Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables

Article

Published Version

RESEARCH ARTICLE

Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables

Hannah C. Bloomfield[1] | David J. Brayshaw[1,2] | Paula L. M. Gonzalez[1,3] | Andrew Charlton-Perez[1]

[1]Department of Meteorology, University of Reading, Reading, UK

[2]National Centre for Atmospheric Science, University of Reading, Reading, UK

[3]International Research Institute for Climate and Society, The Earth Institute, Columbia University, Palisades, New York, USA

Correspondence
Hannah C. Bloomfield, Department of Meteorology, University of Reading, Reading, UK.
Email: h.c.bloomfield@reading.ac.uk

Abstract

Sub-seasonal forecasts are becoming more widely used in the energy sector to inform high-impact, weather-dependent decisions. Using pattern-based methods (such as weather regimes) is also becoming commonplace, although until now an assessment of how pattern-based methods perform compared with gridded model output has not been completed. We compare four methods to predict weekly-mean anomalies of electricity demand and demand-net-wind across 28 European countries. At short lead times (days 0–10) grid-point forecasts have higher skill than pattern-based methods across multiple metrics. However, at extended lead times (day 12+) pattern-based methods can show greater skill than grid-point forecasts. All methods have relatively low skill at weekly-mean national impact forecasts beyond day 12, particularly for probabilistic skill metrics. We therefore develop a method of pattern-based conditioning, which is able to provide windows of opportunity for prediction at extended lead times: when at least 50% of the ensemble members of a forecast agree on a specific pattern, skill increases significantly. The conditioning is valuable for users interested in particular thresholds for decision-making, as it combines the dynamical robustness in the large-scale flow conditions from the pattern-based methods with local information present in the grid-point forecasts.

KEYWORDS
demand, forecasting, pattern forecast, power system, sub-seasonal, weather regimes, wind power

1 | INTRODUCTION

As power systems across the world transition towards low carbon electricity generation, they are becoming increasingly dependent on weather. Consequently, high quality weather forecasts are becoming increasingly important for decision-making days to weeks ahead to maintain a secure and reliable energy system. Examples

of these decisions include: managing reserve generation margins, maintenance scheduling, hydropower scheduling, anticipating winter heating demand requirements and cooling water requirements for conventional generation (White et al., 2017). Each of these decisions rely on accurate forecasts of demand, wind power, solar power or hydro generation, and the consequences of mis-estimation are exacerbated in periods of unusually high

system stress (e.g., low wind, low temperature periods). Accurate forecasts allow users to get extra information relevant for decision-making in the energy market when predicting energy prices. This can ultimately lead to large cost savings and enhanced profits for energy companies.

A power forecast is constructed from two main components: a meteorological component, reliant on the best estimate of the weather on the timescale of interest, and an energy-system component, which explains the conversion from weather to energy (see Cannon et al., 2017 for further discussion). The focus of this work is on the underlying meteorological predictability, rather than the optimal method to convert from meteorological to energy variables. This addresses the overarching question: What is the best meteorological forecasting method that is available at sub-seasonal-to-seasonal (S2S) timescales for creating European national power forecasts?

Several energy-meteorology studies are present in the literature, which demonstrate skill at lead times weeks-to-months ahead at predicting either demand (Bloomfield et al., n.d.; De Felice et al., 2015; Dorrington et al., 2020; Thornton et al., 2019) or wind power generation (Bloomfield et al., n.d.; Clark et al., 2017; Lledó et al., 2019). These studies have typically focused on power forecasts made by converting gridded surface meteorological data (e.g., near-surface winds, surface air temperature, insolation) into quantitative estimates of power generation or demand (typically for individual countries and energy variables). They therefore fail to adopt a comprehensive view of the skill present once forecasts of generation and demand are combined, or across networks and interconnections spanning large regions.

Common themes in the S2S forecasting research are the use of techniques based on large-scale patterns (Robertson & Vitart, 2018), including the so-called 'weather regime' (WR) approaches (Michelangeli et al., 1995). Underlying this research interest are two distinct concepts. It is well established that forecast errors typically grow and saturate more slowly at larger spatio-temporal scales, retaining predictability at longer lead times than individual grid-point forecasts in which error growth and saturation are more rapid (Hoskins, 2013). Suitable patterns are typically identified through a combination of principle-component-based analyses and/or clustering algorithms. The concept of regime-like behaviour has a long history in meteorological research (i.e., the evolution of the atmosphere is viewed as transitioning between a set of somewhat discrete states (Madonna et al., 2017; Michelangeli et al., 1995; Woollings et al., 2010)). Some regimes or regime transitions are inherently more predictable than others (Frame et al., 2011; Matsueda & Palmer, 2018).

Much of the science within the meteorological community has focused on the skill present in forecasts of

Euro-Atlantic teleconnections at seasonal timescales (Lledó et al., 2020) such as the North Atlantic Oscillation (NAO; Baker et al., 2018; Dunstone et al., 2016; Scaife et al., 2014) and Euro-Atlantic WRs at sub-seasonal timescales (Büeler et al., 2020; Charlton-Perez et al., 2018; Ferranti et al., 2015; Matsueda & Palmer, 2018). Forecast skill for the daily occurrence of WRs has, for example, been shown to extend to approximately 2 weeks for sub-seasonal forecasts (Ferranti et al., 2018; Matsueda & Palmer, 2018) with similar results found for WRs over North America (Robertson et al., 2020). Skill extends to 3 months for forecasts of seasonal-mean teleconnection patterns (Lledó et al., 2020), with models showing skill out to a year-ahead for the NAO teleconnection pattern in particular (Dunstone et al., 2016). The level of skill has been shown to be conditional on particular atmospheric states (e.g., Sudden Stratospheric Warmings; Beerli et al., 2017; Büeler et al., 2020; Charlton-Perez et al., 2018) or global teleconnections such as the El Niño–Southern Oscillation and the Madden–Julian oscillation (Lee et al., 2019), supporting the concept that particular atmospheric states offer windows of opportunity for enhanced predictability (Mariotti et al., 2020; Robertson et al., 2020).

Despite the popularity of pattern-based methods to understand and predict the large-scale atmospheric circulation, they are associated with the key caveat that skill in forecasting the large-scale meteorological pattern (typically observed in 500 hPa geopotential height or surface pressure) does not in itself guarantee skill in terms of predicting a desired surface impact response (e.g., near-surface wind speeds or temperatures, wind power generation or electricity demand). Recent work using historic reanalyses has, however, convincingly demonstrated that many Euro-Atlantic weather patterns (with various methods of construction) relate to relevant surface weather conditions (Bloomfield, Brayshaw, & Charlton-Perez, 2020a; Cortesi et al., 2019; Drücke et al., 2020; Garrido-Perez et al., 2020; Grams et al., 2017; Thornton et al., 2017; van der Wiel et al., 2019) and to European electricity demand, wind and solar power generation (Bloomfield, Brayshaw, & Charlton-Perez, 2020a; Grams et al., 2017; Lledó et al., 2020; van der Wiel et al., 2019). This suggests that, given these patterns can be forecast at lead times out to 2 weeks and 3 months for WRs and teleconnections respectively, they could potentially offer benefits over grid-point forecasts. A recently developed pattern-based method (targeted circulation types, TCTs) can further enhance the linkage between surface responses and the large-scale atmospheric circulation by classifying patterns based on energy-system data rather than large-scale meteorological fields (Bloomfield, Brayshaw, & Charlton-Perez, 2020a).

Given the potential of pattern-based forecasts for enhanced predictability and the demonstrated link

between large-scale circulation patterns and surface impacts, pattern-based information is widely seen as desirable for many forecast applications, such as energy-system users and traders wishing to make decisions at extended lead times. A recent survey conducted through the S2S4E climate service project found 60% of the 30 survey participants currently use pattern-based information, with a further 27% seeking to use them in future operations (Project S2S4E, 2020). This may be due to the relatively straightforward interpretation of the pattern-based methods. However, despite this interest, the specific advantages of pattern-based forecasting remain unclear. In particular, questions remain as to whether pattern-based approaches offer: (a) a general increase in skill over comparable use of individual surface grid-point data from the same forecast by identifying the surface response from the more predictable large-scale atmospheric flow, and/or (b) a means to identify variations in predictability over time, such as windows of opportunity where predictability is enhanced or degraded.

In the present study, a grid-point forecast is defined as one that uses grid-point surface meteorological forecasts (e.g., 10 m wind), and applies calibration and conversion models to estimate a relevant power quantity (e.g., national wind power output). In contrast, a pattern-based forecast of the same power quantity first assigns the forecast of the large-scale atmospheric flow into one of a discrete set of pre-identified circulation patterns (WRs or TCTs). Secondly, it applies a conversion model (based on a historical statistical relationship between the pattern and power production) to estimate a surface impact. To the best of our knowledge, there are no examples in the literature of pattern-based forecasts being directly compared with an equivalent grid-point forecast at sub-seasonal timescales. This is despite widespread interest in the climate services community, where pattern-based methods are clearly expected to offer advantages. The research questions for this study are therefore:

1. Do pattern-based forecasts offer more skilful European power predictions than equivalent grid-point based forecasts?
2. Can pattern-based methods help to identify windows of opportunity in which grid-point forecast skill is enhanced/degraded?

This study is structured as follows. The reanalysis data and S2S prediction systems used are described in Section 2. Methods of creating European energy data, and creation and verification of pattern-based forecasts are discussed in Section 3. Section 4 starts with comparisons of the grid-point based forecasts to pattern-based forecasts using WRs and TCTs (Section 4.1), with Sections 4.2 and 4.3 unpacking the reasons for the levels of

pattern-based skill that are present. Following this, the potential for conditional predictability of the grid-point forecasts is investigated (Section 4.4). Conclusions, lessons learned and potential future developments for these methods are highlighted in Section 5.

2 | DATA

2.1 | The ERA5 reanalysis

The process of estimating power system behaviour from meteorological data is complex, due to the importance of both meteorological and non-meteorological factors (see Bloomfield, Gonzalez, et al., 2020b). Due to the added complexities of real-world power data, we choose to follow an idealized model approach, where national demand and wind power generation created from the ERA5 reanalysis (Hersbach et al., 2020) using established models (see Section 3.1) are taken as truth, and the S2S model performance is compared with this. Using this idealized model set-up removes the uncertainty associated with power system operation and human behaviour from the analysis, and allows us to cleanly compare the meteorological skill of the forecast methods. It also removes the uncertainty associated with weather-to-energy conversion models.

The ERA5 reanalysis is available at hourly resolution from 1950 to present at approximately 30 km spatial resolution. The S2S hindcasts are however not available at such high spatial and temporal resolution, so the ERA5 reanalysis is interpolated to match the temporal resolution of the S2S models (see Section 2.2).

2.2 | S2S forecasting systems

The two forecasting systems used in this study are the European Centre for Medium-Range Weather Forecasts (ECMWF) extended range forecasting system (cycle CY41R1) and the National Center for Environmental Prediction Climate Forecast System (NCEP-CFS; version T126L64GFS). Both are available for download from the S2S database (Vitart et al., 2017). In this study, hindcasts are taken from these models covering the period 1996–2015 for ECMWF and 1999–2010 for NCEP-CFS (called NCEP for brevity). ECMWF forecasts are available on Mondays and Thursdays with an 11-member ensemble. The NCEP model produces a forecast every day, but only with four ensemble members. To make the fairest possible comparison between these two models, a lagged-ensemble has been constructed from NCEP, where forecasts from the preceding 2 days of the ECMWF start

dates are grouped together, to provide a 12-member lagged-ensemble twice a week. At short lead times this lagged-ensemble approach will result in reductions in skill of the NCEP forecasts compared with ECMWF (Manrique-Suñén et al., 2020). However, as the main focus of this study is on extended lead times, this effect is thought to be minimal. Only hindcasts initialized in the months December–February are considered in this study, as this is a critical period for weather-dependent power system operation and also a period of relatively high hindcast skill (Bloomfield et al., n.d.). As the focus of this study is on the sub-seasonal predictability, skill assessment is performed on weekly-mean forecasts of which weeks 0–3 are defined as days 0–6, 5–11, 12–18 and 19–25, respectively. The reason for starting at day 5 is that shorter lead times are not the focus of this study, or the intended use of the S2S forecasts. Starting at day 5 also allows for any time needed for forecast acquisition, calibration and conversion to energy variables. An overlap is present between week 0 and week 1 so that week 1 is consistent with the definitions used in the S2S4E project. We note that the data used here correspond to hindcasts, which have smaller ensemble sizes and a worse representation of forecast uncertainty than operational forecasts. This limitation, however, does not affect the robustness of the comparison between the methods presented in this study.

The fields available from the S2S database, which are used for this study, are daily-mean 2 m temperature and midnight (00 UTC) 10 m wind speed at 1.5° spatial resolution. These variables were calibrated using the method of variance inflation (Doblas-Reyes et al., 2005) to a corresponding version of ERA5 interpolated to 1.5° as the reference. The method ensures that the lead-dependent reforecast mean and variance agree with those in ERA5, and also that the correlation between the reforecast and ERA5 is preserved (Doblas-Reyes et al., 2005). Each individual reforecast is calibrated by comparing all the other re-forecasts made on the same calendar date to a reanalysis. This is sometimes known as a leave-one-out approach (e.g., 1996 is corrected using data from 1997 to 2015). The correction to the mean is performed by subtracting a lead-time dependent reanalysis-based climatology. Before calibration, the 10 m hindcast wind speeds were first converted to 100 m wind speeds using a wind power law (to match the available ERA5 wind speeds). Full details on this process are provided in (Bloomfield et al., n.d.), and the corresponding datasets are available from https://doi.org/10.17864/1947.275.

3 | METHODS

In this section, four routes to producing demand and demand-net-wind (DNW) forecasts (shown in Figure 1)

are discussed in detail. These are energy variables from gridded hindcasts (Section 3.1), energy variables from WR hindcasts (Section 3.2), energy variables from TCT hindcasts (Section 3.3) and energy variables from conditional WR and TCT hindcasts (Section 3.4) where the outputs from each method are a time series forecast of national energy variables. Details of relevant skill metrics used in the study are given in Section 3.5.

3.1 | Energy variables from gridded hindcasts

The methods for creating national demand and wind power forecasts discussed below are applied to 1.5° versions of ERA5 and the bias-corrected re-forecasts to give grid-point based reconstructions of national energy variables. Within this study all national energy data are viewed as anomalies, where the mean is subtracted and then normalized by the standard deviation to account for the wide range of magnitudes and variability of national demand and wind energy production across Europe. This makes the magnitude of deviations from 0 equally important in all countries (see Bloomfield, Brayshaw, & Charlton-Perez, 2020a for further discussion). A high-level schematic of this method is given in the blue box in Figure 1, where daily-mean temperatures are converted into weekly-mean demand.

3.1.1 | Electricity demand

Electricity demand is calculated with a national-level multiple linear regression model containing parameters to capture both meteorological and human behaviour. Each country has a unique regression model, which is trained on 2 years of measured demand data (2016–2017) from the ENTSO-E transparency platform (ENTSOE, 2018), and then applied retrospectively to the ERA5 reanalysis from 1980 to 2018 and to the relevant S2S hindcasts. This results in two time series of data, one from the hindcast and one from ERA5, which can be compared. In this study, a weather-dependent demand model is used (which includes only the weather-dependent terms driven by national-average temperatures) to highlight the meteorologically-driven power system variability (see Bloomfield, Brayshaw, & Charlton-Perez, 2020a). Regression coefficients are used for national-average temperatures derived from the native ERA5 grid (~0.3°) rather than from the 1.5° grid to give the best possible representation of national demand. This is particularly important for small countries or those with complex coastlines and orography.
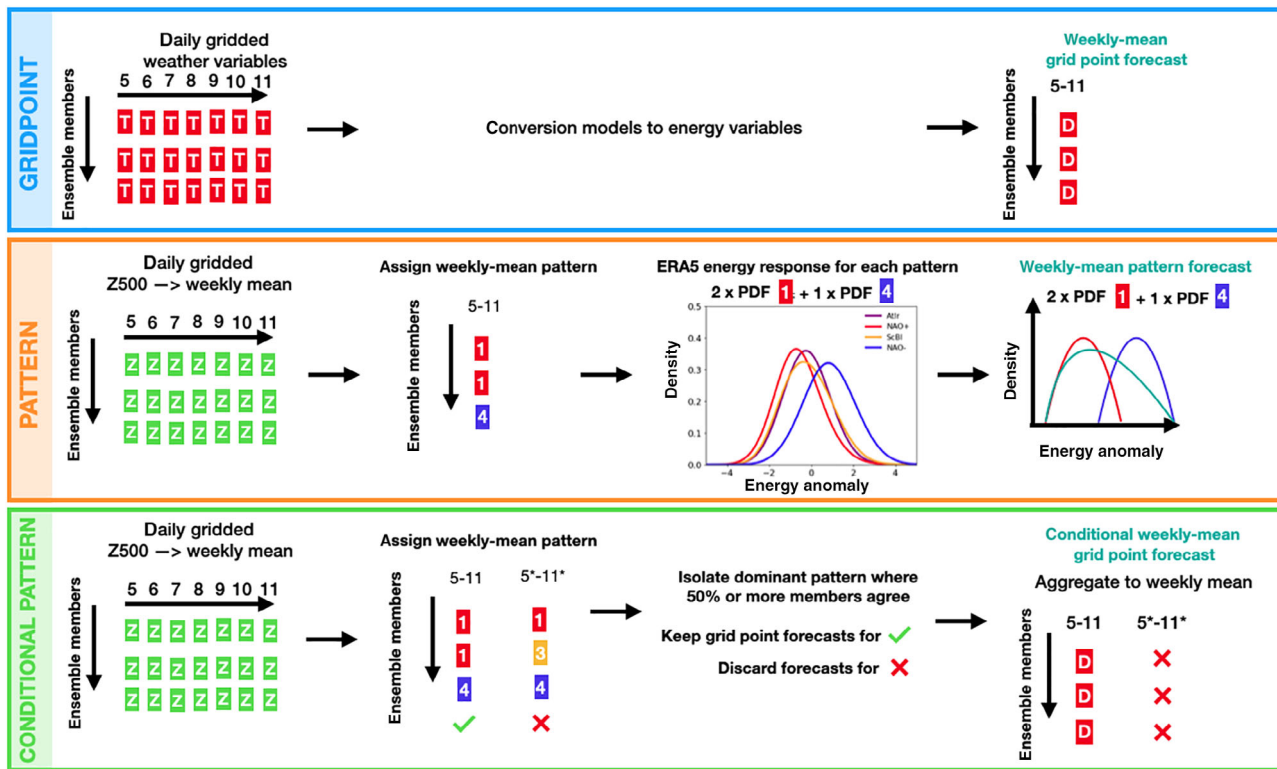
FIGURE 1   A schematic diagram of the three types of method used to forecast national demand and demand-net-wind in this study. Blue box: Grid-point forecasts, where an example of daily-mean temperature (T) can be converted to weekly-mean energy demand (D). Orange box: An example for weather regimes (WRs) where daily geopotential height (Z) can be converted into a probability distribution forecast of demand. Green box: A conditional pattern forecast example for WRs where a weekly-mean pattern (see orange box) is 'kept' if 50% or more ensemble members agree on a pattern. The weekly-mean grid-point energy forecasts for these situations (see blue box) are used as the conditional pattern forecast

3.1.2 | Wind power generation

Wind power capacity factor is calculated based on the methodology of Bloomfield, Brayshaw, & Charlton-Perez (2020a). ERA5 100 m wind speeds are first calibrated to the Global Wind Atlas (GWA), to account for low wind speed biases over regions of complex terrain (GWA, 2018). The national capacity factor is calculated by passing gridded bias-corrected 100 m wind speeds through an appropriately-chosen power curve and aggregating based on the locations of installed turbines taken from thewindpower.net. Further details and validation of the models are given in (Bloomfield, Brayshaw, & Charlton-Perez, 2020a). The wind power model performs well compared with other studies, which follow this now-standard overall approach (Cannon et al., 2015; Lledó et al., 2019; Sharp et al., 2015).

3.2 | Energy variables from WR hindcasts

Daily geopotential height data are taken from the S2S hindcast models, and used as the input to calculate the

WR pattern forecast (see orange box in Figure 1 for a summary of the method). The method of Cassou (2008) is adapted to calculate the weekly-mean WRs from the weekly-mean 500 hPa geopotential height data. The first 14 principal components (explaining 90% of the variance) of December–February area-weighted weekly-mean 500 hPa geopotential height over the Euro-Atlantic region (90 W–30 E, 20 N–80 N) are classified into four circulation types using the k-means clustering algorithm. This method has been used frequently in the literature to classify daily WRs, and previous studies have highlighted the impacts on energy systems (Bloomfield, Brayshaw, & Charlton-Perez, 2020a; van der Wiel et al., 2019). The four patterns are commonly known as the positive and negative phases of the North Atlantic Oscillation (NAO+ and NAO−), Scandinavian blocking (ScBl) and the Atlantic Ridge (Atlr). As shown in the study by van der Wiel et al. (2019), each of the traditional WRs can be related to anomalous energy conditions. For example, the NAO− regime is commonly associated with anomalously high demand and low wind power generation over much of Central and Northern Europe (see Figure 2, left column). We note that there are multiple methods to calculate WRs, which can have

more than four patterns (e.g., Garrido-Perez et al., 2020; Grams et al., 2017). Here a method resulting in four patterns is used for complementary with the definition of the TCTs (which are designed using a variation on the Cassou (2008) method).

To produce a national energy forecast from meteorological patterns, the WRs are first identified in the ERA5 reanalysis using the period 1980–2018. The outputs from this (the definition of the WRs and TCTs, and the surface energy responses during each patterns, see Figure 1) are

then used to create the pattern-based energy forecast. The ERA5 WR definitions (500 hPa geopotential height anomaly centroids) are first used to assign each individual ensemble member of the forecast to a pattern at a given lead time, then the frequency of pattern-occurrence is used to weight a surface response probability distribution (see orange box in Figure 1). For example, if the 11-member ECMWF forecast is assigned such that three members are in pattern 1 and eight members in pattern 2, then the resulting surface impact distribution forecast
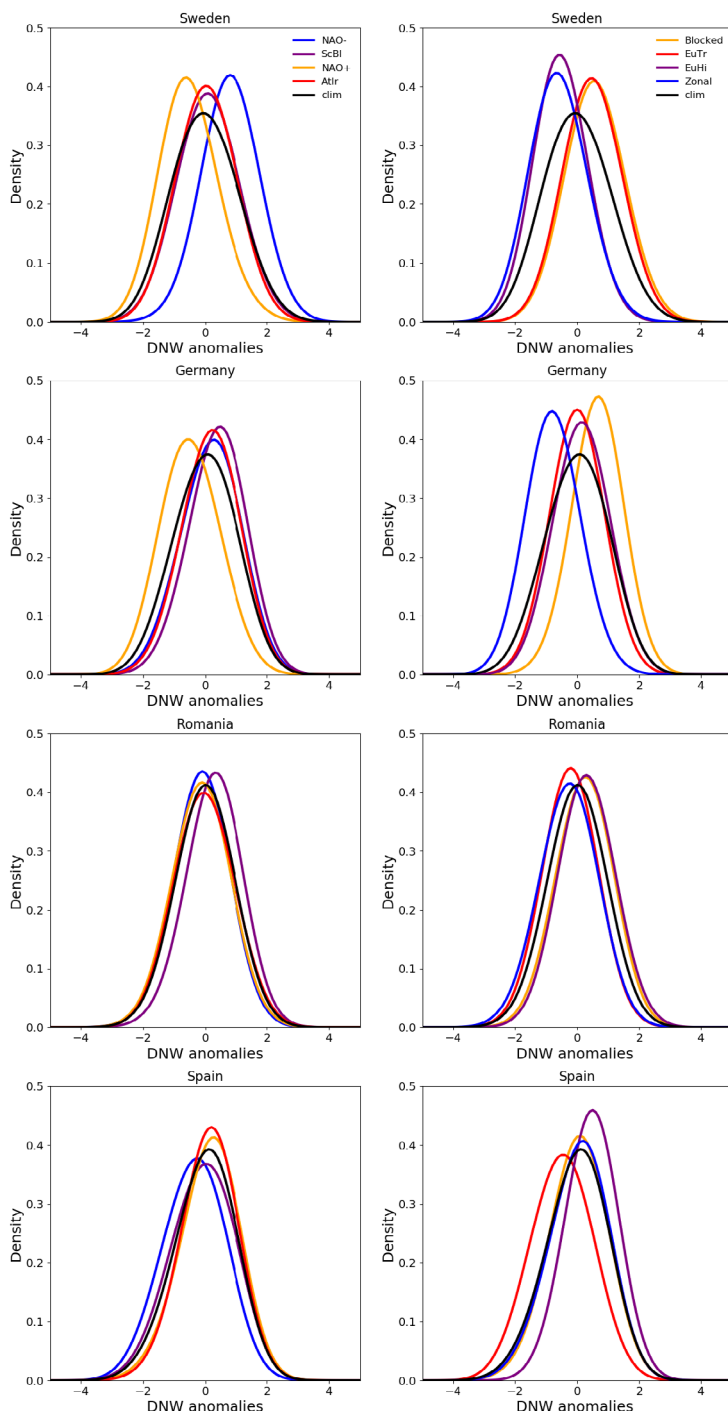

FIGURE 2 Kernel-density estimates of the normalized demand-net-wind (DNW) anomalies present during each of the WRs (left) and TCTs (right) for four European countries. Black lines in all plots represent the climatological distribution of DNW anomalies. Data from December-February, 1999–2010 from ERA5 is used to create these plots (the common period from the S2S models used in this study). Note we have chosen to conduct this work in anomaly space (GW/GW) to make swings in energy indicators comparable between countries (see Bloomfield, Brayshaw, & Charlton-Perez, 2020a for further discussion of this). TCTs, targeted circulation types; WRs, weather regimes

BLOOMFIELD ET AL.

Meteorological Applications
Science and Technology for Weather and Climate
Open Access

RMetS

7 of 16

is three-parts pattern 1, eight-parts pattern 2. We note that this output is constructed different to the output from the grid-point forecast method (which gives a set of 11 possible energy responses), but a probability distribution is required to represent the large range of possible energy responses that can be seen for each meteorological pattern (see Figure 2). For visualization of the geopotential height anomalies associated with each pattern, see the study van der Wiel et al. (2019).

3.3 | Energy variables from TCT hindcasts

TCTs are constructed analogously to the WRs (described above), but by performing the k-means clustering on the first 14 principle components of set of 28 countries' national energy data, rather than using a gridded 500-hPa geopotential height variable (Bloomfield, Brayshaw, & Charlton-Perez, 2020a). The same steps as outlined in the orange box in Figure 1 can still be followed: daily-mean energy variables are converted into a weekly-mean variable and from this a weekly-mean pattern is assigned (based on the energy anomalies from all 28 countries). A weekly-mean pattern forecasts can then be created by combining the probability distributions of the forecast TCTs.

The benefit of using the TCTs is that they are designed to have a stronger relationship to the energy system. TCTs are constructed for both demand and DNW anomalies. An example of the DNW anomalies for both WRs and TCTs over four countries is shown in Figure 2 (similar results are seen for demand, not shown). Comparing the two columns in Figure 2 (left WR, right TCTs) reveals a much clearer separation between the distributions of normalized DNW is seen for the TCTs than for the WRs, suggesting that, given equal skill in predicting the occurrence of WR and TCTs, the TCTs method should produce enhanced skill in predicting the surface response. Particularly clear distinctions are shown between the zonal and blocked TCTs. For more details on the large-scale conditions and energy anomalies associated with each WR and TCT, see the study by Bloomfield, Brayshaw, and Charlton-Perez (2020a).

3.4 | Energy variables conditioned on pattern-based methods

As discussed in the Introduction section, one potential benefit of pattern-based forecasting is the ability to highlight windows of opportunity with enhanced predictive skill based on the occurrence of particular atmospheric

conditions. Previous work by Bloomfield et al. (n.d.) has shown that grid-point forecasts of demand and demand-net-renewables have good skill in week 1 (days 5–12), however this skill rapidly decreases at longer lead times. Windows of opportunity for increased skill are sought by conditioning the grid-point forecasts, based on times when the models are relatively confident in their prediction of future large-scale dynamical conditions. The method used is outlined in the green box of Figure 1, with the example of a conditional WR forecast.

In this study, confidence is defined as 50% or more of ensemble members agreeing on the weather pattern (TCT or WR), which would be present in a given week (i.e., 6 of 11 members for ECMWF or 7 of 12 for NCEP). This threshold was tested through a sensitivity analysis and 50% was identified as a good balance between the skill of the model identifying the dominant pattern, and a large enough number of forecasts available to make decisions (see Appendix 1 for details of the available amount of forecasts within windows of different thresholds).

A conditional pattern forecast consists of the grid-point energy forecasts (weekly-mean energy variables for all ensemble members) for the occasions when the forecast has a dominant pattern (noting that this dominant pattern forecast could be incorrect). The skill of this subset of forecasts can then be compared with the full set of forecasts.

3.5 | Skill metrics

Three verification metrics are used to assess the performance of the ensemble forecasts of energy variables. The first metric assessed is the ensemble-mean correlation (EnsCorr; Wilks, 2011). This gives users a deterministic measure of whether the positive or negative anomalies in forecasts are aligned with observations, and is one of the simplest measures of skill. The second metric is the ranked probability skill score (RPSS) of the terciles of the distribution of the variable. This assesses the performance of the forecast when the continuous variable is reduced to three categories (below normal, normal and above normal; (Epstein, 1969). The third metric is the continuous ranked probability skill score (CRPSS). This assesses the forecast probability distribution of the continuous variable (Brown, 1974). In all cases, we construct skill scores referenced to a climatological forecast (i.e., positive is an improvement on climatology and negative is worse, with unity corresponding to a perfect forecast). The two probabilistic skill scores give the skill of the sub-seasonal forecast relative to always forecasting the climatological probabilities of the events or categories involved. Providing probabilistic skill scores that are

benchmarked against climatology shows how much extra information these methods could provide on a particular forecasts category compared with what is commonly used within many sectors at present (particularly at extended lead times).

An important point to note is that as the hindcasts typically have fewer ensemble members than operational forecasts, this tends to limit the skill of reforecast relative to a true forward-looking forecast. In the case of the S2S ECMWF forecasts, the real-time forecasts have 51 ensemble members, whilst the re-forecasts have only 11 members. To account for this, the method of (Ferro et al., 2008) is applied to the RPSS and CRPSS.

4 | RESULTS

This section begins by answering the first research question: Do pattern-based forecasts offer more skilful European power predictions than equivalent grid-point based forecasts? (Section 4.1). The reasons for the given level of skill are unpacked by further investigating the relationships to the impacted system (Section 4.2) and the predictability of the patterns (Section 4.3). Section 4.4 focuses on the second research question: Can pattern-based methods help to identify windows of opportunity in which grid-point forecast skill is enhanced/degraded? The results from the skill assessment in Section 4.1 are used as a benchmark to look for windows of opportunity in the grid-point forecast methods.

4.1 | Predicting weekly-mean energy variables

Figure 3 shows predictive skill of weekly-mean DNW from the ECMWF model, for weeks 0–3. The grid-point forecasts of weekly-mean DNW in Figure 3 have very high skill in week 0 (the mean of days 0–6; EnsCorr greater than 0.8). This remains high in week 1 but then drops considerably with increased lead time to 0.4 by week 2 (days 12–18). In general, more skill is seen in the grid-point forecasts in Northern and Eastern Europe, this could be due to the weaker relationship between temperature and demand in winter for the Southern European countries, or due to the generally lower amount of installed wind capacity compared with Central and Northern Europe. The largest gains in skill using pattern-based method are seen in weeks 2–3, across Central Europe. Focusing first on week 0, we see that the grid-point forecast method clearly outperforms the two pattern-based methods, using WRs and TCTs (see Figure 3, fourth and fifth columns). In forecast week 2 (third

row of Figure 3), the skill of the grid-point and pattern-based forecasts are quite similar, and often no significant differences are seen. By week 3 (days 19–25), the WRs have statistically significant higher skill than the grid-point forecasts in Central and Northern European countries (Figure 3p). This suggests that, at longer lead times, pattern-based methods are capable of providing slightly more skilful forecasts than the grid-point forecasts. This increase in skill is however limited to Central and Northern Europe for WRs, and to small regions of Central Europe for TCTs.

When comparing the skill of the TCT and WR methods, we find the skill in the TCT-based forecasts is much higher than for the WR-based forecasts in weeks 0 and 1. This is, however, still lower than the skill seen for the grid-point forecasts. In week 2, the TCT-based forecasts perform better than the WR-based forecasts in some regions, but still do not have enough skill to beat the grid-point forecasts. In week 3, the TCT-based forecasts have lower skill than those from WRs.

For brevity, the discussion in this section is limited to one skill metric, energy variable and S2S model. However, we note similar results are seen for energy demand (with slightly higher correlations) and for the more complex probabilistic skill metrics (RPSS and CRPSS) when the comparison between grid-point and pattern-based forecast methods are made. Hence we choose to focus on EnsCorr, as the purpose of this study is to compare forecasting techniques not skill metrics. We include results from a range of metrics in the Appendix 2 for completeness.

The increased skill in demand forecasts when compared with DNW is thought to be due to the models higher skill for forecasting surface temperatures than wind speeds at S2S timescales (Büeler et al., 2020). The differences between demand and DNW forecasts are discussed further in the study by Bloomfield et al. (n.d.). We note the generally lower skill scores seen for the NCEP hindcasts could be due to using the ERA5 reanalysis for verification, which will be more similar in model set-up to the ECMWF hindcasts.

The choice of forecast skill metric will be dependent on the type of decision that is required, therefore we include all of the plots in Appendix 2 to allow for further exploration of this. Generally, the grid-point forecasts from the NCEP model have lower skill than those from ECMWF (see Bloomfield et al., n.d.) and, for the NCEP model, the skill gain from using WR-based forecasts is greater in weeks 2–3 because the quality of the grid-point forecasts degrades more quickly with increasing lead time (see Appendix 2). This degradation of the grid-point forecasts could be influenced by using the ERA5 reanalysis for validation, rather than the NCEP reanalysis.
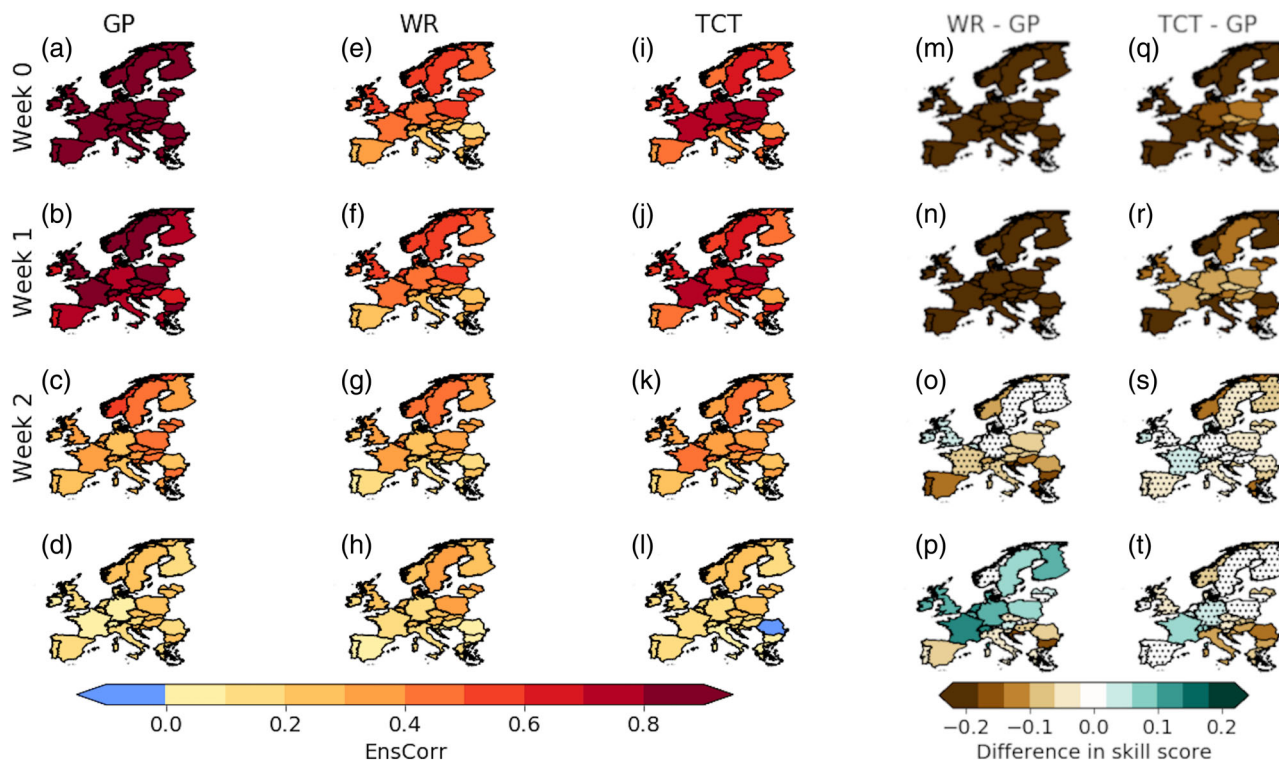
BLOOMFIELD ET AL.

Meteorological Applications
Science and Technology for Weather and Climate
Open Access

RMetS

9 of 16

FIGURE 3 Weekly-mean ensemble-mean correlation of demand-net-wind (DNW) from December-February, using the ECMWF model (1996–2016). (a–d) Grid point (GP) forecasts (e–h) pattern-based forecasts using WRs (i–l) pattern-based forecasts using TCTs (m–p) the difference in GP and WR skill scores (q–t) the difference in GP and TCT skill scores. Stippling in the fourth and fifth column shows no significant difference between the pairs of skill scores (2000 bootstrapped samples from all available forecasts are used to confirm the significance with 95% confidence). Tabulated skill scores are available in Appendix 3. ECMWF, European Centre for Medium-Range Weather Forecasts; TCTs, targeted circulation types; WRs, weather regimes

In summary, we have found that in weeks 0–2 the grid-point forecasts typically equal or outperform the pattern-based approaches considered, whereas in week 3 the WR pattern-based method has a modest advantage. The reasons for the differing levels of performance are investigated in Sections 4.2 and 4.3.

4.2 | Pattern relationship with impact variables

A question that could be raised of the TCT and WR methodology is whether the relationship between the patterns and impact variables (i.e., step 1 in Figure 1, with examples shown in Figure 2) is adequate to produce high quality energy forecasts. To test this assumption, we have assessed the skill present in a perfect model of the pattern-based methods. This term is used because we employ a perfect forecasting strategy where the pattern known to exist in ERA5 on a given date is used as the pattern 'forecast'. Figure 4 shows the skill present in a perfect model is high across multiple skill metrics for the two methods (see Section 3.5 for a definition of these).

Particularly high potential is seen for demand, and for areas of Central and Northern Europe. We note that higher skill is present in the perfect model of TCTs than for WRs, due to their better representation of the surface energy conditions during each pattern (see Figure 2). The relatively low skill of the perfect model seen for RPSS and CRPSS shows that an avenue for improving these pattern-based forecasting methods is to investigate improvements in the statistical relationship between the patterns and the time series of energy data.

From Figure 4 we can conclude that both WRs and TCTs have the potential to provide skill at extended lead times (when skill in grid-point forecasts becomes low) if their occurrence could be forecast accurately. This suggests that the reduced skill present in Figure 3 when compared with Figure 2 is due to the models' ability to correctly assign the WRs and TCTs.

To take this a step further, the skill of a perfect model is compared with the skill of the ECMWF and NCEP weekly-mean grid-point forecasts in Figure 5, showing an example of DNW over Germany. Until day 6 (3), even with a perfect forecast of the weekly-mean TCT, the TCT method is not able to provide more skill than the
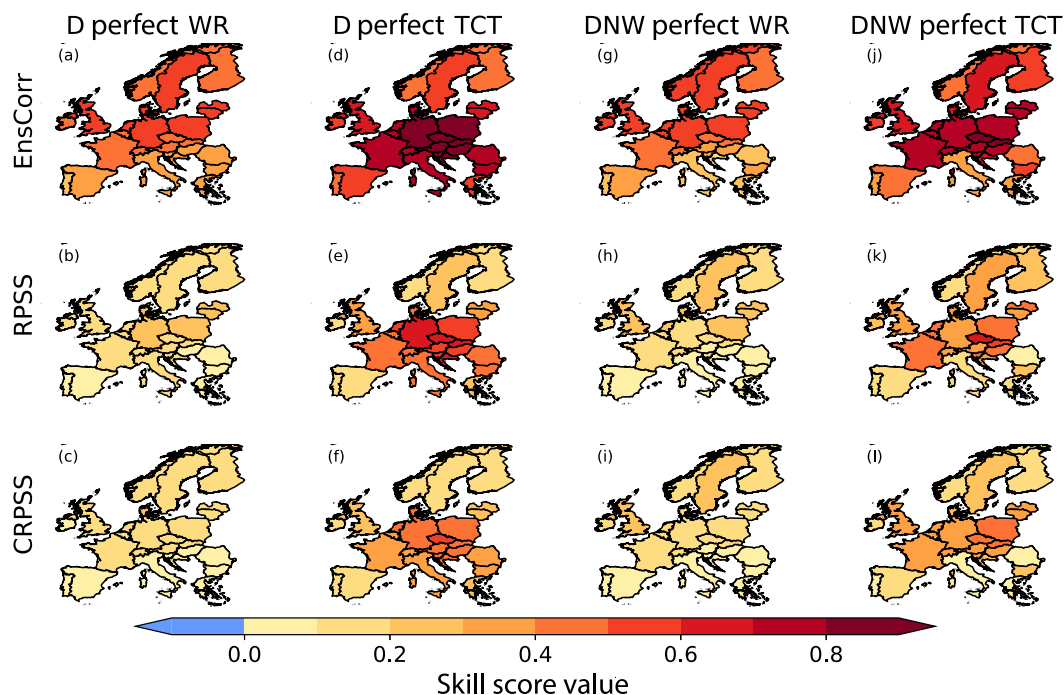
FIGURE 4 The perfect predictability of the WR and TCT pattern-based methods for demand (a–f) and demand-net-wind (DNW; g–l), that is, forecasts using the pattern taken from the ERA5 reanalysis for prediction. Three skill scores are shown, ensemble correlation (top) RPSS (middle) CRPSS (bottom). The period 1996–2016 is used to match the ECMWF hindcast period. Significant differences are not seen if the period is changed. See Section 3.5 for skill score definitions. Tabulated skill scores are available in Appendix 3. CRPSS, Continuous Ranked Probability Skill Score; ECMWF, European Centre for Medium-Range Weather Forecasts; RPSS, Ranked Probability Skill Score; TCTs, targeted circulation types; WRs, weather regimes

weekly-mean grid-point forecast method for the ECMWF (NCEP) model respectively. This shows that at very short lead times we would not expect these pattern-based methods to outperform the grid-point forecasting strategy (due to present limitations in the statistical conversion from pattern to energy variables). Similar results are seen for WRs, but for longer lead times, as the link between the WRs and corresponding impact variables is weaker (see Figure 2 for an example). Figure 5b–d shows the number of days before both TCTs and WRs can potentially beat the grid-point forecasts of RPSS. We see that due to the reduced skill level in the NCEP grid-point forecasts the pattern-based methods can be useful at earlier lead times for all countries (see Bloomfield et al. (n.d.) for further grid-point skill comparisons). If forecast is perfect, then the TCTs are useful at shorter lead times than the WRs for all but a few exceptions (notably Norway, where there is a good relationship between the WRs and surface impacts, see Bloomfield, Brayshaw, & Charlton-Perez, 2020a). This provides further explanation for some of the results shown in Figure 3. We now see that there would not be an expectation of the pattern-based methods to perform better than the grid-point forecasts in week 0. However, if the TCTs and WRs could be assigned accurately, there are potential gains in skill to be had over grid-point forecasts in weeks

1–3, particularly for regions of Central Europe, which have a diverse range of responses to the WR and TCT patterns.

From an operational forecasting perspective, investing effort in pattern-based forecasting has the largest potential for countries in Central Europe, as these countries have a strong response to multiple TCT/WRs (Bloomfield, Brayshaw, & Charlton-Perez, 2020a). This analysis has also shown significant value in developing alternative impact-based forecasting methods to the traditional WRs, which could provide useful information for an average of 3 more days lead-time across the whole of Europe than traditional WRs.

4.3 | Pattern assignment in S2S models

The percentage of both WRs and TCTs, which are assigned correctly in the two hindcasts, are shown in Figure 6. Even in week 0 the forecast models are only assigning the pattern correctly around ~75% (WRs) to ~85% (TCTs for demand) of the time for the ECMWF model, with significantly lower percentages seen for the NCEP model. We note that in lead weeks 0 and 1, the ECMWF model is able to assign the TCTs better than the WRs, but by week 2 the assignment rates are similar
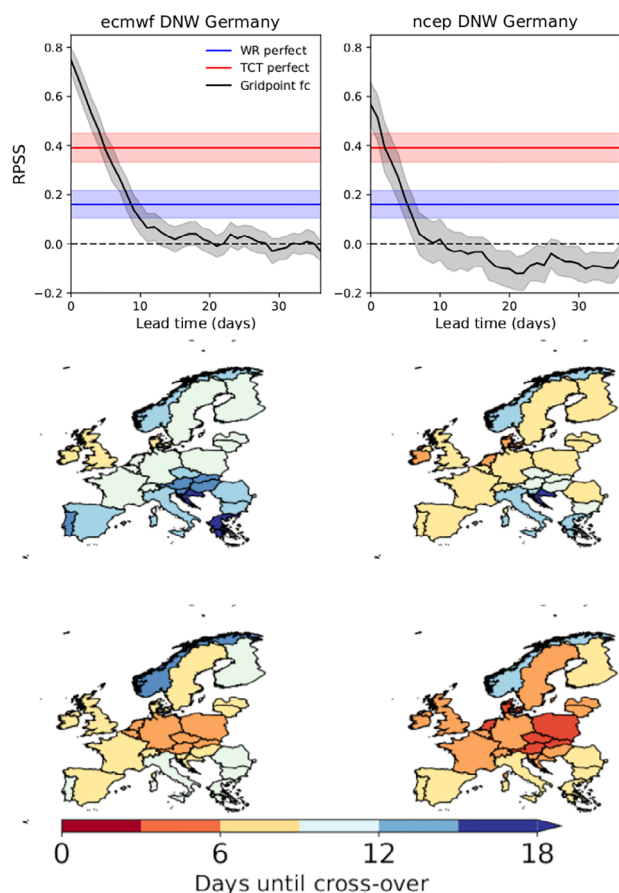
FIGURE 5 First row: An example for Germany of the lead time where WR (green) and TCT (purple) methods have the potential (if perfectly predictable) to provide increased skill compared with the weekly-mean grid-point forecasts (black). Second row: The minimum number of days for the WR method to have the potential to beat grid-point forecasts for the ECMWF model (left) and NCEP model (right). Third row: As second row for the TCT method. ECMWF, European Centre for Medium-Range Weather Forecasts; NCEP, National Center for Environmental Prediction Climate Forecast System; TCTs, targeted circulation types; WRs, weather regimes

(this reflects the higher skill-levels seen in TCT-based forecasts in week 0 and week 1 compared with WR-based forecasts; Figure 3 right hand column). In the NCEP model, there is a higher percentage of WRs assigned correctly than TCTs at all lead times. This could be due to the reduced skill in the NCEP grid-point energy forecasts (see Appendix 2).

From days 10–20, both the ECMWF and NCEP models show an increased number of correct WR assignments (bottom left sub-panel) compared with the demand and DNW TCTs (right sub-panels). The number of correct assignments is also statistically significantly above the climatological number of correct assignments (given by the orange lines in Figure 6). The end of this

period corresponds to the lead time range at which the WRs are able to outperform the grid-point forecasts (Figure 3, bottom row). The increased number of correct WR assignments compared with correct TCT assignments could be due to the models' ability to forecast large-scale conditions more accurately than local phenomena at sub-seasonal lead times (Robertson & Vitart, 2018).

4.4 | Windows of opportunity for grid-point forecasts

In the previous sections, it was demonstrated that pattern-based methods offer the potential for modest enhancements in skill (compared with grid-point forecasts) at longer lead times, typically in week 3. In broad terms, this is consistent with the well-established observation that forecast error-growth saturates more slowly at larger spatial scales (see Section 1). An alternative approach, however, is to use the pattern forecast to condition the grid-point forecast: that is, to only utilize the grid-point forecast when their ensemble members largely agree on the predicted dominant large-scale pattern in the hindcasts.

This section focuses on the NCEP model (where we know grid-point based forecasts to contain a lower level of skill than the ECMWF model and therefore has the most to gain from any pattern-based conditioning). However, similar results are seen for the ECMWF model (not shown for brevity). DNW and RPSS are used as a demonstrative example of skill, though similar results are seen with all previously discussed skill metrics and for energy demand (not shown).

Figure 7 shows the skill difference between the standard grid-point forecast method (see Figure 1) and grid-point forecasts that have been conditioned so that only forecasts where at least 50% of the ensemble members agree on the large-scale pattern that is present are kept. It is first noted that no significant difference in skill is seen in week 0 when conditioned on either WRs or TCTs (first row of Figure 7) as nearly all forecasts fulfil the condition at this early lead time (i.e., every forecast confidently picks a single pattern with at least 50% of the ensemble).

In week 1, skill increases in RPSS of ~0.2 are seen across Europe, maintaining the levels of skill seen in week 0 if all of the grid-point forecasts are used. The conditional forecasts are issued for 95% of the WR forecasts, and 90% of the TCT forecasts. This shows that by only rejecting a small amount of the grid-point forecasts (where the uncertainty in the prediction of the large-scale pattern is very high) a significant gain in the skill can be obtained. This ability to reject very poor forecasts (or at least be aware of the high levels of
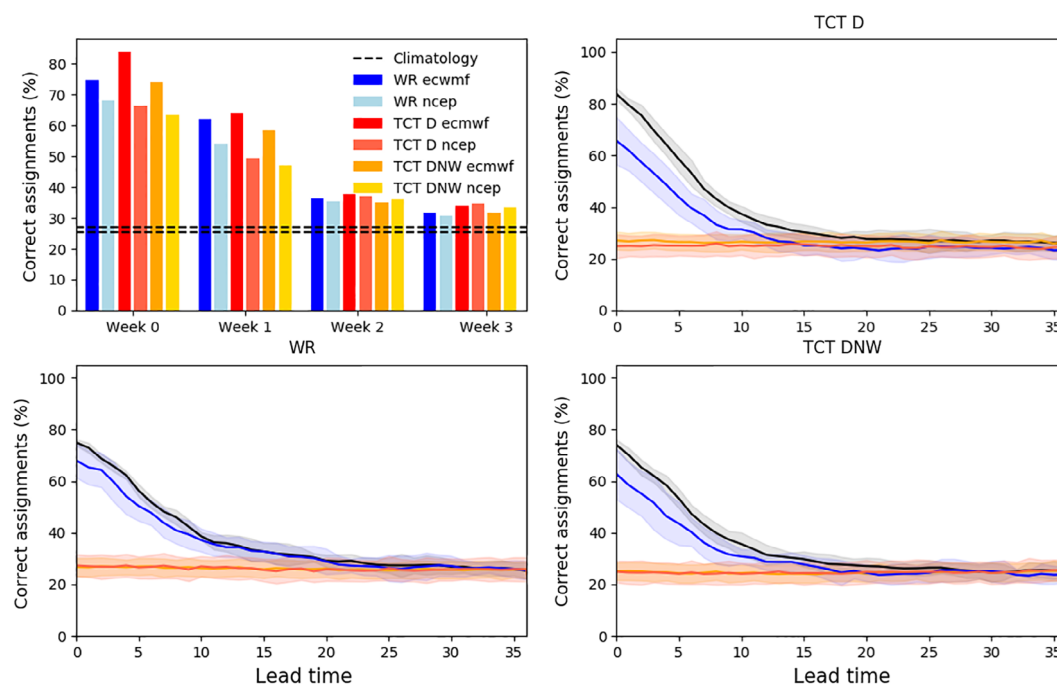
FIGURE 6 Top left: Percentage of correct pattern assignments (hit rate) for the individual members of the ECMWF and NCEP models when compared with the corresponding hit rate in ERA5, for lead weeks 0–3. Dashed black lines give the range of climatological hit rates from ERA5. The percentage of correct weekly-mean pattern assignments against lead time is shown for (top right) demand-only TCTs (bottom left) WRs (bottom right) demand-net-wind (DNW) TCTs. Black, blue and orange lines show the correct assignment rate from ECMWF hindcasts, NCEP hindcasts and the climatological pattern forecasts from each model (shaded regions constructed from 2000 bootstrapped samples). The lead time represents the first day of the weekly-averaging period. For example, day 5 is the weekly-mean pattern averaged from days 5–12. ECMWF, European Centre for Medium-Range Weather Forecasts; NCEP, National Center for Environmental Prediction Climate Forecast System; TCTs, targeted circulation types; WRs, weather regimes

uncertainty) is potentially valuable to users making complex decisions. It is possible for this method to generate a 'forecast bust' if the pattern chosen as dominant is incorrect. However, this is relatively uncommon, and using this pattern conditioning typically improves forecast skill at extended lead times. Although the conditioning is applied here to the grid-point forecast, a similar selection could equally be applied to a pattern-based forecast to increase its skill.

In week 2, the amount of forecasts available for which the condition is met begins to drop rapidly. Seventy-two percent of WR-based forecasts are issued, whereas only 52% of TCT-based forecasts are issued. In these windows of opportunity, however, significant increases in RPSS are present (again similar to the week 0 values of grid-point skill for all forecasts). More conditional forecasts are present for the WR-based method than the TCT-based method for both the ECMWF and NCEP models.

At weekly-mean timescales, the models are best able to identify the NAO+ WR, and the zonal demand-only TCT (with both being assigned correctly greater than 90% of the time when they are the dominant pattern present in week 0). This is in agreement with Ferranti et al. (2018), who also found highest predictability for the

NAO states. This increased predictability could be due to the zonal TCT and NAO+ WR being associated with times when the North Atlantic eddy-driven jet is in the Northern location (Madonna et al., 2017). When the jet is in the northern location, it is less likely to be weakened or split, which have both previously been shown to relate to a reduction in the jet streams persistence and predictability, (Frame et al., 2013).

Both S2S models are able to correctly identify a dominant WR more accurately than a dominant TCT (see Figure 7, fourth column). This is interesting as the WRs themselves have little relationship to the energy system (see fig. 2 of Bloomfield, Brayshaw, & Charlton-Perez, 2020a). This section has therefore shown that although when used in isolation a WR forecast can tell you relatively little about energy-system operation, they can still add value to the energy-system predictions by providing an insight into the flow-dependent predictability of the large-scale circulation, which can condition the grid-point forecasts. An important caveat here is that the TCT patterns are constructed based on the national grid-point forecasts, therefore a dominant TCT forecast is probably associated with relatively low ensemble spread in the grid-point forecasts. This limits the potential gain from conditioning unless the statistical
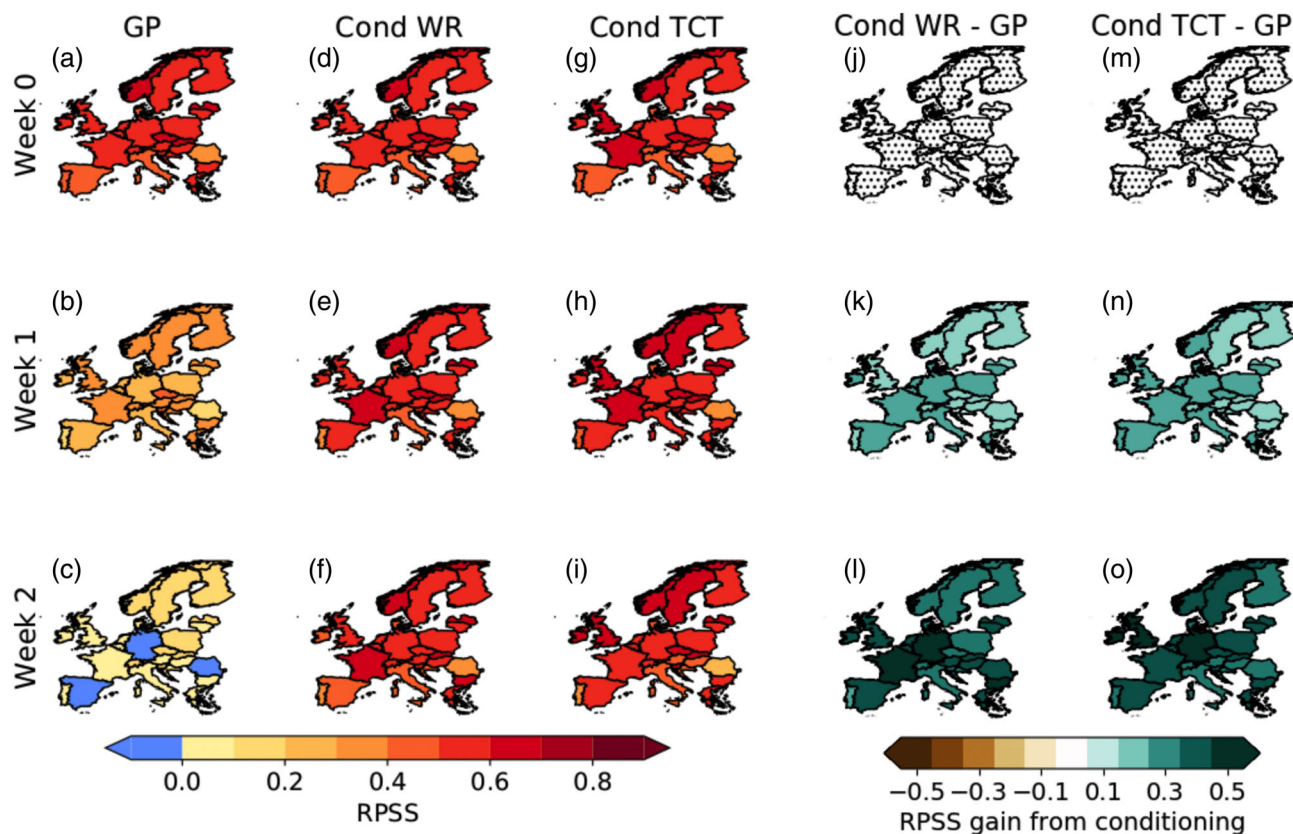
FIGURE 7 Weekly-mean RPSS of demand-net-wind (DNW) from December-February using the NCEP model (1999–2010). (a–c) Grid-point forecasts for weeks 0–2 (d–f) grid-point forecasts where greater than 50% of hindcast ensemble members agree on the dominant WR (g–i), as (d–f) for occasions when a dominant TCT is present in the hindcast. (j–l) Difference in skill between all grid-point forecasts and those conditioned on a dominant WR (m–o) difference in skill between all grid-point forecasts and those conditioned on a dominant TCT. Stippling in the fourth and fifth column shows no significant difference between the pairs of skill scores (2000 bootstrapped samples used to confirm significance). Tabulated skill scores are available in Appendix 3. NCEP, National Center for Environmental Prediction Climate Forecast System; RPSS, Ranked Probability Skill Score; TCTs, targeted circulation types; WRs, weather regimes

method of transferring from a TCT pattern to energy anomaly can be improved (see discussion in Section 4.2).

5 | DISCUSSION AND CONCLUSIONS

Sub-seasonal-to-seasonal (S2S) forecasts are becoming widely used within many impact sectors for decision-making (White et al., 2017). One of the sectors where the use of these forecasts is most sophisticated is the energy sector. Many users are currently using pattern-based techniques to predict future energy-system evolution, due to the encouraging research from the meteorological community showing examples of how models have good skill in predicting patterns (Büeler et al., 2020; Charlton-Perez et al., 2018; Ferranti et al., 2015; Ferranti et al., 2018; Matsueda & Palmer, 2018). However, a quantitative assessment of how pattern-based methods compare to using gridded model output to forecast national

energy variables has not been demonstrated for sub-seasonal timescales within the literature. This was the main aim of this study, comparing two sets of patterns to grid-point forecasts: Euro-Atlantic weather regimes (WRs; Cassou, 2008) and the impact-based targeted circulation type (TCT) method (Bloomfield, Brayshaw, & Charlton-Perez, 2020a).

The potential utility of both pattern-based methods for energy-forecasting was shown to be very promising using perfect forecast experiments, particularly for forecasting the ensemble-mean correlation (Section 3.2). Across all skill metrics, TCTs showed higher potential because of their closer relationship to energy-system behaviour (see Figures 2, 4 and 5). However, when implemented with S2S models, the pattern-based methods skill was lower than grid-point skill at medium-range lead times (week 0 and week 1). But, by weeks 2–3, improvements are seen by using WRs in multiple skill metrics (Figure 3). Central and Northern European countries (particularly, France, Sweden, Finland and the

United Kingdom) show the largest potential benefits from using WR forecasts, whereas for TCT forecasts, the benefits are largest in Central Europe (particularly France, Germany, Belgium and the Netherlands).

The TCT method performs better than WRs at short lead times (week 0 and week 1). However, it is never better than using the grid-point based forecasts. Reasons for this were explored and found to be due to issues with the models' ability to assign the patterns accurately (Figure 6). The skill found for WRs at longer lead times is associated with the window of increased predictability seen in S2S models from days 10–20.

The potential for windows of opportunity within the forecasts has been investigated, defined as times when over half of the ensemble members in a forecast agree on the pattern assignment (Figure 7). At short lead times (week 0), conditioning does not increase the forecast skill, due to low ensemble spread close to the forecast initiation. The skill present in the conditioned grid-point forecasts is significantly higher in week 2. This shows that the combination of pattern forecasts (giving information about the large-scale weather conditions) and grid-point forecasts (able to provide finer details) can be useful at particular times where increased information about the distribution of the forecast can be provided with higher certainty. The conditioning also allows a method to easily reject forecasts of high uncertainty about the large-scale flow conditions, which suggest high uncertainty in the surface energy impacts. This could be vital information for energy modellers taking decisions about plant operation/maintenance or scheduling for potential peak demands.

Interaction with energy users has suggested that more complex, probabilistic skill metrics are more useful for decision-making. However, the models often have low skill at extended lead times with these more complex metrics (see figures in Appendix 2). Here we show how conditioning using either WRs or TCTs can provide information on metrics that are more useful for decision-making at extended lead times, and help to combat the higher uncertainty of forecasts at these time horizons (Soares & Dessai, 2016). We note that the overall performance of the pattern-based methods is limited at present by the statistical link between the chosen pattern and the national energy data. Future developments of this methodology could provide enhanced skill, and potentially show pattern-based forecasts becoming increasingly beneficial. However, at very short lead times it will be challenging to beat the skill seen in grid-point forecasts.

The level of skill found in the grid-point S2S forecasts declines rapidly with lead time. This reduction in useful information for decision makers can be aided through the use of pattern forecasting techniques, particularly if the patterns are used to condition grid-point forecasts. We note that the skill assessment completed in this study is a lower bound for skill. For example, in the European Centre for Medium-Range Weather Forecasts (ECMWF) model, we have used an 11-member ensemble from the hindcasts, rather than the 51-member ensemble from the operational forecasts that are available. The skill assessment for the National Center for Environmental Prediction Climate Forecast System (NCEP) model could also be inhibited by the use of the ERA5 reanalysis. The methods shown in this study could be readily extended to other impact sectors with similar weather dependencies. Future work could continue to develop the presented windows of opportunity to include further meteorological conditioning, such as the state of the stratosphere, or the state of the Madden–Julian oscillation.

AUTHOR CONTRIBUTIONS
Hannah Bloomfield: Conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); software (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **David Brayshaw:** Conceptualization (equal); funding acquisition (lead); methodology (supporting); project administration (lead); supervision (lead); writing – original draft (supporting); writing – review and editing (equal). **Paula Gonzalez:** Conceptualization (equal); investigation (supporting); methodology (supporting); software (supporting); writing – original draft (supporting); writing – review and editing (equal). **Andrew Charlton-Perez:** Conceptualization (equal); data curation (supporting); funding acquisition (lead); methodology (supporting); project administration (lead); software (supporting); supervision (lead); writing – original draft (supporting); writing – review and editing (equal).

DATA AVAILABILITY STATEMENT
The matching hindcast and reanalysis pairs at 1.5° spatial resolution used in this study from 1996–2016 for ECMWF and 1999–2010 for NCEP are currently available at https://doi.org/10.17864/1947.275. Hourly, native resolution ERA5 based reanalysis time series of demand, wind power and solar power

from 1979–2019 are available at https://doi.org/10.17864/1947.272. The meteorological data used to create pattern forecasts and energy data in this study was originally downloaded from the S2S archive: https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/ and the ERA5 data from the climate data store https://cds.climate.copernicus.eu/

ORCID

Hannah C. Bloomfield https://orcid.org/0000-0002-5616-1503

David J. Brayshaw https://orcid.org/0000-0002-3927-4362

Paula L. M. Gonzalez https://orcid.org/0000-0003-0154-0087

Andrew Charlton-Perez https://orcid.org/0000-0001-8179-6220

REFERENCES

Baker, L.H., Shaffrey, L.C., Sutton, R.T., Weisheimer, A. & Scaife, A.A. (2018) An intercomparison of skill and overconfidence/under-confidence of the wintertime North Atlantic oscillation in multimodel seasonal forecasts. *Geophysical Research Letters*, 45(15), 7808–7817.

Beerli, R., Wernli, H. & Grams, C.M. (2017) Does the lower stratosphere provide predictability for month-ahead wind electricity generation in Europe? *Quarterly Journal of the Royal Meteorological Society*, 143(709), 3025–3036.

Bloomfield, H.C., Brayshaw, D.J. & Charlton-Perez, A.J. (2020a) Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. *Meteorological Applications*, 27(1), 1–18.

Bloomfield, H.C., Gonzalez, P.L.M., Lundquist, J.K., Stoop, L.P., Browell, J., Dargaville, R. et al. (2020b) The importance of weather and climate to energy systems: a workshop on next generation challenges in energy-climate modelling. *Bulletin of the American Meteorological Society*, 102(1), E159–E167.

Bloomfield, H.C., Brayshaw, D.J., Gonzalez, P.L.M. & Charlton-Perez, A.J (2021) Determining the bounds of skilful forecast range for probabilistic prediction of system-wide wind power generation. *Earth System Science Data*, 13(5), 2259–2274.

Brown, T.A. (1974) *Admissible scoring systems for continuous distributions*. Santa Monica, California: RAND Corporation.

Büeler, D., Beerli, R., Wernli, H. & 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*, 46(733), 3675–3694.

Cannon, D.J., Brayshaw, D.J., Methven, J., Coker, P.J. & Lenaghan, D. (2015) Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain. *Renewable Energy*, 75, 767–778.

Cannon, D., Brayshaw, D., Methven, J. & Drew, D. (2017) Determining the bounds of skilful forecast range for probabilistic prediction of system-wide wind power generation. *Meteorologische Zeitschrift*, 26(3), 239–252.

Cassou, C. (2008) Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic oscillation. *Nature*, 455(7212), 523–527.

Charlton-Perez, A.J., Ferranti, L. & Lee, R.W. (2018) The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1140–1151.

Clark, R.T., Bett, P.E., Thornton, H.E. & Scaife, A.A. (2017) Skilful seasonal predictions for the European energy industry. *Environmental Research Letters*, 12(2), 024002.

Cortesi, N., Torralba, V., González-Reviriego, N., Soret, A. & Doblas-Reyes, F.J. (2019) Characterization of European wind speed variability using weather regimes. *Climate Dynamics*, 53 (7–8), 4961–4976.

De Felice, M., Alessandri, A. & Catalano, F. (2015) Seasonal climate forecasts for medium-term electricity demand forecasting. *Applied Energy*, 137, 435–444.

Doblas-Reyes, F.J., Hagedorn, R. & Palmer, T.N. (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—ii. Calibration and combination. *Tellus A: Dynamic Meteorology and Oceanography*, 57(3), 234–252.

Dorrington, J., Finney, I., Palmer, T. & Weisheimer, A. (2020) *Beyond skill scores: exploring sub-seasonal forecast value through a case study of French month-ahead energy prediction.* arXiv preprint arXiv:2002.01728.

Drücke, J., Borsche, M., James, P., Kaspar, F., Pfeifroth, U., Ahrens, B. et al. (2020) Climatological analysis of solar and wind energy in Germany using the Grosswetterlagen classification. *Renewable Energy*, 164, 1254–1266.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N. et al. (2016) Skilful predictions of the winter North Atlantic oscillation one year ahead. *Nature Geoscience*, 9 (11), 809–814.

ENTSOE. (2018) *European network of transmission system operators for electricity data platform*. Available at: http://www.entsor.eu/data/Pages/default.aspx [Accessed 1st May 2018].

Epstein, E.S. (1969) A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6), 985–987.

Ferranti, L., Corti, S. & Janousek, M. (2015) Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, 141 (688), 916–924.

Ferranti, L., Magnusson, L., Vitart, F. & 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(715), 1788–1802.

Ferro, C.A.T., Richardson, D.S. & Weigel, A.P. (2008) On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, 15(1), 19–24.

Frame, T.H.A., Ambaum, M.H.P., Gray, S.L. & Methven, J. (2011) Ensemble prediction of transitions of the North Atlantic eddy-driven jet. *Quarterly Journal of the Royal Meteorological Society*, 137(658), 1288–1297.

Frame, T.H.A., Methven, J., Gray, S.L. & Ambaum, M.H.P. (2013) Flow-dependent predictability of the North Atlantic jet. *Geophysical Research Letters*, 40(10), 2411–2416.

Garrido-Perez, J.M., Ordóñez, C., Barriopedro, D., García-Herrera, R. & Paredes, D. (2020) Impact of weather regimes on wind power variability in western Europe. *Applied Energy*, 264, 114731.

Grams, C.M., Beerli, R., Pfenninger, S., Staffell, I. & Wernli, H. (2017) Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7(8), 557–562.

GWA. (2018) *The Global Wind Atlas*. Available at: https://globalwindatlas.info/ [Accessed March 1, 2019].

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.

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(672), 573–584.

Lee, R.W., Woolnough, S.J., Charlton-Perez, A.J. & Vitart, F. (2019) ENSO modulation of MJO teleconnections to the North Atlantic and Europe. *Geophysical Research Letters*, 46(22), 13535–13545.

Lledó, L., Torralba, V., Soret, A., Ramon, J. & Doblas-Reyes, F.J. (2019) Seasonal forecasts of wind power generation. *Renewable Energy*, 143, 91–100.

Lledó, L., Cionni, I., Torralba, V., Bretonnière, P.-A. & Samsó, M. (2020) Seasonal prediction of Euro-Atlantic teleconnections from multiple systems. *Environmental Research Letters*, 15(7), 074009.

Madonna, E., Li, C., Grams, C.M. & Woollings, T. (2017) The link between eddy-driven jet variability and weather regimes in the North Atlantic-European sector. *Quarterly Journal of the Royal Meteorological Society*, 143(708), 2960–2972.

Manrique-Suñén, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N. & Doblas-Reyes, F.J. (2020) Choices in the verification of S2S forecasts and their implications for climate services. *Monthly Weather Review*, 148(10), 3995–4008.

Mariotti, A., Baggett, C., Barnes, E.A., Becker, E., Butler, A., Collins, D.C. et al. (2020) Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, 101(5), E608–E625.

Matsueda, M. & 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(713), 1012–1027.

Michelangeli, P.-A., Vautard, R. & Legras, B. (1995) Weather regimes: recurrence and quasi stationarity. *Journal of the Atmospheric Sciences*, 52(8), 1237–1256.

Project S2S4E. (2020) *S2S4E winter outlook webinar and Euro-Atlantic teleconnections*. Attended: 23rd November 2020.

Robertson, A. & Vitart, F. (2018) *Sub-seasonal to seasonal prediction: the gap between weather and climate forecasting*. Amsterdam, Netherlands: Elsevier.

Robertson, A.W., Vigaud, N., Yuan, J. & Tippett, M.K. (2020) Toward identifying subseasonal forecasts of opportunity using north American weather regimes. *Monthly Weather Review*, 148(5), 1861–1875.

Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N. et al. (2014) Skillful long-range prediction of European and north American winters. *Geophysical Research Letters*, 41(7), 2514–2519.

Sharp, E., Dodds, P., Barrett, M. & Spataru, C. (2015) Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information. *Renewable Energy*, 77, 527–538.

Soares, M.B. & Dessai, S. (2016) Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. *Climatic Change*, 137(1–2), 89–103.

Thornton, H.E., Scaife, A.A., Hoskins, B.J. & Brayshaw, D.J. (2017) The relationship between wind power, electricity demand and winter weather patterns in Great Britain. *Environmental Research Letters*, 12(6), 064017.

Thornton, H.E., Scaife, A.A., Hoskins, B.J., Brayshaw, D.J., Smith, D.M., Dunstone, N. et al. (2019) Skilful seasonal prediction of winter gas demand. *Environmental Research Letters*, 14(2), 024009.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C. et al. (2017) The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98(1), 163–173.

White, C.J., Carlsen, H., Robertson, A.W., Klein, R.J.T., Lazo, J.K., Kumar, A. et al. (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, 24(3), 315–325.

van der Wiel, K., Bloomfield, H.C., Lee, R.W., Stoop, L.P., Blackport, R., Screen, J.A. et al. (2019) The influence of weather regimes on European renewable energy production and demand. *Environmental Research Letters*, 14(9), 094010.

Wilks, D.S. (2011) *Statistical methods in the atmospheric sciences*, Vol. 100. Oxford, United Kingdom: Academic Press.

Woollings, T., Hannachi, A. & Hoskins, B. (2010) Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, 136(649), 856–868.

SUPPORTING INFORMATION

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