireanalysis, CMIP5 and CMIP6 ensembles, with the x-axis marking the central year of the window. Thick lines show the ensemble mean, while shading indicates the full ensemble spread. Dashed red and black lines show the 1900-2010 average occurrence for reanalysis and CMIP6 respectively.

will not be coherent across an average of coupled models – is a major driver of decadal regime variability. Analysing AMIP
395   regime statistics could validate this claim, and is left for future work. This emphasises the importance of evaluating model
regimes against an appropriately long historical benchmark. For example, looking only at the years 1980-2010, CMIP6 would
seem to have a positive NAO- occurrence 'bias', while between 1955 and 1985 it would seem to have a large negative bias of
5%. Comparing centennial averages however, there is no NAO- occurrence bias at all.

Regime persistence and its variability is shown in Figure 10. Here the disagreements between reanalysis estimates of his-
400   torical regime persistence are even larger than for occurrence, and worryingly divergent spread in NAO- persistence is seen
between reanalyses in the first half of the 20th century. It is quite possible that this spread is a result of poor historical data
constraints, which are particularly severe in the higher latitudes around Greenland. CMIP6 shows a bias towards reduced per-
sistence for both BLK and NAO- regimes of approximately 2%, while the AR regime shows no persistence bias, and there is
a tendency for CMIP6 to produce too persistent Neutral events. This is consistent with the known biases in climate models
405   towards overly zonal flows, and underpersistent regimes, as is the lack of any substantive improvement in these biases between
CMIP5 and CMIP6 (Davini and D'Andrea, 2016).

The variability of the CMIP ensemble means is of course far smaller than of the reanalysis mean, and it is also valuable to
consider whether the individual models are able to represent the correct levels of slow variability in variability occurrence and
persistence, as well as their mean values. Figure 11 shows the standard deviation of the occurrence and persistence metrics
410   for a range of smoothing timescales. e.g. the 1-year smoothing timescale captures the interannual variability while the 30-year
smoothing timescale captures multidecadal variability. This approach was found to be more robust than a more-conventional
Fourier spectra analysus, due to the non-stationarity in some of the reanalysis time-series, and the presence of model time
series of different lengths.

The interannual variability of occurrence and persistence is well captured in models for all regimes, but biases appear on
415   longer timescales of variability. In particular, models underestimate the inter- and multi-decadal variability of the Neutral and
NAO- regime occurrence, and to a lesser extent for BLK occurrence as well. Low-frequency variability in regime persistence
however is too high in CMIP6, representing a notable departure from CMIP5, which has little bias in persistence variability.

5   Physical predictors of regime representation

In the previous sections we have explored the spatial and temporal representation of regime dynamics across the CMIP en-
420   sembles. We have seen that there is considerable variability between models in the fidelity, occurrence and persistence of
regimes, and this naturally prompts us to ask *why* some models have more realistic regimes than others. In this section, we
use the CMIP models as an ensemble of opportunity in order to relate regime representation to other specific features of the
model climate. By considering how the climate mean-state of six potential predictors, detailed in section 2.4, covary with
different aspects of regime representation, we can gain physical insight into what factors shape the characteristics of regime
425   systems. To further increase the amount of data available to us, and to broaden the inter-model spread, we also now include
the PRIMAVERA simulations detailed in SI Table 2, which were produced as part of HighResMip, and so many of which
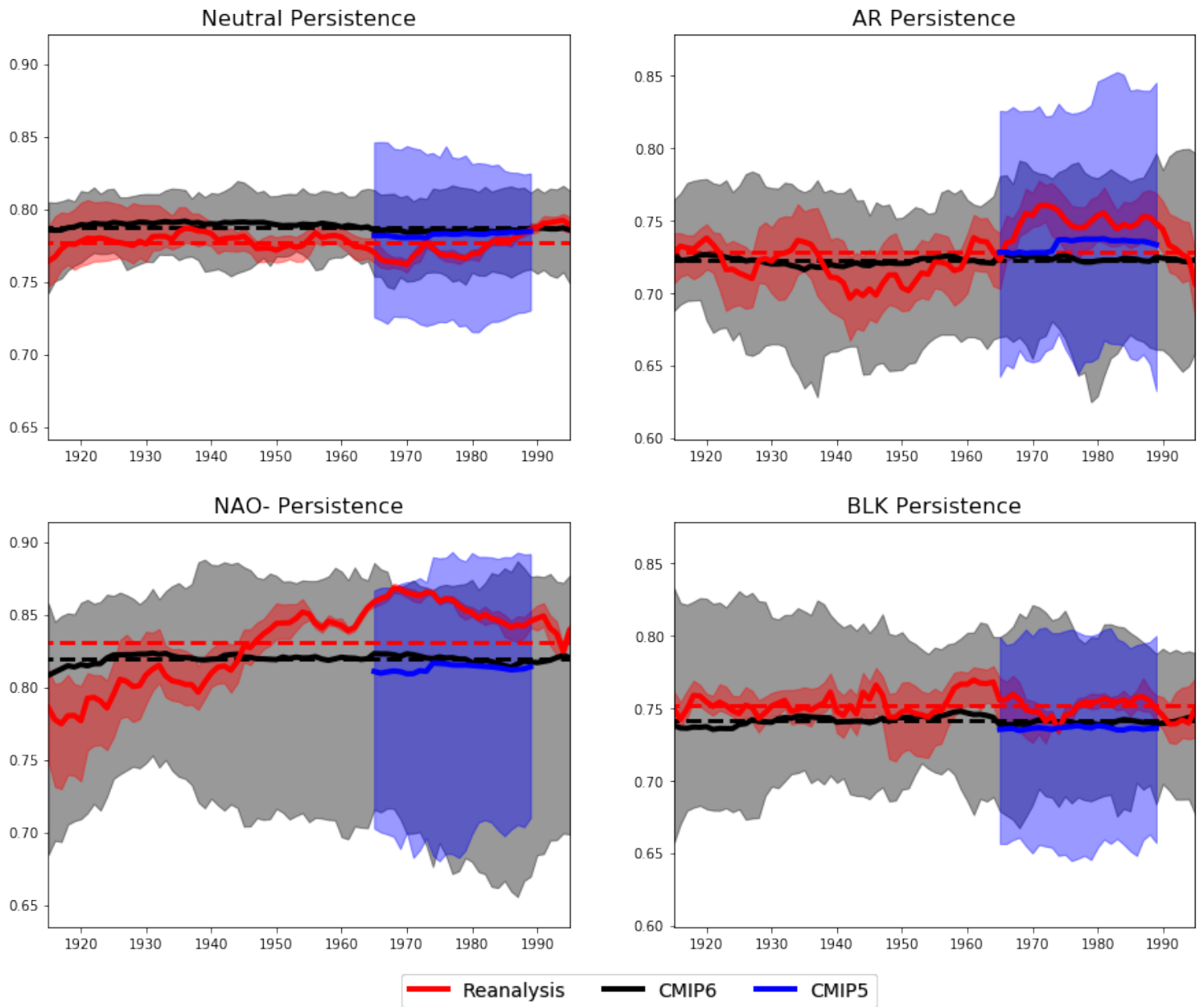
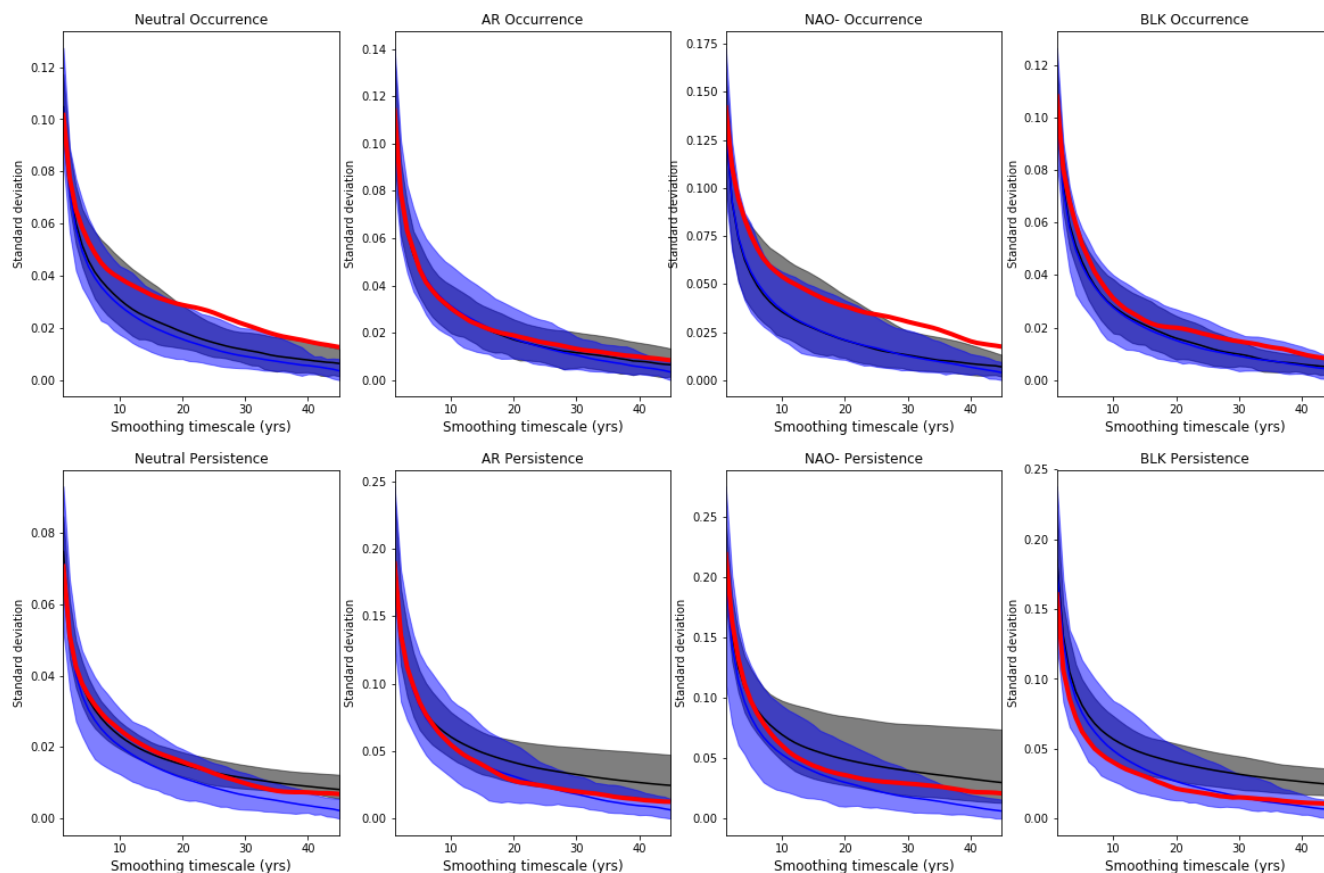**Figure 10.** As in Figure 9 but showing regime persistence.

Figure 11. Standard deviation of seasonal regime occurrence and persistence as a function of smoothing timescale, showing the variability of the regime statistics on a range of timescales. Thick lines indicate ensemble means of each models' standard deviation, while shading shows the model spread. Red, black and blue lines indicate mulit-reanalysis mean, CMIP6 and CMIP5 standard deviations respectively.

have much higher resolutions than found within the CMIP6 ensemble. Six underlying models were used, each run at a number of different resolutions: CMCC-CM2 (Cherchi et al., 2019), CNRM-CM6 (Voldoire et al., 2019), EC-Earth3 (Haarsma et al., 2020), ECMWF-IFS (Roberts et al., 2018), HadGEM3-GC31 (Williams et al., 2018), MPI-ESM1-2 (Gutjahr et al., 2019), and
430   AWI-CM-1.0 (Sein et al., 2017).

Of the predictors we consider, EF-JetLat-Regr and Jet Speed are both local properties of the Euro-Atlantic troposphere, while QBO-std10, Arctic Nov-mean, and NA SST-grad are remote properties of the stratosphere, Arctic sea ice and North Atlantic ocean respectively, which have known teleconnections to the tropospheric circulation, as discussed in Section 2.4. Resolution accounts for both the more detailed representation of orography and small scale eddies, but will also invariably
435   capture general inter-generational modelling improvements found in increasingly sophisticated and high resolution climate

22

models. This 'generic' contribution to the impact of resolution is to some extent mitigated in the PRIMAVERA models, where models of the same generation were run at varying resolution, and without re-tuning of high-resolution model versions.

This approach is of course imperfect: the limitations of treating CMIP datasets as true independent ensembles has been noted in Knutti et al. (2013), and there are many more differences between climate models than the small number of predictors

440 we consider here. Further, the diagnostic approach we take here cannot establish causal relationships, which would require targeted single-model experiments with controlled parameter variations. Nevertheless, the dozens of coupled climate model integrations in the CMIP and PRIMAVERA datasets provide tens of thousands of years of associated data, which would be computationally prohibitive to replicate in a purer fashion. The suggestive links between predictors and regime representation we detect here can then provide direction for more focused modelling work in the future.

445 We train a regression model to predict 11 different aspects of regime representation. These are the occurrence and persistence probabilities of the 3 geopotential-jet regimes and the Neutral state, and three spatial regime metrics: the regime stability, the regime fidelity evaluated against ERA20C, and the variance ratio. The variance ratio, as used in Dorrington and Strommen (2020); Fabiano et al. (2020), quantifies the multimodality of the underlying phase space, with high values indicating very pronounced multimodality, and so assesses how 'regime-like' the model's circulation is, without necessarily requiring it to

450 resemble the regimes seen in reanalysis.

Figure 13 shows the regression coefficients for the predictive model, while Figure 12 shows the amount of intermodel variance in the regime metrics that the regression model can explain. We also assess the impact of using fewer than 6 predictors, and while in some cases there is little additional explanatory power gained from introducing more than 3 predictors, the spatial regime metrics in particular benefit from including all 6 predictors. An analysis of the adjusted R^2 confirms that significant

455 explanatory value is being introduced by including all predictors (cf. Supplementary Figure 6). Some of the most striking correlations between predictors and regime behaviour are also presented as scatter plots. Figure 14 shows the correlation of EF-JetLat-Regr with the variance ratio and active regime persistence, Figure 15 shows the relation of QBO standard deviation and jet speed on spatial regime structure, Figure 16 shows the correlation of jet speed with regime persistence, and Figure 17 shows the impact of horizontal resolution on both regime spatial structure and persistence. The results were found to be

460 qualitatively insensitive to using all model ensemble members, instead of the ensemble mean, as shown in Supplementary Figures 8-11.

5.1 Spatial regime structure

The variance ratio is the most predictable of the spatial regime metrics, followed by regime fidelity. Regime stability is not as well explained by the regression model, conjecturally because many models have stability close to 1, and so intermodel

465 variability is lower overall (as in Figure 5). All three spatial regime metrics, however, are robustly increased by improving model resolution, which can be seen most strongly for fidelity. This builds upon previous work showing the beneficial impact of increased resolution on regime structure (Dawson and Palmer, 2015), and on blocking as a result of improved orography (White et al., 2019; Davini et al.), now showing it holds, essentially linearly, across the very wide range of resolutions (between 25 and 400 km) spanning PRIMAVERA, CMIP6 and CMIP5, in a multimodel context with multiple ensemble members.

470   We also see that strong eddy feedbacks on the jet are linked to an increased variance ratio. Climate models tend to have overly weak eddy feedbacks as well as too low variance ratios, as shown by Figure 14, and we see that when this bias is reduced (i.e. eddy feedbacks are stronger), models' variance ratios are closer to reanalysis. Both fidelity and variance ratio are significantly higher in models with lower jet speed and a more variable QBO (both associated with a reduction in model bias, cf. Figure 15). Low jet speeds are a natural result of increased stirring of momentum and wave activity associated with
475   blocking (Nakamura and Huang, 2018), and so this observed correlation aligns well with theory.

Figure 15b) and d) show a clear group of models with very low QBO variability, primarily from CMIP5, and which upon visual inspection were seen to be models with essentially no QBO at all. While some models without a QBO have reasonably high regime fidelity, all models with a fidelity below 0.8 lacked a QBO. This is in agreement with the well known importance of stratospheric dynamics for tropospheric regime structure (Baldwin and Dunkerton, 2001; Beerli and Grams, 2019), both due

480   to direct forcing of the QBO on the subtropical jet, and through a forcing on the NAO modulated by the polar vortex (Gray et al., 2018).

Finally we see a weaker Gulf Stream meridional SST gradient (note that this is a negative quantity) is correlated with higher fidelity and stability in models. The link is particularly strong for regime stability, and model SST gradients are in fact the main source of predictive skill for the stability. We can understand this, somewhat conjecturally, as weak meridional SST gradients

485   being a consequence of unresolved Gulf Stream eddies and thus a sign of weak ocean-atmosphere coupling (Bellucci et al., 2021a; Zhang et al., 2021). If the coupling is weak, then there will be less forced tropospheric variability on decadal timescales, and so less interdecadal variation in regime structure. This might be expected to result in higher regime stability. However, we are not able to assess the validity of this conjecture in this work.

5.2   Regime occurrence

490   The different occurrence rates of regimes between models can also be partially explained by the predictors considered here, excepting variability in BLK occurrence. The strongest relationship is that fast jet speeds are correlated to increased AR occurrence. This was also hinted at in Barnes and Polvani (2013a), which found models with faster jet speeds were biased towards a more Northerly jet latitude. Improving model resolution increases the number of Neutral days at the expense of the number of NAO- days, consistent with the elimination of the CMIP5 NAO- occurrence bias in CMIP6. Increasing the strength

495   of eddy feedbacks has the opposite impact, increasing the amount of NAO- days, as well as also AR days, by reducing the number of Neutral days. Given that CMIP6 models have weak eddy feedbacks compared to reanalysis, we might expect CMIP6 to show underoccurrence biases for the NAO- and AR, and too many Neutral days. While we do indeed see a bias towards too many Neutral days, there are no corresponding biases in the AR and NAO- regimes. Instead it is BLK that is underoccurrent in models. This may suggest the existence of a compensating model error, that we have not managed to isolate in our analysis

500   here.

5.3 Regime persistence

Variability in regime persistence can be more fully explained than regime occurrence, although again, variations in BLK persistence are the least explicable. We see that low jet speed is highly correlated with high regime persistence in all regimes, including the Neutral regime, suggesting a link to longer but sparser blocking events. This can be viewed as a regime inter-
pretation of the result highlighted in Woollings et al. (2018), where low jet speeds were shown to be associated with increased latitudinal variability. A priori, increased latitudinal variability might be associated with increases in regime occurrence (the jet meandering more between different regimes) or persistence (the jet wobbling more within distinct regimes), and our result therefore adds additional clarity by demonstrating that the extra variability primarily projects onto regime persistence. Strong eddy feedbacks are linked to high regime persistence in the three blocking regimes, in line with the hypothesised role of eddy
vorticity fluxes as an important mechanism of blocking maintenance (Shutts, 1983). The fact that the clearest relationship between eddy forcing and persistence is seen for the NAO- regime can be understood in terms of earlier studies (Lorenz and Hartmann, 2003; Barnes and Hartmann, 2010), which show that equatorward-shifted jets are more persistent than poleward-shifted jets, owing to the fact that eddy feedbacks are stronger closer to the equator. Lower levels of November Arctic sea ice also seem linked to high AR persistence. While resolution only significantly explains variability in BLK persistence, all
regimes show a correlation between improved resolution and *decreased* blocking persistence, as seen in Figure 17, which serves to exacerbate model bias.

6 Discussion and Conclusion

In this paper we have analysed how well the two most important aspects of the wintertime Euro-Atlantic circulation – the variability of the eddy-driven jet and transitions between anticyclonic blocking regimes – are represented in current-state-
of-the-art climate models, as captured by the CMIP6 ensemble. Through comparisons with the previous generation of CMIP5 models, and consideration of the high-resolution PRIMAVERA ensemble, we have carried out the most comprehensive analysis of model regime and jet behaviour to date, including more than 70 separate models. We have made use of the *geopotential-jet regime* framework, introduced recently in Dorrington and Strommen (2020), which filters out the linear variability of jet speed prior to clustering in order to focus on the impact of fundamentally nonlinear jet-latitude variations on the geopotential height
field. With the aid of this large dataset and this hybrid methodology, we have made several key points:

- The spatial structure of anticyclonic blocking regimes, and the distributions of daily jet latitude and speed, have significantly improved between CMIP5 and CMIP6. This is primarily due to a reduction in the number of poorly performing CMIP6 models: almost all models have a multimodal jet latitude distribution, and a majority capture all three peaks, while the 3 geopotential-jet regime patterns observed in reanalysis –associated with blocking anomalies–are more accurately observed, with average regime fidelity increasing by 6%.

- The average CMIP model has a faster jet speed and weaker eddy feedbacks on the jet latitude than found in reanalysis - related variables which indicate insufficient stirring of momentum in the jet core. Models with lower jet speeds and strong
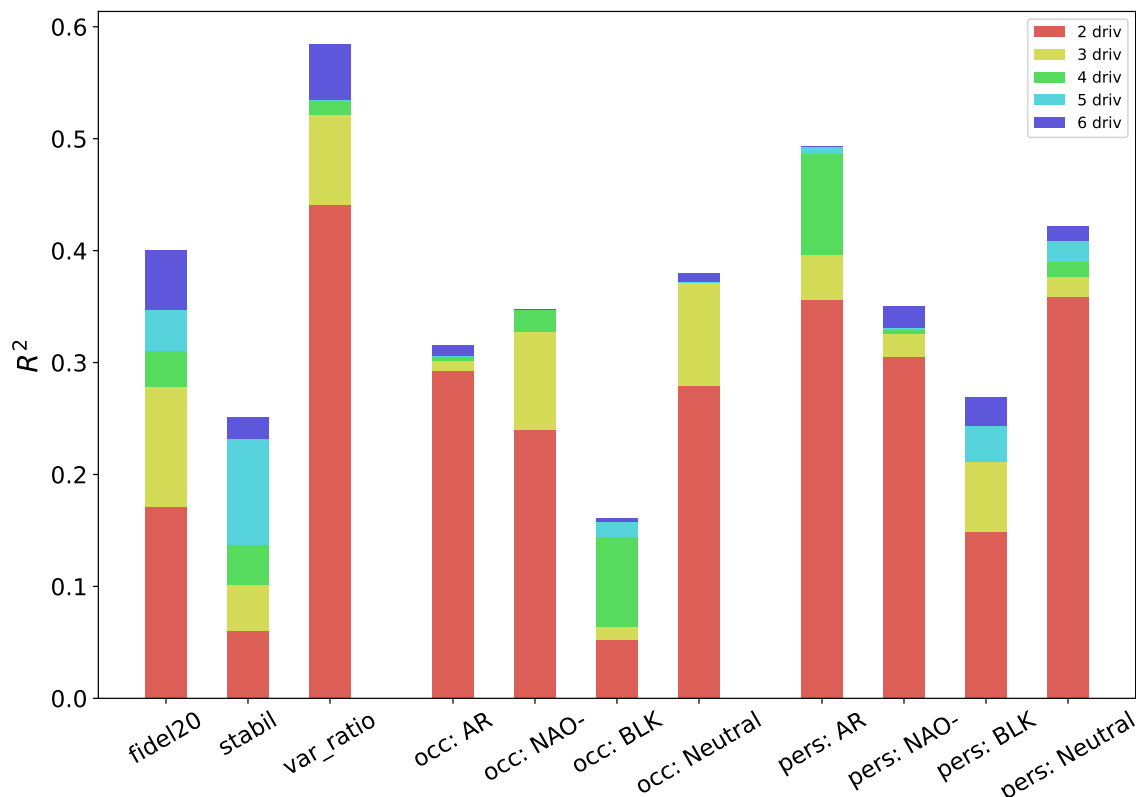
Figure 12. The fraction of intermodel variance in regime metrics explained by the multilinear ridge regression model, as a function of the number of predictors included.

eddy-jet feedbacks were found to have stronger and more persistent regimes, indicating that these biases are probably key issues in the underpersistence of the NAO- and BLK regimes. Importantly, while models with improved horizontal resolution exhibit stronger, more stable, and more realistic regime behaviour, they tend to have slightly reduced regime persistence, and high resolution models do not have significantly reduced biases in jet speed and eddy forcing. It is clear then that resolution upgrades alone will not eliminate the biases in blocking regimes.

– Model biases in the occurrence and persistence of regimes have not improved in general, moving from CMIP5 to CMIP6. The average climate model visits the BLK regime too infrequently, and BLK and NAO- events do not persist for long enough. Neutral conditions, without a clear regime anomaly, are too prevalent. The sign of these biases is completely consistent with the documented under-occurrence of European blocking, and underpersistent Greenland and European blocks, as identified with blocking indices in other work.
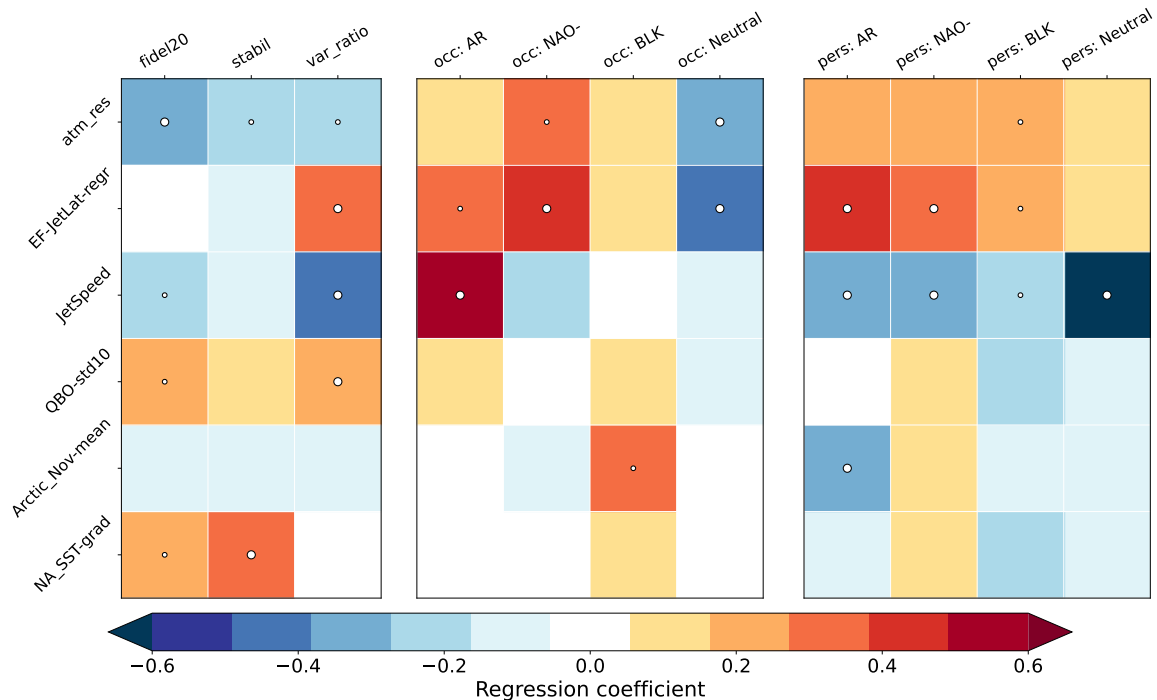
Figure 13. Regression coefficients from the six-predictor multilinear ridge regression model, trained on CMIP5, CMIP6 and PRIMAVERA models, for spatial regime representation (left), regime occurrence fractions (centre) and regime persistence probabilities (right). Where the explanatory power of a predictor for a metric is significant at the 95th (99th) percentile, it is marked with a small (large) dot.

– Reanalysis shows considerable low-frequency variability in the occurrence and persistence of regimes, especially in the NAO- regime. Coupled climate models are unable to reproduce these precise patterns of variability, suggesting they are

545   low-frequency 'weather': chaotic variability driven by the specific oceanic initial conditions of the real world. To assess bias in uninitialised models' regime statistics, it is therefore important to compare against a suitably long historical record, or use an initialised/prescribed ocean state.

– The 3 geopotential-jet regime patterns introduced in DS20 were reproduced in 5 reanalysis products and almost all models considered. They were also found to be more stable, and better captured by models, than any other choice of

550   cluster number, in either of the regime frameworks considered. This makes them particularly suitable for, e.g., inter-model comparisons or analysis of slow atmospheric variability, and suggests that they are capturing a genuine multi-modality of the underlying circulation.

The cross-correlations between model errors in jet and regime dynamics that we have shown in this paper clearly indicate that the jet stream and blocking events are strongly interacting phenomena, and should be understood as parts of a larger coupled

555   system of non-linear variability in the Euro-Atlantic. That is - accurately resolving blocking regimes, and the trimodality
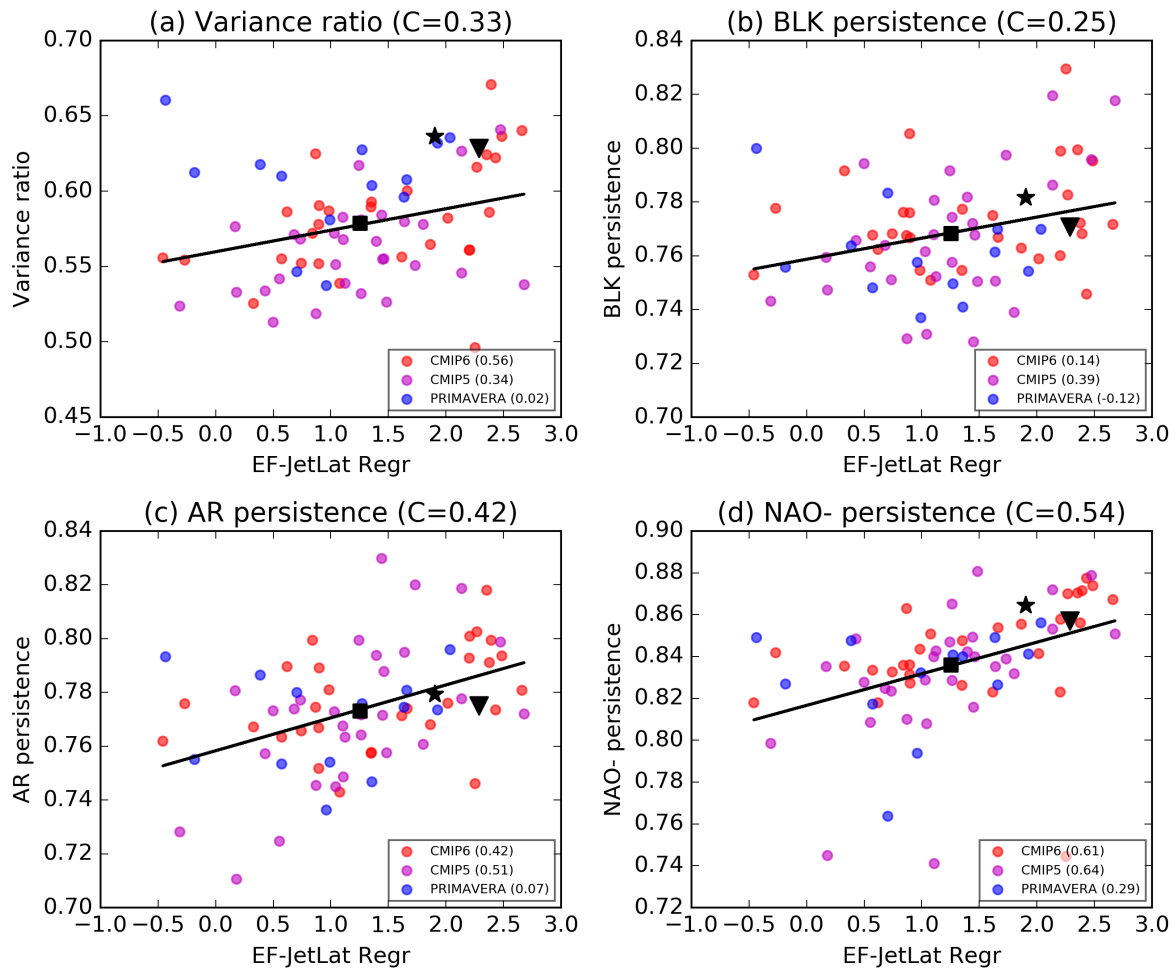
27

Figure 14. The covariation between eddy feedback strength and variance ratio (a) and regime persistence (b-d) in CMIP5, CMIP6 and PRIMAVERA datasets. Dots mark the value for each climate model, while a square indicates the multimodel mean value. A star and triangle mark the values estimated from the ERA5 and ERA20C reanalyses respectively. For models with multiple ensemble members, the ensemble mean values have been used. Trend lines shows the best fit linear relationship for all models, with the Spearman correlation indicated as the value C in each subplot title. The correlation between variables for each individual model dataset is shown in the subplot captions.
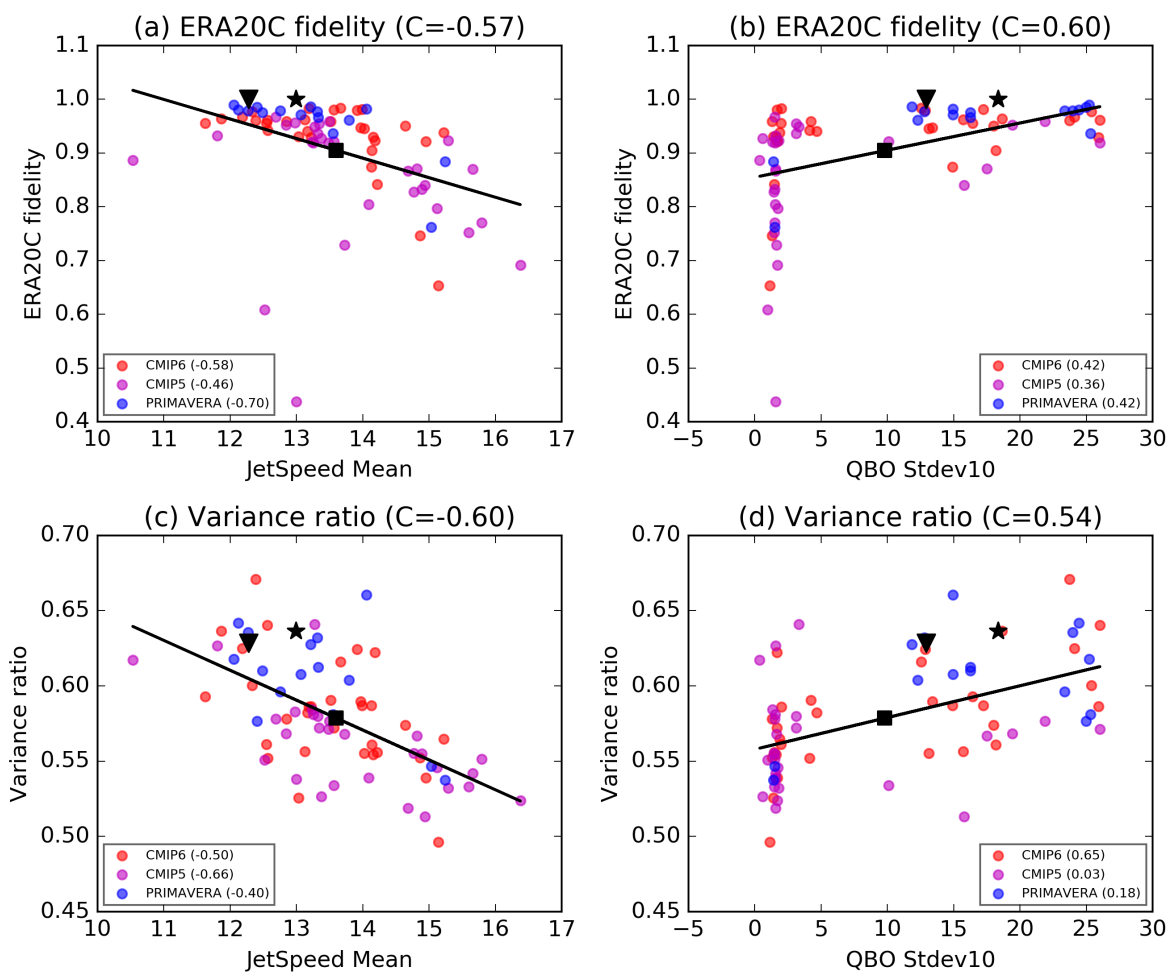
Figure 15. Left: As in Figure 14, but showing the covariation between mean DJF jet speed and regime fidelity and variance ratio. Right: The same, but showing relationship with QBO standard deviation measured at 10Hpa.
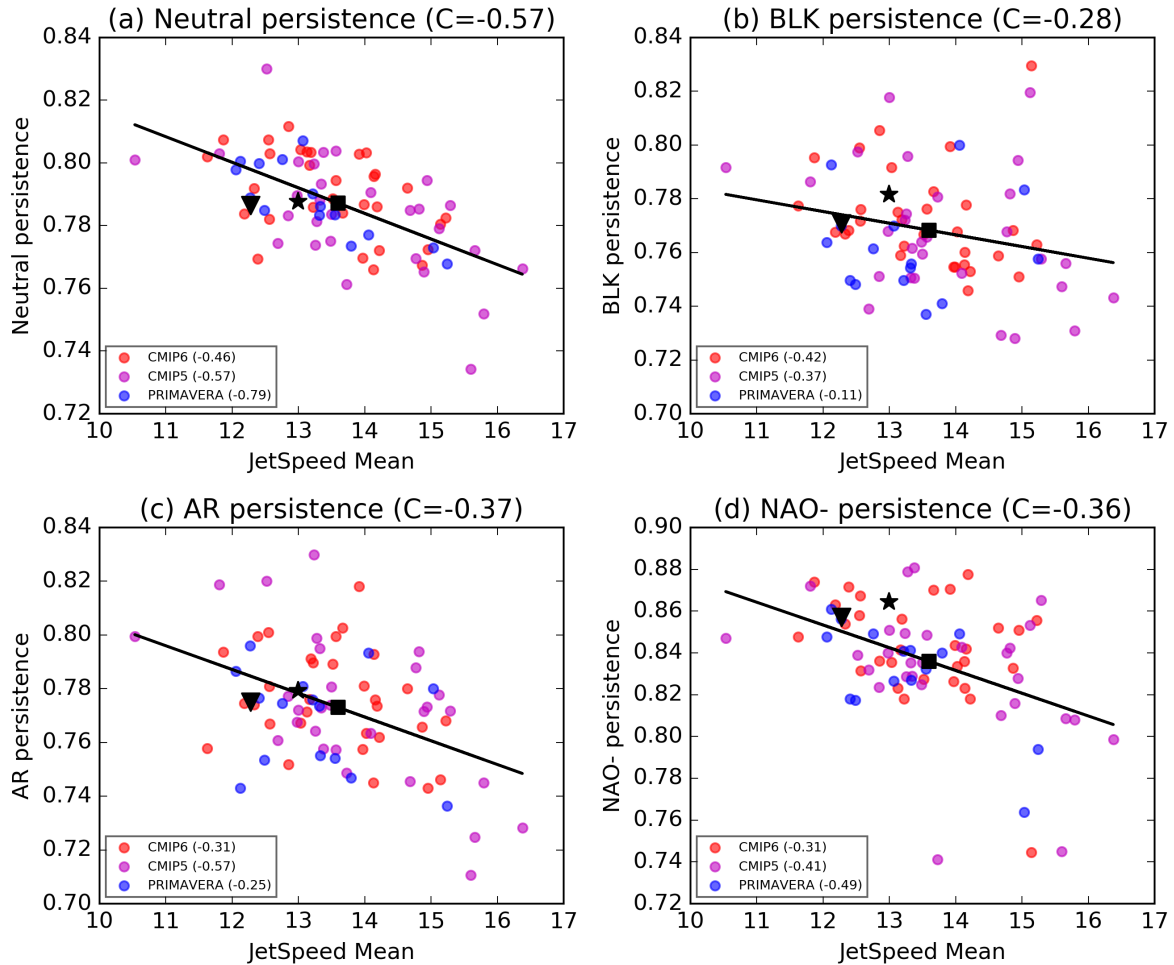
Figure 16. As in Figure 14 but showing the covariation between mean DJF jet speed and regime persistence.

of the jet are two sides of the same coin. However much historical work on the Euro-Atlantic has focused either on blocking regimes or on jet variability individually, with notable exceptions such as Madonna et al. (2017), which leaves many potentially enlightening areas of research – such as how meridional jet shifts causally relate to the onset of blocking and vice versa– under-explored.

560   By adopting the hybrid geopotential-jet regime framework in this paper, we have shown that stable, well-reproduced regimes can be found in climate models, which connect to both shifts in the jet latitude and to anticyclonic blocking events. The increased stability of the regimes reduces the sampling error in regime variability, and so we have been able to obtain confident estimates of model regime bias, and potential causal factors, in a statistically significant fashion. Another benefit of the geopotential-jet framework is that by considering regime variability only along the latitudinal variability of the jet, our results
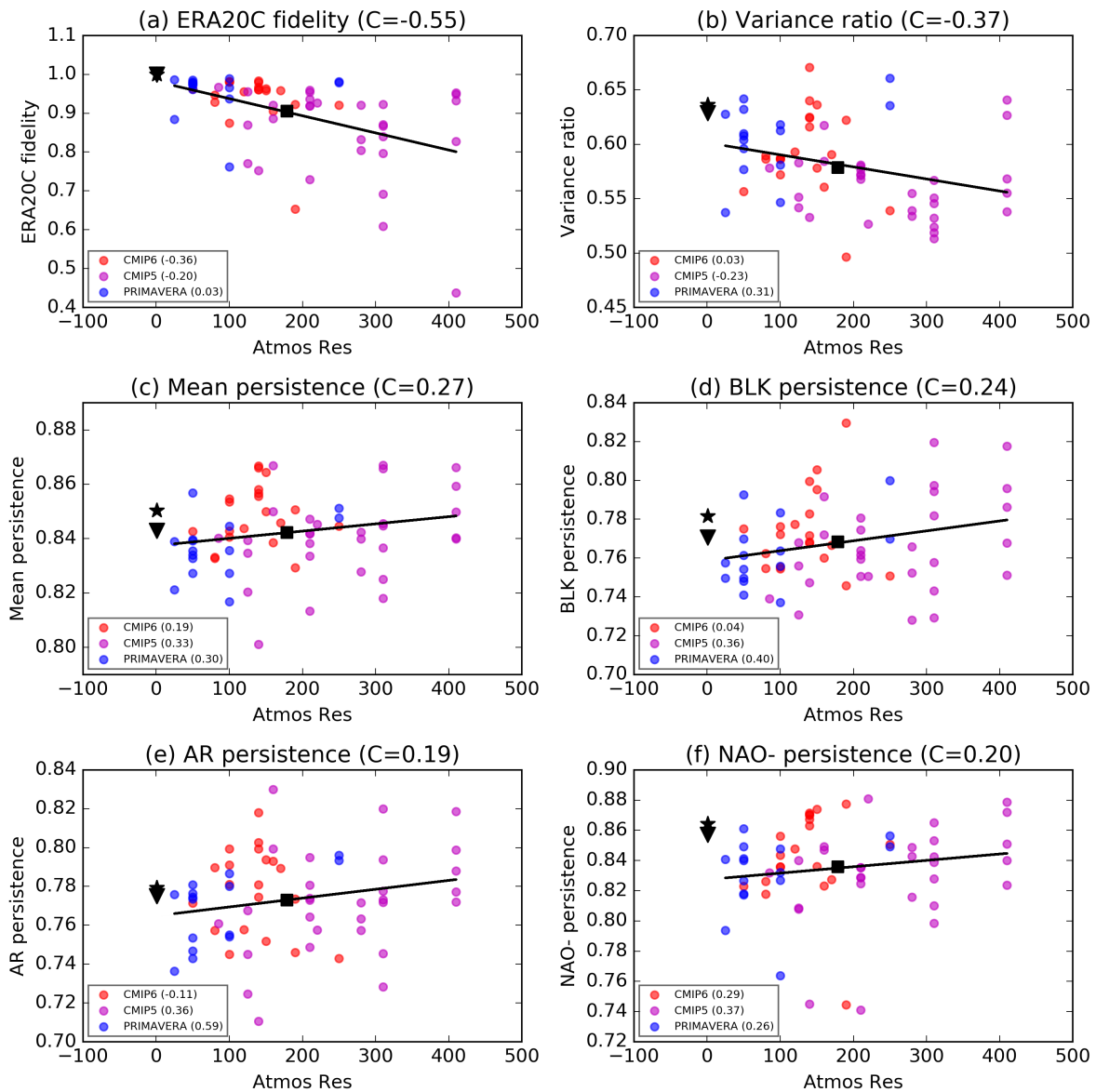
Figure 17. As in Figure 14 but showing the covariation between horizontal model resolution and regime fidelity (a), variance ratio (b) and regime persistence (c-f).

565   can be more easily related to studies considering the jet in a non-regime context. For example, the result of Woollings et al. (2018) relating the latitudinal jet variability and the mean jet speed finds a natural interpretation in our framework: no clear such interpretation seems apparent using the classical circulation regimes. More fundamentally, our framework can be viewed as a natural extension to that of Barnes and Polvani (2013b), with the latitudinal variability of the jet (the 'wobbling') interpreted as regime variability. Because the jet latitude is multimodal, we posit that the use of regime techniques may be crucial

570   to properly account for this variability.

The improvements in regime structure we observe are consistent with improvement in classical regimes in CMIP6 (Fabiano et al., 2020). While some prior regime studies (Dawson et al., 2012; Cattiaux et al., 2013), have suggested climate models often struggle to reproduce the regime patterns found in reanalysis, we find models do fairly well at reproducing geopotential-jet regime structure. Potentially, much of the previously observed poor performance may be a combined result of the higher

575   interdecadal variability present for classical regime patterns, and the use of relatively short reanalysis periods to compare against. When using classical circulation regimes computed with 80+ years of data, or geopotential-jet regimes computed with 30+ years of data, climate models typically achieve very high regime fidelity. The documented bias towards underoccurrence of the BLK regime, and the underpersistence of both BLK and NAO- regimes, are consistent with biases in blocking frequency statistics (Davini and D'Andrea, 2020; Schiemann et al., 2020), demonstrating that these model errors emerge from very

580   different methods of analysis. As long as these persistent model errors remain, there will be a corresponding degradation in the ability of models to accurately represent wintertime extremes. Concretely, with too few persistent Greenland and European blocking events in CMIP6 models, and an overestimation of non-blocked days, we might expect biases towards too few extreme cold events (Brunner et al., 2018), and too many extreme rainfall events (Sousa et al., 2016).

While the impacts of resolution, eddy feedbacks, and mean jet speed on regime structure were particularly clear, we also

585   gained some suggestive insights into stratospheric and oceanic influences on regimes. The variability of a model's QBO correlated significantly with stronger, more realistic, regime structure, and models without a clearly resolved QBO often - but not always - had major deficiencies in their regimes. This suggests that regime structure is amplified by stratospheric teleconnections, either directly through induced changes in the subtropical circulation, or via forcing on the polar vortex - an atmospheric feature we did not examine in this work. That some models with a poor QBO still showed good regime structure could be a

590   result of compensating model errors, or a sign that stratospheric forcings on the troposphere can be adequately parameterised. While some significant impacts of mean November Arctic sea ice and the meridional Gulf Stream SST gradient were seen on regime metrics, and we have conjectured possible mechanisms, the connections are still unclear and more focused work is required in the future to reveal if these are robust indeed correlations, and if so to understand the causal pathways involved. We found that our predictive model could explain more than 30% of the inter-model variance in regime representation for most

595   metrics. The exceptions are regime stability, which was close to 1 for most models, and so did not exhibit a very large amount of inter-model variability, and the persistence and occurrence of the BLK regime. The BLK regime shows the most pronounced model biases in CMIP6, and so understanding the origin of these biases is of key importance. Clearly, the predictive variables we considered are missing some key driver of European blocking dynamics. Given that European blocking is impacted much more by the land-processes and continental orography of Western Europe than the Greenland blocking and Atlantic ridge, it is

600   perhaps not surprising that its behaviour is more complex to understand. One major influence on blocking we have not examined here is the influence of diabatic processes, which are increasingly being recognised as important for blocking development (Pfahl et al., 2015; Steinfeld et al., 2020), and so this may also contribute to the low explicability of BLK dynamics.

There are many avenues for future exploration, and extensions of this current paper. Having obtained a detailed understanding of how well climate models capture historical regimes, a natural next step is to consider future regime changes under
605   anthropogenic warming, and we will analyse this in forthcoming work. Another important open question is whether the exact pattern of regime occurrence and persistence variability observed in reanalysis can be reproduced by AMIP models or fully initialised multidecadal forecasts. Whether such interdecadal fluctuations are chaotic or predictable may have important consequences for decadal forecasting and near-term climate change impacts on Western Europe. The causes of the BLK regime bias need to be isolated before they can be improved, and the reasons for the overly weak eddy feedbacks on the jet - respon-
610   sible for much regime underpersistence - need to be understood and eliminated. There is also work needed on the analysis of surface impacts associated with the geopotential-jet regimes used in this paper, and whether models capture those impacts. Finally, as mentioned, it is likely that to achieve a deeper understanding of Euro-Atlantic variability, it is important to view the eddy-driven jet stream and persistent blocking events as interconnected phenomenon, and there is still much work needed to develop a holistic understanding of how the jet and blocking covary and influence one another.

References

Anstey, J. A., Davini, P., Gray, L. J., Woollings, T. J., Butchart, N., Cagnazzo, C., Christiansen, B., Hardiman, S. C., Osprey, S. M., and Yang, S.: Multi-model analysis of Northern Hemisphere winter blocking: Model biases and the role of resolution, Journal of Geophysical
635   Research: Atmospheres, 118, 3956–3971, https://doi.org/10.1002/JGRD.50231, 2013.

Athanasiadis, P. J., Yeager, S., Kwon, Y.-O., Bellucci, A., Smith, D. W., and Tibaldi, S.: Decadal predictability of North Atlantic blocking and the NAO, NPJ Climate and Atmospheric Science, 3, 1–10, 2020.

Baker, H. S., Woollings, T., and Mbengue, C.: Eddy-Driven Jet Sensitivity to Diabatic Heating in an Idealized GCM, Journal of Climate, 30, 6413 – 6431, https://doi.org/10.1175/JCLI-D-16-0864.1, 2017.

640   Baldwin, M. P. and Dunkerton, T. J.: Stratospheric Harbingers of Anomalous Weather Regimes, Science, 294, 581–584, https://doi.org/10.1126/SCIENCE.1063315, 2001.

Barnes, E. A. and Hartmann, D. L.: Testing a theory for the effect of latitude on the persistence of eddy-driven jets using CMIP3 simulations, Geophysical Research Letters, 37, https://doi.org/https://doi.org/10.1029/2010GL044144, 2010.

Barnes, E. A. and Polvani, L.: Response of the Midlatitude Jets, and of Their Variability, to Increased Greenhouse Gases in the CMIP5
645   Models, Journal of Climate, 26, 7117–7135, https://doi.org/10.1175/JCLI-D-12-00536.1, 2013a.

Barnes, E. A. and Polvani, L.: Response of the Midlatitude Jets, and of Their Variability, to Increased Greenhouse Gases in the CMIP5 Models, Journal of Climate, 26, 7117 – 7135, https://doi.org/10.1175/JCLI-D-12-00536.1, 2013b.

Barnes, E. A. and Screen, J. A.: The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it?, Wiley Interdisciplinary Reviews: Climate Change, 6, 277–286, https://doi.org/10.1002/wcc.337, 2015.

650   Beerli, R. and Grams, C. M.: Stratospheric modulation of the large-scale circulation in the Atlantic-European region and its implications for surface weather events, Quarterly Journal of the Royal Meteorological Society, 145, 3732–3750, https://doi.org/10.1002/qj.3653, 2019.

Bellucci, A., Athanasiadis, P. J., Scoccimarro, E., Ruggieri, P., Gualdi, S., Fedele, G., Haarsma, R. J., Garcia-Serrano, J., Castrillo, M., Putrahasan, D., Sanchez-Gomez, E., Moine, M.-P., Roberts, C. D., Roberts, M. J., Seddon, J., and Vidale, P. L.: Air-Sea interaction over the Gulf Stream in an ensemble of HighResMIP present climate simulations, Climate Dynamics 2021 56:7, 56, 2093–2111,
655   https://doi.org/10.1007/S00382-020-05573-Z, 2021a.

Bellucci, A., Athanasiadis, P. J., Scoccimarro, E., Ruggieri, P., Gualdi, S., Fedele, G., Haarsma, R. J., Garcia-Serrano, J., Castrillo, M., Putrahasan, D., Sanchez-Gomez, E., Moine, M. P., Roberts, C. D., Roberts, M. J., Seddon, J., and Vidale, P. L.: Air-Sea interaction over the Gulf Stream in an ensemble of HighResMIP present climate simulations, Climate Dynamics, 56, 2093–2111, https://doi.org/10.1007/s00382-020-05573-z, 2021b.

660   Brayshaw, D. J., Hoskins, B., and Blackburn, M.: The Basic Ingredients of the North Atlantic Storm Track. Part I: Land–Sea Contrast and Orography, Journal of the Atmospheric Sciences, 66, 2539–2558, https://doi.org/10.1175/2009JAS3078.1, 2009.

Brunner, L., Schaller, N., Anstey, J., Sillmann, J., and Steiner, A. K.: Dependence of Present and Future European Temperature Extremes on the Location of Atmospheric Blocking, Geophysical Research Letters, 45, 6311–6320, https://doi.org/10.1029/2018GL077837, 2018.

Butchart, N., Anstey, J. A., Hamilton, K., Osprey, S., McLandress, C., Bushell, A. C., Kawatani, Y., Kim, Y.-H., Lott, F., Scinocca, J.,
665   Stockdale, T. N., Andrews, M., Bellprat, O., Braesicke, P., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Dobrynin, M., Garcia, R. R., Garcia-Serrano, J., Gray, L. J., Holt, L., Kerzenmacher, T., Naoe, H., Pohlmann, H., Richter, J. H., Scaife, A. A., Schenzinger, V., Serva, F., Versick, S., Watanabe, S., Yoshida, K., and Yukimoto, S.: Overview of experiment design and comparison of models participating in phase 1 of the

SPARC Quasi-Biennial Oscillation initiative (QBOi), Geoscientific Model Development, 11, 1009–1032, https://doi.org/10.5194/gmd-11-1009-2018, 2018.

670   Cassou, C.: Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation, Nature, 455, 523–527, https://doi.org/10.1038/nature07286, 2008.

Cattiaux, J., Douville, H., and Peings, Y.: European temperatures in CMIP5: origins of present-day biases and future uncertainties, Climate Dynamics 2013 41:11, 41, 2889–2907, https://doi.org/10.1007/S00382-013-1731-Y, 2013.

Charlton-Perez, A. J., Ferranti, L., and Lee, R. W.: The influence of the stratospheric state on North Atlantic weather regimes, Quarterly

675   Journal of the Royal Meteorological Society, 144, 1140–1151, https://doi.org/10.1002/qj.3280, 2018.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model, Journal of Advances in Modeling Earth Systems, 11, 185–209, https://doi.org/10.1029/2018MS001369, 2019.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin,

680   P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, https://doi.org/10.1002/QJ.776, 2011.

Corti, S., Molteni, F., and Palmer, T. N.: Signature of recent climate change in frequencies of natural atmospheric circulation regimes, Nature, 398, 799–802, https://doi.org/10.1038/19745, 1999.

685   Crouse, D. F.: On implementing 2D rectangular assignment algorithms, IEEE Transactions on Aerospace and Electronic Systems, 52, 1679–1696, https://doi.org/10.1109/TAES.2016.140952, 2016.

Davini, P. and D'Andrea, F.: Northern Hemisphere atmospheric blocking representation in global climate models: Twenty years of improvements?, Journal of Climate, 29, 8823–8840, https://doi.org/10.1175/JCLI-D-16-0242.1, 2016.

Davini, P. and D'Andrea, F.: From CMIP3 to CMIP6: Northern hemisphere atmospheric blocking simulation in present and future climate,

690   Journal of Climate, 33, 10 021–10 038, https://doi.org/10.1175/JCLI-D-19-0862.1, 2020.

Davini, P., Fabiano, F., and Sandu, I.: Orographic resolution driving the improvements associated with horizontal resolution increase in the Northern Hemisphere winter mid-latitudes, https://doi.org/10.5194/wcd-2021-51.

Dawson, A. and Palmer, T. N.: Simulating weather regimes: impact of model resolution and stochastic parameterization, Climate Dynamics, 44, 2177–2193, https://doi.org/10.1007/s00382-014-2238-x, 2015.

695   Dawson, A., Palmer, T. N., and Corti, S.: Simulating regime structures in weather and climate prediction models, Geophysical Research Letters, 39, 1–6, https://doi.org/10.1029/2012GL053284, 2012.

Delworth, T. L. and Zeng, F.: The impact of the North Atlantic Oscillation on climate through its influence on the Atlantic meridional overturning circulation, Journal of Climate, 29, 941–962, https://doi.org/10.1175/JCLI-D-15-0396.1, 2016.

Delworth, T. L., Zeng, F., Zhang, L., Zhang, R., Vecchia, G. A., and Yang, X.: The central role of ocean dynamics in connecting the

700   North Atlantic oscillation to the extratropical component of the Atlantic multidecadal oscillation, Journal of Climate, 30, 3789–3805, https://doi.org/10.1175/JCLI-D-16-0358.1, 2017.

Deser, C., Tomas, R. A., and Peng, S.: The transient atmospheric circulation response to North Atlantic SST and sea ice anomalies, Journal of Climate, https://doi.org/10.1175/JCLI4278.1, 2007.

Domeisen, D. I. V., Grams, C. M., and Papritz, L.: The role of North Atlantic–European weather regimes in the surface impact of sudden

705   stratospheric warming events, Weather and Climate Dynamics, 1, 373–388, https://doi.org/10.5194/wcd-1-373-2020, 2020.

Dorrington, J. and Strommen, K. J.: Jet Speed Variability Obscures Euro-Atlantic Regime Structure, Geophysical Research Letters, 47, https://doi.org/10.1029/2020gl087907, 2020.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M., and Knight, J.: Skilful predictions of the winter North Atlantic Oscillation one year ahead, Nature Geoscience 2016 9:11, 9, 809–814, https://doi.org/10.1038/ngeo2824, 2016.

710 Eichelberger, S. J. and Hartmann, D. L.: Zonal Jet Structure and the Leading Mode of Variability, Journal of Climate, 20, 5149–5163, https://doi.org/10.1175/JCLI4279.1, 2007.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/GMD-9-1937-2016, 2016.

715 Fabiano, F., Christensen, ·. H. M., Strommen, ·. K., Athanasiadis, ·. P., Baker, ·. A., Schiemann, ·. R., and Corti, ·. S.: Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance, 54, 5031–5048, https://doi.org/10.1007/s00382-020-05271-w, 2020.

Fabiano, F., Meccia, V. L., Davini, P., Ghinassi, P., and Corti, S.: A regime view of future atmospheric circulation changes in northern mid-latitudes, Weather and Climate Dynamics, 2, 163–180, https://doi.org/10.5194/wcd-2-163-2021, 2021.

720 Franzke, C., Woollings, T., and Martius, O.: Persistent Circulation Regimes and Preferred Regime Transitions in the North Atlantic, Journal of the Atmospheric Sciences, 68, 2809–2825, https://doi.org/10.1175/JAS-D-11-046.1, 2011.

Gerber, E. P. and Vallis, G. K.: On the Zonal Structure of the North Atlantic Oscillation and Annular Modes, Journal of the Atmospheric Sciences, 66, 332–352, https://doi.org/10.1175/2008JAS2682.1, 2009.

Ghil, M.: Climate Change: Multidecadal and Beyond, vol. 6, World Scientific Publ. Co./Imperial College Press, 725 https://doi.org/10.1142/9789814579933_0002, 2015.

Gray, L. J., Anstey, J. A., Kawatani, Y., Lu, H., Osprey, S., and Schenzinger, V.: Surface impacts of the Quasi Biennial Oscillation, Atmospheric Chemistry and Physics, 18, 8227–8247, https://doi.org/10.5194/ACP-18-8227-2018, 2018.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., Haak, H., and Stössel, A.: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP), Geoscientific Model 730 Development, 12, 3241–3281, https://doi.org/10.5194/gmd-12-3241-2019, 2019.

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.-A. B., Caron, L.-P., 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, F. M., Rodrigo, M., Ruprich-Robert, Y., Le Sager, P., Tourigny, E., Wang, S., van Weele, M., and Wyser, K.: HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR. Description, model performance, data handling and validation, Geoscientific 735 Model Development Discussions, 2020, 1–37, https://doi.org/10.5194/gmd-2019-350, 2020.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

740 Harvey, B. J., Cook, P., Shaffrey, L. C., and Schiemann, R.: The Response of the Northern Hemisphere Storm Tracks and Jet Streams to Climate Change in the CMIP3, CMIP5, and CMIP6 Climate Models, Journal of Geophysical Research: Atmospheres, 125, https://doi.org/10.1029/2020JD032701, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hoerl, A. E. and Kennard, R. W.: Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, 55–67, https://doi.org/10.1080/00401706.1970.10488634, 1970.

Hoskins, B. J., James, I. N., and White, G. H.: The shape, propagation and mean-flow interaction of large-scale weather systems., Journal of the Atmospheric Sciences, https://doi.org/10.1175/1520-0469(1983)040<1595:TSPAMF>2.0.CO;2, 1983.

Jiménez-Esteve, B. and Domeisen, D. I.: The tropospheric pathway of the ENSO-North Atlantic teleconnection, Journal of Climate, 31, 4563–4584, https://doi.org/10.1175/JCLI-D-17-0716.1, 2018.

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., and Gray, L. J.: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather, Nature Geoscience, 8, 433–440, https://doi.org/10.1038/ngeo2424, 2015.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophysical Research Letters, 40, 1194–1199, https://doi.org/10.1002/GRL.50256, 2013.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J. R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H., Kosaka, Y., Martin, M., Poli, P., Rayner, N., Rustemeier, E., and Schepers, D.: CERA-20C: A Coupled Reanalysis of the Twentieth Century, Journal of Advances in Modeling Earth Systems, 10, 1172–1195, https://doi.org/10.1029/2018MS001273, 2018.

Lee, S. and Kim, H. K.: The dynamical relationship between subtropical and eddy-driven jets, Journal of the Atmospheric Sciences, https://doi.org/10.1175/1520-0469(2003)060<1490:TDRBSA>2.0.CO;2, 2003.

Li, C. and Wettstein, J. J.: Thermally driven and eddy-driven jet variability in reanalysis, Journal of Climate, 25, 1587–1596, https://doi.org/10.1175/JCLI-D-11-00145.1, 2012.

Lorenz, D. J. and Hartmann, D. L.: Eddy-zonal flow feedback in the Northern Hemisphere winter, Journal of Climate, https://doi.org/10.1175/1520-0442(2003)16<1212:EFFITN>2.0.CO;2, 2003.

Madonna, E., Li, C., Grams, C. M., and Woollings, T.: The link between eddy-driven jet variability and weather regimes in the North Atlantic-European sector, Quarterly Journal of the Royal Meteorological Society, 143, 2960–2972, https://doi.org/10.1002/qj.3155, 2017.

Michelangeli, P.-A., Vautard, R., Legras, B., Michelangeli, P.-A., Vautard, R., and Legras, B.: Weather Regimes: Recurrence and Quasi Stationarity, Journal of the Atmospheric Sciences, 52, 1237–1256, https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2, 1995.

Nakamura, N. and Huang, C. S. Y.: Atmospheric blocking as a traffic jam in the jet stream, Science, 361, 42–47, https://doi.org/10.1126/SCIENCE.AAT0721, 2018.

O'Reilly, C. H., Minobe, S., and Kuwano-Yoshida, A.: The influence of the Gulf Stream on wintertime European blocking, Climate Dynamics, 47, 1545–1567, https://doi.org/10.1007/s00382-015-2919-0, 2016.

Oudar, T., Cattiaux, J., and Douville, H.: Drivers of the Northern Extratropical Eddy-Driven Jet Change in CMIP5 and CMIP6 Models, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086695, 2020.

Parker, T., Woollings, T., Weisheimer, A., O'Reilly, C., Baker, L., and Shaffrey, L.: Seasonal Predictability of the Winter North Atlantic
780 Oscillation From a Jet Stream Perspective, Geophysical Research Letters, 46, 10 159–10 167, https://doi.org/10.1029/2019GL084402,
2019.

Pfahl, S., Schwierz, C., Croci-Maspoli, M., Grams, C. M., and Wernli, H.: Importance of latent heat release in ascending air streams for
atmospheric blocking, Nature Geoscience 2014 8:8, 8, 610–614, https://doi.org/10.1038/ngeo2487, 2015.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet,
785 Y., Hólm, E. V., Bonavita, M., Isaksen, L., Fisher, M., Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F.,
Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C:
An Atmospheric Reanalysis of the Twentieth Century, Journal of Climate, 29, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1,
2016.

Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., and Simpson, I. R.: Progress in Simu-
790 lating the Quasi-Biennial Oscillation in CMIP Models, Journal of Geophysical Research: Atmospheres, 125, e2019JD032 362,
https://doi.org/https://doi.org/10.1029/2019JD032362, e2019JD032362 10.1029/2019JD032362, 2020.

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., and Keeley, S. P. E.: Climate model configurations of the ECMWF
Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP, Geoscientific Model Development, 11, 3681–3712,
https://doi.org/10.5194/gmd-11-3681-2018, 2018.

795 Robin, Y., Yiou, P., and Naveau, P.: Detecting changes in forced climate attractors with Wasserstein distance, Nonlinear Processes in Geo-
physics, 24, 393–405, https://doi.org/10.5194/NPG-24-393-2017, 2017.

Robinson, W. A.: Does eddy feedback sustain variability in the zonal index?, Journal of the Atmospheric Sciences,
https://doi.org/10.1175/1520-0469(1996)053<3556:DEFSVI>2.0.CO;2, 1996.

Rodríguez-Fonseca, B., Suárez-Moreno, R., Ayarzagüena, B., López-Parages, J., Gómara, I., Villamayor, J., Mohino, E., Losada,
800 T., and Castaño-Tierno, A.: A Review of ENSO Influence on the North Atlantic. A Non-Stationary Signal, Atmosphere, 7, 87,
https://doi.org/10.3390/atmos7070087, 2016.

Scaife, A. A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M., MacLachlan, C., Martin, N., Nie, Y., Ren, H.-L., Roberts, M.,
Robinson, W., Smith, D., and Vidale, P. L.: Does increased atmospheric resolution improve seasonal climate predictions?, Atmospheric
Science Letters, 20, e922, https://doi.org/10.1002/ASL.922, 2019.

805 Schiemann, R., Athanasiadis, P., Barriopedro, D., Doblas-Reyes, F., Lohmann, K., Roberts, M. J., Sein, D. V., Roberts, C. D., Terray, L., and
Vidale, P. L.: Northern Hemisphere blocking simulation in current climate models: evaluating progress from the Climate Model Intercom-
parison Project Phase 5 to 6 and sensitivity to resolution, Weather and Climate Dynamics, 1, 277–292, https://doi.org/10.5194/WCD-1-
277-2020, 2020.

Sein, D. V., Koldunov, N. V., Danilov, S., Wang, Q., Sidorenko, D., Fast, I., Rackow, T., Cabos, W., and Jung, T.: Ocean Modeling
810 on a Mesh With Resolution Following the Local Rossby Radius, Journal of Advances in Modeling Earth Systems, 9, 2601–2614,
https://doi.org/10.1002/2017MS001099, 2017.

Shutts, G. J.: The propagation of eddies in diffluent jetstreams: Eddy vorticity forcing of 'blocking' flow fields, Quarterly Journal of the
Royal Meteorological Society, 109, 737–761, https://doi.org/10.1002/qj.49710946204, 1983.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer,
815 L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D.,
Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T. C., Lorrey,

A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3, Journal of Climate, 34, 1417–1438, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

820 Small, R. J., Msadek, R., Kwon, Y.-O., Booth, J. F., and Zarzycki, C.: Atmosphere surface storm track response to resolved ocean mesoscale in two sets of global climate model experiments, Climate Dynamics, 52, 2067–2089, 2019.

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M. M., Ramos, A. M., and Liberato, M. L. R.: Responses of European precipitation distributions and regimes to different blocking locations, Climate Dynamics 2016 48:3, 48, 1141–1160, https://doi.org/10.1007/S00382-016-3132-5, 2016.

825 Steinfeld, D., Boettcher, M., Forbes, R., and Pfahl, S.: The sensitivity of atmospheric blocking to upstream latent heating – numerical experiments, Weather and Climate Dynamics, 1, 405–426, https://doi.org/10.5194/WCD-1-405-2020, 2020.

Straus, D. M. and Molteni, F.: Circulation regimes and SST forcing: Results from large GCM ensembles, Journal of Climate, https://doi.org/10.1175/1520-0442(2004)017<1641:CRASFR>2.0.CO;2, 2004.

Strommen, K.: Jet latitude regimes and the predictability of the North Atlantic Oscillation, Quarterly Journal of the Royal Meteorological
830 Society, n/a, https://doi.org/10.1002/qj.3796, 2020.

Strommen, K., Mavilia, I., Corti, S., Matsueda, M., Davini, P., Hardenberg, J., Vidale, P., and Mizuta, R.: The Sensitivity of Euro-Atlantic Regimes to Model Horizontal Resolution, Geophysical Research Letters, 46, 7810–7818, https://doi.org/10.1029/2019GL082843, 2019a.

Strommen, K., Mavilia, I., Corti, S., Matsueda, M., Davini, P., von Hardenberg, J., Vidale, P. L., and Mizuta, R.: The Sensitivity of Euro-Atlantic Regimes to Model Horizontal Resolution, Geophysical Research Letters, 46, 7810–7818, https://doi.org/10.1029/2019GL082843,
835 2019b.

Strong, C. and Magnusdottir, G.: Dependence of NAO variability on coupling with sea ice, Climate Dynamics, https://doi.org/10.1007/s00382-010-0752-z, 2011.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

840 Vissio, G., Lembo, V., Lucarini, V., and Ghil, M.: Evaluating the Performance of Climate Models Based on Wasserstein Distance, Geophysical Research Letters, 47, https://doi.org/10.1029/2020GL089385, 2020.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., 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,
845 E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, Journal of Advances in Modeling Earth Systems, 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Wang, L., Ting, M., and Kushner, P. J.: A robust empirical seasonal prediction of winter NAO and surface climate, Scientific Reports, 7, 279, https://doi.org/10.1038/s41598-017-00353-y, 2017.

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., and Palmer, T.: Atmospheric seasonal forecasts of the twentieth century: multi-
850 decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution, Quarterly Journal of the Royal Meteorological Society, 143, 917–926, https://doi.org/10.1002/qj.2976, 2017.

White, R. H., Hilgenbrink, C., and Sheshadri, A.: The Importance of Greenland in Setting the Northern Preferred Position of the North Atlantic Eddy-Driven Jet, Geophysical Research Letters, 46, 14 126–14 134, https://doi.org/10.1029/2019GL084780, 2019.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder,
855   P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann,
R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: 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, 357–380,
https://doi.org/10.1002/2017MS001115, 2018.

Woollings, T. and Blackburn, M.: The North Atlantic Jet Stream under Climate Change and Its Relation to the NAO and EA Patterns, Journal
860   of Climate, 25, 886 – 902, https://doi.org/10.1175/JCLI-D-11-00087.1, 2012.

Woollings, T., Hoskins, B., Blackburn, M., and Berrisford, P.: A New Rossby Wave-Breaking Interpretation of the North Atlantic Oscillation,
https://doi.org/10.1175/2007JAS2347.1, 2008.

Woollings, T., Hannachi, A., and Hoskins, B.: Variability of the North Atlantic eddy-driven jet stream, Quarterly Journal of the Royal
Meteorological Society, 136, 856–868, https://doi.org/10.1002/qj.625, 2010.

865   Woollings, T., Czuchnicki, C., and Franzke, C.: Twentieth century North Atlantic jet variability, Quarterly Journal of the Royal Meteorolog-
ical Society, 140, 783–791, https://doi.org/10.1002/qj.2197, 2014.

Woollings, T., Barnes, E., Hoskins, B., Kwon, Y. O., Lee, R. W., Li, C., Madonna, E., McGraw, M., Parker, T., Rodrigues, R., Spensberger,
C., and Williams, K.: Daily to decadal modulation of jet variability, Journal of Climate, 31, 1297–1314, https://doi.org/10.1175/JCLI-D-
17-0286.1, 2018.

870   Zhang, W., Kirtman, B., Siqueira, L., Clement, A., and Xia, J.: Understanding the signal-to-noise paradox in decadal climate predictability
from CMIP5 and an eddying global coupled model, Climate Dynamics, 56, 2895–2913, https://doi.org/10.1007/s00382-020-05621-8,
2021.