

**RESEARCH ARTICLE**

# Year-round sub-seasonal forecast skill for Atlantic–European weather regimes

Dominik Büeler<sup>1</sup>  | Laura Ferranti<sup>2</sup>  | Linus Magnusson<sup>2</sup>  |  
Julian F. Quinting<sup>1</sup>  | Christian M. Grams<sup>1</sup> 

<sup>1</sup>Department Troposphere Research, Institute of Meteorology and Climate Research (IMK-TRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

<sup>2</sup>European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

**Correspondence**

D. Büeler, Institute of Meteorology and Climate Research (IMK-TRO), Karlsruhe Institute of Technology, Postfach 3640, 76021 Karlsruhe, Germany.  
Email: dominik.bueeler@kit.edu

**Funding information**

Helmholtz-Gemeinschaft, Grant/Award Number: VH-NG-1243

**Abstract**

Weather regime forecasts are a prominent use case of sub-seasonal prediction in the midlatitudes. A systematic evaluation and understanding of year-round sub-seasonal regime forecast performance is still missing, however. Here we evaluate the representation of and forecast skill for seven year-round Atlantic–European weather regimes in sub-seasonal reforecasts from the European Centre for Medium-Range Weather Forecasts. Forecast calibration improves regime frequency biases and forecast skill most strongly in summer, but scarcely in winter, due to considerable large-scale flow biases in summer. The average regime skill horizon in winter is about 5 days longer than in summer and spring, and 3 days longer than in autumn. The Zonal Regime and Greenland Blocking tend to have the longest year-round skill horizon, which is driven by their high persistence in winter. The year-round skill is lowest for the European Blocking, which is common for all seasons but most pronounced in winter and spring. For the related, more northern Scandinavian Blocking, the skill is similarly low in winter and spring but higher in summer and autumn. We further show that the winter average regime skill horizon tends to be enhanced following a strong stratospheric polar vortex (SPV), but reduced following a weak SPV. Likewise, the year-round average regime skill horizon tends to be enhanced following phases 4 and 7 of the Madden–Julian Oscillation (MJO) but reduced following phase 2, driven by winter but also autumn and spring. Our study thus reveals promising potential for year-round sub-seasonal regime predictions. Further model improvements can be achieved by reduction of the considerable large-scale flow biases in summer, better understanding and modeling of blocking in the European region, and better exploitation of the potential predictability provided by weak SPV states and specific MJO phases in winter and the transition seasons.

**KEYWORDS**

blocking, Europe, European Centre for Medium-Range Weather Forecasts, Madden–Julian Oscillation, stratospheric polar vortex, sub-seasonal forecast, weather regimes, year-round

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

## 1 | INTRODUCTION

Numerical weather prediction has been improving substantially during the last decades (Bauer *et al.*, 2015). This is the result of a continuous increase in computational power, improvements in operational ensemble modeling and data assimilation systems, and a better understanding of atmospheric processes driving predictability (Vitart, 2014; Bauer *et al.*, 2015). Nevertheless, sub-seasonal weather forecasts (>15 days ahead) for the extratropics—particularly the Atlantic–European region—still have only moderate to weak skill on average (e.g., Buizza and Leutbecher, 2015; Son *et al.*, 2020). This is due primarily to the theoretical intrinsic predictability limit on synoptic scales of about two weeks, which results from the chaotic nature of the atmosphere even with near-perfect forecast models and initial and boundary conditions (Lorenz, 1963). The fast upscale error growth in the imperfect state-of-the-art models thus reduces this (practical) predictability limit further (e.g., Zhang *et al.*, 2019). However, the predictability limit does not necessarily apply to various lower frequency planetary-scale phenomena. These can modulate the large-scale circulation in the midlatitudes substantially and thus provide intrinsic predictability well into sub-seasonal lead times (e.g., Palmer, 1993; Hoskins, 2013). Translating this intrinsic sub-seasonal predictability into forecast skill is one of today's major challenges of numerical weather prediction, because it requires identification and filling of the gaps between intrinsic and practical predictability limits on these different coupled spatiotemporal scales.

The extratropical variability on these larger atmospheric scales can be depicted by weather regimes, which are quasistationary, persistent, and recurrent large-scale flow patterns in the midlatitudes (Vautard, 1990; Michelangeli *et al.*, 1995). Weather regimes modulate surface weather strongly on continental and multiday to weekly scales and thus have substantial socio-economic impacts. For instance, particularly persistent regimes regularly lead to cold spells in winter and heatwaves in summer (e.g., Yiou and Nogaj, 2004; Ferranti *et al.*, 2018; Schaller *et al.*, 2018; Spensberger *et al.*, 2020), which are often associated with enhanced mortality (e.g., Huang *et al.*, 2020). Their modulation of near-surface wind, temperature, or solar irradiation affects the energy industry further via fluctuations in electricity production, demand, and prices (e.g., Grams *et al.*, 2017; Beerli and Grams, 2019; van der Wiel *et al.*, 2019). On the other hand, the categorization of the large-scale circulation into weather regimes is helpful for better understanding and improving of sub-seasonal predictability—not just from a physical point of view, but also because it condenses the large amount of data generated by sub-seasonal models.

A comprehensive knowledge of the sub-seasonal forecast skill of state-of-the-art models in predicting weather regimes is thus inevitable both for operational forecasters and model developers.

Regarding the Atlantic–European region, previous studies have primarily investigated forecast skill for the well-established, classic set of four weather regimes during winter (e.g., Ferranti *et al.*, 2015; 2018; Matsueda and Palmer, 2018): the positive and negative phases of the North Atlantic Oscillation (NAO+ and NAO–), a blocking anticyclone over Northern Europe (commonly called European, Scandinavian, or Euro-Atlantic Blocking), and a blocking anticyclone over the North Atlantic (called Atlantic Ridge). A common finding of all these studies is the higher skill for the two NAO phases than for the blocking-type regimes, for both medium-range (<15 days ahead) and sub-seasonal lead times. The longest skill horizon has been found for the NAO–, which results from its relatively high persistence and thus high intrinsic predictability (Ferranti *et al.*, 2018; Matsueda and Palmer, 2018; Lin, 2020), but also from the frequent and well-modeled transition from a blocking over Scandinavia into NAO– via cyclonic Rossby wave breaking (Michel and Rivière, 2011; Ferranti *et al.*, 2018). On the other hand, the lower forecast skill for the blocking-type regimes (particularly over Europe) is a result of their lower intrinsic predictability (Faranda *et al.*, 2016; Hochman *et al.*, 2021), but might also emerge from some of their underlying physical processes, which models still struggle to capture properly. These processes occur on various spatial scales, ranging from latent heat release in meso- to synoptic-scale systems (Rodwell *et al.*, 2013; 2018; Grams and Archambault, 2016; Grams *et al.*, 2018) to Rossby-wave propagation on larger scales (Quinting and Vitart, 2019). The errors associated with these processes lead to biases in the transitions into blocking-type regimes, and, once they are active, to biases in their persistence (Ferranti *et al.*, 2015; 2018; Matsueda and Palmer, 2018). Further research on blocking dynamics on sub-seasonal timescales is thus inevitable to overcome some of these problems and improve sub-seasonal forecasts for blocking over Europe.

The differences in sub-seasonal forecast skill are caused by not only differences in the internal dynamics of the regimes but also their sensitivity to lower-frequency phenomena governing sub-seasonal predictability to first order. For the Atlantic–European region, two important such phenomena are the winter stratospheric polar vortex (SPV) and the Madden–Julian Oscillation (MJO: Madden and Julian, 1971; 1972). Their strong modulation primarily of the NAO is thus a further reason for the better model performance regarding the NAO regimes: anomalously strong states of the SPV are often followed by relatively persistent large-scale states resembling the NAO+,

while anomalously weak SPV states tend to be followed by persistent large-scale states resembling the NAO– (e.g., Baldwin and Dunkerton, 2001; Ambaum and Hoskins, 2002; Domeisen, 2019). This stratosphere–troposphere coupling generally enhances sub-seasonal forecast skill for the NAO (e.g., Tripathi *et al.*, 2015; Scaife *et al.*, 2016; Charlton-Perez *et al.*, 2018; Feng *et al.*, 2021). Nevertheless, models still struggle to predict the correct surface weather response over Europe, particularly following weak SPV states (e.g., Büeler *et al.*, 2020; Kolstad *et al.*, 2020; Domeisen *et al.*, 2020a). Likewise, enhanced MJO convection in the tropical Western Pacific (phases 6–7 of the real-time multivariate MJO index by Wheeler and Hendon, 2004) is statistically followed by NAO–, and an enhanced MJO convection in the Indian Ocean (phases 3–4) by NAO+ (e.g., Cassou, 2008; Lin *et al.*, 2009). The MJO thus also enhances sub-seasonal forecast skill for the NAO (e.g., Vitart and Molteni, 2010; Ferranti *et al.*, 2018; Feng *et al.*, 2021). Other studies have investigated how further regimes, beside the NAO, are modulated by the SPV and MJO (Cassou, 2008; Charlton-Perez *et al.*, 2018; Lee *et al.*, 2019; 2020a). Some of these studies considered the modulation of a higher number of regimes as investigated in our article (see below; Klaus, 2017; Beerli and Grams, 2019; Domeisen *et al.*, 2020b). Apart from the SPV and MJO, further lower-frequency phenomena such as the El Niño–Southern Oscillation (e.g., Toniazzo and Scaife, 2006; Jiménez-Estève and Domeisen, 2018; Yamagami and Matsueda, 2020), the Quasi-Biennial Oscillation (e.g., Anstey and Shepherd, 2014), and variations in sea-surface temperature (e.g., Rodwell *et al.*, 1999), soil moisture (e.g., Koster *et al.*, 2010), snow cover (e.g., Orsolini *et al.*, 2016), and sea-ice cover (e.g., Alexander *et al.*, 2004) can also modulate regime evolution. The role of most of these phenomena, however, has mainly been investigated for winter but is not well understood for the other seasons.

Despite their prevalence in atmospheric research, the four Atlantic–European weather regimes by construction only account for a part of atmospheric variability (Grams *et al.*, 2020). Their usability for predicting surface-weather-related parameters on a regional scale for socio-economic sectors such as the energy industry can thus be limited (e.g., Bloomfield *et al.*, 2020). Furthermore, the four regimes change their characteristics considerably when defined for different seasons, which is why they are often investigated for individual seasons in isolation. In this study, we thus investigate the sub-seasonal forecast performance in predicting a novel set of seven year-round Atlantic–European weather regimes by Grams *et al.* (2017), which have been shown to offer certain benefits compared with the four classic regimes: they can explain sub-seasonal surface weather modulation in Europe better in situations in which the four regimes are too coarse

to do so (Beerli and Grams, 2019; Grams *et al.*, 2020; Domeisen *et al.*, 2020b). Quantifying their forecast skill will thus provide us with a refined view of the problems (and strengths) of state-of-the-art sub-seasonal models. The year-round definition of the regimes will allow further for a much more systematic analysis of the large-scale flow throughout the year, which is crucial to fill the surprisingly sparsely investigated but highly important sub-seasonal forecast skill in summer and in the transition seasons (see Cortesi *et al.*, 2021, as one recent study addressing this gap). Despite these advantages, the higher number of regimes comes with the inevitable trade-off of reduced sample sizes per regime (see also, e.g., Neal *et al.*, 2016) and the possibility of a slightly lower intrinsic sub-seasonal predictability compared with a lower number of regimes. A systematic investigation of these trade-offs would be important, but goes beyond the scope of our study.

Our article addresses these objectives as follows: Section 2 briefly describes the sub-seasonal reforecast data from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the verifying reanalysis used in this study (Section 2.1), introduces the set of seven Atlantic–European weather regimes and describes how they are identified in the forecasts (Section 2.2), and describes the forecast verification scores and statistical tests used (Section 2.3). Section 3 contains a multifaceted verification of these regime forecasts, focusing on four different research questions: how do large-scale flow biases and their removal (i.e., forecast calibration) affect regime occurrence in the forecast (Section 3.1); how can the regime frequency biases remaining in the calibrated forecasts be explained by biases in regime life-cycle duration, number, and transitions (Section 3.2); what is the sub-seasonal forecast skill for the different seasons and regimes (Section 3.3); and to what extent do a modification of the verified lead-time window (Section 3.4.1) as well as lower-frequency phenomena such as the SPV and MJO (Section 3.4.2) serve as windows of opportunity for enhanced sub-seasonal regime forecast skill? The article ends with Section 4, in which we summarize and conclude the main findings and provide ideas for further research.

## 2 | DATA AND METHODS

### 2.1 | Model and reanalysis

We analyse 21 years (1997–2017) of sub-seasonal reforecasts (i.e., forecasts recomputed from an initial date in the past and initialized with reanalysis data, hereafter just denoted “forecasts”) from the ECMWF provided through the Subseasonal-to-Seasonal (S2S) Prediction Project Database (Vitart *et al.*, 2017; note that we plan to extend our analysis to further S2S models in a future study). The

forecasts have been initialized from ERA-Interim (Dee *et al.*, 2011) twice per week with 11 ensemble members (1 control and 10 perturbed forecasts) and run up to a lead time of 46 days. We increase this initialisation frequency by including different model versions (CY43R1, CY43R3, and CY45R1, implemented on November 22, 2016, July 11, 2017, and June 6, 2018, respectively), which add forecasts starting from additional calendar days and yield a total of 4,080 forecasts (first initialisation on January 2, 1997, last initialization on December 13, 2017). The horizontal grid spacing of the atmosphere (16 km before and 32 km after a lead time of 15 days), the number of vertical levels (91), and the horizontal grid spacing of the ocean ( $0.25^\circ$ ) are the same throughout these model versions. The forecast data, more specifically daily instantaneous (0000 UTC) geopotential height at 500 hPa, has been retrieved from the database with a horizontal grid spacing of  $1^\circ$  (the remapping to this grid is done automatically during the retrieval process). As the reforecasts have been initialized from ERA-Interim, we use this dataset with the same horizontal grid spacing (also remapped during the retrieval) as a verification dataset. Using the new successor reanalysis dataset ERA5 (Hersbach *et al.*, 2020) instead would very likely not affect the results of our study, because the two reanalyses should largely be similar with respect to the midtropospheric large-scale patterns investigated in our study.

## 2.2 | Weather regimes

As mentioned in Section 1, our study is based on a novel set of seven Atlantic–European weather regimes (Grams *et al.*, 2017). This section first explains how the climatological mean patterns of these seven regimes are defined based on ERA-Interim and, second, how each time step in the forecast (and the corresponding ERA-Interim time step for verification) is assigned to one of these regimes.

The climatological mean weather regime patterns are defined based on the full ERA-Interim period (1979–2018) as follows (slightly adapted from Grams *et al.*, 2017): we compute six-hourly 500-hPa geopotential height anomalies with respect to the corresponding 91-day running mean calendar date climatologies (i.e.,  $\pm 45$  days centered around each 6-hr time step). The anomalies are filtered with a five-day low-pass filter and seasonally normalized.<sup>1</sup> The seasonal normalization is the key step for

the year-round regime definition, because it overcomes the substantially weaker anomalies in summer than in winter. We then apply an empirical orthogonal function (EOF) analysis to the filtered and seasonally normalized anomalies within the North-Atlantic–European domain from  $80^\circ\text{W}$  to  $40^\circ\text{E}$  and  $30^\circ$  to  $90^\circ\text{N}$  (this domain is consistent with other studies: e.g., Michelangeli *et al.*, 1995; Ferranti *et al.*, 2015). Finally, a *k*-means clustering is applied to the anomalies in the phase space spanned by the first seven EOFs (explaining approximately 70% of the variance), which yields an optimal number of seven cluster means representing the seven weather regimes. Figure 1 shows the mean 500-hPa geopotential height anomalies corresponding to these cluster means: there are three “cyclonic regimes”, the Atlantic Trough (AT), the Zonal Regime (ZO), and the Scandinavian Trough (ScTr), in which a negative geopotential height anomaly associated with enhanced cyclonic activity dominates. They correlate with the positive phase of the NAO to different degrees, with the ZO being most similar (see figure 2a in Beerli and Grams, 2019). The residual four regimes, the Atlantic Ridge (AR), European Blocking (EuBL), Scandinavian Blocking (ScBL), and Greenland Blocking (GL), are referred to as “blocking regimes” with a dominating positive geopotential height anomaly. AR largely corresponds to the equally named regime in the classic regime definition (e.g., Michelangeli *et al.*, 1995; Ferranti *et al.*, 2015), GL strongly resembles the negative phase of the NAO (Beerli and Grams, 2019), and EuBL and ScBL can be seen as two different variations of the classic blocking regime (e.g., Michelangeli *et al.*, 1995; Ferranti *et al.*, 2015).

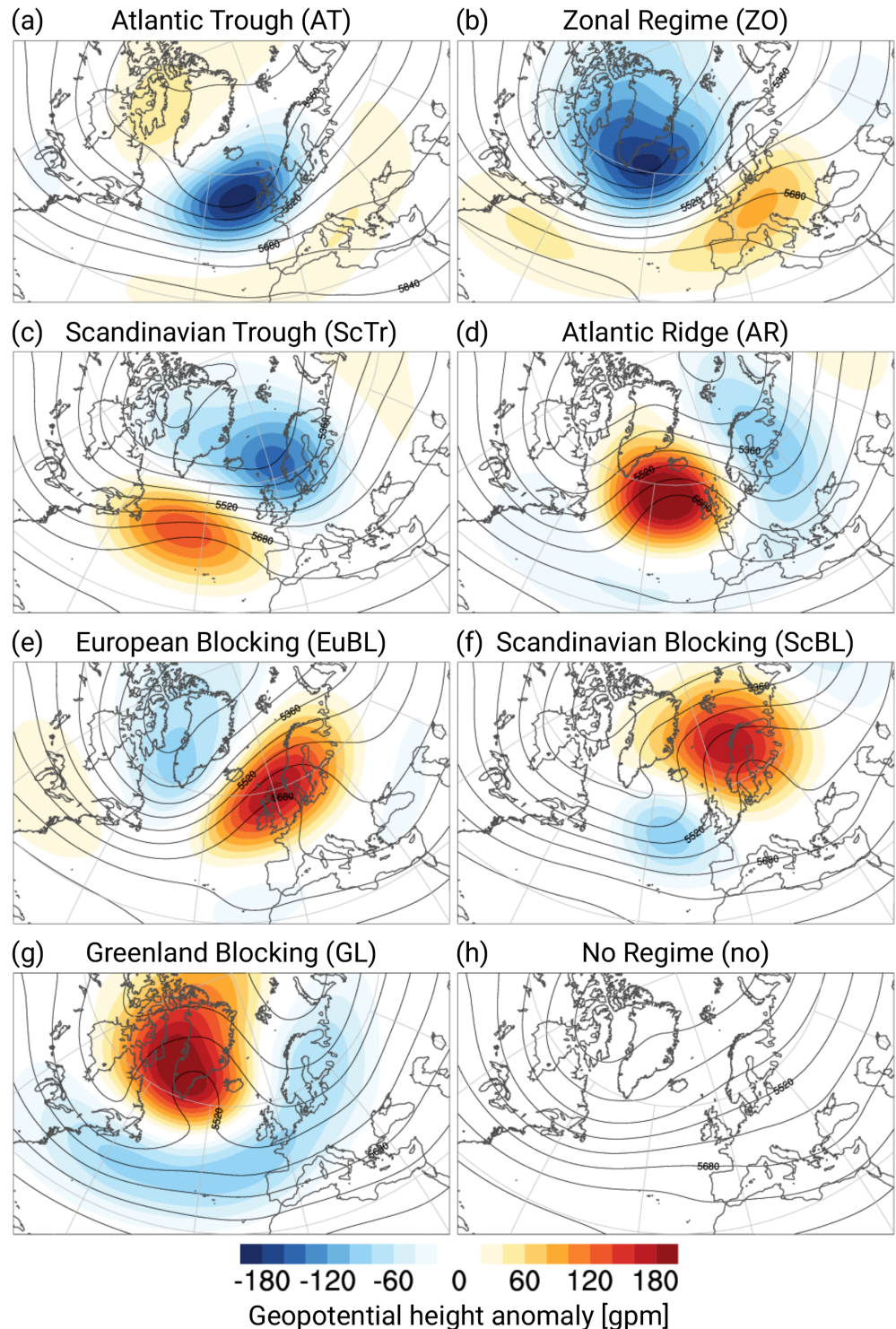
Following Grams *et al.* (2017), we then identify the active weather regime life cycle in the forecast objectively: pursuing the same principle as for the regime definition (cf. above), we first compute the instantaneous low-pass-filtered and seasonally normalized 500-hPa geopotential height anomalies for each ensemble member and at each lead time step. We do this for two sets of forecasts—calibrated and noncalibrated.

To obtain the calibrated forecasts, we remove the forecast biases from the geopotential height anomalies by computing the underlying geopotential height calendar day climatology over the set of forecasts (as a 91-day running mean over all ensemble members of all 21 years between 1997 and 2017, i.e., over a reduced period compared with the one the climatological mean regime patterns are based on) for each of the 46 lead time steps separately (to account for any kind of model drift). This yields a 46-day-long climatology “vector” for each calendar day for which a forecast initialization is available,

<sup>1</sup>The seasonal normalization is achieved by dividing the low-pass-filtered geopotential height anomaly at each grid point by a calendar-day-dependent scalar that quantifies the climatological variability of geopotential height anomalies at the corresponding calendar day. This scalar is computed as the spatial average (over all grid

points in the investigated domain; cf. later) of the temporal 31-day running standard deviation over all anomalies between 1979 and 2018.

**FIGURE 1** Cluster mean 500-hPa geopotential height (contours; gpm, i.e., geopotential meters) and corresponding anomalies (shading; gpm) of the seven year-round Atlantic–European weather regimes (defined based on ERA-Interim data between 1979 and 2018) and the “no regime” category (see Section 2.2 for details) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



which in principle is consistent with other studies based on the S2S dataset (e.g., Vitart, 2017; Büeler *et al.*, 2020). Consistent with the removal of the geopotential height bias, we also remove the bias in the scalar for the seasonal normalization (cf. footnote 1) by computing it based on the (low-pass-filtered) geopotential height anomalies of the calibrated forecasts in the reduced forecast period and for each lead time instead of ERA-Interim.

In contrast, the noncalibrated forecasts are obtained by computing the anomalies based on the geopotential height calendar day climatology over the ERA-Interim fields (over the reduced period between 1997 and 2017 as well), and, consistently, by normalizing them seasonally with the scalar based on ERA-Interim in the reduced period. Although the calibrated forecasts are the basis for most of the regime verification in this article (Sections 3.2, 3.3,

and 3.4), we also discuss how the geopotential height biases in the noncalibrated forecasts are characterized, how they are linked to regime frequency biases, and how regime frequency biases are reduced in the calibrated forecasts with the geopotential height biases being removed.

Once the instantaneous filtered and normalized geopotential height anomalies (calibrated and noncalibrated) have been computed for each forecast, we project something onto something the seven cluster mean geopotential height anomalies (Figure 1) following the method of Michel and Rivière (2011):

$$P_{wr}(t) = \frac{1}{\sum_{(\lambda,\phi)} \cos \phi} \sum_{(\lambda,\phi)} \Phi(\lambda, \phi, t) \Phi_{wr}(\lambda, \phi) \cos \phi. \quad (1)$$

$P_{wr}(t)$  is a scalar measure for the spatial correlation of the instantaneous anomaly field  $\Phi(\lambda, \phi, t)$  at lead time  $t$  (at each grid point with latitude  $\lambda$  and longitude  $\phi$  within the EOF domain) with the cluster mean anomaly field  $\Phi_{wr}(\lambda, \phi)$  for the regime  $wr$  (i.e., the cluster mean anomalies shown in Figure 1). Following Michel and Rivière (2011), we then compute a nondimensional regime index  $I_{wr}(t)$  for each regime and forecast based on anomalies of the projections  $P_{wr}(t)$  (with respect to the climatological mean projection  $\overline{P_{wr}}$ ) that are normalized with the climatological standard deviation of the projection (over all available forecasts  $i$ ):

$$I_{wr}(t) = \frac{P_{wr}(t) - \overline{P_{wr}}}{\sqrt{\frac{1}{N} \sum_{i=1}^N [P_{wr}(t) - \overline{P_{wr}}]^2}}. \quad (2)$$

Note that both the climatological mean projection  $\overline{P_{wr}}$  and the standard deviation of the projection in the denominator of Equation 2 are computed based on the set of forecasts for the calibrated forecasts and on ERA-Interim for the noncalibrated forecasts (to be consistent with the bias corrections described above). Finally, to determine the active weather regime at each lead time step  $t$ , we apply a set of so-called life-cycle criteria to the evolution of  $I_{wr}(t)$  (see Grams *et al.*, 2017 for further details): a regime is active if its  $I_{wr}(t)$  is maximum among all seven  $I_{wr}(t)$  and equal to or above 0.9<sup>2</sup> for five consecutive days or longer. These life-cycle criteria consequently introduce a “no regime” category for those time steps at which none of the seven regimes fulfils the criteria. Compared with previous studies, in which the active regime is determined at each lead time step (based on

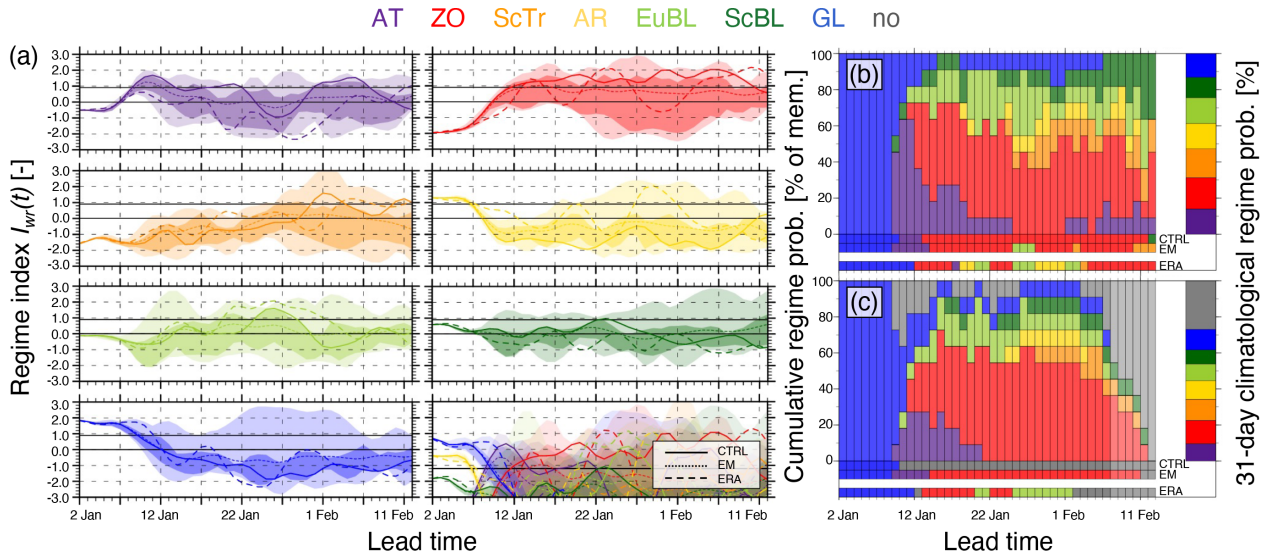
either minimum distances between principal components within the EOF space or maximum agreement between patterns in physical space: e.g., Ferranti *et al.*, 2015; Neal *et al.*, 2016; Matsueda and Palmer, 2018), our life-cycle definition has the following advantages: the persistence criterion prevents sudden jumps in the regime attribution. Together with the minimum regime index threshold, the method allows us further to define sufficiently strong and physically meaningful life-cycle objects with objectively identified onset, maximum, and decay stages. This enables an in-depth analysis of regime life-cycle characteristics such as duration, number, and transitions.

Figure 2 illustrates the forecast products that can be generated from these different intermediate steps for an example forecast initialized on January 2, 1997: Figure 2a shows the synchronous evolution of  $I_{wr}(t)$  for the seven regimes in the ensemble. It gives an overview of the evolution of the dominating and suppressed regimes with lead time, as well as the associated forecast spread. Building upon Figure 2a, Figure 2b shows the ensemble forecast probability for a certain  $I_{wr}(t)$  to be maximum. As a final forecast product, Figure 2c indicates the ensemble forecast probability for a certain regime to be active after applying the aforementioned life-cycle criteria. This latter forecast product is the basis for most of the forecast verification presented in this study. Note that the verification will only be done up to a lead time of 32 days, due to the loss of data at the end of each forecast, which results from both the low-pass filtering and a convergence to the “no regime” category due to the life-cycle persistence criterion (see light-shaded lead times in Figure 2c). This lead time is still enough, considering the weak skill on these timescales.

Apart from the forecast calibration used in this study and described above, we have tested another more flow-dependent calibration technique, which removes the climatological regime index biases from each of the seven regime indices  $I_{wr}(t)$  (instead of removing only one mean 500-hPa geopotential height bias) before determining the regime life cycles. However, this more sophisticated technique does not change the effect of forecast calibration on the regime forecasts, which is why we stick with the widely used and more easily applicable standard calibration technique in this study.

To verify the forecasts, we additionally identify the weather regime evolution in their corresponding 46-day ERA-Interim periods, following the same principle as above (this means we treat ERA-Interim like an additional ensemble member, i.e., the perfect model forecast, against which we can verify the forecast). This ensures a fair verification of the forecasts because ERA-Interim is truncated by the same amount of data at the end of the 46-day period due to the low-pass filtering. The only difference

<sup>2</sup>This threshold is the only difference compared with the life-cycle criteria in Grams *et al.* (2017) and has been defined such that the year-round climatological frequency of the “no regime” category equals approximately 30%, which is the residual amount of variance that is not explained by the first seven EOFs.

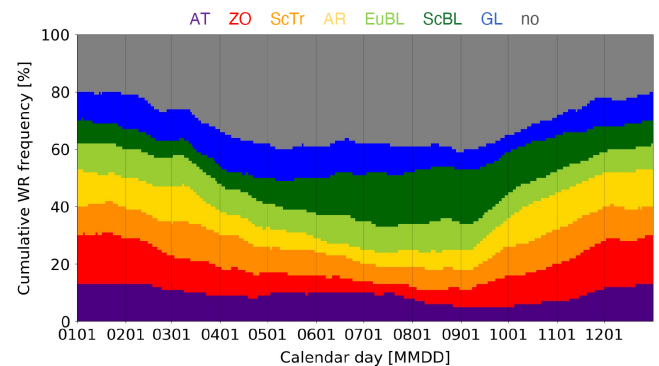


**FIGURE 2** (a) Weather regime index  $I_{wr}(t)$  (Equation 2) distribution, (b) cumulative probability of life cycles based on maximum regime index  $I_{wr}(t)$ , and (c) cumulative probability of life cycles based on life-cycle criteria (bottom) as a function of lead time ( $x$ -axes) for an example ensemble reforecast initialized on January 2, 1997, with the colors representing the eight regime categories shown in the legend (see text for details). The ensemble spread in (a) is divided into the interquartile range (dark shading) and the minimum and maximum, respectively (light shading). The bottom right of (a) further shows the overlay of the seven regime indices. The control forecast (CTRL), ensemble mean forecast (EM), and corresponding ERA-Interim period (ERA) are further indicated by the solid, dotted, and dashed lines in (a) and in the three bottom rows of (b) and (c). The 31-day running mean calendar day frequency in ERA-Interim at the end of the forecast is indicated on the right of (b) and (c). Details about the light-shaded lead times at the end of the forecast in (c) are given in the text [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

for the regime identification in ERA-Interim is that the underlying 500-hPa geopotential height climatology, the scalar for normalizing the geopotential height anomalies seasonally, and the climatological mean projection  $\overline{P_{wr}}$  to obtain  $I_{wr}(t)$  (cf. above) are all based on ERA-Interim data between 1997 and 2017 (i.e., the same as for the non-calibrated forecast) instead of forecast data. The resulting year-round climatological frequency of the seven regimes (including the “no regime” category) in ERA-Interim is shown in Figure 3: the relative frequencies are distributed more homogeneously among the regimes in winter compared with summer. Nevertheless, the cyclonic regimes tend to dominate in winter, whereas the blocked regimes are prevalent in summer. Among the cyclonic regimes, the Zonal Regime is the dominant one in winter, whereas the Atlantic Trough is dominant in summer. Likewise, the European Blocking dominates among the continental blocking regimes in winter, whereas the Scandinavian Blocking is the most frequent continental blocking regime in summer. The “no regime” category is more frequent in summer than in winter.

### 2.3 | Skill scores

To verify the categorical ensemble forecast probability of weather regime life-cycle occurrence at each lead time



**FIGURE 3** 91-day running mean calendar day climatological cumulative relative frequency of weather regime life cycles in ERA-Interim (defined over the investigated period of 1997–2017 only) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(cf. Figure 2c), we use the fair Brier score (BS; Ferro *et al.*, 2008; equation 3 in Ferro, 2014):

$$BS = \frac{1}{N} \sum_{k=1}^N \sum_{wr \in WR} \left( (y_k^{wr} - o_k^{wr})^2 - \frac{m y_k^{wr} (m - m y_k^{wr})}{m^2 (m - 1)} \right). \quad (3)$$

The fair BS is the classic BS (Brier, 1950; Wilks, 2011), which is the squared difference between the predicted probability  $y_k^{wr}$  (between 0 and 1) for regime  $wr$  of forecast  $k$  and the corresponding observed dichotomous value  $o_k^{wr}$

(0 or 1), minus a correction term that accounts for the relatively small ensemble member size  $m = 11$ . This whole term is then averaged over the  $N$  forecasts ( $N = 4,080$  if all forecasts are verified) to obtain the fair BS. Note that Equation 3 can be used to compute the single-category BS for an individual regime (i.e., with  $WR = AT, ZO, ScTr, AR, EuBL, ScBL, GL, \text{ or } no$ ) but also the multicategory BS for all regimes together (i.e., with  $WR = \{AT, ZO, ScTr, AR, EuBL, ScBL, GL, no\}$ ). Finally, we compute the fair Brier skill score (hereafter just referred to as BSS; Wilks, 2011) to relate the fair BS of the numerical model forecast to the BS of a climatological reference forecast ( $BS_{ref}$ ; note that  $BS_{ref}$  is not corrected):

$$BSS = \frac{BS - BS_{ref}}{0 - BS_{ref}}. \quad (4)$$

As a reference forecast, we use the 91-day running mean climatological calendar day regime frequency in ERA-Interim (Figure 3) at each lead time step. Using the fair instead of the classic BSS has a substantial effect on the skill horizon when verifying reforecasts from the S2S database with relatively low numbers of ensemble members (see Figure S1 in the Supporting Information showing the difference in year-round average skill for the regimes investigated here). A disadvantage of the fair BS or BSS, respectively, is that no decomposition into reliability, resolution, and uncertainty has been defined yet (personal communication by Christopher Ferro, University of Exeter), as exists for the classic BS (Wilks, 2011).

We also verify the continuous weather regime index  $I_{wr}$  at each lead time of the forecast, for which we use the fair continuous ranked probability score (CRPS; Ferro *et al.*, 2008; equation 4 in Fricker *et al.*, 2013):

$$CRPS = \frac{1}{N|WR|} \sum_{k=1}^N \sum_{wr \in WR} \int_{-\infty}^{+\infty} \left( P_{fc}^{k,wr}(x) - P_{obs}^{k,wr}(x) \right)^2 dx - \frac{1}{2m^2(m-1)} \sum_{i \neq j} \left| x_i^{k,wr} - x_j^{k,wr} \right|. \quad (5)$$

The fair CRPS is the classic CRPS (first term after the integral), which is the squared difference between the predicted and observed (empirical) cumulative distribution functions  $P_{fc}^{k,wr}(x)$  and  $P_{obs}^{k,wr}(x)$  of a forecast variable  $x$  (of forecast  $k$  and regime  $wr$ ), respectively, minus a correction term that accounts for the small ensemble size  $m = 11$  (second term after the integral). The cumulative distribution functions can both be expressed as Heaviside functions  $H(x)$  (with  $H(x) = 0$  for  $x < 0$  and  $H(x) = 1$  for  $x \geq 0$ ), namely  $P_{fc}^{k,wr}(x) = \frac{1}{m} \sum_{i=1}^m H(x - x_i^{k,wr})$  and  $P_{obs}^{k,wr}(x) = H(x - x_{obs}^{k,wr})$ . In our case,  $x_i^{k,wr}$  is the predicted regime

index  $I_{wr}$  of member  $i$  and forecast  $k$ , and  $x_{obs}^{k,wr}$  is the corresponding verifying observation in ERA-Interim. Note that the members  $i$  are sorted according to their value  $x_i^{k,wr}$  prior to computing  $P_{fc}^{k,wr}(x)$ . Similarly to the BS, we can use Equation 5 to compute either the single-category CRPS for one regime or the multicategory CRPS for all regimes together, with the latter simply being the average over the single-category CRPSs of the individual regimes.

Finally, the fair continuous ranked probability skill score (CRPSS) is obtained by relating the CRPS of the forecast to the  $CRPS_{ref}$  of the corresponding reference forecast in the same way as for the fair BSS (Equation 4). As a basis to compute the  $CRPS_{ref}$ , we use a “climatological ensemble” consisting of all  $I_{wr}$  values (i.e., “members”) in ERA-Interim within a 91-day running window centered around the calendar day of the verified lead time step (the same window definition as for obtaining the  $BS_{ref}$ ; cf. above).

Evaluating the forecast performance for specific flow situations can lead to relatively small forecast samples. To account for the robustness of the skill scores associated with these small samples, we apply a bootstrapping to all skill-score computations in this study (in addition to computing the actual skill score for each forecast sample). More specifically, we randomly resample (with replacement)  $10^4$  times a set of forecasts of the same size as the evaluated set of forecasts and compute the skill score for each of these random samples. We then define the actual skill scores of two forecast groups to be significantly different at the 5% level if their confidence intervals between the 5th and 95th percentiles—derived from these skill score distributions—do not overlap. Following the same principle, we also determine whether biases in climatological regime occurrence frequencies or transition frequencies are significant. The bias itself is computed as the difference between the regime occurrence (or transition) frequency in the forecast and in ERA-Interim. If the confidence interval between the 5th and 95th percentiles of the regime occurrence (or transition) frequency obtained with the bootstrapping in the forecast does not overlap with the confidence interval in ERA-Interim, the bias is defined to be significant at the 5% level.

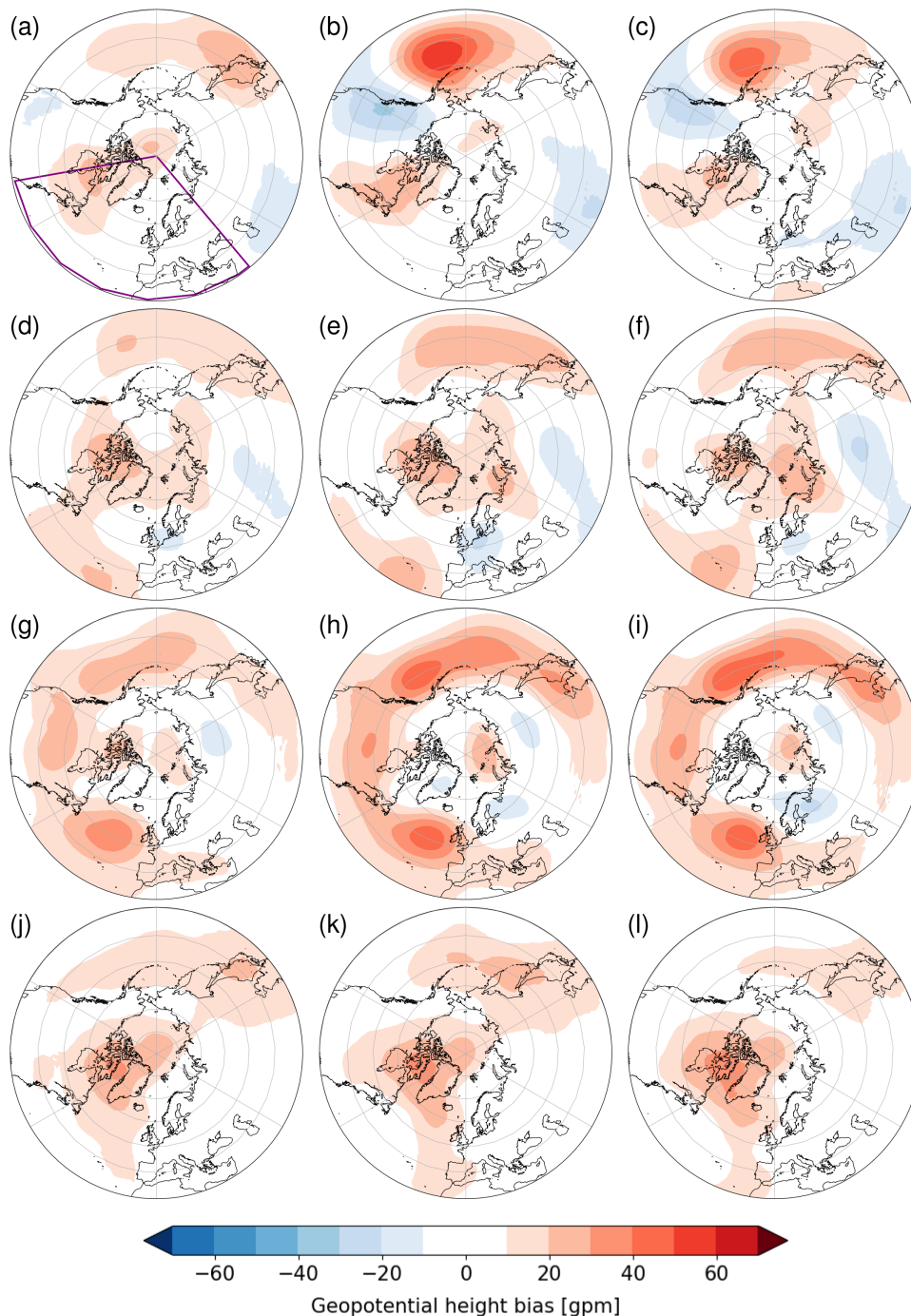
## 3 | RESULTS

### 3.1 | Role of forecast calibration for weather-regime frequency biases and forecast skill

We first demonstrate the link between large-scale circulation biases and the representation of weather regimes

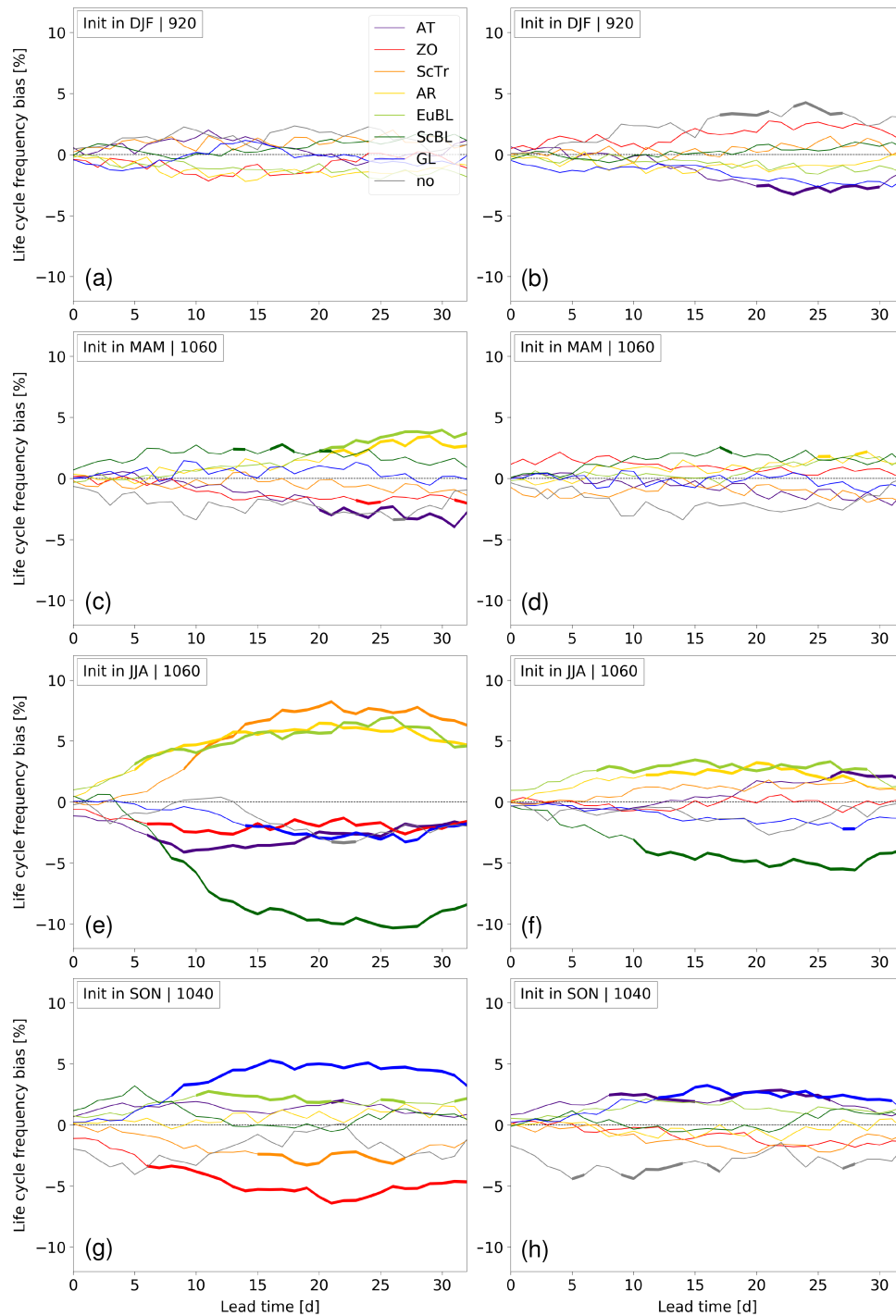


**FIGURE 4** 500-hPa geopotential height model climatology biases (gpm; of noncalibrated forecasts) in the Northern Hemisphere for forecasts initialized on (a–c) January 1, (d–f) April 2, (g–i) July 2, and (j–l) October 1 at 10 (left), 20 (middle), and 30 days lead time (right). The purple box indicates the EOF domain in which the weather regimes are defined (see text for details) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



in the forecasts. Figure 4 shows the lead-time-dependent 500-hPa geopotential height biases of the forecast climatology with respect to the ERA-Interim climatology on a calendar day centered in each season (i.e., the biases of the noncalibrated forecast, which are removed later on to obtain the calibrated forecast; cf. Section 2.2 for details). In general, the biases increase in the medium range (left column) and tend to saturate at sub-seasonal lead times (middle and right columns). There are substantial differences between the seasons: in the Atlantic–European region, the biases are smallest in winter, with weak positive values

over the North American east coast and parts of Greenland. The positive biases increase slightly in spring and cover most of the polar cap and the Atlantic region around the Azores. The biases maximize in summer with substantial positive values (up to 40 geopotential meters (gpm)) in the central North Atlantic. In autumn, they become similar to those in spring but with smaller magnitudes over the Pacific and Atlantic. The Atlantic–European domain, in which our weather regimes are defined (box in Figure 4a, cf. Section 2.2), is thus affected by the biases primarily in summer and scarcely in winter. Substantial biases can



**FIGURE 5** Seasonal weather regime life-cycle frequency biases (%) y-axis) in the noncalibrated (left) and calibrated forecasts (right) with respect to ERA-Interim as a function of lead time (d; x-axis). Bold lines indicate significant biases. The seasons and the corresponding available numbers of forecasts are indicated in the boxes. Note that ERA-Interim is treated like a "perfect ensemble member" to obtain the bias: the life-cycle occurrence in each ensemble forecast is compared against the occurrence in its corresponding 46-day ERA-Interim period, which means that the same date in ERA-Interim appears several times but at different lead times for different forecasts [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

also be found in the North Pacific and North American regions, with the highest values in winter and summer. Although this upstream region might play a crucial role in the dynamics of Atlantic–European weather regimes (e.g., Rivière and Orlanski, 2007; Michel and Rivière, 2011; Michel *et al.*, 2012; Rivière and Drouard, 2015), the role of the corresponding biases is not discussed here and remains a topic for further research.

The geopotential height biases in the Atlantic–European domain are mutually linked to the

weather-regime frequency biases: Figure 5 shows the seasonal life-cycle frequency biases as a function of lead time in the noncalibrated forecasts (i.e., without the geopotential height biases removed) and calibrated forecasts (i.e., with the geopotential height biases removed) with respect to ERA-Interim for each regime. The frequency biases in the noncalibrated forecasts (Figure 5, left column) correspond closely to the behavior of the geopotential height biases (Figure 4): the frequency biases are negligible in winter, with nonsignificant values of

a few percent (Figure 5a), and largest in summer, with significant absolute values of up to 10% (Figure 5e). In spring and autumn, the biases range in between the values for winter and summer (Figures 5c,g). Furthermore, the biases tend to saturate beyond 15–20 days, which is consistent with the geopotential height biases. To give an example, the largest frequency biases of the anticorrelated ScTr and ScBL in summer (Figure 5e) can be understood from the positive geopotential height bias in the central North Atlantic (Figures 4g,i). The latter is co-located with the positive cluster mean geopotential height anomaly of the ScTr (Figure 1c) and the negative anomaly of the ScBL (Figure 1f). Similarly, the large frequency biases of the anticorrelated ZO and GL (Figures 1b,g) in autumn (Figure 5g) are linked to the positive geopotential height bias over Greenland (Figures 4j,l). As shown in the right column of Figure 5, calibrating the forecasts removes almost all significant frequency biases. The remaining significant biases are thus related to biases in model variability, the potential origin of which is investigated in more detail in Section 3.2: these are the smaller but still significant positive EuBL and AR and negative ScBL biases in summer (Figure 5f) and the still significant positive GL bias in autumn (Figure 5h). Forecast calibration can also increase frequency biases slightly, such as for the AT and “no regime” in winter and autumn (Figures 5b,h). This indicates that correcting for a mean forecast error does not improve every flow situation, likely because the mean forecast error is dominated by errors in particular flow situations.

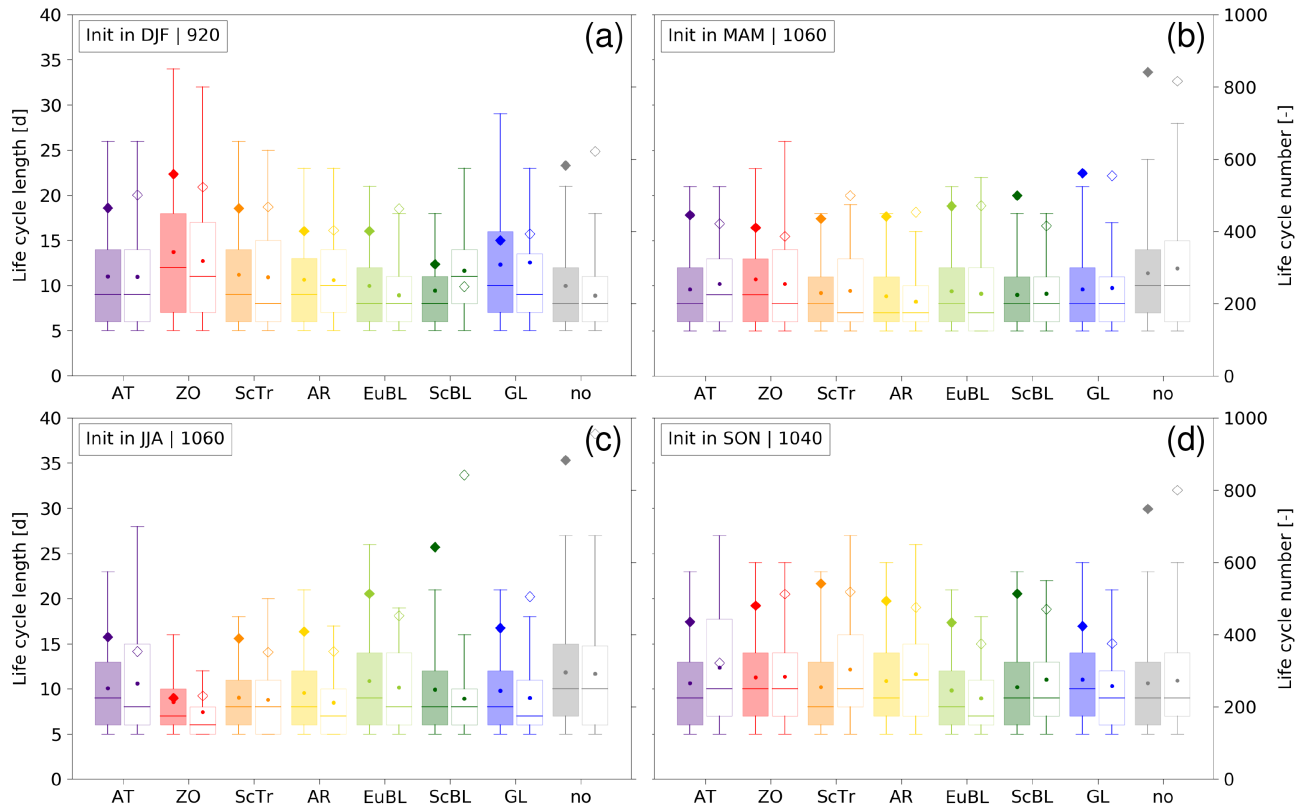
Another way of quantifying the effect of forecast calibration on regime occurrence is to count the number of calibrated forecasts in which any of the ensemble members yields a regime (life-cycle) assignment at a specific lead time different from their noncalibrated counterparts. In winter, the percentage of forecasts yielding any changes in regime assignment after calibration increases from 37% after 5 days to 71% after 20 days lead time. In summer, these numbers increase from 53% after 5 days to 94% after 20 days lead time, whereas in spring and autumn they range somewhere in between. This analysis thus clearly reflects the seasonal effects of forecast calibration illustrated by Figure 5 and discussed above, with the strongest effect of forecast calibration in summer and the weakest effect in winter, and a stronger effect at sub-seasonal than medium-range lead times.

The effects of forecast calibration on forecast skill are more subtle and barely significant (see Figures S2 and S3 in the Supporting Information; note that a detailed discussion of forecast skill is provided in Section 3.3): the average weather-regime skill horizon increases by 1–3 days in summer (depending on whether a BSS of 0.1 or 0.0 is defined as “no skill”), but this is not significant. Moreover, the slightly negative BSS in summer at sub-seasonal lead times

is partly removed by the calibration. In the other seasons, the effect is negligible (see Figure S2). For some individual regimes such as the EuBL, ScTr, or ScBL, the year-round skill horizon also increases by a few days but not significantly (again depending on the definition of the skill horizon; see Figure S3), which is likely driven by the reduction of large frequency biases in summer (Figure 5). In summary, sub-seasonal forecast calibration is most important in summer, least important in winter, and moderately important in the transition seasons. The improvements achieved with forecast calibration manifest primarily in reduced weather-regime frequency biases, but less so in improved forecast skill. In the following, we will thus use the calibrated forecasts as a basis for further analysis.

### 3.2 | Verification of weather regime duration, number, and transitions

To reveal potential sources of the aforementioned lead-time-dependent weather-regime frequency biases in the calibrated forecasts (Figure 5, right column), we now analyse how well the same forecasts reproduce climatologies of regime life-cycle duration, number of regime life-cycle objects, and transitions between regime life cycles. We thereby focus on life-cycle objects as consecutive periods in which a regime is active (i.e., time between onset and decay; cf. Section 2.2). Figure 6 shows the seasonal duration and total number of regime life-cycle objects in the forecasts and in ERA-Interim (throughout all lead times). In addition, Figure 7 illustrates the seasonal frequencies of transitions between these regime life cycles (i.e., from the decay of one regime life cycle to the onset of another within at most 4 days) in ERA-Interim (shading; adding up to 100% along the horizontal) and the associated significant biases in the forecasts (numbers). First, the significant positive and negative lead-time-dependent frequency biases for the EuBL and ScBL in summer (Figure 5f) can partly be explained by too many EuBL and too few ScBL life cycles, respectively (diamonds in Figure 6c), but less so by considerable biases in their duration (box-and-whiskers in Figure 6c). Furthermore, they can be explained by biases in transitions into these two regimes (Figure 7c): in summer, many regimes have frequent transitions into the ScBL. Three of these transitions (from AT, ZO, and AR) are strongly and significantly underestimated, which indicates that the negative lead-time-dependent ScBL frequency bias might partly be caused by too few transitions into ScBL. Likewise, the forecast overestimates the (rarely observed) transitions from AT, ScTr, and AR into EuBL, which might partly explain the positive lead-time-dependent EuBL frequency bias. There is also a high and strongly underestimated

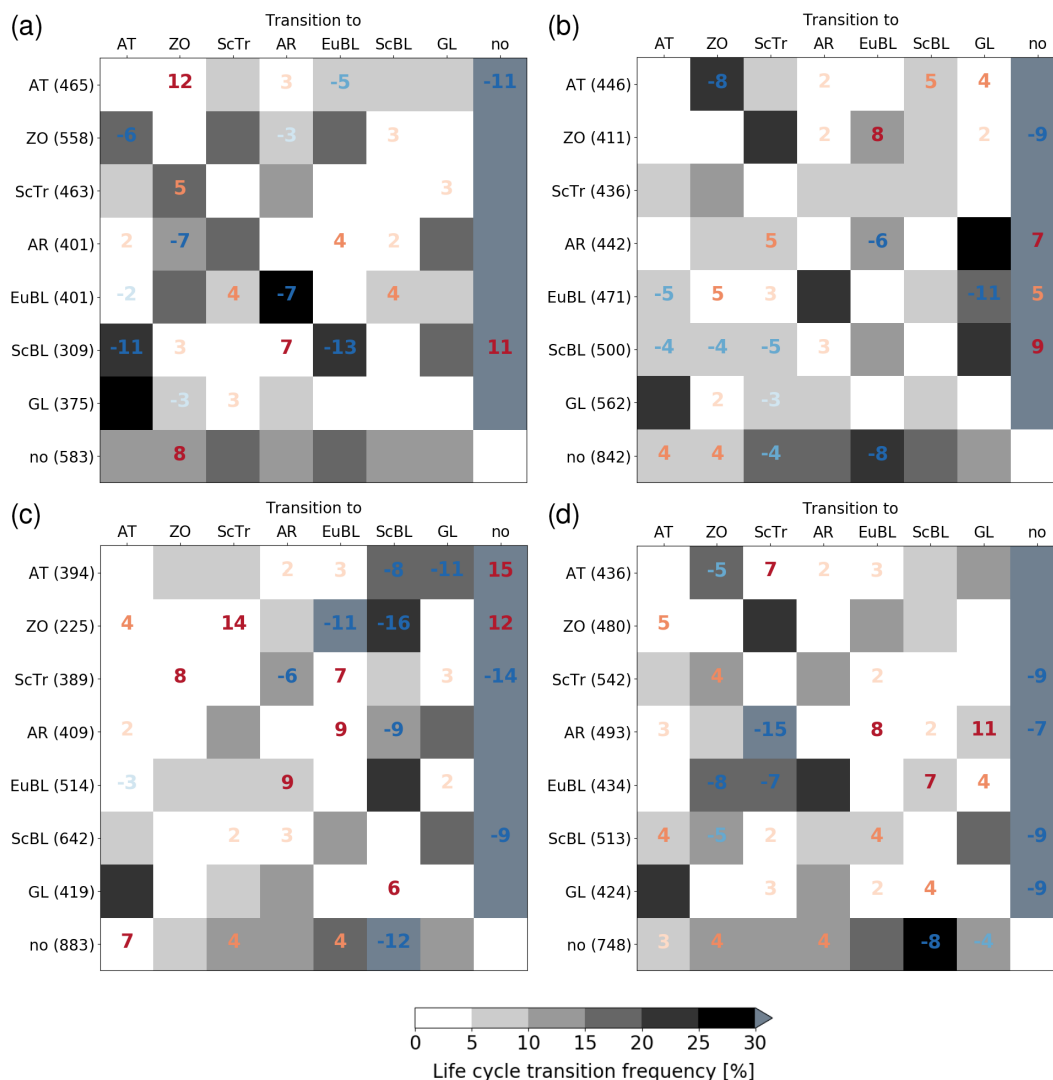


**FIGURE 6** Seasonal life-cycle duration (box-and-whiskers with the 5th and 95th percentiles indicated by the whiskers, the interquartile range by the box, the median by the line, and the mean by the filled circle; in days; left y-axis) and total life-cycle number (diamond symbols; right y-axis) for the individual regimes (x-axis) in the forecast (filled) and in ERA-Interim (blank). The seasons and the corresponding available numbers of forecasts are indicated in the boxes. Note that the life cycles cannot be shorter than 5 days by construction and not longer than 46 days as the maximum lead time. The statistics here refer to the number (and duration) of life-cycle objects (i.e., from onset to decay), in contrast to the frequency biases in Figure 5 being simply based on the active life cycle at each day. Also here, the statistics is computed over ERA-Interim being treated like a “perfect ensemble member” (cf. Figure 5) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

transition frequency from ZO into EuBL, which might counteract the positive lead-time-dependent EuBL frequency bias but should be interpreted with caution due to the very rare ZO occurrence in summer (Figure 3). Moreover, the relatively well-captured transitions in summer from EuBL into ScBL and vice versa indicate that the opposite lead-time-dependent ScBL and EuBL frequency biases do not seem to be caused by erroneous transitions between the two regimes. The third significant lead-time-dependent frequency bias in summer—the overestimation of the AR (Figure 5f)—can partly be explained by too many and too persistent AR life cycles (Figure 6c), whereas the transition biases into AR do not appear to be a reason (Figure 7c). Finally, the significant positive lead-time-dependent frequency bias of the GL in autumn (Figure 5h) might be related to too many and too persistent GL life cycles (Figure 6d) and too many transitions from AR and EuBL into the GL in the forecast (Figure 7d). Similarly, the positive lead-time-dependent frequency bias of the AT in autumn (Figure 5h) can be related to the strong overestimation in the number of AT

life cycles (which likely overcomes their underestimated duration) and too many transitions from ZO, AR, and ScBL into the AT (Figures 6d and 7d).

Apart from explaining some of the lead-time-dependent regime frequency biases in summer and autumn (shown in Figure 5), Figures 6 and 7 reveal further interesting aspects: the duration of life cycles strongly differs for the different regimes and seasons. In winter, the ZO and GL are the most persistent and the EuBL and “no regime” the least persistent regimes on average (Figure 6a). During summer, however, the AT, EuBL, and “no regime” are the longest and ZO the shortest regimes (Figure 6c). In spring and autumn, life-cycle duration is much more similar among the regimes (Figures 6b,d). The most striking mismatches in regime duration between the forecasts and ERA-Interim appear for the ScBL in winter (underestimation) and the ZO in summer (overestimation). In summary, Figure 6 thus demonstrates substantial differences in regime duration in winter and summer, which are likely related to differences in intrinsic predictability. The fact that the model captures some



**FIGURE 7** Seasonal relative life-cycle transition frequencies (%) from a specific weather regime (y-axis) to each weather regime (x-axis) in ERA-Interim (shading) and the corresponding significant-only transition frequency biases (relative frequencies in forecasts (%) minus relative frequencies in ERA-Interim (%); blueish and reddish numbers) in (a) DJF, (b) MAM, (c) JJA, and (d) SON. The transitions are computed based on the same life-cycle objects as in Figure 6, with a transition being counted if the decay of one regime (y-axis) is followed by the onset of another regime (x-axis) within at most 4 days. The numbers in brackets along the y-axis indicate the total number of life cycles of the corresponding regime. The frequencies along the horizontal sum up to 100% [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

of these differences better than others will thus be useful to understand some of the differences in forecast skill for the different regimes (cf. Section 3.3).

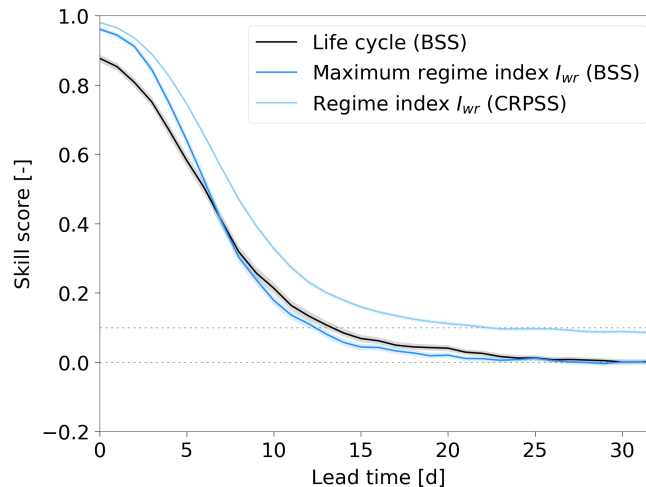
Considering the life-cycle transitions in more detail reveals a set of climatologically preferred pathways between regimes (Figure 7, with a focus on just those transitions that occur at least 10% more often than all the others for a considered regime): the most striking transition in all seasons is from GL to AT. This indicates that the decay of the blocking over Greenland (GL) typically manifests as an intensification and northward shift of the jet stream over the North Atlantic, likely going along with a shift from the negative to the positive state of the leading EOF (e.g., Ferranti *et al.*, 2018). There are no significant

biases associated with this transition, which indicates that the model captures this important pathway remarkably well. Another frequent transition in winter is from EuBL to AR, indicating an upstream propagation of the Central European anticyclone to be most common. This transition is significantly underestimated by the forecasts (Figure 7a). In spring, frequent transitions are from AT to ZO and from AR to GL, reflecting the specific pathways with which extreme positive and negative states of the leading EOF, respectively, typically develop (e.g., Ferranti *et al.*, 2018). The former of these transitions is significantly underestimated by the forecasts (Figure 7b). In summer, an important and well-captured transition is the one from EuBL to ScBL, indicating that the northward progression

of a Central European high-pressure anomaly toward Scandinavia is a typical fate of EuBL life cycles (Figure 7c). In summary, the transition verification in Figure 7 thus reveals that the model captures some of the climatologically preferred transition paths remarkably well (most prominently the one from GL to AT). This information can be useful from an operational forecasting perspective, because a climatologically frequent transition predicted by a large fraction of ensemble members could be seen as an indicator for a physically meaningful behavior of the forecast. Vice versa, nonexistent transitions also provide useful information: for instance, there are hardly any direct transitions from ZO to GL in any of the seasons (Figure 7), which indicates that the shift from a positive to a negative phase of the NAO (i.e., along the leading EOF) occurs only through intermediate steps that can only be captured with a higher number of regimes considered here. Forecasts that would provide strong evidence for such a transition would thus have to be interpreted with caution.

### 3.3 | Verification of weather regime forecast skill

After evaluating the representation of climatological weather regime characteristics in the forecast, we now investigate forecast skill for the regimes. The year-round multicategory skill (BSS: Equation 4) horizon for predicting the life cycle (i.e., the regime with the maximum regime index  $I_{wr}(t)$  above 0.9 for at least 5 days; cf. Section 2.2 and Figure 2c) of all regimes (excluding the “no regime”) is approximately 20–25 days (black line in Figure 8). However, although a BSS slightly larger than zero implies “skill” by definition (which is the case between 20 and 25 days), this might not be a useful level of skill any more from a forecasting perspective. For this reason, we focus primarily on the arbitrary but reasonable level of BSS=0.1 as a reference level to compare the different flow-dependent skill horizons in the following. Considering this level, the skill horizon for the life cycle is about 14 days, which is in the range of other studies’ results (e.g., Buizza and Leutbecher, 2015; Neal *et al.*, 2016; Ferranti *et al.*, 2018; Son *et al.*, 2020). To see the effect of our life-cycle definition on the overall skill horizon, Figure 8 additionally shows the BSS for the active regime just defined based on the maximum regime index  $I_{wr}(t)$  (i.e., without the life-cycle criteria being applied; cf. Section 2.2 and Figure 2b) and the CRPSS (Equation 5) for the continuous  $I_{wr}(t)$  (cf. Equation 2 and Figure 2a). The BSS horizon for the maximum  $I_{wr}(t)$  is about 1–2 days shorter than the BSS horizon for the life cycle. This demonstrates the added value of including a life cycle (i.e., persistence and projection threshold) criterion for predicting the

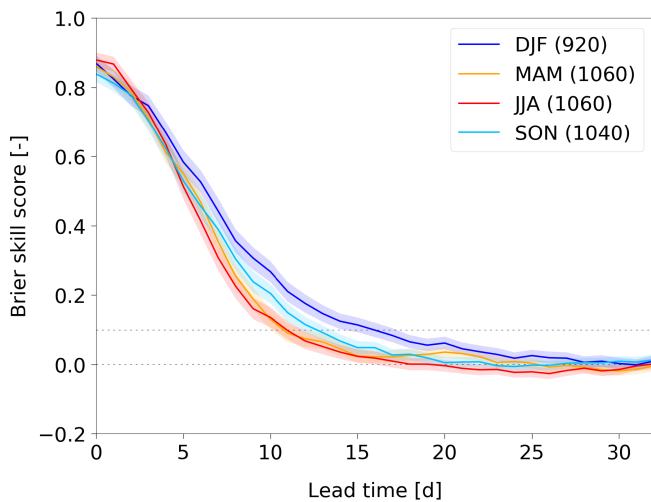


**FIGURE 8** Year-round multicategory BSS for life cycle (black), multicategory BSS for maximum regime index  $I_{wr}$  (gray blue), and multicategory CRPSS for regime index  $I_{wr}$  (light blue) for all weather regimes (y-axis; see Section 2.3 for details) as a function of lead time (x-axis). The BSS for the life cycle (black) is computed without including the “no regime” category to allow for a fair comparison with the other two skill scores, which do not contain this category by definition. In addition to the actual skill score over all forecasts (lines), the range between the 5th and 95th percentiles of the bootstrapped skill score distribution is shown by the shading, which aims to assess whether differences between skill scores are significant (see Section 2.3 for details; the same is shown in all subsequent skill score figures) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

individual weather regimes.<sup>3</sup> Interestingly, the asymptotic level of the CRPSS is higher at about 0.1, which is likely linked to how the underlying climatological reference forecast is created. Nevertheless, this level is also reached at about 20–25 days, which indicates that the continuous  $I_{wr}(t)$  information does not provide skilful information at longer lead times. In the following, we thus focus on the BSS for the regime life cycle.

Figure 9 stratifies the multicategory skill for all regimes according to the four seasons (note that the multicategory skill for the life cycle of all regimes always includes the

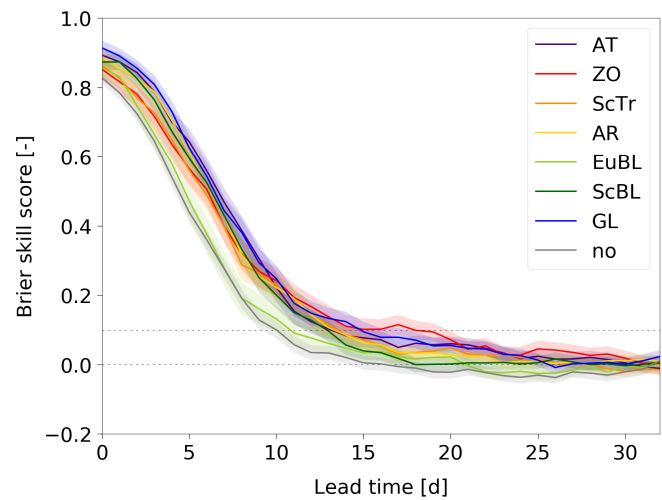
<sup>3</sup>Note that the lower BSS for the life cycle than for the maximum  $I_{wr}(t)$  at forecast initialization results from the fact that the life-cycle definition depends on the regime index evolution over several days. The rare cases in which the regime index evolution throughout the first few lead time steps is dominated by two regimes with similarly high indices can result in a majority of ensemble members favoring one of the two regimes slightly over the other (and thus erroneously causing this regime to fulfill the life-cycle criteria), in contrast to ERA-Interim doing the opposite. Although the model forecasts do not actually perform badly in these cases, the erroneous regime life-cycle assignments are excessively punished by the categorical BSS, which is reflected in the overall BSS not being 1.



**FIGURE 9** Seasonal multicategory BSS for all weather regimes (life cycle; including the “no regime” category) as a function of lead time. The numbers in the legend show the number of forecasts in the respective season. The stratification is done according to whether the forecast initial date is in the corresponding season [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

“no regime” category hereafter, unless something else is stated). The differences in the skill horizon (referring to the level of 0.1) are substantial and significant: the skill in winter is about 5 days longer than in summer and spring, and about 3 days longer than in autumn (cf. also Neal *et al.*, 2016). To what proportion these differences are caused by differences in intrinsic predictability and differences in model errors is an interesting and important question, but goes beyond the scope of this study. For instance, Dalcher and Kalnay (1987) showed theoretically that the intrinsic predictability is higher in winter than in summer because of the combination of a higher error growth rate but smaller saturation error in summer. Nevertheless, it is likely that the larger regime frequency biases in summer—even in the calibrated forecasts (Figure 5f)—indicate a large potential for model improvements.

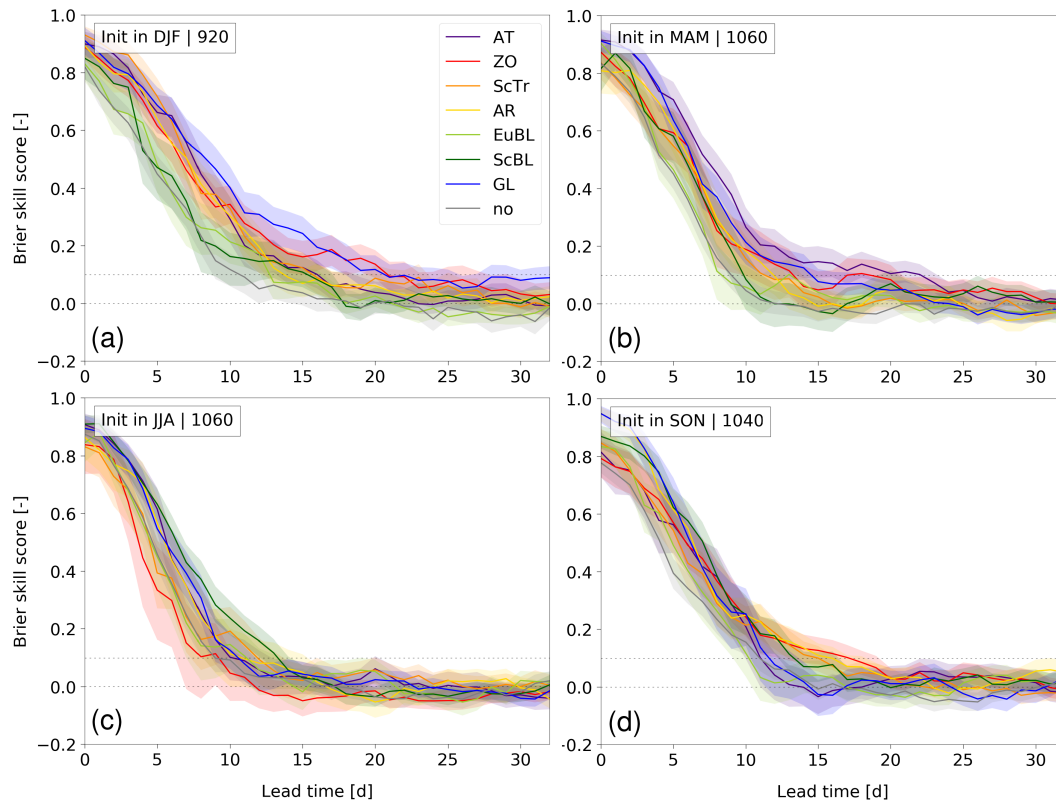
Figure 10 shows the year-round single-category skill for the individual regimes, revealing some significant differences: most importantly, the skill horizon (referring to the 0.1 level) for the EuBL is about 11 days and thus 3–5 days shorter than for the other regimes, including the ScBL. This is remarkable and indicates that the well-known difficulties in predicting blocking, as found by previous studies using the four classic Atlantic–European regimes (e.g., Ferranti *et al.*, 2015; 2018; Matsueda and Palmer, 2018), is caused primarily by those blocking types located over Central Europe rather than the ones over Scandinavia. A better understanding of the dynamical processes associated with these two blocking types will thus help to improve (sub-seasonal) blocking forecasts.



**FIGURE 10** Year-round single-category BSS for individual weather regimes (life cycle) as a function of lead time [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

The skill horizon for the ZO and GL tends to be longest (15–20 days referring to the 0.1 level), although this is not significant compared with many regimes. Nevertheless, it is striking that the skill for the ZO is significantly larger than zero for up to about 30 days, which demonstrates that remarkable sub-seasonal predictability and windows of opportunity must exist for specific ZO phases (see discussion below for individual seasons). Finally, Figure 10 indicates a relatively low skill for the “no regime” category. This highlights the difficulty in predicting phases that lack persistence and do not fit clearly into one of the distinct large-scale patterns in the low-dimensional phase space. Forecasts thus benefit from introducing a “no regime” category as a “window of low sub-seasonal predictability”. We have also computed regime-specific skill scores for the maximum  $I_{wr}(t)$  (BSS: Figure S4a) and the continuous  $I_{wr}(t)$  (CRPSS: Figure S4b). Although the skill differences tend to become smaller, the relatively low skill for the EuBL is still apparent. Furthermore, the relative differences in skill change when considering  $I_{wr}(t)$  (i.e., CRPSS), with the lowest skill found for the AT (together with the EuBL) and the highest skill for the GL.

Stratifying the single-category skill for individual regimes after the four seasons reveals several important aspects, although the skill scores are less robust due to the smaller sample size (Figure 11): first of all, the aforementioned low year-round skill for the EuBL (Figure 10) is evident in all seasons but most pronounced in winter and spring. However, the large differences in year-round skill between the EuBL and ScBL (Figure 10) are primarily a result of summer and autumn (in summer, the skill for the ScBL tends to be largest among all regimes), whereas their skill is similarly low in winter and spring. Furthermore, the relatively high year-round skill for the ZO and



**FIGURE 11** Seasonal single-category BSS for individual weather regimes (life cycle) as a function of lead time. The seasons and the corresponding available numbers of forecasts are indicated in the boxes [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

GL (Figure 10) is driven primarily by the high skill in winter, where these two regimes clearly stand out: they have significantly larger than zero skill for up to 30 days and reach the skill level of 0.1 almost 10 days later than the EuBL with the lowest skill. This indicates important windows of sub-seasonal predictability in winter. In contrast, the skill for the ZO is very low in summer (likely related to the rare ZO occurrence; Figure 3) and for the GL relatively low in autumn. Another interesting aspect is the high skill for the AT in spring (it reaches the skill level of 0.1 almost 15 days later than the EuBL). Figure 11 thus reveals a variety of—in some cases substantial—skill differences in the individual seasons, which are relevant from both an operational forecasting and a model development perspective. Some of the differences might be explained by the differences in intrinsic predictability caused by differences in persistence (Figure 6): for instance, the high skill for the ZO and GL in winter is likely driven by their relatively large and well forecast life-cycle duration. The prolonged duration, in turn, is probably related to phases of anomalous states of the SPV in winter, which are known to be statistically followed by persistent positive and negative NAO phases that correlate strongly with ZO and GL (cf. Section 1; note that a more detailed analysis of the effect of anomalous SPV states on regime skill will follow in Section 3.4.2). Likewise, the

low skill for the EuBL in winter or for the ZO in summer might be linked to their short duration (and, in the case of ZO, the rare occurrence; Figure 6). On the other hand, the significant lead-time-dependent regime frequency biases in summer and autumn (Figures 5f,h) appear much more vaguely in the skill differences: for instance, the positive frequency biases for the EuBL in summer and the GL and AT in autumn might indeed co-occur with relatively low forecast skill in these two seasons. In contrast, the forecast skill for the ScBL in summer is remarkably high despite having the largest negative frequency bias. It thus appears promising that a reduction of the ScBL frequency bias in summer might extend the ScBL skill horizon even further, which is crucial for predicting heat waves on sub-seasonal timescales (e.g., Schaller *et al.*, 2018; Spensberger *et al.*, 2020).

We have further computed the year-round (Figure S5) and seasonal (Figure S6) multicategory skill for all regimes depending on the active regime at the forecast initial time. This stratification, however, strongly reduces forecast sample size and hence the robustness of the associated skill scores. We thus find only few robust differences in skill depending on the regime at initial time. Some of these worth mentioning are the tendency towards enhanced skill at medium-range lead times of the winter and spring forecasts starting with GL (consistent with Ferranti *et al.*,

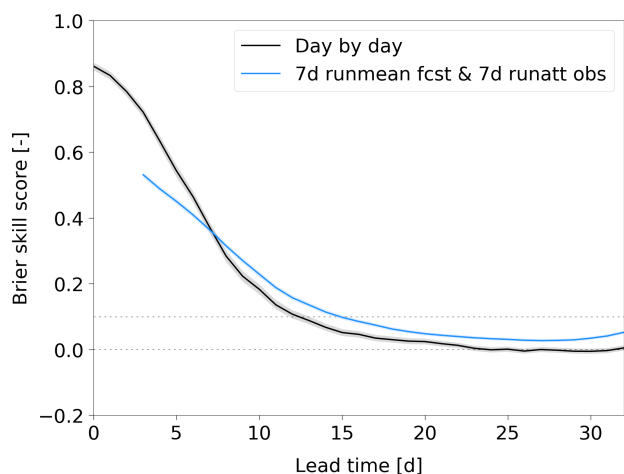


2015; Matsueda and Palmer, 2018; Lin, 2020). This might be influenced by the relatively high persistence of the GL in winter (Figure 6a) and possibly also by its very frequent and well-captured transition into AT (Figure 7a). Furthermore, the skill tends to drop relatively fast in the spring forecasts starting with ScTr, which is interesting considering the fact that the transitions from ScTr in spring are not associated with any significant biases (Figure 7b).

### 3.4 | Windows of opportunity for enhanced sub-seasonal weather regime forecast skill

#### 3.4.1 | Role of verification window

Our analysis so far has verified how well the sub-seasonal forecasts can predict the active weather regime each day in different flow situations. Beyond the medium range, however, this is both a physically limited and, beyond certain lead times, intrinsically impossible prediction problem, and from an operational forecasting perspective not even of primary interest. Increasing the lead-time window for which we would like to extract useful forecast information can thus be a meaningful way to improve state-of-the-art sub-seasonal regime forecasts (cf. also Zhu *et al.*, 2014; Buizza and Leutbecher, 2015). We thus investigate the sub-seasonal forecast skill horizon for predicting the regimes within a running window of 7 days (instead of day-by-day; Figure 12). More specifically, we verify the running mean regime frequency in the ensemble against the running attribution of regime occurrence



**FIGURE 12** Year-round multicategory BSS for all weather regimes (life cycle) computed at each lead-time day (black) and for a running lead-time window of 7 days centered around the evaluated lead time (blue; running mean for the forecasts and running attribution for the observation; see text for details) as a function of lead time [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

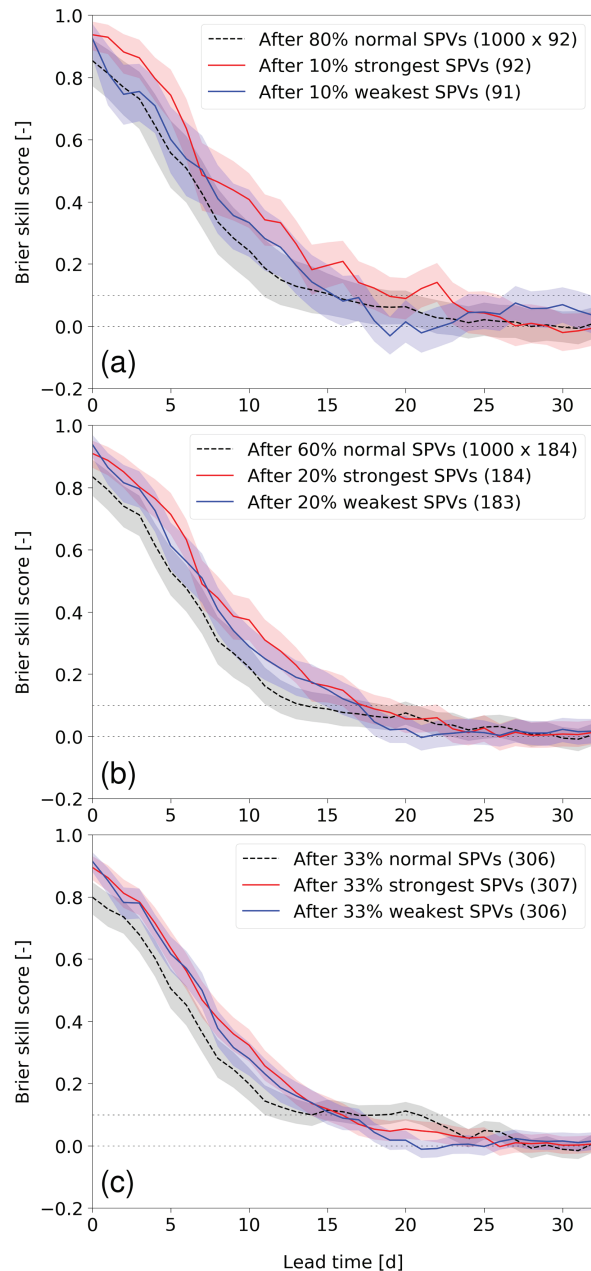
in ERA-Interim. The running attribution means that a regime is defined to occur ( $o_k^{wr} = 1$ ; cf. Equation 3) if it is active on at least one day within the running window. Regarding the year-round multicategory skill for all seven regimes, this running window approach extends the sub-seasonal skill horizon significantly by about 3 days compared with the classic day-by-day approach (blue compared with black line in Figure 12). The single-category skill horizon for the individual regimes, however, reacts differently (Figure S7): it increases substantially for some regimes such as the EuBL (by up to 5 days) but scarcely changes for other regimes such as the ScBL. The increase of the skill horizon for the EuBL might indicate that the relatively low day-by-day EuBL skill (Figure 10) is partly caused by forecast errors in the timing of the life cycles (i.e., onset or decay). We have also applied this running window approach either in the forecasts only or in ERA-Interim only, but the increase of the skill horizon is largest when the running window is applied to both (i.e., the approach presented here). Furthermore, the results are almost identical when changing the running window to 5 or 9 days. Finally, it is important to note that such an approach substantially reduces skill in the medium range, where day-by-day predictions are still highly skilful (Figure 12). At these lead times, the verification window would thus have to be reduced, for instance by defining the verification window as a function of lead time (Zhu *et al.*, 2014). Moreover, it is likely that the effect of a running mean in the forecast space might be smaller in the operational ECMWF forecasting system, in which the higher number of ensemble members (51) should account for this to some extent by construction. Nevertheless, our analysis demonstrates the potential for extracting more skilful forecast information on sub-seasonal forecast ranges beyond two weeks by means of modifying the temporal aggregation of forecast products.

#### 3.4.2 | Role of lower-frequency phenomena

As introduced in Section 1, the midlatitude sub-seasonal forecast skill horizon can also be extended significantly by lower-frequency climate phenomena exerting a dynamical forcing on the extratropical flow via planetary-scale teleconnections. For the Atlantic–European region, two important such forcings come from the winter SPV and the tropical MJO. We thus investigate the influence of the SPV intensity and MJO state at the forecast initial time on the multicategory forecast skill for all regimes together. The SPV intensity at each forecast initial date is defined by the instantaneous geopotential height anomaly at 100 hPa (i.e., in the lower stratosphere) averaged over the polar cap north of 60°N in ERA-Interim (a negative polar cap

anomaly corresponds to a strong SPV and vice versa). Defining the SPV intensity in the lower stratosphere has been shown to be meaningful to investigate the tropospheric impact from the stratosphere (e.g., Baldwin *et al.*, 2003; Karpechko, 2015; Beerli *et al.*, 2017; Charlton-Perez *et al.*, 2018; Büeler *et al.*, 2020). The state of the MJO at each forecast initial date is defined based on the multivariate MJO index RMMI (Wheeler and Hendon, 2004) provided operationally by the Australian Bureau of Meteorology,<sup>4</sup> with an active MJO if the RMMI  $\geq 1$ .

Figure 13a shows the skill in winter of those forecasts initialized with the 10% strongest and 10% weakest SPV states in comparison with the skill of the residual forecasts (see captions for details on how the skill of the residual forecasts is obtained). The extreme SPV states modify forecast skill compared with normal conditions, but with a distinct effect of strong and weak SPV states: strong SPV states tend to enhance skill for both medium-range and sub-seasonal lead times (significantly at some lead times) and thus extend the skill horizon by up to about 5 days (referring to the 0.1 level). Weak SPV states, in contrast, tend to increase skill only in the medium range but decrease beyond. Interestingly, the pattern changes when comparing the 20% or 33% strongest and weakest SPV states (Figures 13b,c): the increase in skill for medium-range lead times becomes more pronounced and significant after both strong and weak SPV states. On sub-seasonal lead times, the skill still tends to be higher after strong compared with weak SPV states, but the forecasts initialized with normal SPV states tend to perform better in a relative sense and even outperform the ones initialized with anomalous SPV states for the tercile definition (Figure 13c). As the NAO tends to be most sensitive to the SPV (cf. Section 1), we have done the same SPV sensitivity analysis as in Figure 13 but separately for two groups of regimes (Figure S8): those that correlate most strongly with the NAO (ZO, ScTr, and GL) and those that correlate least strongly with the NAO (AT, AR, EuBL, and ScBL). This reveals that the generally higher skill for all regimes following the 10% and 20% strongest compared with the 10% and 20% weakest SPV states (Figure 13a and 13b) is driven largely by those regimes not related to the NAO, particularly at medium-range lead times when the skill following strong SPV states is enhanced remarkably and significantly (Figure S8b,d). On the other hand, the lower skill following weak compared with strong SPV states at sub-seasonal lead times appears for both groups of regimes and thus seems to be independent of particular regime types (Figure S8). In summary, the distinct skill



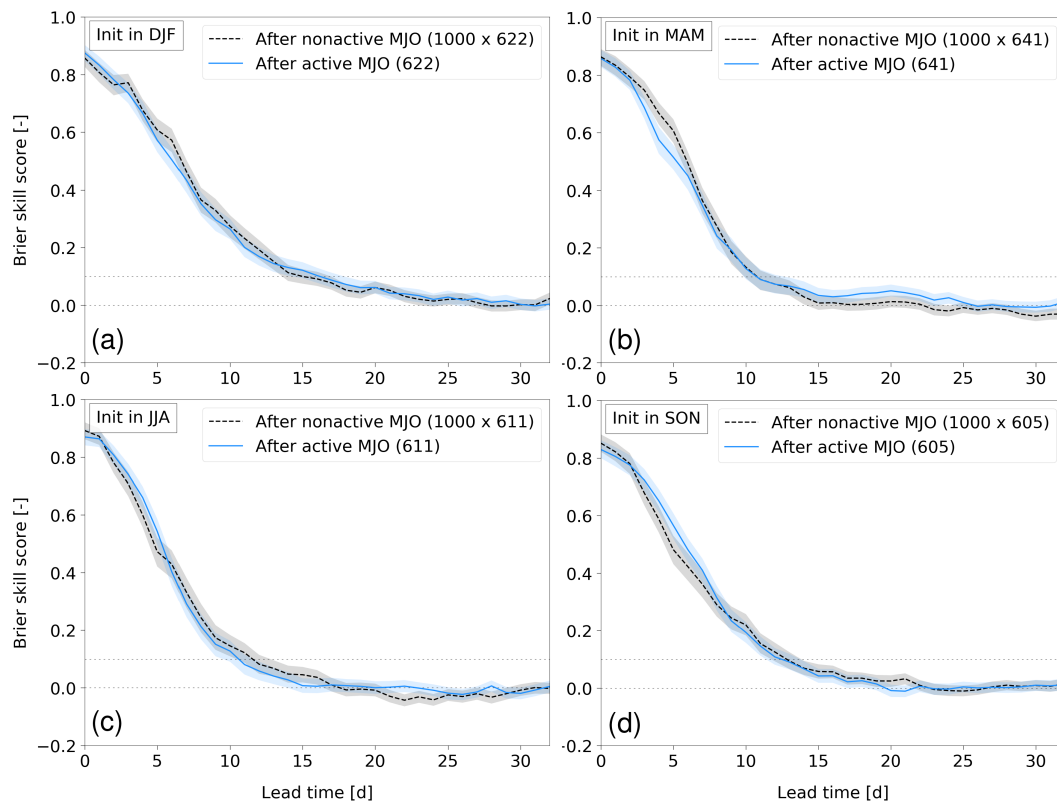
**FIGURE 13** Winter (DJF) multicategory BSS for all weather regimes (life cycle) depending on the stratospheric polar vortex (SPV) intensity at forecast initial time: (a) 10% strongest (red), 10% weakest (blue), and 80% normal (black) SPV intensities; (b) 20% strongest (red), 20% weakest (blue), and 60% normal (black) SPV intensities; (c) 33% strongest (red), 33% weakest (blue), and 33% normal (black) SPV intensities. The BSS for the normal SPV intensities in (a) and (b) is based on a distribution of 1000 random forecast samples (with replacement) of the same size as the 10%/20% bins drawn from all winter forecasts initialized with an SPV intensity other than the 10%/20% strongest and 10%/20% weakest intensities, respectively (allowing for a statistically robust comparison). The black lines in (a) and (b) thus show the mean over these distributions and the black shading indicates the range between the corresponding 5th and 95th percentiles. The numbers in the legend show the number of forecasts in the respective forecast groups [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

<sup>4</sup>Retrieved from <http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt> on February 18, 2021.

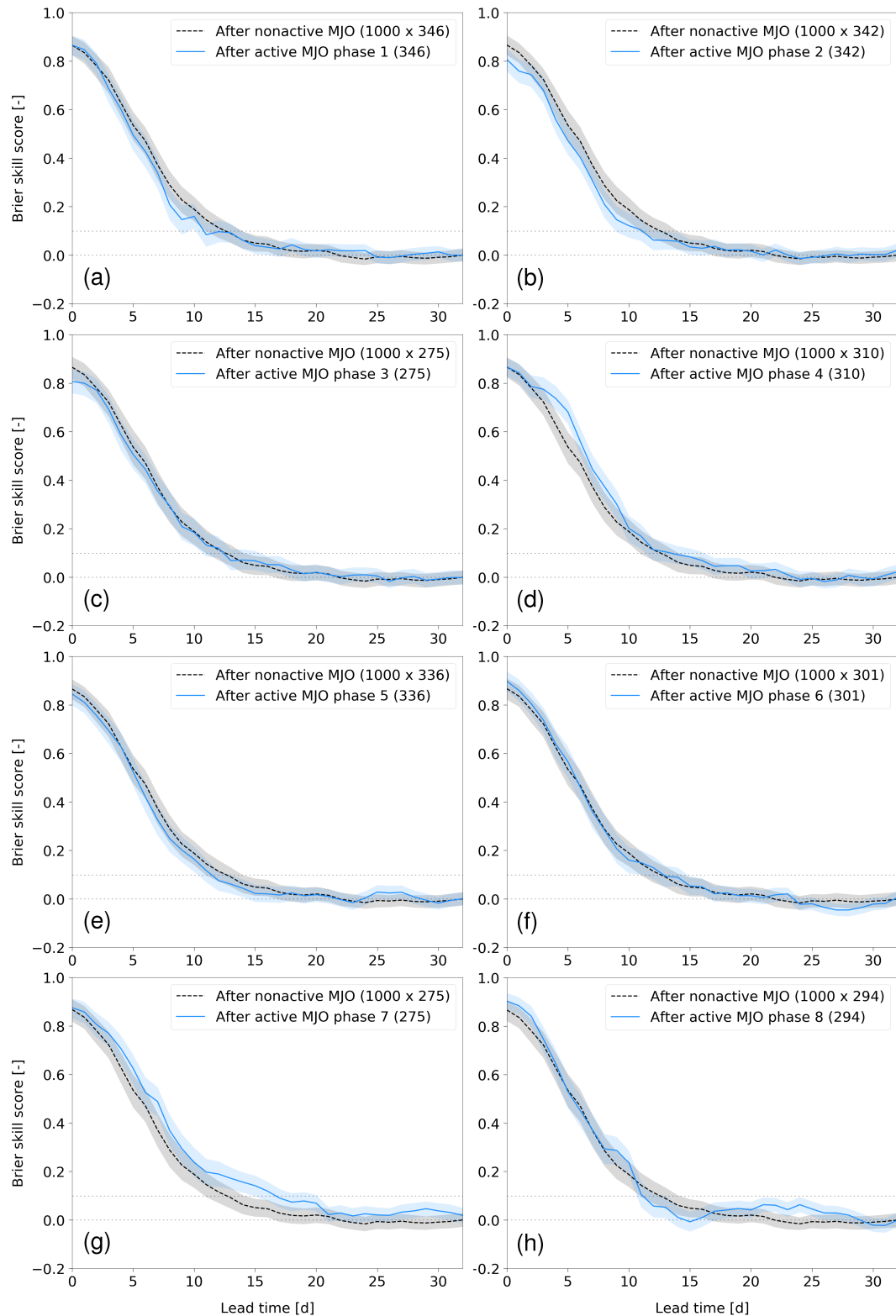
modifications by strong and weak SPVs are interesting, considering the fact that previous studies have often pointed out the enhanced sub-seasonal predictability following weak SPV states, particularly so-called sudden stratospheric warmings (SSWs; Scherhag, 1952; Baldwin *et al.*, 2021). At the same time, they are in line with Büeler *et al.* (2020), who showed that strong SPV states tend to increase and weak SPV states to decrease sub-seasonal forecast skill for near-surface temperature in large parts of Europe. The reduction in skill following weak SPV states likely results from the large case-to-case variability of both the weak SPV states themselves (e.g., Mitchell *et al.*, 2013; Butler *et al.*, 2015) and the associated subsequent tropospheric large-scale response (e.g., Beerli and Grams, 2019; Büeler *et al.*, 2020; Domeisen *et al.*, 2020b). Models thus need to capture this variability better to exploit fully the potentially enhanced sub-seasonal predictability following weak SPV states. Analysing the observed and modeled frequency of the seven regimes following weak SPV states might be a promising way to achieve this, because the refined regime definition likely reveals biases that would not be captured by the classic set of four regimes. We plan

to investigate this in the future. Apart from these differences between strong and weak SPV states, our analysis further demonstrates that the enhanced sub-seasonal forecast skill following extreme stratospheric states seems to be given only when the initial SPV intensity is truly extreme (such as in a recent event described by, e.g., Lee *et al.*, 2020b) and not just above or below normal. In other words, the already high regime forecast skill in winter at sub-seasonal lead times (Figure 9) would be even higher if the forecasts initialized with the upper and lower thirds of SPV intensities were neglected. This is an interesting finding, the physical reason for which should be investigated further. Finally, we show that the increase in skill following strong SPV states tends to be much stronger for those regimes that are not related to the NAO than those that are, particularly in the medium range. This is surprising and again highlights the added value of considering our refined set of regimes compared with the NAO only.

Figure 14 shows how an active compared with a non-active MJO at forecast initial time modifies the regime forecast skill during the different seasons. The differences in skill are rather small and hardly significant, even in



**FIGURE 14** Seasonal multicategory BSS for all weather regimes (life cycle) of the forecasts initialized with an active Madden–Julian Oscillation (MJO; blue) compared with a nonactive MJO (black). The seasons are indicated in the boxes. The BSS for the nonactive MJO is based on a distribution of 1000 random forecast samples (with replacement) of the same size as the corresponding active-MJO forecast bins drawn from all forecasts initialized with a nonactive MJO (allowing for a statistically robust comparison). The black lines thus show the mean over these distributions and the black shading indicates the range between the corresponding 5th and 95th percentiles. The numbers in the legend show the number of forecasts in the respective forecast groups [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]



**FIGURE 15** Year-round multicategory BSS for all weather regimes (life cycle) of the forecasts initialized with a specific active MJO phase (blue) compared with a nonactive MJO (black). The BSS for the nonactive MJO is based on a distribution of 1000 random forecast samples (with replacement) of the same size as the corresponding active-MJO-phase forecast bins drawn from all forecasts initialized with a nonactive MJO (allowing for a statistically robust comparison). The black lines thus show the mean over these distributions and the black shading indicates the range between the corresponding 5th and 95th percentiles. The numbers in the legend show the number of forecasts in the respective forecast groups [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

winter when the dynamical forcing from the MJO tends to be strongest (or at least understood best: e.g., Zhang and Dong, 2004; Stan *et al.*, 2017). Stratifying the active MJO into its specific phases, however, reveals that this, on average small, skill modification results from a balance between enhanced skill after some MJO phases and reduced skill after others (Figure 15): averaged over the year, the strongest (and partly significant) increase in skill occurs after phases 7 and 4 and the strongest decrease after phase 2 (Figure 15b,d,g, respectively). The skill increase for phase 7 appears primarily in winter and spring (see also, e.g., Feng *et al.*, 2021) and that for phase 4 in winter and autumn (see Figures S9–S12). The skill decrease after phase 2 is caused primarily by reduced skill during spring and autumn (see Figures S9–S12). Furthermore, the enhanced skill after phase 4 appears primarily in the medium range, whereas phase 7 instead increases skill in the sub-seasonal range and can thus extend the skill horizon substantially (by up to around 5 days). Certain MJO phases also modify regime forecast skill in summer, although not significantly (see Figure S11). Like the SPV, the MJO has also been shown to modulate primarily the NAO regimes (cf. Section 1). We have thus analysed further how the MJO modifies the multicategory skill for the aforementioned “NAO-related” (ZO, ScTr, and GL) and “NAO-unrelated” regimes (AT, AR, EuBL, and ScBL) separately (Figures S13, S14, and S15): the subtle and nonsignificant modifications in skill for all regimes following an active MJO (Figure 14) are similarly small and mostly nonsignificant for both the NAO-related and NAO-unrelated regimes (Figure S13). The most prominent exception is the significant increase in sub-seasonal skill for the NAO-related regimes following an active MJO in spring (Figure S13c). On the other hand, the modifications of the year-round skill for all regimes following MJO phases 2, 4, and 7 (Figure 15b,d,g) are driven largely by the NAO-related regimes (Figures S14 and S15), which for instance exhibit a striking and significant extension of the year-round skill horizon by more than 5 days following phase 7 (Figure S15e). In contrast, the year-round skill for the NAO-unrelated regimes is less sensitive to specific MJO phases (Figures S14 and S15), with the exception of the significant increase of the medium-range skill following phase 4 (Figure S14h). Furthermore, the year-round skill for the NAO-related regimes is reduced most substantially following phase 1 (Figure S14a). The NAO-unrelated regimes partly balance this reduction, which is why the year-round skill for all regimes is barely reduced following phase 1 (Figure 15a). In summary, knowledge about the MJO being in the specific phases discussed can thus provide important windows of opportunity for enhanced sub-seasonal regime forecast skill. This is the case not just in winter but also in the two

transition seasons. Furthermore, it tends to be dominated by the regimes related to the NAO, although the skill of the regimes not related to the NAO is also modified by specific individual MJO phases. This highlights the need to distinguish between individual (categories of) regimes when using the MJO as a source of sub-seasonal predictability. At the same time, our analysis demonstrates that there is likely still room for model improvements with respect to the large-scale atmospheric response over the Atlantic–European region following other specific MJO phases.

## 4 | SUMMARY AND CONCLUSIONS

We analyse the year-round sub-seasonal ECMWF reforecast skill for a novel set of seven Atlantic–European weather regimes. In the first part of our article, we demonstrate that forecast calibration (i.e., removing the seasonally and lead-time-dependent 500-hPa geopotential height bias) generally improves regime forecasts most strongly in summer, followed by spring and autumn, but hardly at all in winter. This can be explained by the substantially larger geopotential height biases in summer than in winter over the Atlantic–European region, the dynamical sources of which should be investigated further. The calibration-induced improvements manifest in significant reductions of lead-time-dependent regime frequency biases, but only a small improvement of regime skill.

In the second part of the article, we analyse how the remaining significant lead-time-dependent regime frequency biases in the calibrated forecasts might be explained by biases in regime life-cycle duration, number, and transitions: the positive frequency bias of the European Blocking in summer results partly from too many life cycles and too many transitions into the regime. Vice versa, the negative frequency bias of the Scandinavian Blocking in summer is linked partly to too few life cycles and too few transitions into the regime. Similarly, the positive Atlantic Ridge frequency bias in summer is linked to too many and too long-lasting life cycles. The positive Greenland Blocking and Atlantic Trough frequency biases in autumn coincide with too many and, for the former, too long-lasting life cycles and too many transitions into the two regimes. Apart from this, we reveal considerable and relatively well-predicted differences in average regime life-cycle duration, which indicate potential differences in intrinsic predictability: in winter, the Zonal Regime and Greenland Blocking are the most persistent and the European Blocking the least persistent regimes. For the Greenland Blocking, this is in line with persistent negative NAO phases (e.g., Matsueda and Palmer, 2018). In summer, the Atlantic Trough and European Blocking are the most

persistent and the Zonal Regime the least persistent. Furthermore, there are a number of climatologically frequent and well-forecast regime transitions, most prominently from the Greenland Blocking into the Atlantic Trough throughout the year, which might be useful for judging the performance of operational forecasts in advance.

The third part of the article demonstrates that sub-seasonal forecast skill varies substantially for different seasons and regimes: the average useful regime skill horizon (defined based on  $BSS = 0.1$ ) amounts to approximately 14 days over the whole year. In winter, however, it is about 5 days longer than in summer and spring, and about 3 days longer than in autumn. Considering the individual regimes over the whole year, the skill horizon for the European Blocking is 3–5 days shorter than for all the other regimes—including the related Scandinavian Blocking. Stratifying into seasons reveals that the reduced skill for the European Blocking exists in all four seasons, but is most pronounced in winter and spring. However, the remarkable difference in skill between the European and Scandinavian Blocking appears primarily in summer and autumn but not in winter and spring. In contrast, the year-round skill horizon for the Zonal Regime and Greenland Blocking tends to be longest, which is driven primarily by the high skill in winter when these two regimes are most persistent. Finally, the low year-round skill for “no regime” demonstrates the benefit of introducing this category in identifying “windows of low sub-seasonal predictability”.

As a last step, we investigate various windows of opportunity for enhanced sub-seasonal regime forecast skill. First, we demonstrate that the year-round sub-seasonal skill horizon can be increased by several days if we task the model to predict the regime occurrence within multi-day lead-time windows rather than day-by-day (similarly to the approach of Zhu *et al.*, 2014). The fact that this skill modification varies considerably among the regimes points towards potential problems in predicting the timing (i.e., onset and decay) of specific regime life cycles. Second, we investigate how specific states of two important lower-frequency phenomena, namely the winter SPV and the year-round MJO, modify sub-seasonal regime forecast skill: an anomalously strong SPV at the forecast initial time tends to extend the skill horizon in winter by several days, whereas an anomalously weak SPV increases skill only slightly in the medium range but tends to decrease skill beyond. This is in line with similar asymmetries in sub-seasonal skill modifications for European near-surface temperature found by Büeler *et al.* (2020) and Domeisen *et al.* (2020a) and highlights the need to improve the tropospheric response in sub-seasonal models following weak SPV states. Biases in this response are likely related to the fact that the model struggles to capture

the relatively variable set of regimes that can follow weak SPV states (e.g., Beerli and Grams, 2019; Domeisen *et al.*, 2020b). Furthermore, a rather surprising finding is the nonlinear relationship between the skill modification beyond the medium range following strong SPV states and the way strong SPV states are defined: the increase in skill compared with normal SPV states only holds for the 10% strongest SPV states but vanishes and even turns into a reduction in skill for the 20% and 33% strongest SPV states, respectively. Further research should investigate whether this is a sampling issue or whether it reflects a kind of threshold behavior in the sense that the SPV itself or the closely linked refraction of the vertical wave propagation from the troposphere (e.g., Ambaum and Hoskins, 2002; Polvani and Waugh, 2004) needs to have a certain intensity to enter a persistent (and thus more predictable) phase of a strong stratosphere–troposphere coupling.

Furthermore, forecasts initialized during an active MJO do not exhibit significantly higher skill than forecasts initialized during a nonactive MJO, both throughout the year and in the individual seasons. This is caused by the fact that specific MJO phases have both positive and negative effects on regime skill, balancing each other out: the strongest increase of the skill horizon by up to 5 days appears after phase 7 (primarily in winter and spring), followed by a moderate increase in medium-range skill after phase 4 (primarily in winter and autumn). The strongest decrease in skill follows phase 2 (primarily in spring and autumn). Some of these skill modifications might become more pronounced and significant if the active MJO state is defined based on a larger amplitude than the standard one ( $RMMI \geq 1$ ). Nevertheless, the balancing effect on regime skill between individual phases might still remain in this case.

It is important to mention that the modification of forecast skill by the SPV and MJO likely differs between individual regimes. Our preliminary investigations, however, have shown that the robustness of the skill scores becomes small when stratifying into individual seasons, regimes, and SPV and MJO states. Conclusions about robust skill modifications in such flow situations thus need to be made with caution. Nevertheless, we have taken a first step in this direction and investigated how specific SPV and MJO states modify the skill separately for those regimes that are strongly related to the NAO and those that are not. Although various previous studies pointed out that primarily the NAO is sensitive to these lower-frequency phenomena (cf. Section 1), we provide strong evidence that the skill can also be significantly modified by the SPV and MJO for those regimes not related to the NAO. This is particularly the case for the SPV, which for instance substantially increases medium-range skill for the NAO-unrelated regimes when in an anomalously strong state. In the

future, we thus plan to analyse in more detail how the (observed and modeled) occurrence and ultimately skill of the seven individual regimes is modulated by the specific SPV and MJO states. The higher number of regimes will thereby help to detect potential problems and biases in the model associated with the response to lower-frequency phenomena, which might be more difficult to achieve with a coarser regime definition.

Our study is associated with two caveats that are worth mentioning because they might slightly affect some of our findings: first, we use ECMWF reforecasts with a reduced set of 11 instead of the full operational set of 51 ensemble members. It is likely that the spread of the full ensemble might improve the skill in certain flow situations, because even the fair BSS used in this study can correct the skill with an estimator based on the range of physical pathways provided by the reduced ensemble only. Second, the latest cycle of the model versions used (CY45R1; Section 2.1) reproduces the spread of the MJO index, for instance, as well as the MJO amplitude better than previous cycles (partly due to improvements in the stochastic perturbation scheme).<sup>5</sup> The forecast performance might thus improve for certain flow situations if using only this last model version, but, at the same time, the smaller reforecast sample size would make a robust verification more challenging.

By using a novel set of seven year-round Atlantic–European weather regimes, we provide new and important insight complementing the findings of previous studies: first, the notorious problems of sub-seasonal weather models in predicting continental blocking (cf. Section 1) tend to occur year-round. With the higher number of regimes, we can reveal that they are caused primarily by those blockings located over Central Europe (i.e., European Blocking), but, in winter and spring, additionally by those located over Northern Europe (i.e., Scandinavian Blocking). This indicates that these two related blocking types might be driven by different dynamical mechanisms, which are captured differently well by the model depending on the season. Differences in these mechanisms can be related to different contributions from lower-frequency planetary-scale processes compared with synoptic-scale processes. The role of synoptic-scale processes and their intrinsic predictability limit for sub-seasonal forecasts is thus a subject of our current research (Quinting and Grams, 2021; Wandel *et al.*, 2021). Second, the year-round regime definition enables a systematic and comparable skill analysis in all four seasons, which has hardly been done so far (e.g., Cortesi *et al.*, 2021). The revealed lowest regime skill in summer might be improved by reducing

the largest large-scale flow biases in summer that remain even in the calibrated forecasts. At the same time, the highest skill in winter might be improved further by exploiting better the potential predictability provided by lower-frequency phenomena such as the SPV and MJO, the dynamical forcing of which is generally strongest in winter. More specifically, such improvements should focus on the model response following weak SPVs and following specific phases of the MJO. Overall, however, sub-seasonal model improvements should go hand in hand with improving our understanding of flow-dependent intrinsic predictability: for instance, there is surprisingly little research on why the intrinsic predictability limit in winter is higher than in summer (cf., e.g., Dalcher and Kalnay, 1987). A better understanding of how different processes contribute to error growth in these two seasons might be an important way forward in this context. Along the same lines, it is important to understand the intrinsic predictability of atmospheric blocking better, particularly of the related European and Scandinavian Blockings investigated in this study. Improving our understanding of intrinsic predictability will ultimately help to reveal the seasons and flow situations with the largest potential for model improvements. Last but not least, it is important to assess critically the benefits of improved sub-seasonal regime predictions from an end-user perspective. Various studies have shown that sub-seasonal predictions of more impact-oriented surface weather parameters, tailored to the specific needs of end users (particularly from the energy industry), can be as useful as (or even more useful than) predictions of regimes (e.g., Bloomfield *et al.*, 2020; 2021; Mariotti *et al.*, 2020; Torralba *et al.*, 2021). Although our higher number of regimes likely outperforms the classic four regimes in terms of regional surface weather imprint, it will thus be important to develop these regimes and their related forecast products further in close collaboration with operational forecasters.

Apart from the potential for model improvement, our study highlights that important windows of enhanced sub-seasonal predictability already exist in state-of-the-art models: in winter, the Zonal Regime and Greenland Blocking, which are closely related to the positive and negative NAO, can be predicted well beyond 20 days with reasonable skill. A similar skill horizon exists for the Atlantic Trough in spring. This remarkable skill is likely influenced by the dynamical forcing from lower-frequency phenomena such as the SPV and MJO investigated here, but also from others such as the El Niño–Southern Oscillation, the Quasi-Biennial Oscillation, or variations in sea-surface temperature, soil moisture, and snow and sea-ice cover. Uniting the knowledge of these different windows of opportunity for enhanced sub-seasonal predictability in sophisticated statistical post-processing tools might thus

<sup>5</sup><https://www.ecmwf.int/en/forecasts/documentation-and-support/evolution-ifs/cycles/summary-cycle-45r1>, retrieved on April 21, 2021.

be fruitful to improve operational sub-seasonal forecast skill further.

## ACKNOWLEDGEMENTS

We thank the two anonymous reviewers for assessing and thus helping to improve this publication. We further thank Mio Matsueda, Akio Yamagami, and Takumi Matsunobu for various fruitful discussions on forecast calibration and verification techniques, Christopher Ferro and Sebastian Lerch for helpful input on skill score correction and decomposition, and the SPREADOUT group at KIT, particularly Jan Wandel, for fruitful input throughout the project. Furthermore, the S2S Prediction Project team is acknowledged for providing convenient access to reforecast data. The contributions of D.B., J.F.Q., and C.M.G. are funded by the Helmholtz Association as part of the Young Investigator Group “Sub-seasonal Predictability: Understanding the Role of Diabatic Outflow” (SPREADOUT, grant VH-NG-1243). Open access funding is enabled and organized by Projekt DEAL.


## AUTHOR CONTRIBUTIONS

**Dominik Büeler:** conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review and editing. **Laura Ferranti:** conceptualization; writing – review and editing. **Linus Magnusson:** conceptualization; writing – review and editing. **Julian F. Quinting:** conceptualization; methodology; writing – review and editing. **Christian M. Grams:** conceptualization; data curation; funding acquisition; methodology; project administration; resources; supervision; writing – review and editing.

## ORCID

**Dominik Büeler**  <https://orcid.org/0000-0002-9904-6281>

**Laura Ferranti**  <https://orcid.org/0000-0002-2317-3292>

**Linus Magnusson**  <https://orcid.org/0000-0003-4707-2231>

**Julian F. Quinting**  <https://orcid.org/0000-0002-8409-2541>

**Christian M. Grams**  <https://orcid.org/0000-0003-3466-9389>

## REFERENCES

- Alexander, M.A., Bhatt, U.S., Walsh, J.E., Timlin, M.S., Miller, J.S. and Scott, J.D. (2004) The atmospheric response to realistic Arctic sea ice anomalies in an AGCM during winter. *Journal of Climate*, 17, 890–905. [https://doi.org/10.1175/1520-0442\(2004\)017<0890:TARTRA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0890:TARTRA>2.0.CO;2)
- Ambaum, M.H. and Hoskins, B.J. (2002) The NAO troposphere–stratosphere connection. *Journal of Climate*, 15, 1969–1978. [https://doi.org/10.1175/1520-0442\(2002\)015%3C1969:TNTSC%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C1969:TNTSC%3E2.0.CO;2)
- Anstey, J.A. and Shepherd, T.G. (2014) High-latitude influence of the quasi-biennial oscillation. *Quarterly Journal of the Royal Meteorological Society*, 140, 1–21. <https://doi.org/10.1002/qj.2132>
- Baldwin, M.P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A.H., Charlton-Perez, A.J., Domeisen, D.I.V., Garfinkel, C.I., Garny, H., Gerber, E.P., Hegglin, M.I., Langematz, U. and Pedatella, N.M. (2021) Sudden stratospheric warmings. *Reviews of Geophysics*, 59, e2020RG000708. <https://doi.org/10.1029/2020RG000708>
- Baldwin, M.P. and Dunkerton, T.J. (2001) Stratospheric harbingers of anomalous weather regimes. *Science*, 294, 581–584. <https://doi.org/10.1126/science.1063315>
- Baldwin, M.P., Stephenson, D.B., Thompson, D.W.J., Dunkerton, T.J., Charlton, A.J. and O’Neill, A. (2003) Stratospheric memory and skill of extended-range weather forecasts. *Science*, 301, 636–640. <https://doi.org/10.1126/science.1087143>
- Bauer, P., Thorpe, A. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55. <https://doi.org/10.1038/nature14956>
- Beerli, R. and Grams, C.M. (2019) 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>
- Beerli, R., Wernli, H. and Grams, C. (2017) Does the lower stratosphere provide predictability for month-ahead wind electricity generation in Europe? *Quarterly Journal of the Royal Meteorological Society*, 143, 3025–3036. <https://doi.org/10.1002/qj.3158>
- Bloomfield, H.C., Brayshaw, D.J. and Charlton-Perez, A.J. (2020) Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. *Meteorological Applications*, 27, e1858. <https://doi.org/10.1002/met.1858>
- Bloomfield, H.C., Brayshaw, D.J., Gonzalez, P.L.M. and Charlton-Perez, A. (2021) Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables. *Meteorological Applications*, 28, e2018. <https://doi.org/10.1002/met.2018>
- Brier, G.W. (1950) Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- Büeler, D., Beerli, R., Wernli, H. and Grams, C.M. (2020) Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quarterly Journal of the Royal Meteorological Society*, 146, 3675–3694. <https://doi.org/10.1002/qj.3866>
- Buizza, R. and Leutbecher, M. (2015) The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, 141, 3366–3382. <https://doi.org/10.1002/qj.2619>
- Butler, A.H., Seidel, D.J., Hardiman, S.C., Butchart, N., Birner, T. and Match, A. (2015) Defining sudden stratospheric warmings. *Bulletin of the American Meteorological Society*, 96, 1913–1928. <https://doi.org/10.1175/BAMS-D-13-00173.1>
- Cassou, C. (2008) Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature*, 455, 523–527. <https://doi.org/10.1038/nature07286>
- Charlton-Perez, A.J., Ferranti, L. and Lee, R.W. (2018) The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 144, 1140–1151. <https://doi.org/10.1002/qj.3280>
- Cortesi, N., Torralba, V., Lledó, L., Manrique-Suñén, A., Gonzalez-Reviriego, N., Soret, A., and Doblas-Reyes F.J. (2021)



- Yearly evolution of Euro-Atlantic weather regimes and of their sub-seasonal predictability. *Climate Dynamics*, 56, 3933–3964. <https://doi.org/10.1007/s00382-021-05679-y>
- Dalcher, A. and Kalnay, E. (1987) Error growth and predictability in operational ECMWF forecasts. *Tellus A: Dynamic Meteorology and Oceanography*, 39, 474–491. <https://doi.org/10.3402/tellusa.v39i5.11774>
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. <https://doi.org/10.1002/qj.828>.
- Domeisen, D.I.V. (2019) Estimating the frequency of sudden stratospheric warming events from surface observations of the North Atlantic Oscillation. *Journal of Geophysical Research: Atmospheres*, 124, 3180–3194. <https://doi.org/10.1029/2018JD030077>
- Domeisen, D.I.V., Butler, A.H., Charlton-Perez, A.J., Ayarzagüena, B., Baldwin, M.P., Dunn-Sigouin, E., Furtado, J.C., Garfinkel, C.I., Hitchcock, P., Karpechko, A.Y., Kim, H., Knight, J., Lang, A.L., Lim, E., Marshall, A., Roff, G., Schwartz, C., Simpson, I.R., Son, S. and Taguchi, M. (2020a) The role of the stratosphere in sub-seasonal to seasonal prediction: 2. Predictability arising from stratosphere–troposphere coupling. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD030923. <https://doi.org/10.1029/2019JD030923>
- Domeisen, D.I.V., Grams, C.M. and Papritz, L. (2020b) The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events. *Weather Climate Dynamics*, 1, 373–388. <https://doi.org/10.5194/wcd-1-373-2020>
- Faranda, D., Masato, G., Moloney, N., Sato, Y., Daviaud, F., Dubrulle, B. and Yiou, P. (2016) The switching between zonal and blocked mid-latitude atmospheric circulation: a dynamical system perspective. *Climate Dynamics*, 47, 1587–1599. <https://doi.org/10.1007/s00382-015-2921-6>
- Feng, P.-N., Lin, H., Derome, J. and Merlis, T.M. (2021) Forecast skill of the NAO in the sub-seasonal-to-seasonal prediction models. *Journal of Climate*, 34, 4757–4769. <https://doi.org/10.1175/JCLI-D-20-0430.1>
- Ferranti, L., Corti, S. and Janousek, M. (2015) Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, 141, 916–924. <https://doi.org/10.1002/qj.2411>
- Ferranti, L., Magnusson, L., Vitart, F. and Richardson, D.S. (2018) How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe? *Quarterly Journal of the Royal Meteorological Society*, 144, 1788–1802. <https://doi.org/10.1002/qj.3341>
- Ferro, C.A.T. (2014) Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140, 1917–1923. <https://doi.org/10.1002/qj.2270>
- Ferro, C.A.T., Richardson, D.S. and Weigel, A.P. (2008) On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, 15, 19–24. <https://doi.org/10.1002/met.45>
- Fricker, T.E., Ferro, C.A.T. and Stephenson, D.B. (2013) Three recommendations for evaluating climate predictions. *Meteorological Applications*, 20, 246–255. <https://doi.org/10.1002/met.1409>
- Grams, C., Ferranti, L. and Magnusson, L. (2020) How to make use of weather regimes in extended-range predictions for Europe. *ECMWF Newsletter*, 165, 14–19. <https://doi.org/10.21957/mlk72gj183>
- Grams, C.M. and Archambault, H.M. (2016) The key role of diabatic outflow in amplifying the midlatitude flow: a representative case study of weather systems surrounding western North Pacific extratropical transition. *Monthly Weather Review*, 144, 3847–3869. <https://doi.org/10.1175/MWR-D-15-0419.1>
- Grams, C.M., Beerli, R., Pfenninger, S., Staffell, I. and Wernli, H. (2017) Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7, 557–562. <https://doi.org/10.1038/nclimate3338>.
- Grams, C.M., Magnusson, L. and Madonna, E. (2018) An atmospheric dynamics perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. *Quarterly Journal of the Royal Meteorological Society*, 144, 2577–2591. <https://doi.org/10.1002/qj.3353>
- 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. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. <https://doi.org/10.1002/qj.3803>
- Hochman, A., Messori, G., Quinting, J.F., Pinto, J.G. and Grams, C.M. (2021) Do Atlantic–European weather regimes physically exist? *Geophysical Research Letters*, 48, <https://doi.org/10.1029/2021GL095574>
- Hoskins, B. (2013) The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Quarterly Journal of the Royal Meteorological Society*, 139, 573–584. <https://doi.org/10.1002/qj.1991>
- Huang, W.T.K., Charlton-Perez, A., Lee, R.W., Neal, R., Sarran, C. and Sun, T. (2020) Weather regimes and patterns associated with temperature-related excess mortality in the UK: a pathway to sub-seasonal risk forecasting. *Environmental Research Letters*, 15, 124052. <https://doi.org/10.1088/1748-9326/abcbb>
- Jiménez-Esteve, B. and Domeisen, D.I.V. (2018) 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>
- Karpechko, A.Y. (2015) Improvements in statistical forecasts of monthly and two-monthly surface air temperatures using a stratospheric predictor. *Quarterly Journal of the Royal Meteorological Society*, 141, 2444–2456. <https://doi.org/10.1002/qj.2535>.
- Klaus, S.T. (2017) The connection between the Madden–Julian Oscillation and European weather regimes and its modulation by low frequency forcings. MSc Thesis, ETH Zurich, 79 pp., <https://doi.org/10.3929/ethz-b-000238793>
- Kolstad, E.W., Wulff, C.O., Domeisen, D.I.V. and Woollings, T. (2020) Tracing North Atlantic Oscillation forecast errors to stratospheric

- origins. *Journal of Climate*, 33, 9145–9157. <https://doi.org/10.1175/JCLI-D-20-0270.1>
- Koster, R.D., Mahanama, S.P., Yamada, T.J., Balsamo, G., Berg, A.A., Boisserie, M., Dirmeyer, P.A., Doblas-Reyes, F.J., Drewitt, G., Gordon, C.T., Guo, Z., Jeong, J.H., Lawrence, D.M., Lee, W.S., Li, Z., Luo, L., Malyshev, S., Merryfield, W.J., Seneviratne, S.I., Stanelle, T., Van Den Hurk, B.J., Vitart, F. and Wood, E.F. (2010) Contribution of land surface initialization to sub-seasonal forecast skill: first results from a multi-model experiment. *Geophysical Research Letters*, 37, L02402. <https://doi.org/10.1029/2009GL041677>.
- Lee, J.C.K., Lee, R.W., Woolnough, S.J. and Boxall, L.J. (2020a) The links between the Madden–Julian Oscillation and European weather regimes. *Theoretical and Applied Climatology*, 141, 567–586. <https://doi.org/10.1007/s00704-020-03223-2>
- Lee, R.W., Woolnough, S.J., Charlton-Perez, A.J. and Vitart, F. (2019) ENSO modulation of MJO teleconnections to the North Atlantic and Europe. *Geophysical Research Letters*, 46, 13535–13545. <https://doi.org/10.1029/2019GL084683>
- Lee, S.H., Lawrence, Z.D., Butler, A.H. and Karpechko, A.Y. (2020b) Seasonal forecasts of the exceptional Northern Hemisphere winter of 2020. *Geophysical Research Letters*, 47, e2020GL090328. <https://doi.org/10.1029/2020GL090328>
- Lin, H. (2020) Subseasonal forecast skill over the northern polar region in boreal winter. *Journal of Climate*, 33, 1935–1951. <https://doi.org/10.1175/JCLI-D-19-0408.1>
- Lin, H., Brunet, G. and Derome, J. (2009) An observed connection between the North Atlantic Oscillation and the Madden–Julian Oscillation. *Journal of Climate*, 22, 364–380. <https://doi.org/10.1175/2008JCLI2515.1>
- Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- Madden, R.A. and Julian, P.R. (1971) Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, 28, 702–708. [https://doi.org/10.1175/1520-0469\(1971\)028<0702:DOADOI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2)
- Madden, R.A. and Julian, P.R. (1972) Description of global-scale circulation cells in the tropics with a 40–50 day period. *Journal of the Atmospheric Sciences*, 29, 1109–1123. [https://doi.org/10.1175/1520-0469\(1972\)029<1109:DOGSCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2)
- Mariotti, A., Baggett, C., Barnes, E.A., Becker, E., Butler, A., Collins, D.C., Dirmeyer, P.A., Ferranti, L., Johnson, N.C., Jones, J., Kirtman, B.P., Lang, A.L., Molod, A., Newman, M., Robertson, A.W., Schubert, S., Waliser, D.E. and Albers, J. (2020) Windows of opportunity for skilful forecasts sub-seasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, 101, E608–E625. <https://doi.org/10.1175/BAMS-D-18-0326.1>
- Matsueda, M. and Palmer, T.N. (2018) Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, 144, 1012–1027. <https://doi.org/10.1002/qj.3265>
- Michel, C. and Rivière, G. (2011) The link between Rossby wave breakings and weather regime transitions. *Journal of the Atmospheric Sciences*, 68, 1730–1748. <https://doi.org/10.1175/2011JAS3635.1>
- Michel, C., Riviere, G., Terray, L. and Joly, B. (2012) The dynamical link between surface cyclones, upper-tropospheric Rossby wave breaking and the life cycle of the Scandinavian blocking. *Geophysical Research Letters*, 39, L10806. <https://doi.org/10.1029/2012GL051682>
- Michelangeli, P.-A., Vautard, R. and Legras, B. (1995) Weather regimes: recurrence and quasi stationarity. *Journal of the Atmospheric Sciences*, 52, 1237–1256. [https://doi.org/10.1175/1520-0469\(1995\)052<1237:WRRASQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1237:WRRASQ>2.0.CO;2)
- Mitchell, D.M., Gray, L.J., Anstey, J., Baldwin, M.P. and Charlton-Perez, A.J. (2013) The influence of stratospheric vortex displacements and splits on surface climate. *Journal of Climate*, 26, 2668–2682. <https://doi.org/10.1175/JCLI-D-12-00030.1>
- Neal, R., Fereday, D., Crocker, R. and Comer, R.E. (2016) A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, 23, 389–400. <https://doi.org/10.1002/met.1563>
- Orsolini, Y.J., Senan, R., Vitart, F., Balsamo, G., Weisheimer, A. and Doblas-Reyes, F.J. (2016) Influence of the Eurasian snow on the negative North Atlantic Oscillation in sub-seasonal forecasts of the cold winter 2009/2010. *Climate Dynamics*, 47, 1325–1334. <https://doi.org/10.1007/s00382-015-2903-8>
- Palmer, T.N. (1993) Extended-range atmospheric prediction and the Lorenz model. *Bulletin of the American Meteorological Society*, 74, 49–65. [https://doi.org/10.1175/1520-0477\(1993\)074<0049:ERAPAT>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<0049:ERAPAT>2.0.CO;2)
- Polvani, L.M. and Waugh, D.W. (2004) Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *Journal of Climate*, 17, 3548–3554. [https://doi.org/10.1175/1520-0442\(2004\)017<3548:UWAFAA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2)
- Quinting, J.F. and Grams, C.M. (2021) Towards a systematic evaluation of warm conveyor belts in numerical weather prediction and climate models. Part I: predictor selection and logistic regression model. *Journal of the Atmospheric Sciences*, 78, 1465–1485. <https://doi.org/10.1175/JAS-D-20-0139.1>
- Quinting, J.F. and Vitart, F. (2019) Representation of synoptic-scale Rossby wave packets and blocking in the S2S Prediction Project Database. *Geophysical Research Letters*, 46, 1070–1078. <https://doi.org/10.1029/2018GL081381>
- Rivière, G. and Drouard, M. (2015) Dynamics of the Northern Annular Mode at weekly time scales. *Journal of the Atmospheric Sciences*, 72, 4569–4590. <https://doi.org/10.1175/JAS-D-15-0069.1>
- Rivière, G. and Orlanski, I. (2007) Characteristics of the Atlantic storm-track eddy activity and its relation with North Atlantic Oscillation. *Journal of the Atmospheric Sciences*, 64, 241–266. <https://doi.org/10.1175/JAS3850.1>
- Rodwell, M.J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., Diamantakis, M., Earnshaw, P., Garcia-Mendez, A., Isaksen, L., Källén, E., Klocke, D., Lopez, P., McNally, T., Persson, A., Prates, F. and Wedi, N. (2013) Characteristics of occasional poor medium-range weather forecasts for Europe. *Bulletin of the American Meteorological Society*, 94, 1393–1405. <https://doi.org/10.1175/BAMS-D-12-00099.1>
- Rodwell, M.J., Richardson, D.S., Parsons, D.B. and Wernli, H. (2018) Flow-dependent reliability: a path to more skilful ensemble forecasts. *Bulletin of the American Meteorological Society*, 99, 1015–1026. <https://doi.org/10.1175/BAMS-D-17-0027.1>
- Rodwell, M.J., Rowell, D.P. and Folland, C.K. (1999) Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature*, 398, 320–323. <https://doi.org/10.1038/18648>

- Scaife, A.A., Karpechko, A.Y., Baldwin, M.P., Brookshaw, A., Butler, A.H., Eade, R., Gordon, M., MacLachlan, C., Martin, N., Dunstone, N. and Smith, D. (2016) Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, 17, 51–56. <https://doi.org/10.1002/asl.598>.
- Schaller, N., Sillman, J., Anstey, J., Fischer, E.M., Grams, C.M. and Russo, S. (2018) Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles. *Environmental Research Letters*, 13, 054015. <https://doi.org/10.1088/1748-9326/aaba55>
- Scherhag, R. (1952) Die explosionsartigen Stratosphärenwärmungen des Spätwinter 1951/1952 (The explosive warmings in the stratosphere of the late winter 1951/1952). *Berichte des Deutschen Wetterdienstes U.S. Zone*, 38, 51–63.
- Son, S.W., Kim, H., Song, K., Kim, S.W., Martineau, P., Hyun, Y.K. and Kim, Y. (2020) Extratropical prediction skill of the sub-seasonal-to-seasonal (S2S) prediction models. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031273. <https://doi.org/10.1029/2019JD031273>
- Spensberger, C., Madonna, E., Boettcher, M., Grams, C.M., Papritz, L., Quinting, J.F., Röthlisberger, M., Sprenger, M. and Zschenderlein, P. (2020) Dynamics of concurrent and sequential Central European and Scandinavian heatwaves. *Quarterly Journal of the Royal Meteorological Society*, 146, 2998–3013. <https://doi.org/10.1002/qj.3822>
- Stan, C., Straus, D.M., Frederiksen, J.S., Lin, H., Maloney, E.D. and Schumacher, C. (2017) Review of tropical-extratropical teleconnections on intraseasonal time scales. *Reviews of Geophysics*, 55, 902–937. <https://doi.org/10.1002/2016RG000538>.
- Toniazzo, T. and Scaife, A.A. (2006) The influence of ENSO on winter North Atlantic climate. *Geophysical Research Letters*, 33, L24704. <https://doi.org/10.1029/2006GL027881>.
- Torralba, V., Gonzalez-Reviriego, N., Cortesi, N., Manrique-Suñén, A., Lledó, L., Marcos, R., Soret, A. and Doblas-Reyes, F.J. (2021) Challenges in the selection of atmospheric circulation patterns for the wind energy sector. *International Journal of Climatology*, 41, 1525–1541. <https://doi.org/10.1002/joc.6881>
- Tripathi, O.P., Charlton-Perez, A., Sigmond, M. and Vitart, F. (2015) Enhanced long-range forecast skill in boreal winter following stratospheric strong vortex conditions. *Environmental Research Letters*, 10, 104007. <https://doi.org/10.1088/1748-9326/10/10/104007>.
- Vautard, R. (1990) Multiple weather regimes over the North Atlantic: analysis of precursors and successors. *Monthly Weather Review*, 118, 2056–2081. [https://doi.org/10.1175/1520-0493\(1990\)118<2056:MWRONTN>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<2056:MWRONTN>2.0.CO;2)
- Vitart, F. (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140, 1889–1899. <https://doi.org/10.1002/qj.2256>.
- Vitart, F. (2017) Madden–Julian Oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, 143, 2210–2220. <https://doi.org/10.1002/qj.3079>.
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A.W., Rutti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D.J., Xiao, H., Zaripov, R. and Zhang, L. (2017) The sub-seasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98, 163–173. <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Vitart, F. and Molteni, F. (2010) Simulation of the Madden–Julian Oscillation and its teleconnections in the ECMWF forecast system. *Quarterly Journal of the Royal Meteorological Society*, 136, 842–855. <https://doi.org/10.1002/qj.623>
- Wandel, J., Quinting, J.F. and Grams, C.M. (2021) Towards a systematic evaluation of warm conveyor belts in numerical weather prediction and climate models. Part II: verification of operational reforecasts. *Journal of the Atmospheric Sciences*. <https://doi.org/10.1175/JAS-D-20-0385.1>
- Wheeler, M.C. and Hendon, H.H. (2004) An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Monthly Weather Review*, 132, 1917–1932. [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2)
- van der Wiel, K., Bloomfield, H.C., Lee, R.W., Stoop, L.P., Blackport, R., Screen, J.A. and Selten, F.M. (2019) The influence of weather regimes on European renewable energy production and demand. *Environmental Research Letters*, 14, 094010. <https://doi.org/10.1088/1748-9326/ab38d3>.
- Wilks, D. (2011) *Statistical Methods in the Atmospheric Sciences*. San Diego, CA: Academic Press.
- Yamagami, A. and Matsueda, M. (2020) Subseasonal forecast skill for weekly mean atmospheric variability over the Northern Hemisphere in winter and its relationship to midlatitude teleconnections. *Geophysical Research Letters*, 47, e2020GL088508. <https://doi.org/10.1029/2020GL088508>
- Yiou, P. and Nogaj, M. (2004) Extreme climatic events and weather regimes over the North Atlantic: when and where? *Geophysical Research Letters*, 31, L07202. <https://doi.org/10.1029/2003GL019119>
- Zhang, C. and Dong, M. (2004) Seasonality in the Madden–Julian Oscillation. *Journal of Climate*, 17, 3169–3180. [https://doi.org/10.1175/1520-0442\(2004\)017<3169:SITMO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3169:SITMO>2.0.CO;2)
- Zhang, F., Qiang Sun, Y., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H. and Emanuel, K. (2019) What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, 76, 1077–1091. <https://doi.org/10.1175/JAS-D-18-0269.1>
- Zhu, H., Wheeler, M.C., Sobel, A.H. and Hudson, D. (2014) Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Monthly Weather Review*, 142, 1556–1569. <https://doi.org/10.1175/MWR-D-13-00222.1>

## SUPPORTING INFORMATION

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

**How to cite this article:** Büeler, D., Ferranti, L., Magnusson, L., Quinting, J.F. & Grams, C.M. (2021) Year-round sub-seasonal forecast skill for Atlantic–European weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 147(741), 4283–4309. Available from: <https://doi.org/10.1002/qj.4178>