

RESEARCH ARTICLE

Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries

Dominik Büeler^{1,3}  | Remo Beerli^{2,3}  | Heini Wernli³  | Christian M. Grams^{1,3} 

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

²Axpo Solutions AG, Baden, Switzerland

³Institute for Atmospheric and Climate Science, ETH Zürich, Zurich, Switzerland

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

Axpo Solutions AG;
Helmholtz-Gemeinschaft, Grant/Award Number: VH-NG-1243

Abstract

Meteorologists in the energy industry increasingly draw upon the potential for enhanced sub-seasonal predictability of European surface weather following anomalous states of the winter stratospheric polar vortex (SPV). How the link between the SPV and the large-scale tropospheric flow translates into forecast skill for surface weather in individual countries – a spatial scale that is particularly relevant for the energy industry – remains an open question. Here we quantify the effect of anomalously strong and weak SPV states at forecast initial time on the probabilistic extended-range reforecast skill of the European Centre for Medium-Range Weather Forecasts (ECMWF) in predicting country- and month-ahead-averaged anomalies of 2 m temperature, 10 m wind speed, and precipitation. After anomalous SPV states, specific surface weather anomalies emerge, which resemble the opposing phases of the North Atlantic Oscillation. We find that forecast skill is, to first order, only enhanced for countries that are entirely affected by these anomalies. However, the model has a flow-dependent bias for 2 m temperature (T2M): it predicts the warm conditions in Western, Central and Southern Europe following strong SPV states well, but is overconfident with respect to the warm anomaly in Scandinavia. Vice versa, it predicts the cold anomaly in Scandinavia following weak SPV states well, but struggles to capture the strongly varying extent of the cold air masses into Central and Southern Europe. This tends to reduce skill (in some cases significantly) for Scandinavian countries following strong SPV states, and most pronounced, for many Central, Southern European, and Balkan countries following weak SPV states. As most of the weak SPV states are associated with sudden stratospheric warmings (SSWs), our study thus advises particular caution when interpreting sub-seasonal regional T2M forecasts following SSWs. In contrast, it suggests that the model benefits from enhanced predictability for a considerable part of Europe following strong SPV states.

KEYWORDS

energy, European countries, polar vortex, stratosphere, sub-seasonal forecast skill, sudden stratospheric warming, surface weather

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.

© 2020 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

1 | INTRODUCTION

The energy industry is strongly affected by weather on hourly to seasonal time-scales via influences on electricity supply, demand and prices (e.g., Troccoli *et al.*, 2014; Thornton *et al.*, 2019). With the continuously increasing use of renewable energy sources such as wind or sunlight, this weather dependence has and will become even more critical (e.g., Grams *et al.*, 2017; Bloomfield *et al.*, 2018; van der Wiel *et al.*, 2019). For instance, the daily-averaged wind power generation in Germany can vary up to 30 GW¹ and even more for Europe depending on the prevailing synoptic situation (e.g., Beerli *et al.*, 2017; Grams *et al.*, 2017). Consulting operational numerical weather forecasts for specific countries is thus an important component of the energy industry's daily business. The steady increase in computational power combined with a continuous improvement of numerical weather prediction (NWP) systems in the last decades (Bauer *et al.*, 2015) has allowed us to extend the operational forecast horizon to sub-seasonal time-scales (10–60 days; Vitart and Robertson, 2019). This time-scale closes the gap toward seasonal climate predictions, which have already been run operationally for more than a decade (e.g., Palmer *et al.*, 2004; Kirtman *et al.*, 2014), and thus advances the worldwide efforts toward seamless weather and climate prediction systems (Vitart and Robertson, 2019). Forecasting European weather for lead times beyond two weeks is still only weakly to moderately skilful, with a strong dependence on the season and the initial and predicted large-scale flow situation (e.g., Weigel *et al.*, 2008; Vitart, 2014; Ferranti *et al.*, 2018), and therefore is an area of active research (Vitart *et al.*, 2017). Nevertheless, the European energy industry has already become one of the most important end-users of operational sub-seasonal forecasts. This requires a profound understanding on if and how important sources of sub-seasonal predictability enhance forecast skill for energy-industry-tailored parameters. Beerli *et al.* (2017) tackled this question by investigating the role of the stratosphere for the predictability of country-aggregated wind power generation across Europe on monthly time-scales in reanalysis data (see below for more details). Here we build upon this study and investigate how the predictability from the stratosphere found by Beerli *et al.* (2017) translates into sub-seasonal numerical forecast skill for energy-industry-relevant European surface weather on a country scale. We continue with a brief discussion of the current knowledge about sources of sub-seasonal predictability and forecast skill for European surface weather,

before we discuss the scope of the study at hand in more detail.

Predictability of the large-scale flow over Europe on sub-seasonal time-scales is generally gained from the state of low-frequency climate modes such as the winter stratospheric polar vortex (SPV; e.g., Baldwin and Dunkerton, 1999; Ambaum and Hoskins, 2002), the Madden–Julian Oscillation (e.g., Cassou, 2008; Lin *et al.*, 2009), the El Niño–Southern Oscillation (e.g., Huang *et al.*, 1998; Toniazzo and Scaife, 2006), tropical rainfall in general (e.g., Scaife *et al.*, 2017; Stan *et al.*, 2017), the Quasi-Biennial Oscillation (e.g., Holton and Tan, 1980; Anstey and Shepherd, 2014; Andrews *et al.*, 2019), or variations in North Atlantic sea surface temperature (e.g., Rodwell *et al.*, 1999) and soil moisture (e.g., Koster *et al.*, 2010; Ardilouze *et al.*, 2017). It is now established that amongst those, the SPV is potentially the most important source of sub-seasonal predictability for European weather during mid- and late winter (Kidston *et al.*, 2015; Butler *et al.*, 2019). This is because anomalously strong SPV states are statistically followed by persistent (up to 60 days) positive phases of the North Atlantic Oscillation (NAO), and anomalously weak SPV states by persistent negative phases of the NAO (Baldwin and Dunkerton, 2001; Charlton-Perez *et al.*, 2018). As the NAO accounts for a major part of the large-scale flow variability over Europe (Hurrell, 1996) and thus strongly influences energy-industry-relevant surface weather (e.g., Brayshaw *et al.*, 2011; Clark *et al.*, 2017; Zubiate *et al.*, 2017), it acts as a connecting link between the stratosphere and the surface. Typically, a positive phase of the NAO is associated with stronger (weaker) than normal near-surface wind and precipitation in Northern (Southern) Europe and higher than normal near-surface temperatures in Central to Northern Europe, and *vice versa* during a negative phase of the NAO. The mechanisms behind the downward influence from the stratosphere to the troposphere are manifold and partly still debated (Butler *et al.*, 2019). Most likely, the tropospheric conditions before and during the anomalous SPV state (e.g., Ambaum and Hoskins, 2002; Attard *et al.*, 2016; Kodera *et al.*, 2016; Karpechko *et al.*, 2017; Schneidereit *et al.*, 2017; White *et al.*, 2019; Domeisen *et al.*, 2020b) as well as internal dynamics of the SPV itself (e.g., Holton and Mass, 1976; Scott and Haynes, 2000; Albers and Birner, 2014; Jucker, 2016) influence the anomalous SPV event and its subsequent surface impact. A particular type of anomalously weak SPV state is the so-called sudden stratospheric warming (SSW; Scherhag, 1952; Matsuno, 1971; Butler *et al.*, 2015), during which the stratosphere warms abruptly and the westerlies associated with the SPV temporarily reverse. As with weak SPV states as a whole, SSWs can trigger negative phases of the NAO leading to extreme cold waves

¹https://www.energy-charts.de/power_de.htm?source=solar-wind&year=2019&month=3 (accessed 24 January 2020)

particularly over Northern Europe and Northern Russia (e.g., Kidston *et al.*, 2015; Kautz *et al.*, 2020). However, due to the large case-to-case variability in the dynamics of SSWs (e.g., Mitchell *et al.*, 2013; Kretschmer *et al.*, 2018), the tropospheric weather regime response can also differ from the classical negative phase of the NAO and for instance yield opposite temperature conditions for Europe (Beerli and Grams, 2019; Domeisen *et al.*, 2020b). SSWs are thus particularly delicate events for the European energy industry.

The potential predictability from the stratosphere has triggered various studies quantifying the effect of anomalous SPV states, in particular of the critical SSWs, on sub-seasonal forecast skill. Most of these studies found an increase in skill both for statistical models (e.g., Baldwin *et al.*, 2003; Charlton *et al.*, 2003; Christiansen, 2005; Karpechko, 2015) and numerical models (e.g., Charlton *et al.*, 2004; Mukougawa *et al.*, 2009; Sigmond *et al.*, 2013; Tripathi *et al.*, 2015a; 2015b; Scaife *et al.*, 2016) in predicting a particular tropospheric weather regime response or anomalous large-scale surface weather conditions directly. However, so far only Beerli *et al.* (2017) have investigated how anomalous SPV states affect the skill for predicting country-averaged surface weather all over Europe, although this is the scale the energy industry strongly relies on. They used a simple statistical model based on the SPV state in ERA-Interim to predict month-ahead-averaged anomalies of country-aggregated wind power generation in Europe. They obtained considerable skill for Northern European countries, but only moderate skill for Central European and no skill for Southern European countries. The skill was shown to be associated with both anomalously strong and weak SPV states at initial time, with the exception of Southern Europe, where the weak SPV states led to negative skill and thus to no skill overall. They showed that the NAO-related near-surface wind anomaly dipoles that follow the anomalous SPV states and affect primarily Northern and Southern Europe were mainly responsible for this skill pattern across Europe.

In this study, we investigate to what degree the influence of an anomalously strong and weak SPV on statistical forecast skill for wind power generation, as found by Beerli *et al.* (2017), applies to the probabilistic surface weather reforecast skill of the ECMWF (European Centre for Medium-Range Weather Forecasts) sub-seasonal NWP ensemble. We relate the SPV state at forecast initial time to the skill in predicting country- and month-ahead-averaged 2 m temperature, 10 m wind speed, and precipitation in Europe. These parameters are directly relevant for the energy industry and other end users of sub-seasonal forecasts (such as national weather

services).² We analyze the three surface parameters separately, although they often need to be considered in combination when predicting country-specific energy demand, for instance. Investigating the skill from such a multivariate perspective would thus be important as well but this goes beyond the scope of this study. Section 2 starts with an overview of the used datasets as well as the applied skill score and statistical tests. The results are given in Section 3. They demonstrate how the predicted anomaly composites compare to ERA-Interim and how this translates into the skill for individual countries. Section 4 discusses and summarizes the most important findings.

2 | DATA AND METHODS

2.1 | Model and observations

We investigate the skill of ECMWF sub-seasonal reforecasts (hereafter just denoted “forecasts”), obtained from the Subseasonal-to-Seasonal (S2S) Prediction Project Database (Vitart *et al.*, 2017), for 2 m temperature $T2M$ (daily-averaged), 10 m wind speed $U10M$ (instantaneous at 0000 UTC), and total precipitation P (daily accumulated) during 22 winters between December 1995 and February 2017 (initial dates in DJF, i.e., the lead times of some forecasts extend into March). The reforecasts had been initialized from ERA-Interim twice a week, run with 11 ensemble members (1 control and 10 perturbed forecasts) for a lead time of 46 days, and interpolated to a $1.5^\circ \times 1.5^\circ$ grid after computation. Similarly to other studies (e.g., Schiraldi and Roundy, 2017; Deflorio *et al.*, 2018), we use more than two reforecasts per week to increase sample size (to a total of 1,280), which comes with the caveat of mixing different model versions (in our case those are CY41R1, CY41R2, CY43R1, and CY43R3). Some of these versions differ with respect to horizontal resolution of the atmosphere (i.e., from 32 km up to day 10 and 64 km beyond to 15 km up to day 15 and 31 km beyond) and ocean (i.e., from 1 to 0.25°) and the inclusion of active sea ice. The number of vertical levels (91), which is critical for the representation of stratosphere–troposphere coupling (e.g., Manney *et al.*, 2017; Kawatani *et al.*, 2019), is the same in all model versions. It is further important to note that the reforecast dataset used here likely yields

²Analyzing 100 m (instead of 10 m) wind speed would be ideal with regard to wind power applications, but this parameter is not available in the Subseasonal-to-Seasonal Prediction Project Database (Vitart *et al.*, 2017) used for this study. However, a comparison of daily 10 m and 100 m wind speed in 40 years of reanalysis data averaged over Germany yields a very high correlation (not shown), which makes us confident in using 10 m wind speed instead.

lower levels of skill than the operational sub-seasonal forecasting system of the ECMWF would have, particularly during the first two weeks, due to the better initial conditions from data assimilation and a higher number of ensemble members in the operational set-up (cf., e.g., Vitart, 2014). Nevertheless, investigating reforecast data enables us to elucidate systematic model behaviour. As an observational reference, we use ERA-Interim reanalysis data (Dee *et al.*, 2011) for the same winters between 1995 and 2017, also retrieved on a $1.5^\circ \times 1.5^\circ$ grid. The three surface weather parameters in ERA-Interim are postprocessed to be consistent with the model data: $T2M$ is calculated as an average over the 0000, 0600, 1200, and 1800 UTC values, $U10M$ from the instantaneous 10 m u - and v -components at 0000 UTC, and total P as the sum over the 6-hourly accumulated fields at 0600, 1200, 1800, and 0000 UTC.

2.2 | Definition of country- and month-ahead-averaged anomalies

We evaluate the model skill for country- and month-ahead-averaged anomalies of $T2M$, $U10M$, and P , which are defined as follows. First, we calculate the model climatology as a function of the calendar day at forecast initial time and the lead time. For example, the model climatology for the forecast initialized on 1 December 2000 is obtained by averaging over all 11 ensemble members of all the forecasts initialized on the same calendar day (i.e., on 1 December 1995, 1996, 1997, ..., 2000, ...) separately for every lead time. The daily anomalies at every lead time of a certain forecast are then calculated with respect to the corresponding lead-time-dependent model climatology. A similar principle for calculating calibrated model anomalies has been used in various studies analyzing data from the S2S database (e.g., Vitart, 2017; de Andrade *et al.*, 2019; Wulff and Domeisen, 2019). Second, these daily anomalies are averaged spatially over every European country and, third, temporally over the first 31 days of lead time (including forecast initial time). This yields one value per forecast for every country and parameter. The corresponding country- and month-ahead-averaged anomalies in ERA-Interim are obtained in the same way, but with respect to the ERA-Interim climatology, which is calculated as a 31-day running mean over the calendar-day average of all available years (1995–2017). Although the approach for calculating the model climatology is different from the one used for the ERA-Interim climatology (using ensemble mean instead of running mean), it allows for a similarly strong smoothing and eliminates (lead-time-dependent) model drifts. In this study, we focus on month-ahead anomalies as a particularly

relevant time-scale for the energy industry (e.g., Dorrington *et al.*, 2020). In addition, predictability on sub-seasonal time-scales is generally higher for temporally and spatially aggregated fields compared to instantaneous fields (e.g., Buizza and Leutbecher, 2015). Nevertheless, it is important to keep in mind that the first two weeks typically contribute most to forecast skill on sub-seasonal time-scales (e.g., Weigel *et al.*, 2008; Vitart, 2014). For this reason, we also briefly discuss forecast skill for weeks 1 to 4.

2.3 | Definition of stratospheric polar vortex strength

To investigate the model skill dependent on stratospheric conditions, we additionally calculate the strength of the polar vortex in the lower stratosphere (for simplicity just referred to as stratospheric polar vortex, SPV, hereafter) at forecast initial time (Beerli *et al.*, 2017). First, we calculate the instantaneous geopotential height anomalies at 100 hPa for every ensemble member (the 150 hPa level, used by Beerli *et al.*, 2017, is not available in the S2S database) with respect to model climatologies computed as for the surface parameters. The ensemble mean of these anomalies is then spatially averaged over the polar cap north of 60°N . Defining the SPV strength on a level in the lower stratosphere instead of the middle stratosphere (~ 10 hPa), where the SPV is typically strongest, has shown to be most meaningful for evaluating stratospheric surface weather impacts (Baldwin *et al.*, 2003; Karpechko, 2015; Beerli *et al.*, 2017; Charlton-Perez *et al.*, 2018). This is likely because the lower stratospheric circulation is a good indicator for stratosphere–troposphere coupling. To investigate the influence of anomalously weak and strong SPV states on forecast skill, we define percentile bins of the SPV anomalies of all forecast initial times and, based on those, select the forecasts initialized in the 2% strongest, 10% strongest, 10% weakest, and 2% weakest SPV states (weak SPV states correspond to positive polar-cap-averaged geopotential height anomalies, and *vice versa*).

2.4 | Ranked probability skill score (RPSS)

We use a three-category ranked probability skill score (RPSS; Wilks, 2011) to quantify how well a set of n model forecasts predicts above-normal, near-normal, or below-normal $T2M$, $U10M$, and P in a country (i.e., the upper, middle, and lower terciles of the corresponding month-ahead anomalies, respectively) compared to

a climatological reference forecast. In our case, the climatological reference forecast assumes an equal occurrence probability of one third for each tercile. The RPSS is calculated by relating the mean ranked probability score (RPS) of the n model forecasts, $\overline{RPS} = \frac{1}{n} \sum_{i=1}^n RPS_i$, to the mean ranked probability score of the corresponding climatological reference forecasts, $\overline{RPS}_{\text{clim}}$ (Wilks, 2011):

$$RPSS = 1 - \frac{\overline{RPS}}{\overline{RPS}_{\text{clim}}}. \quad (1)$$

The RPS for a single forecast is calculated as (Wilks, 2011)

$$RPS = \sum_{m=1}^3 (Y_m - O_m)^2 = \sum_{m=1}^3 \left[\left(\sum_{j=1}^m y_j \right) - \left(\sum_{j=1}^m o_j \right) \right]^2. \quad (2)$$

Here, y_j and o_j are the cumulative probability vector components of the forecast and the observation with $j = 1, 2, 3$ corresponding to the three terciles. For example, a forecast, in which 30% of the ensemble members predict an anomaly belonging to the first tercile, 60% to the second tercile, and 10% to the third tercile, yields a cumulative forecast probability vector $Y = (0.3, 0.9, 1.0)$. Assuming the observed anomaly would belong to the second tercile (i.e., the second component of the observation vector would have a probability of 1 and the other two components one of 0) would yield a cumulative observation vector $O = (0, 1, 1)$. And, in our case, the cumulative climatological reference forecast vector required for calculating $\overline{RPS}_{\text{clim}}$ would consequently be $Y_{\text{clim}} = (1/3, 2/3, 1)$. The RPS is thus 0 for a perfect forecast and positive for a non-perfect forecast. And, consequently, the RPSS is 1 for a perfect forecast, 0 if the forecast is equally good as the climatological reference forecast, and negative if the forecast is worse than the climatological reference forecast. In this study, the terciles for the forecast vector Y for a specific country and parameter are defined based on the corresponding month-ahead anomalies from all ensemble members in all available forecasts between 1995 and 2017, whereas the terciles for the observation vector O are defined based on the daily month-ahead anomalies in ERA-Interim but over the same years.

2.5 | Statistical testing

Grouping the forecasts into those initialized in the extreme SPV states (Section 2.3), which is a key approach of this study, yields a relatively small sample size by nature. For instance, the two groups associated with the 10% strongest and 10% weakest SPV states consist of 129 and 127 individual forecasts, respectively. In addition, several

of these individual forecasts are initialized only some days apart and thus most likely during the same extreme SPV period. The two aforementioned forecast groups consist of only 12 such periods each (mainly representing individual winters), which reduces the samples to a small number of truly independent forecasts. To account for these sample sizes and autocorrelation problems as well as to test whether forecast groups are significantly different from climatological (i.e., SPV-independent) conditions (see below for details), we create two different bootstrapped distributions for each of these forecast groups: a (group-internal) resampled distribution and a corresponding climatological distribution. To obtain the resampled distribution, we randomly resample 1,000 times 80% of all the winters in a certain forecast group (for instance, 80% of the 12 winters in the case of the aforementioned forecast groups) without replacement. To obtain the climatological distribution corresponding to a certain forecast group, we randomly sample 1,000 times a set of forecasts (with the same size as well as a similar calendar day sequence of initial times as the original forecast group) from all available forecasts independent of the SPV state.

On the one hand, the resampled and corresponding climatological distributions are used to quantify robustness, as defined by Papritz (2020), and significance of the mean month-ahead anomalies (i.e., the composite maps) of a forecast group (Figures 3, 5, and 6 below in Section 3). Robustness indicates to what degree the mean anomaly is representative for the underlying individual anomalies (i.e., how little the individual anomalies differ from the mean anomaly). To this end, we calculate the mean anomaly for each of the 1,000 resamples of the resampled distribution, which yields a distribution of mean anomalies. Based on this distribution and following Papritz (2020), we define the actual mean anomaly to be robust on the first level, if the absolute value of the mean anomaly is greater than the interquartile range of the resampled distribution, and on the second level, if it is greater than the difference between the 90 and 10% percentiles (Figure 1). As a second measure, significance quantifies how strongly the anomaly differs from climatology. We determine this by comparing the resampled distribution of mean anomalies of a forecast group to the corresponding climatological distribution of mean anomalies. We define the mean anomaly to be significant if the 12.5% percentile of the resampled distribution is greater than the 87.5% percentile of the corresponding climatological distribution for positive anomalies and inverse for negative anomalies (Figure 1; the numbers 12.5 and 87.5% result from the arbitrary but reasonable condition that the centred 75% of the two distributions should not overlap). Although these two statistical characteristics are related, the robustness additionally helps us to understand the RPSS for a certain

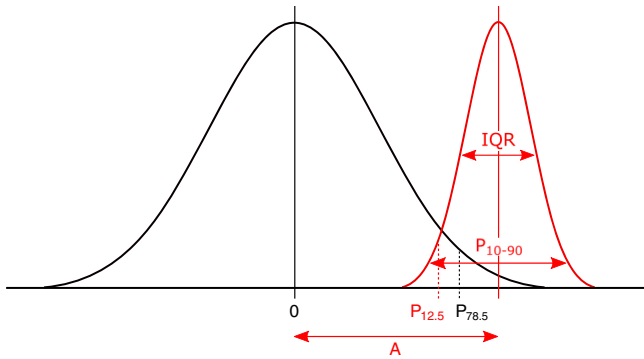


FIGURE 1 Principle for assessing robustness and significance demonstrated with a synthetic example: the mean anomaly A of the forecast group is robust on the first level because A is larger than the interquartile range IQR of the resampled distribution (red curve), and on the second level because it is larger than the range between the 10 and 90%iles P_{10-90} . However, the anomaly is not significant as the 12.5%ile $P_{12.5}$ of the resampled distribution is not larger than the 78.5%ile $P_{78.5}$ of the climatological distribution (black curve)

mean anomaly: for instance, a robust modelled anomaly but non-robust observed anomaly could indicate that the model mostly predicts an anomaly of the same tercile in contrast to the observed anomalies falling into different terciles (due to their larger spread around the mean anomaly), which would yield a low RPSS.

On the other hand, we use the resampled and corresponding climatological distributions to determine whether the RPSS of a forecast group is significantly different from climatological (i.e., SPV-independent) conditions (Figures 4 and 7 below in Section 3). This is achieved by following the same principle as used for testing the significance of the mean anomalies: we calculate the RPSS for each of the 1,000 resamples in the resampled distribution and for each of the 1,000 samples in the corresponding climatological distribution. Based on the resulting two RPSS distributions, we define the RPSS of a forecast group to be significantly different from climatological conditions on the 25, 10, or 5% level, if the 25/75%, 10/90%, or 5/95% percentiles of the two distributions do not overlap.

3 | RESULTS

3.1 | Overall month-ahead forecast skill

Figures 2a–c show the RPSS for month-ahead 2 m temperature ($T2M$), 10 m wind speed ($U10M$), and precipitation (P) of all forecasts between 1995 and 2017 for most European countries. On average, the RPSS is higher for $T2M$ than for $U10M$ and P , and ranges from 0 up to 0.4. There are large regional variations of skill with a characteristic

pattern for all of the three parameters: for $T2M$, the RPSS tends to be higher in Central to Eastern Europe, Sweden, and Italy than in Western to Southwestern Europe as well as the UK, Norway, Finland, and parts of the Balkans. For $U10M$, the RPSS is highest in Northern Europe (particularly Sweden) but also in the UK, Baltics, and Iberian Peninsula and much lower in Central to Eastern Europe and the Balkans. A similar pattern occurs for P , although the highest values in Northern Europe are limited to Norway.

Figures 2d–f provide a first qualitative test of whether and how anomalous SPV states at forecast initial time – independently of whether they are anomalously strong or weak – might influence this regional variation in the overall RPSS in Figures 2a–c: it summarizes the ERA-Interim anomaly patterns for the three parameters after the 10% strongest and 10% weakest SPV states (see caption for more details; note that the separate anomalies following these two extreme SPV states will be shown and discussed later in Figures 3d,f, 5d,f, and 6d,f). In addition, the mean absolute month-ahead sea level pressure (SLP) of the two forecast groups is indicated with two different contours. Figures 2d–f thus reflect the NAO-related patterns that typically follow anomalous SPV states (e.g., Tripathi *et al.*, 2015b; Beerli *et al.*, 2017): after the 10% strongest SPV states, a strong meridional SLP gradient occurs over the North Atlantic and Europe (solid contours). This likely indicates the positive-NAO-like weather pattern characterized by a strong zonal jet stream. In contrast, the meridional SLP gradient is much weaker after the 10% weakest SPV states (dashed contours), which likely reflects the negative-NAO-like weather pattern with a weaker, wavier, and more southward-shifted jet stream. These circulation patterns lead to the characteristic NAO surface weather anomalies, indicated by the shading, which particularly affect Central to Northeastern Europe in terms of $T2M$ and Northern and Southern Europe in terms of $U10M$ and P . The regions affected by strong anomalies (Figures 2d–f) are to first order co-located with the regions experiencing a relatively high RPSS (Figures 2a–c). This points toward a potential influence of anomalous SPV states on the overall forecast skill via the modification of the NAO. However, it does not indicate the separate contributions from the anomalies following strong and weak SPV states. In the next section, we thus investigate this correlation, and potential causality, in more detail by quantifying the model skill after anomalously strong and weak SPV states separately. We first analyze how well the model reproduces the anomaly patterns after strong and weak SPV states compared to ERA-Interim. This helps us to understand the RPSS for individual countries after anomalous SPV states, which is discussed subsequently. Our main focus is on $T2M$

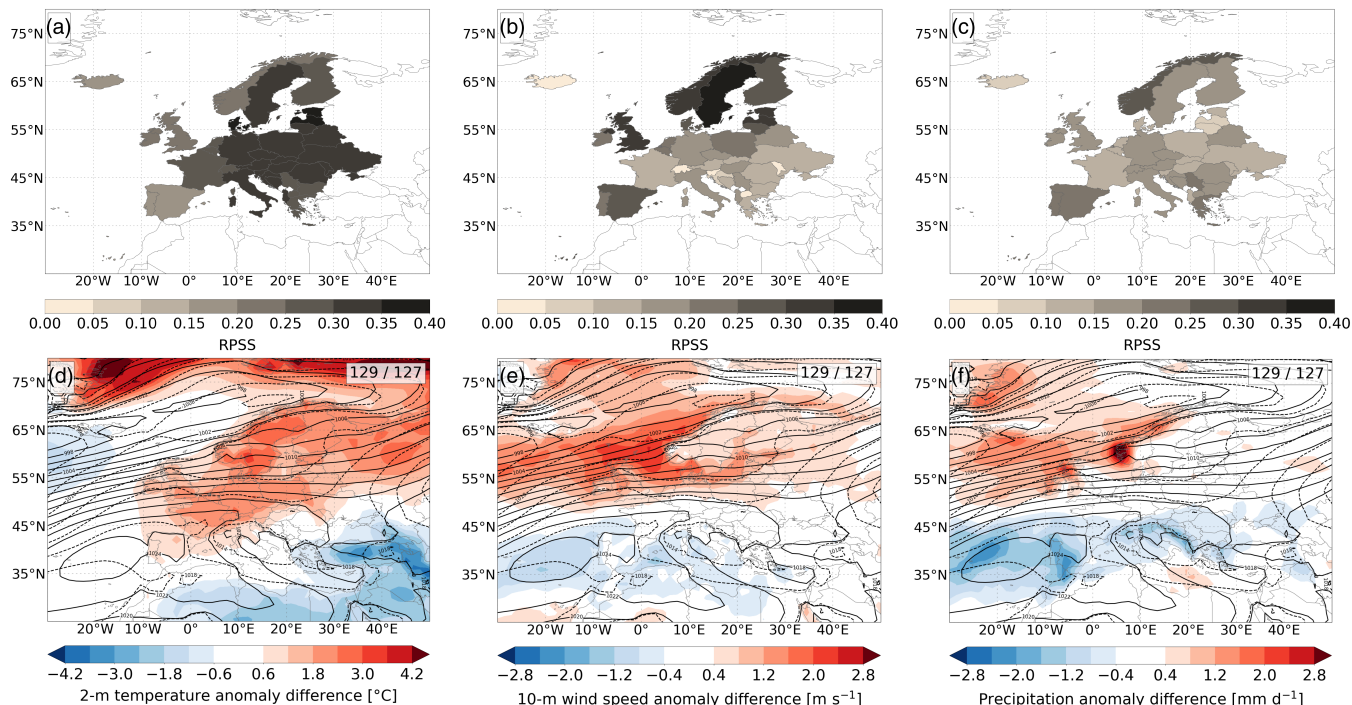


FIGURE 2 (a–c) RPSS of all 1,280 forecasts for country- and month-ahead-averaged (a) $T2M$, (b) $U10M$, and (c) P . Note that the RPSS is only calculated for the countries marked by the respective brownish colour shade. (d–f) Difference (shading) between the mean month-ahead anomalies in ERA-Interim for the dates initialized in the 10% strongest SPV states minus the mean month-ahead anomalies for the dates initialized in the 10% weakest SPV states for, again, (d) $T2M$, (e) $U10M$, and (f) P . Note that the anomalies following the 10% strongest and weakest SPV states are not exactly opposite (cf. Figures 3d,f, 5d,f, and 6d,f) and thus contribute differently to the magnitude of the anomaly differences shown here. The mean sea level pressure for the dates initialized in the 10% strongest (weakest) SPV states is indicated by the solid (dashed) contours. The values in the box in the upper right corners indicate the numbers of forecasts in these two forecast groups

(Section 3.2.1); $U10M$ and P are discussed more briefly afterwards (Section 3.2.2).

3.2 | Month-ahead anomalies and forecast skill after anomalous SPV states

3.2.1 | 2 m temperature

Figure 3 shows the mean modelled month-ahead $T2M$ anomalies in the forecasts initialized in the 2% strongest, 10% strongest, 10% weakest, and 2% weakest SPV states in comparison to the corresponding mean anomalies in ERA-Interim. In addition, robustness and significance of the anomalies are indicated (Section 2.5 gives details). After the 2% strongest SPV states (Figures 3a,b), the model predicts a strong and widely robust positive-NAO-like warm anomaly (of up to 2°C) over large parts of Europe. This is relatively consistent with ERA-Interim in Central, Western, and Southwestern Europe despite differences in the magnitudes. However, there is a strong mismatch in Northern Europe, particularly Scandinavia, where only weak and non-robust anomalies appear in

ERA-Interim. Although most anomalies in the model and in ERA-Interim are robust, they are hardly significant. A comparable anomaly structure with a similarly good agreement between the model and ERA-Interim occurs after the 10% strongest SPV states (Figures 3c,d), but larger parts of the anomalies are significant and the mismatch in Northern Europe is smaller compared to the 2% strongest SPV states. A similarly structured but cold negative-NAO-like anomaly is predicted after the 10% weakest SPV states in good agreement with ERA-Interim (Figures 3e,f). The most pronounced difference compared to the 10% strongest SPV states is that the anomaly tends to extend less into Southwestern Europe. After the 2% weakest SPV states (Figures 3g,h), the model predicts a cold anomaly (of up to -2°C) over a wide range of Europe, which is weaker and much less robust than the warm anomaly after the 2% strongest SPV states. In ERA-Interim, however, the cold anomaly is even weaker than in the model and hardly robust over most of Europe, except for Sweden. In summary, the model thus tends to have a north–south asymmetry in model performance after the most extreme SPV states: it predicts the warm conditions over Central, Western, and

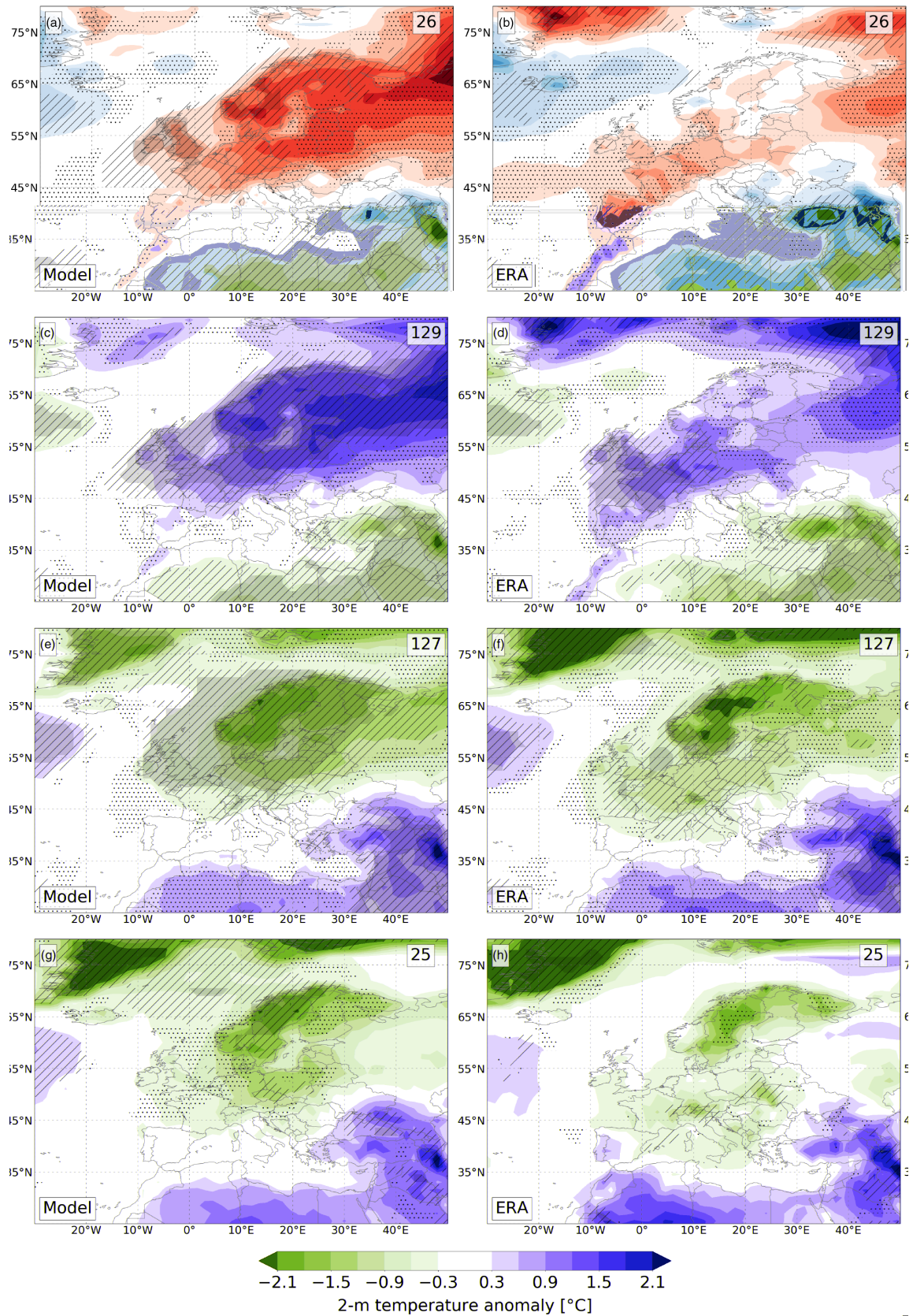


FIGURE 3 (a, c, e, g) Composites of month-ahead $T2M$ anomalies of the forecasts with different SPV states at initialization time: (a) 2% strongest, (c) 10% strongest, (e) 10% weakest, and (g) 2% weakest SPV states. (b, d, f, h) show the corresponding ERA-Interim anomalies. In addition, the robustness is indicated by the point hatching (first, weaker level) and the line hatching (second, stronger level) and the significance is indicated by the black transparent shading (Section 2.5 gives details). The values in the upper-right corners indicate the number of forecasts in the corresponding forecast groups

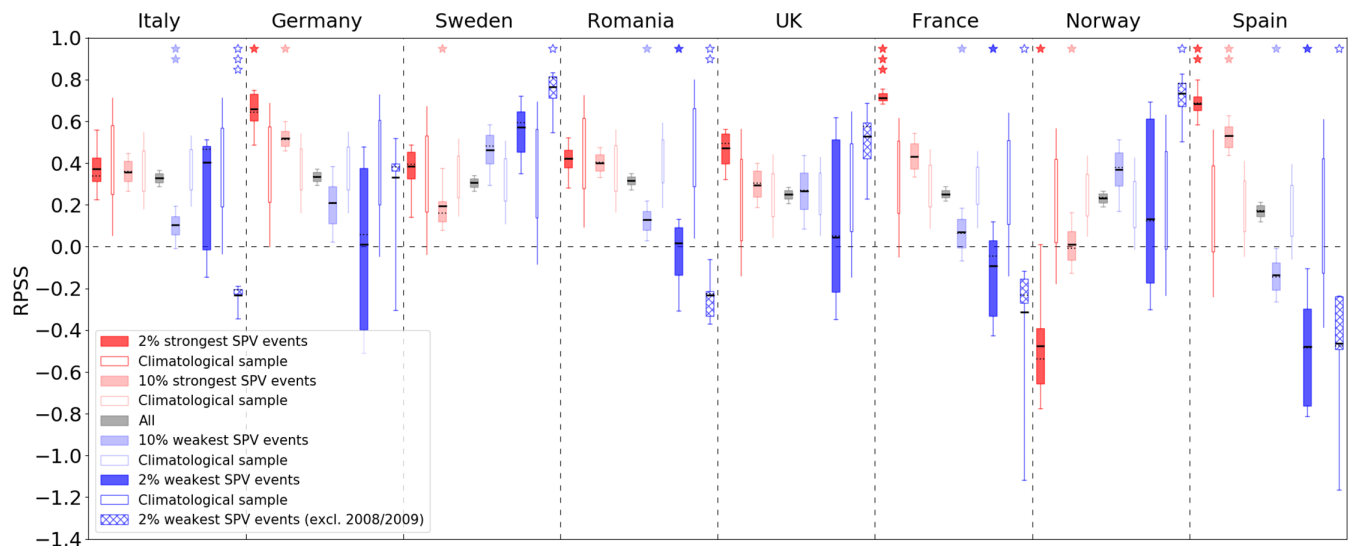


FIGURE 4 RPSS distributions for T_{2M} in selected European countries of all forecasts (filled grey), of the forecasts initialized in the 2% strongest (filled dark red), the 10% strongest (filled light red), the 10% weakest (filled light blue), the 2% weakest SPV states (filled dark blue), and the 2% weakest SPV states if the winter 2008/2009 is excluded from the dataset (hatched dark blue; text gives details). As described in detail in Section 2.5, the distributions of the filled boxes (including the hatched box) are based on the bootstrapped resampled distributions of the corresponding forecast group. The blank narrow boxes show the bootstrapped climatological distributions corresponding to each forecast group. One, two or three stars above a box indicate that the RPSS (filled boxes) is significantly different from its corresponding climatological RPSS (blank boxes) at the 25, 10, or 5% level. The boxes span the interquartile range, the whiskers indicate the 10 and 90%iles, and the horizontal dashed lines show the median. The horizontal solid lines indicate the RPSS calculated over all forecasts of the corresponding bin, independent of the bootstrapped distribution. Figure S1 shows the same analysis but for all European countries

Southern Europe relatively well but tends to be overconfident with respect to the warm anomalies over Scandinavia after the 2% strongest SPV states. In contrast, it predicts the cold anomalies over Scandinavia relatively well but struggles to capture the correct extent of the cold air masses into Western, Central, and Southern Europe after the 2% weakest SPV states. The overall weaker robustness of both the modelled and observed anomalies after the 2% weakest SPV states may further indicate the stronger case-to-case variability in the large-scale response (or in how far the cold air masses extend into Europe) following weak SPV states as found by Beerli and Grams (2019) and Domeisen *et al.* (2020b). After normal SPV states, the model predicts hardly any systematic anomalies, which is consistent with ERA-Interim (not shown). In the following, we demonstrate how the discussed model performance after the different SPV states on the large scale translates into the RPSS for T_{2M} in individual countries.

Figure 4 shows the RPSS for T_{2M} for a selection of countries for all forecasts (grey boxes) and for the forecast groups initialized in the specific SPV states (coloured boxes) which we have already discussed in Figure 3. We show distributions of the RPSS based on the bootstrapped resampled distribution introduced in Section 2.5 (filled boxes) in addition to the actual RPSS for a specific

forecast group (horizontal black solid lines inside the boxes). To test whether the RPSS of a specific forecast group is significantly different from the climatological (i.e., SPV-independent) RPSS, we compare its resampled RPSS distribution (filled boxes) to the corresponding climatological RPSS distribution (blank narrow boxes) which is based on randomly selected forecast groups of equal size (Section 2.5 gives details). Three significance levels are indicated with one (25% level), two (10% level), or three stars (5% level) above the corresponding boxes. For brevity, results are shown for subjectively selected countries from the different European regions (which, however, does not necessarily mean that they are representative in terms of skill; Figure S1 in the Supporting Information shows all countries). Before focusing on the RPSS conditioned on the SPV states, it is worth noting that the RPSS of all forecasts in Figure 4 (grey boxes; as in the RPSS maps in Figure 2) is significantly different on the 10% level between the countries with the highest skill, such as Italy or Germany, and the countries with the lowest skill, such as Spain (because the 10/90% percentiles of the corresponding grey RPSS distributions do not overlap). This result itself has implications for energy meteorology, considering for instance the dependence of the energy demand of Germany and Spain on near-surface temperature during winter (e.g., Bessec and Fouquau, 2008).

Our main objective is to understand to what degree the forecasts initialized in anomalously strong and weak SPV states relate to and possibly influence the strong regional skill variation shown in Figure 2 (and by the grey boxes in Figure 4). The RPSS after the 2% strongest SPV states (filled dark red boxes) for Germany, France, Spain, and the UK is higher compared to climatological conditions (blank dark red boxes; significant for all except the UK). Also the RPSS after the 10% strongest SPV states (light red boxes) is enhanced for some of these countries (significantly for Germany and Spain). For Italy, Sweden, and Romania, the RPSS is unchanged after the 2% and 10% strongest SPV states (with the exception of a significant reduction after the 10% weakest SPV states for Sweden). In contrast, the RPSS is significantly reduced for Norway after both the 2% and 10% strongest SPV states. Note that these modifications in skill are, if at all, significant on very different levels, as indicated by the stars in Figure 4. Nevertheless, there are some outstanding and highly significant modifications of the RPSS such as the increase for France or Spain. The reason for this skill pattern can be understood from the anomaly maps in Figures 3a–d: the countries with a significantly enhanced RPSS (Germany, France, and Spain) are affected by the well-predicted warm anomaly over Europe, which is mostly robust both in the model and in ERA-Interim. This consistency in the robustness implies that the positive mean anomaly is representative for most individual anomalies. Therefore, most of the individual anomalies are consistently assigned to the upper tercile both in the model and in ERA-Interim, which yields a relatively high RPSS. In contrast, the model predicts a robust warm anomaly for Norway and Sweden after the 2% strongest SPV states, which hardly appears in ERA-Interim (particularly for Norway). This implies that the model predicts a positive anomaly (upper tercile) in most cases with a strong SPV, which is in contrast to the observed individual anomalies varying either between negative and positive values (all three terciles) or around the climatological state (middle tercile). The erroneous prediction of the upper tercile in most cases thus leads to the relatively low, in the case of Norway even significantly reduced, RPSS. Italy, as a third example, is largely unaffected by any (robust) anomaly both in the model and in ERA-Interim, which explains its unchanged RPSS compared to climatological conditions. Note that the consistency in robustness between the model and ERA-Interim thus appears as a better measure to understand differences in the RPSS than the consistency in significance. Therefore, we mainly focus on robustness hereafter.

The skill pattern looks substantially different after the 2% weakest SPV states (filled dark blue boxes in Figure 4): only Sweden tends to have an increased RPSS

compared to climatological conditions, although this is not significant. For all the other countries, the RPSS tends to stay unchanged (Italy, UK, and Norway) or to be reduced (Germany, Romania, France, and Spain; significantly for all except Germany). Similar but smaller RPSS changes emerge after the 10% weakest SPV states (light blue boxes) with a few exceptions (a significant decrease for Italy and a non-significant increase for Norway). However, it is important to note that the spread in the RPSS after the 2% weakest SPV states is remarkably large for many countries. This indicates a strong case-to-case variability in the model performance and will be analyzed in more detail subsequently. Figures 3e–h provide an explanation for this skill pattern: after the 2% weakest SPV states (Figures 3g,h), Sweden is the only country where the robust cold anomaly in the model agrees well with ERA-Interim and thus tends to increase the RPSS. For most other countries, the cold anomalies are rather small and only weakly robust (both in the model and in ERA-Interim). This indicates a strong case-to-case variability in the anomalies following the 2% weakest SPV states, which probably explains the non-robust and thus statistically indistinguishable RPSS for some countries (such as Germany) compared to climatological conditions. However, for some Central, Southern European, and Balkan countries such as Romania, France, and Spain, the RPSS is significantly reduced (and in some cases even negative). We hypothesize that this is due to their location at the transition zone from the colder conditions over Northern Europe to the warmer conditions over North Africa and Western Asia. Whenever the model has a subtle error in the prediction of this transition zone (which is at least the case on average as indicated by Figures 3e–h), it can lead to the wrong anomaly (middle or even opposite tercile) and thus a reduced RPSS for these countries. A similar hypothesis has been suggested by Domeisen *et al.* (2020a), who also found an average mismatch in this transition zone for $T2M$ following weak SPV states between different models and reanalysis. After the 10% weakest SPV states (Figures 3e,f), the model response agrees much better with the observed response, particularly over Northern Europe. Therefore, the RPSS for Scandinavian countries (Sweden and Norway) tends to increase. Nevertheless, there are still mismatches between the model and ERA-Interim with respect to the extension of the cold anomalies into Central and Southern Europe. This explains the significant reduction in the RPSS for Italy, Romania, France, and Spain even after the 10% weakest SPV states and gives further indications that the conditions in the transition zone between cold and warm anomalies are challenging to predict. Figure S1 (same as Figure 4 but for all countries) shows a similar reduction in the RPSS for most Southern European and Balkan countries and thus corroborates the systematic problem in the

model in predicting the $T2M$ response in these regions after weak SPV states.

As discussed in section 2.5, our bootstrapping approach of resampling subsets of winters (i.e., distinct periods of anomalous SPV states) accounts for the small sample size of the forecast groups initialized in the 2% most extreme SPV states. Nevertheless, the resulting RPSS distributions (Figure 4) are biased toward few periods that are represented by many individual forecasts (this is because the RPSS of every resample including this period is dominated by the forecasts of this period). In our dataset, the most obvious period of this kind is the persistent phase of an extremely weak SPV in February 2009, which accounts for around 50% of all the 25 forecasts initialized in the 2% weakest SPV states. To test the sensitivity of our skill analysis to this period, we have recalculated the RPSS distribution after the 2% weakest SPV states after removing the whole winter 2008/2009 from the dataset (hatched blue boxes in Figure 4). Indeed, the RPSS spread becomes much smaller and reveals a clearer skill pattern for many countries (compared to the 2% weakest SPV states including the winter 2008/2009, i.e., the filled blue boxes): the RPSS improves for Norway (significantly) and the UK, which is consistent with Sweden and thus corroborates the argument of enhanced skill for Northern Europe after weak SPV states. On the other hand, the RPSS is significantly reduced and negative for Italy, which aligns with the still poor model performance for other Central, Southern European, and Balkan countries (such as Romania, France, and Spain). Therefore, the weak SPV event in winter 2008/2009 indeed influences the skill pattern in Figure 4. Nevertheless, removing this outlier demonstrates that the north–south contrast in skill for $T2M$ following the weakest SPV states is a robust finding. The reason for this is that, even without the dominant winter 2008/2009, the anomaly composites (corresponding to Figures 3e,f, but not shown here) indicate a model bias in the extent of the cold air masses into Central and Southern Europe compared to ERA-Interim.

Aside from determining whether the skill after a specific extreme SPV state is significantly different from climatological conditions, Figure 4 further allows detecting significant differences in skill between strong and weak SPV states directly. For instance, the (enhanced) RPSS for France and Spain after the strongest SPV states is significantly higher than their (reduced) RPSS after the weakest SPV states (as the 10 and 90% percentiles of the equally sized dark red and dark blue RPSS distributions in Figure 4 do not overlap). The same, but *vice versa*, applies to Norway if the winter 2008/2009 is removed. These country-specific findings can be an important guidance for end-users of sub-seasonal weather forecasts such as energy meteorologists.

TABLE 1 Pearson correlation coefficients (r) between the month-ahead $T2M$ anomalies of the ensemble mean and of ERA-Interim for the countries and specific forecast groups shown in Figure 4

Country	All	2% strongest SPV states	2% weakest SPV states
Italy	0.71	0.82	0.54
Germany	0.69	0.87	0.71
Sweden	0.68	0.69	0.81
Romania	0.71	0.69	0.41
UK	0.68	0.77	0.72
France	0.65	0.87	0.31
Norway	0.66	0.66	0.77
Spain	0.61	0.86	0.10

To test the sensitivity of our results for $T2M$ to the used skill score, we have performed a similar analysis as summarized in Figure 4 but by determining correlation instead of the RPSS. More specifically, we have computed the Pearson correlation coefficient between the month-ahead $T2M$ anomalies of the ensemble mean and of ERA-Interim for the same countries and forecast groups shown in Figure 4. The results, summarized in Table 1, confirm the results from the RPSS analysis for many countries at least in a qualitative sense: countries with a significantly enhanced or reduced RPSS after extreme SPV states compared to climatological conditions have a correspondingly enhanced or reduced correlation coefficient after the same extreme SPV states. For instance, the correlation coefficients for France are 0.87 after the 2% strongest SPV states, 0.31 after the 2% weakest SPV states, and 0.65 for all forecasts, which confirms the asymmetric skill pattern shown in Figure 4. Similarly, Sweden has a correlation coefficient of 0.69 after the 2% strongest SPV states, 0.81 after the 2% weakest SPV states, and 0.68 for all forecasts, which is rather inverse to the skill pattern for France, as shown in Figure 4. To interpret these results in a statistically sound way, the correlation coefficients would have to be calculated for the same bootstrapped distributions as in Figure 4 and not only for the ensemble mean. Nevertheless, this simple correlation analysis supports our interpretation of at least the qualitative pattern of country-specific skill modification for $T2M$ after extreme SPV states based on the thorough analysis using the RPSS.

3.2.2 | 10 m wind speed and precipitation

Figures 5a–d demonstrate remarkably well that, on the large scale, the model predicts the positive-NAO-like

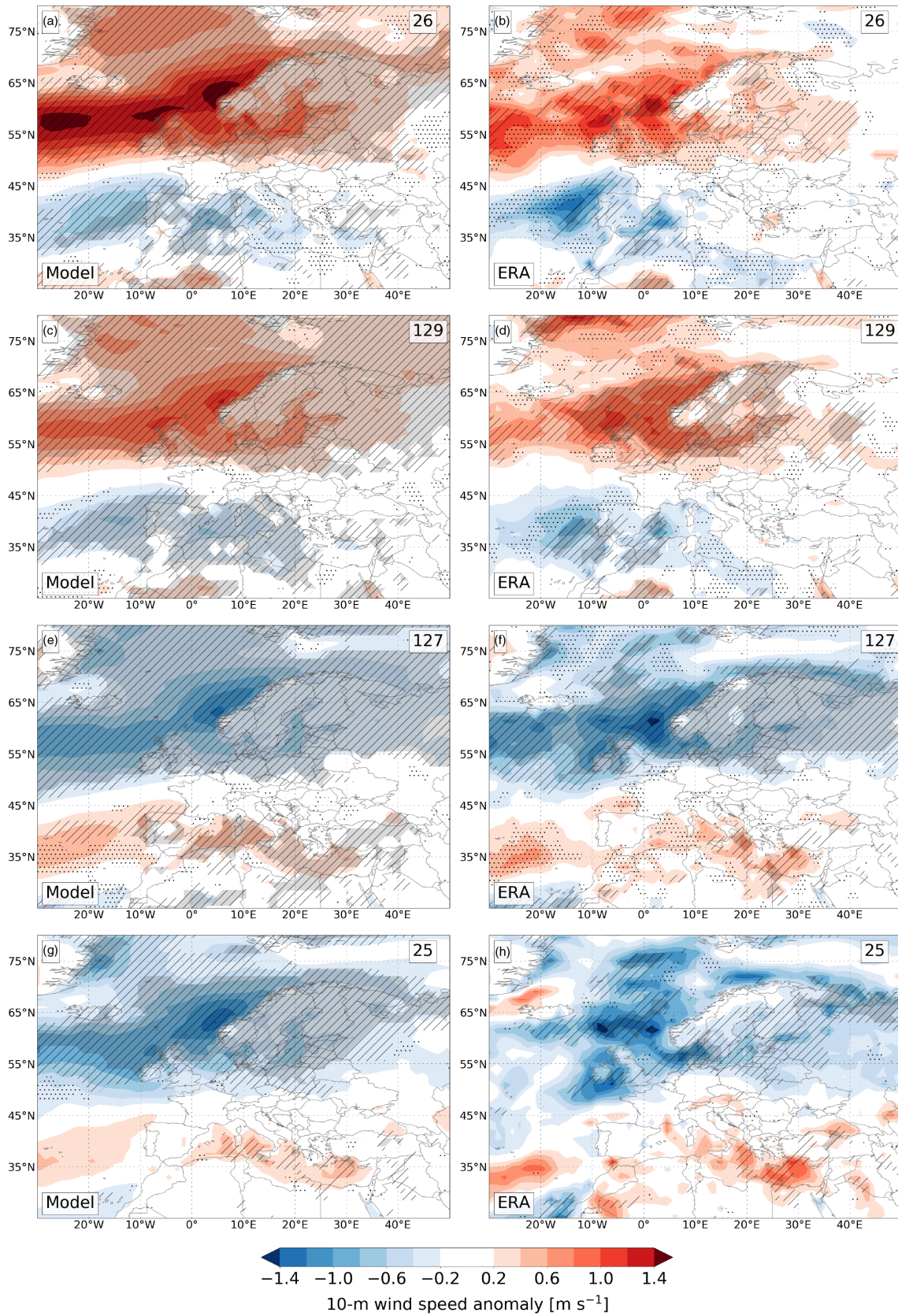


FIGURE 5 As Figure 3, but for U_{10M}

anomaly dipole structure with stronger than normal $U10M$ over Northern Europe and weaker than normal $U10M$ over Mediterranean Europe after the 2 and 10% strongest SPV states. The same applies to the inverse, negative-NAO-like anomaly dipole after the 2 and 10% weakest SPV states (Figures 5e–h). Most of the anomalies are consistently robust both in the model and in ERA-Interim and they reach values of more than $\pm 1.4 \text{ m}\cdot\text{s}^{-1}$. The most pronounced mismatches occur on the regional scale: after the 2% and 10% strongest SPV states (Figures 5a–d), the modelled anomalies extend slightly too far to the east. After the 2% weakest SPV states, the transition zone between negative and positive anomalies tends to be slightly displaced in the model compared to ERA-Interim. Furthermore, the positive anomalies in Mediterranean Europe are rather weak, patchy, and hardly robust, both in the model and in ERA-Interim. The large-scale structure and robustness of the P anomalies, shown in Figure 6, strongly correlate with the one for $U10M$ (Figure 5) because of the common dependency of the two parameters on cyclone activity. The most important difference is that the maximum P anomalies, reaching values of more than $\pm 1.4 \text{ mm}\cdot\text{day}^{-1}$, are concentrated along the coast of Norway (mainly the mountainous part) and the western coasts of the Mediterranean, which are primarily affected by the NAO-related meridional excursions of the storm track. Consistent with the mismatches found for $U10M$ (Figure 5), the model tends to predict the extent of the negative P anomaly in the Mediterranean after the 2% and 10% strongest SPV states too far to the east into the Balkans. Likewise, there is a meridional shift in the position of the rather patchy anomaly dipole after the (particularly 2%) weakest SPV states in the model compared to ERA-Interim.

Figure 7 shows that the RPSS of all forecasts (grey boxes) for $U10M$ (Figure 7a) and P (Figure 7b) tends to be lower than for $T2M$ (Figure 4) on average. This indicates an overall lower predictability for $U10M$ and P compared to $T2M$ at least in the model. Considering the anomalous SPV states, Figure 7a shows that the skill pattern for $U10M$ is qualitatively similar to the skill pattern for $T2M$ (Figure 4) in the sense that most displayed countries experience either an enhanced or unchanged RPSS after strong SPV states (although on a lower significance level on average). However, the key difference is that Southern European and Balkan countries such as Italy, Spain, or Romania do not experience a reduced but rather unchanged or even enhanced RPSS also after weak SPV states (particularly if the winter 2008/2009 is removed). This is because the $U10M$ anomaly dipole (Figures 5e–h) is shifted more poleward compared to $T2M$ (Figures 3e–h) and thus affects Southern Europe and the Balkans more directly. However, more Central European countries such as Germany or France still tend to experience reduced skill

after weak SPV states, which results from their location in the transition zone between the positive and negative anomalies (as discussed above for $T2M$). The slight positional error of the predicted anomaly dipole and thus this transition zone in the meridional direction after the 2% weakest SPV states (Figures 5g,h) gives evidence for this. In contrast to $U10M$, for P the symmetric skill pattern with a (in some cases significantly) enhanced RPSS both after strong and weak SPV states is limited to Italy, Norway, and Spain (Figure 7b; again, particularly if the winter 2008/2009 is removed). This is because the considerable P anomalies mainly occur in these regions, as discussed above (Figure 6). For most other countries, the RPSS after the 2% weakest SPV states is either unchanged or tends to be reduced, which is again due to the positional error of the predicted anomaly dipole (for instance over Germany and France; Figures 6g,h). An exception is Romania with an inverse asymmetric skill pattern: the skill tends to be reduced (non-significantly) after the 2% strongest SPV states due to the overconfident dry anomaly of the model over the Balkans (Figures 6a,b), but enhanced (significantly if the winter 2008/2009 is removed) after the 2% weakest SPV states due to the relatively well-predicted and robust wet anomaly concentrated over the Balkans and the Black Sea region (Figures 6g,h).

4 | DISCUSSION AND CONCLUSIONS

We have demonstrated that the ECMWF sub-seasonal reforecast skill (three-category RPSS) for predicting energy-industry-relevant, country- and month-ahead-averaged surface weather anomalies substantially varies among different European regions: for $T2M$, the skill is generally higher for Central, Eastern, and Northeastern Europe compared to Southwestern Europe, whereas for $U10M$ and P , the skill is higher for Northern and Southern Europe compared to Central Europe. The strong regional variation in forecast skill primarily reflects the anomaly patterns associated with the NAO, which dominates European surface weather variability on sub-seasonal time-scales. Beside other climate modes (Section 1), anomalous SPV states can lead to persistent NAO-like surface weather anomalies and thus considerably influence skill. In this study, we have demonstrated that the changes in skill of the forecasts initialized during anomalous SPV states are not always positive but can be negative when focussing on a country level: for $T2M$, many Western, Central, Southern European and Balkan countries tend to have enhanced skill after the strongest SPV states (compared to SPV-independent, climatological conditions), but reduced and in some cases even negative skill after the

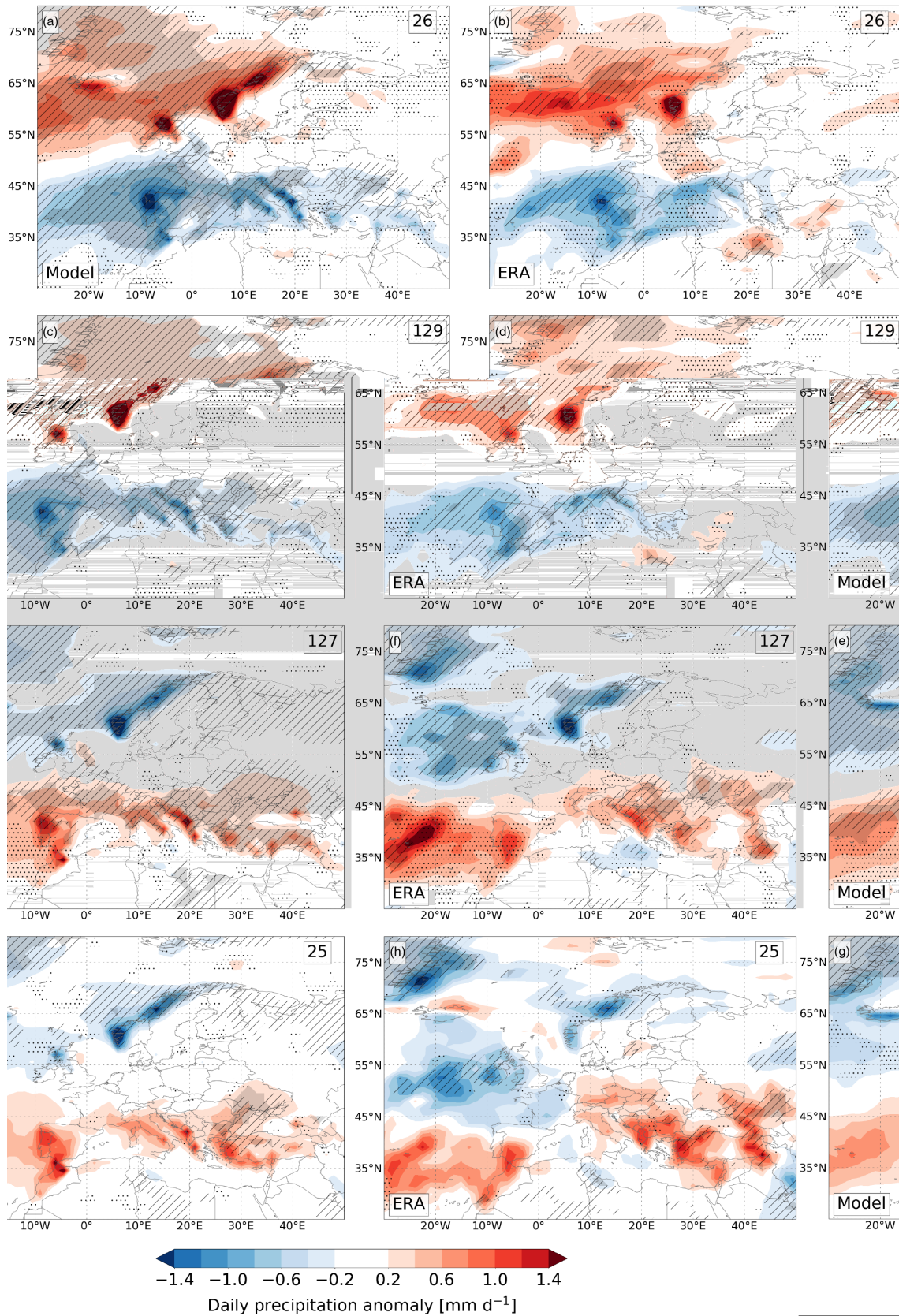


FIGURE 6 As Figure 3, but for P



FIGURE 7 As Figure 4, but for (a) U_{10M} and (b) P . Figures S2 and S3 show the same analysis but for all European countries

weakest SPV states. The asymmetry tends to be *vice versa* for Scandinavian countries, with enhanced skill after the weakest but reduced skill after the strongest SPV states. This can be explained by a north–south asymmetry in model performance after the most extreme SPV states on average: the model predicts the positive-NAO-like warm anomalies in Western, Central, and Southern Europe relatively well but tends to be overconfident with respect to the warm conditions in Northern Europe, particularly Scandinavia, after the strongest SPV states. In contrast, it predicts the negative-NAO-like cold anomalies in Northern Europe centred over Scandinavia considerably well but struggles to correctly capture the strongly case-dependent extent of the anomalously cold air masses into Central and Southern Europe. It is important to note that a substantial case-to-case variability occurs in the mean anomaly patterns following extreme SPV states, which

can critically affect the robustness of such an analysis. However, by applying a thorough significance test, we can show that this SPV-dependent north–south asymmetry in skill for T_{2M} is remarkably robust for various countries such as France, Spain, or Norway (Figure S1 gives an overview of all countries). This is a key finding of our study and suggests a flow-dependent T_{2M} model bias in the corresponding regions following extreme SPV states. Furthermore, it aligns with the enhanced (reduced) skill for T_{2M} in Central to Southern Europe after strong (weak) SPV states found by Domeisen *et al.* (2020a, their figures 5 and 6), and demonstrates how their findings translate into a country-scale perspective. In contrast to T_{2M} , the model skill for U_{10M} and P tends to increase for both Northern and Southern European countries after both strong and weak SPV states (for P , this increase is mainly limited to Norway and the Iberian Peninsula). This results from

the NAO-like north–south dipole in the anomaly pattern, which is predicted relatively well at least on the larger scale. Nevertheless, the skill still tends to be reduced or at least unchanged for certain Central European countries such as France following extreme SPV states. We hypothesize that this is because Central Europe is located at the transition zone between the NAO-induced positive and negative $U10M$ and P anomalies, which is slightly displaced in the model in the meridional direction and thus yields erroneous anomalies on average. Our analysis thus provides indications that a good forecast skill for the NAO index (and thus the large-scale flow pattern) does not necessarily imply a good forecast skill for surface weather in those European regions that are located at the edge of the NAO-related surface weather anomalies and thus are particularly sensitive to a correct forecast of their position. The differences in skill between Northern and Central Europe for $U10M$ are similar to the findings of Beerli *et al.* (2017) for wind power generation (their figure 8a). However, their purely statistical approach does not have skill for Southern European countries after weak SPV states, which is in contrast to our NWP-based findings and thus demonstrates the added value of using a dynamical model for this region.

When investigating forecast skill on sub-seasonal time-scales, it is important to keep in mind that the first two weeks typically contribute most to the skill (e.g., Weigel *et al.*, 2008; Vitart, 2014). To test this hypothesis, we have performed the same RPSS analysis as summarized in Figure 4 but for the $T2M$ anomalies averaged separately for weeks 1 to 4 after forecast initial time (Figures S4 and S5). Indeed, the RPSS decreases roughly exponentially from week 1 to 4, independent of the SPV state at forecast initial time. In most cases, there is no or even negative skill left for week 4. The most striking exception is the RPSS for weeks 3 and 4 after the 2% and 10% strongest SPV states, which is still considerably high for many Central to Southern European countries and thus substantially contributes to their high RPSS for the month-ahead anomalies. The remarkable skill for $T2M$ for weeks 3 and 4 is useful for the energy industry, which also relies on weather forecasts on weekly time-scales for their planning and decision-making.

Despite remarkable correlations between the initial SPV state and surface weather forecast skill, our results do not tackle the question of causality. Various studies have shown that the large-scale tropospheric flow regime (and thus the surface weather response) following anomalous SPV states does not only depend on the stratosphere itself but also on the tropospheric state before and during the stratospheric event (e.g., Ambaum and Hoskins, 2002; Karpechko *et al.*, 2017). For instance, it is possible that the enhanced skill for $T2M$ for Western, Central, and Southern Europe following strong SPV states, as found

in this study, partly results from a relatively predictable positive-NAO-like flow regime that is already active before and during the strong SPV states. This would be in line with the SPV–NAO coupling found by Ambaum and Hoskins (2002) and indicate that the strong SPV is rather an integrator than a trigger of the positive phase of the NAO. Similarly, the tropospheric conditions before and during the weak SPV states investigated here might be co-responsible for how well the model predicts the extent of the cold anomalies, centred over Scandinavia, into Central and Southern Europe (cf., e.g., Domeisen *et al.*, 2020b). Although this is important from a dynamical perspective, it does not diminish the usefulness of the SPV state as an indicator for modified sub-seasonal forecast skill.

The strong regional contrasts in the modification of country-scale model skill after extreme SPV states is a key finding of this study. Two main implications result from this: on the one hand, end-users of sub-seasonal forecasts such as energy meteorologists or national weather services have to be aware of this regional skill modification when making decisions based on country-scale surface weather predictions after extreme SPV states. In particular, the reduced skill for $T2M$ for many Central, Southern European, and Balkan countries following weak SPV states might be of critical importance for the energy industry, which requires correct forecasts of European cold air outbreaks potentially following weak SPV states. On the other hand, our results for $T2M$ reveal two problems in the model: predicting the average temperature conditions in Scandinavia following strong SPV states and in Central to Southern Europe following weak SPV states. The sources of these problems can be manifold. For instance, the model can have errors in the representation of the stratosphere–troposphere coupling, of the large-scale weather regime response following extreme SPV states, or of the regional surface weather imprint of these large-scale weather regimes. We believe that a detailed analysis of the large-scale weather regime response following extreme SPV states would be an important starting point to better understand these regional model biases. The first problem (reduced skill for $T2M$ in some Scandinavian countries following strong SPV states) could be associated with the misrepresentation of variations in the positive-NAO-like weather regime response, which determine if and how far the anomalously warm air extends into Scandinavia. Using a higher number of Atlantic-European weather regimes than just the bimodal NAO, Beerli and Grams (2019) showed that the positive phase of the NAO indeed manifests itself mainly in the zonal regime with a warm anomaly centred over Scandinavia, or the Scandinavian trough regime with a warm anomaly shifted more to the east over Western Russia. Regarding the second problem (reduced skill for $T2M$ in

many Central to Southern European countries following weak SPV states), it is worthwhile relating our results to the widely studied sudden stratospheric warmings (SSWs) as particularly extreme cases of weak SPV states. Comparing the forecasts after the 2% weakest SPV states to the SSW compendium by Butler *et al.* (2017) reveals that all except one forecast in this category were initialized ten or more days after (and, in one case during) an SSW. This indicates that the erroneous extent of the cold anomaly into Central to Southern Europe, which results in reduced skill for many countries, are at least an indirect consequence of SSWs. Therefore, we require a better understanding of the dynamics of weak SPV states, in particular SSWs, their subsequent surface weather response, and how these two aspects are represented in the model. Beerli and Grams (2019) and Domeisen *et al.* (2020b) showed that weak SPV states or SSWs, respectively, tend to be followed by either the mostly mild and windy Atlantic trough regime or the cold and calm Greenland blocking regime which corresponds to the negative phase of the NAO. Whether the model has a systematic bias in (one of) these regime responses following both strong and weak SPV states is a subject of our future work.

ACKNOWLEDGEMENTS

We thank the two anonymous reviewers for critically assessing and thus helping to improve this publication. We further acknowledge Lukas Papritz for inputs on the statistical testing and the colleagues from the SPREAD-OUT group at KIT for input throughout the project. The operators of the WWRP/WCRP Subseasonal-to-Seasonal Prediction Project database are acknowledged for providing access to the model data and ECMWF for providing access to the ERA-Interim reanalysis data. Axpo Solutions AG is acknowledged for supporting the contributions of DB and RB and the Helmholtz Young Investigator Group grant VH-NG-1243 for supporting the contributions of DB and CMG.

ORCID

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

Remo Beerli  <https://orcid.org/0000-0002-0897-9650>

Heini Wernli  <https://orcid.org/0000-0001-9674-4837>

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

REFERENCES

- Albers, J.R. and Birner, T. (2014) Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, 71, 4028–4054. <https://doi.org/10.1175/JAS-D-14-0026.1>
- 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)
- de Andrade, F.M., Coelho, C.A.S. and Cavalcanti, I.F.A. (2019) Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dynamics*, 52, 5451–5475. <https://doi.org/10.1007/s00382-018-4457-z>
- Andrews, M.B., Knight, J.R., Scaife, A.A., Lu, Y., Wu, T., Gray, L.J. and Schenzinger, V. (2019) Observed and simulated teleconnections between the stratospheric quasi-biennial oscillation and Northern Hemisphere winter atmospheric circulation. *Journal of Geophysical Research: Atmospheres*, 124, 1219–1232. <https://doi.org/10.1029/2018JD029368>
- 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>
- Ardilouze, C., Batté, L., Bunzel, F., Decremier, D., Déqué, M., Doblas-Reyes, F.J., Douville, H., Fereday, D., Guemas, V., MacLachlan, C., Müller, W. and Prodhomme, C. (2017) Multi-model assessment of the impact of soil moisture initialization on mid-latitude summer predictability. *Climate Dynamics*, 49, 3959–3974. <https://doi.org/10.1007/s00382-017-3555-7>
- Attard, H.E., Rios-Berrios, R., Guastini, C.T. and Lang, A.L. (2016) Tropospheric and stratospheric precursors to the January 2013 sudden stratospheric warming. *Monthly Weather Review*, 144, 1321–1339. <https://doi.org/10.1175/MWR-D-15-0175.1>
- Baldwin, M.P. and Dunkerton, T.J. (1999) Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *Journal of Geophysical Research*, 104, 30937–30946. <https://doi.org/10.1029/1999JD900445>
- 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>
- Bessec, M. and Fouquau, J. (2008) The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach. *Energy Economics*, 30, 2705–2721. <https://doi.org/10.1016/j.eneco.2008.02.003>
- Bloomfield, H.C., Brayshaw, D.J., Shaffrey, L.C., Coker, P.J. and Thornton, H.E. (2018) The changing sensitivity of power systems to meteorological drivers: a case study of Great Britain. *Environmental Research Letters*, 13, 054028. <https://doi.org/10.1088/1748-9326/aabff9>
- Brayshaw, D.J., Troccoli, A., Fordham, R. and Methven, J. (2011) The impact of large scale atmospheric circulation patterns on wind power generation and its potential predictability: a case study

- over the UK. *Renewable Energy*, 36, 2087–2096. <https://doi.org/10.1016/j.renene.2011.01.025>
- 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., Charlton-Perez, A., Domeisen, D.I.V., Garfinkel, C., Gerber, E.P., Hitchcock, P., Karpechko, A.Y., Maycock, A.C., Sigmund, M., Simpson, I. and Son, S.-W. (2019). Sub-seasonal predictability and the stratosphere, Chapter 11 in *Sub-Seasonal to Seasonal Prediction*. Elsevier, Amsterdam, Netherlands.
- 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>
- Butler, A.H., Sjöberg, J.P., Seidel, D.J. and Rosenlof, K.H. (2017) A sudden stratospheric warming compendium. *Earth System Science Data*, 9, 63–76. <https://doi.org/10.5194/essd-9-63-2017>
- 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, A.J., O'Neill, A., Lahoz, W.A. and Massacand, A.C. (2004) Sensitivity of tropospheric forecasts to stratospheric initial conditions. *Quarterly Journal of the Royal Meteorological Society*, 130, 1771–1792. <https://doi.org/10.1256/qj.03.167>
- Charlton, A.J., O'Neill, A., Stephenson, D.B., Lahoz, W.A. and Baldwin, M.P. (2003) Can knowledge of the state of the stratosphere be used to improve statistical forecasts of the troposphere?. *Quarterly Journal of the Royal Meteorological Society*, 129, 3205–3224. <https://doi.org/10.1256/qj.02.232>
- 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>
- Christiansen, B. (2005) Downward propagation and statistical forecast of the near-surface weather. *Journal of Geophysical Research: Atmospheres*, 110(D14104). <https://doi.org/10.1029/2004JD005431>
- Clark, R.T., Bett, P.E., Thornton, H.E. and Scaife, A.A. (2017) Skilful seasonal predictions for the European energy industry. *Environmental Research Letters*, 12, 024002. <https://doi.org/10.1088/1748-9326/aa57ab>
- 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>
- Deflorio, M.J., Waliser, D.E., Guan, B., Lavers, D.A., Martin Ralph, F. and Vitart, F. (2018) Global assessment of atmospheric river prediction skill. *Journal of Hydrometeorology*, 19, 409–426. <https://doi.org/10.1175/JHM-D-17-0135.1>
- 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 subseasonal to seasonal prediction: 2. Predictability arising from stratosphere–troposphere coupling. *Journal of Geophysical Research: Atmospheres*, 125. <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 and Climate Dynamics*. <https://doi.org/10.5194/wcd-2019-16>
- Dorrington, J., Finney, I., Palmer, T.N. and Weisheimer, A. (2020) Beyond skill scores: exploring sub-seasonal forecast value through a case-study of French month-ahead energy prediction. *Quarterly Journal of the Royal Meteorological Society*. <https://doi.org/10.1002/qj.3863>
- 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>
- 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>
- Holton, J.R. and Mass, C. (1976) Stratospheric vacillation cycles. *Journal of the Atmospheric Sciences*, 33, 2218–2225. [https://doi.org/10.1175/1520-0469\(1976\)033%3C2218:SVC%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033%3C2218:SVC%3E2.0.CO;2)
- Holton, J.R. and Tan, H.-C. (1980) The influence of the equatorial Quasi-Biennial Oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sciences*, 37, 2200–2208. [https://doi.org/10.1175/1520-0469\(1980\)037%3C2200:tioteq%3E2.0.co;2](https://doi.org/10.1175/1520-0469(1980)037%3C2200:tioteq%3E2.0.co;2)
- Huang, J., Higuchi, K. and Shabbar, A. (1998) The relationship between the North Atlantic Oscillation and El Niño–Southern Oscillation. *Geophysical Research Letters*, 25, 2707–2710. <https://doi.org/10.1029/98GL01936>
- Hurrell, J.W. (1996) Influence of variations in extratropical winter-time teleconnections on Northern Hemisphere temperature. *Geophysical Research Letters*, 23, 665–668. <https://doi.org/10.1029/96GL00459>
- Jucker, M. (2016) Are sudden stratospheric warmings generic? Insights from an idealized GCM. *Journal of the Atmospheric Sciences*, 73, 5061–5080. <https://doi.org/10.1175/JAS-D-15-0353.1>
- 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>
- Karpechko, A.Y., Hitchcock, P., Peters, D.H.W. and Schneider, A. (2017) Predictability of downward propagation of major sudden stratospheric warmings. *Quarterly Journal of the Royal Meteorological Society*, 143, 1459–1470. <https://doi.org/10.1002/qj.3017>
- Kautz, L.-A., Polichtchouk, I., Birner, T., Garny, H. and Pinto, J.G. (2020) Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, 146, 1040–1055. <https://doi.org/10.1002/qj.3724>
- Kawatani, Y., Hamilton, K., Gray, L.J., Osprey, S.M., Watanabe, S. and Yamashita, Y. (2019) The effects of a well-resolved stratosphere on the simulated boreal winter circulation in a climate model. *Journal of the Atmospheric Sciences*, 76, 1203–1226. <https://doi.org/10.1175/JAS-D-18-0206.1>

- Kidston, J., Scaife, A.A., Hardiman, S.C., Mitchell, D.M., Butchart, N., Baldwin, M.P. and Gray, L.J. (2015) Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, 8, 433–440. <https://doi.org/10.1038/NCEO2424>
- Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., Paolino, D.A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M.P., Becker, E., Peng, P., Tripp, P., Huang, J., Dewitt, D.G., Tippett, M.K., Barnston, A.G., Li, S., Rosati, A., Schubert, S.D., Rienecker, M., Suarez, M., Li, Z.E., Marshak, J., Lim, Y.K., Tribbia, J., Pegion, K., Merryfield, W.J., Denis, B. and Wood, E.F. (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95, 585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1>
- Kodera, K., Mukougawa, H., Maury, P., Ueda, M. and Claud, C. (2016) Absorbing and reflecting sudden stratospheric warming events and their relationship with tropospheric circulation. *Journal of Geophysical Research: Atmospheres*, 121, 80–94. <https://doi.org/10.1002/2015JD023359>
- Koster, R.D., Mahanama, S.P., Yamada, T.J., Balsamo, G., Berg, A.A., Boissarie, 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 subseasonal forecast skill: first results from a multi-model experiment. *Geophysical Research Letters*, 37(L02402). <https://doi.org/10.1029/2009GL041677>
- Kretschmer, M., Cohen, J., Matthias, V., Runge, J. and Coumou, D. (2018) The different stratospheric influence on cold-extremes in Eurasia and North America. *npj Climate and Atmospheric Science*, 1, 44. <https://doi.org/10.1038/s41612-018-0054-4>
- 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>
- Manney, G.L., Hegglin, M.I., Lawrence, Z.D., Wargan, K., Millán, L.F., Schwartz, M.J., Santee, M.L., Lambert, A., Pawson, S., Knosp, B.W., Fuller, R.A. and Daffer, W.H. (2017) Reanalysis comparisons of upper tropospheric–lower stratospheric jets and multiple tropopause. *Atmospheric Chemistry and Physics*, 17, 11541–11566. <https://doi.org/10.5194/acp-17-11541-2017>
- Matsuno, T. (1971) A dynamical model of the stratospheric sudden warming. *Journal of the Atmospheric Sciences*, 28, 1479–1494. [https://doi.org/10.1175/1520-0469\(1971\)028%3C1479:admots%3E2.0.co;2](https://doi.org/10.1175/1520-0469(1971)028%3C1479:admots%3E2.0.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>
- Mukougawa, H., Hirooka, T. and Kuroda, Y. (2009) Influence of stratospheric circulation on the predictability of the tropospheric Northern Annular Mode. *Geophysical Research Letters*, 36, L08814. <https://doi.org/10.1029/2008GL037127>
- Palmer, T.N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délecluse, P., Déqué, M., Diez, E., Doblas-Reyes, F.J., Feddersen, H., Graham, R., Gualdi, S., Guérémy, J.-F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A.P., Orfila, B., Rogel, P., Terres, J.-M. and Thomson, M.C. (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, 86, 853–872. <https://doi.org/10.1175/BAMS-85-6-853>
- Papritz, L. (2020) Arctic lower tropospheric warm and cold extremes: horizontal and vertical transport, diabatic processes, and linkage to synoptic circulation features. *Journal of Climate*, 33, 993–1016. <https://doi.org/10.1175/JCLI-D-19-0638.1>
- 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>
- Scaife, A.A., Comer, R.E., Dunstone, N.J., Knight, J.R., Smith, D.M., MacLachlan, C., Martin, N., Peterson, K.A., Rowlands, D., Carroll, E.B., Belcher, S. and Slingo, J. (2017) Tropical rainfall, Rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, 143, 1–11. <https://doi.org/10.1002/qj.2910>
- 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 in der U.S. Zone*, 38, 51–63
- Schiraldi, N.J. and Roundy, P.E. (2017) Seasonal-to-subseasonal model forecast performance during agricultural drought transition periods in the U.S. Corn Belt. *Monthly Weather Review*, 145, 3687–3708. <https://doi.org/10.1175/MWR-D-17-0026.1>
- Schneidereit, A., Peters, D.H.W., Grams, C.M., Quinting, J.F., Keller, J.H., Wolf, G., Teubler, F., Riemer, M. and Martius, O. (2017) Enhanced tropospheric wave forcing of two anticyclones in the pre-phase of January 2009 major stratospheric sudden warming event. *Monthly Weather Review*, 145, 1797–1815. <https://doi.org/10.1175/MWR-D-16-0242.1>
- Scott, R.K. and Haynes, P.H. (2000) Internal vacillations in stratosphere-only models. *Journal of the Atmospheric Sciences*, 57, 3233–3250. [https://doi.org/10.1175/1520-0469\(2000\)057%3C3233:IVISOM%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057%3C3233:IVISOM%3E2.0.CO;2)
- Sigmond, M., Scinocca, J.F., Kharin, V.V. and Shepherd, T.G. (2013) Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, 6, 98–102. <https://doi.org/10.1038/ngeo1698>
- 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>
- Thornton, H.E., Scaife, A.A., Hoskins, B.J., Brayshaw, D.J., Smith, D.M., Dunstone, N., Stringer, N. and Bett, P.E. (2019) Skilful seasonal prediction of winter gas demand. *Environmental Research Letters*, 14(024009). <https://doi.org/10.1088/1748-9326/aaf338>
- 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>
- Tripathi, O.P., Baldwin, M., Charlton-Perez, A., Charron, M., Eckermann, S.D., Gerber, E., Harrison, R.G., Jackson, D.R., Kim, B.-M., Kuroda, Y., Lang, A., Mahmood, S., Mizuta, R., Roff, G., Sigmond, M. and Son, S.-W. (2015a) The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quarterly Journal of the*

- Royal Meteorological Society*, 141, 987–1003. <https://doi.org/10.1002/qj.2432>
- Tripathi, O.P., Charlton-Perez, A., Sigmond, M. and Vitart, F. (2015b) 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>
- Troccoli, A., Dubus, L. and Haupt, S.E. (2014) *Weather Matters for Energy*. Springer, Berlin.
- 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>
- 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. and Robertson, A.W. (2019). Introduction: why sub-seasonal to seasonal prediction (S2S)?, Chapter 1 in *Sub-Seasonal to Seasonal Prediction*. Elsevier, Amsterdam, Netherlands.
- 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., Mas-trangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A.W., Ruti, 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 subseasonal 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>
- Weigel, A.P., Baggenstos, D., Liniger, M.A., Vitart, F. and Appenzeller, C. (2008) Probabilistic verification of monthly temperature forecasts. *Monthly Weather Review*, 136, 5162–5182. <https://doi.org/10.1175/2008MWR2551.1>
- White, I., Garfinkel, C.I., Gerber, E.P., Jucker, M., Aquila, V. and Oman, L.D. (2019) The downward influence of sudden stratospheric warmings: association with tropospheric precursors. *Journal of Climate*, 32, 85–108. <https://doi.org/10.1175/JCLI-D-18-0053.1>
- Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, CA.
- Wulff, C.O. and Domeisen, D.I.V. (2019) Higher subseasonal predictability of extreme hot European summer temperatures as compared to average summers. *Geophysical Research Letters*, 46, 11520–11529. <https://doi.org/10.1029/2019GL084314>
- Zubiate, L., McDermott, F., Sweeney, C. and O'Malley, M. (2017) Spatial variability in winter NAO–wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns. *Quarterly Journal of the Royal Meteorological Society*, 143, 552–562. <https://doi.org/10.1002/qj.2943>

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, Beerli R, Wernli H, Grams CM. Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *QJR Meteorol Soc.* 2020;1–20. <https://doi.org/10.1002/qj.3866>