ARTICLE OPEN

How do North American weather regimes drive wind energy at the sub-seasonal to seasonal timescales?

Ye Liu [1✉], Sha Feng [1✉], Yun Qian[1], Huilin Huang[1] and Larry K. Berg[1]

There has been an increasing need for forecasting power generation at the subseasonal to seasonal (S2S) timescales to support the operation, management, and planning of the wind-energy system. At the S2S timescales, atmospheric variability is largely related to recurrent and persistent weather patterns, referred to as weather regimes (WRs). In this study, we identify four WRs that influence wind resources over North America using a universal two-stage procedure approach. These WRs are responsible for large-scale wind and power production anomalies over the CONUS at the S2S timescales. The WR-based reconstruction explains up to 40% of the monthly variance of power production over the western United States, and the explanatory power of WRs generally increases with the increase of timescales. The identified relationship between WRs and power production reveals the potential and limitations of the regional WR-based wind resource assessment over different regions of the CONUS across multiple timescales.

INTRODUCTION

A tremendous effort to decarbonize the energy sector is underway with the massive installation of renewable energy capacity in the United States, in which wind energy with 136 GW wind power capacity accounts for 9.1% of total electricity generation and 32% of capacity additions in 2021[1]. In 2021, the Biden administration announced a goal of 30 GW offshore wind generation by 2030 that will unlock a pathway to 110 GW by 2050[2]. The growth of wind power generation increases the exposure of the power systems to variable weather and climate, which requires accurate forecasts at multiple timescales for many actions in the operation of the grid[3,4]. Forecasting the wind power generation over time periods ranging from hours to several days ahead has had tremendous improvement[5], while the skill of forecasts beyond 2 weeks remains poor. In recent years, there has been an increasing need for forecasting power generation at the subseasonal to seasonal (S2S, 2 weeks to one season) timescales to support the operation, management, and planning of the wind-energy system[6–8]. Knowledge of the fluctuations of power production at the S2S timescales can also help guide deployment pathways that balance power generation from wind and other renewables[6].

The skillful forecasting lead time of a predicted phenomenon largely depends on its spatial scale. The surface wind speed normally has a spatial scale of less than 1000 km with a skillful forecast less than 10 days, while large-scale and low-frequency atmospheric circulations extending up to 10,000 km have forecasting horizons of beyond 2 weeks[9]. As a result, the global climate models have better performance in simulating and projecting large-scale atmospheric circulations than the local surface wind speed. These atmospheric circulations can be classified into a set of recurrent and quasi-stationary (persistent) patterns, commonly referred to as weather regimes (WRs)[10–12]. Building a solid WR and wind resource relationship could potentially help improve the model skill in simulating and predicting the hub-height wind speed at longer timescales, such as the S2S timescales. Resulting from interactions between synoptic-scale and planetary-scale atmospheric waves[13,14], WRs can persist for 1–2 weeks, which is beyond the lifetimes of individual weather disturbances[15]. The persistence and/or transitioning of WRs are often

used in S2S variability and predictability studies[16]. Meanwhile, the interactions of WRs and atmospheric teleconnections, such as the Pacific-North American (PNA) teleconnection pattern and North Atlantic Oscillation (NAO) for North America, suggest additional predictability[16,17]. Analysis based on NCEP CFSv2 showed that the climate model can predict the WR up to 30 days ahead at certain geographic location[18], depending on WR's persistence and interactions with teleconnections. Many studies demonstrated an improvement of the forecasting skill by linking the variability of WRs with local weather and extreme events[19,20] and found that the changes of WRs often determine most variations of surface variables, such as temperature, precipitation, and surface winds, at S2S timescales. For these reasons, a growing number of studies have used WRs for power production assessment in response to the emerging needs of energy management[3,6,7,21,22] in the Europe-Atlantic sector and India with a focus on cold seasons. However, as far as we know, no such studies have focused on wind resources (wind speed and power production) in warm seasons, or, what is worse, none for the United States sector.

The United States experiences complex weather patterns, resulting in various responses to WRs with distinct seasonal variability at the S2S timescales. In this study, we provide the assessment of the relationship of WRs and the variability of local wind resources for different seasons over the CONUS. We aim to address two questions: (1) How do the WRs drive hub-height wind speed and power production at the S2S timescales over the CONUS? (2) To what degree can WRs explain the variability of hub-height wind and power production? The results associated with the second question are organized based on the geographic spans of U.S. isolated system operators (ISOs, Supplementary Fig. 1) to directly benefit the wind-energy industry. It is worth mentioning that, compared to the existing studies, this work quantifies the impact of the WRs on the variability of local wind resources.

Previous studies have mainly used K-means clustering to identify the WRs that impact a given region of interest[12,15,18,23]. However, this approach fails to preserve the topological relationship between the WRs, which is particularly beneficial in reconstructing the topological surface in the WR space. In this

[1]Pacific Northwest National Laboratory, Richland, WA 99352, USA. ✉email: ye.liu@pnnl.gov; sfeng@pnnl.gov
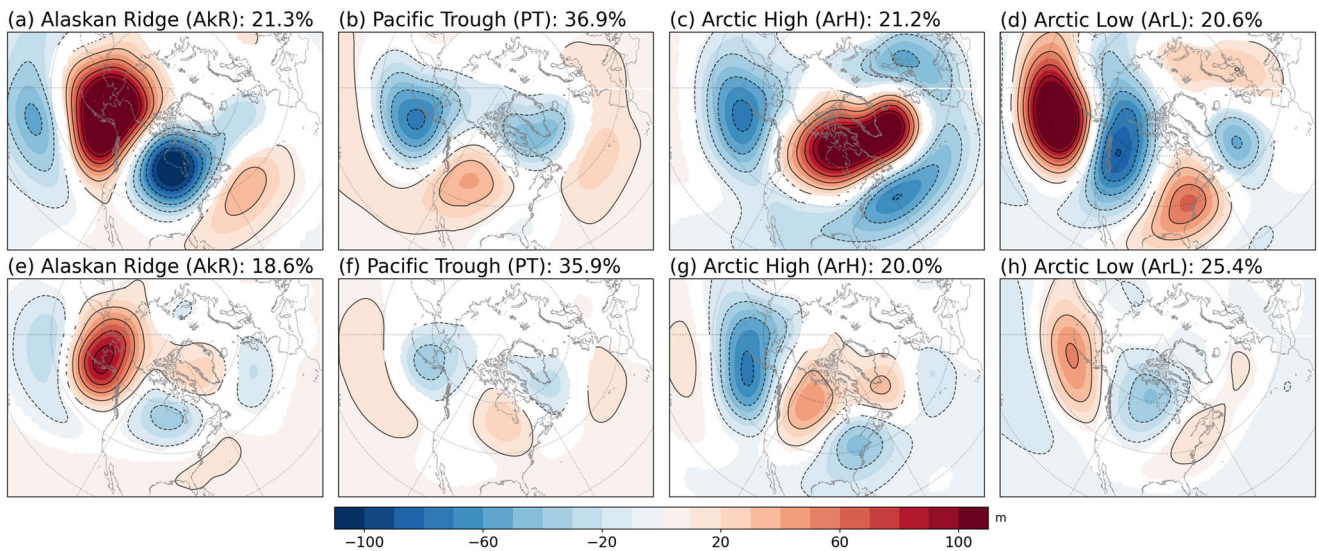
Fig. 1 Daily ERA5 Z500 mean anomalies (m) with respect to the daily climatology for each regime over 1981–2020. a–d cold (October–March) and **e–h** warm (April–September) seasons. Only significant anomalies at the 5% level according to a two-sided bootstrap resampling test are shown. The numbers at the top-right of each panel indicate the frequency of the regimes.

study, we have adapted a universal two-stage procedure[24] that incorporates both SOM and K-means to replicate the North American WRs identified in the previous studies while maintaining the topological relationship for reconstruction. SOM is a widely used clustering analysis[11,25–29] that performs a topology-preserving mapping. However, when the number of nodes is small, SOM tends to cluster the input vectors into symmetrically paired nodes, whereas those from K-means are independent of each other. As a result, this two-stage procedure approach combines the strengths of both SOM and K-means while addressing their individual shortcomings[25].

RESULTS

The dominant weather regimes over North America

We apply a two-stage clustering procedure to the daily 500-hPa geopotential height (Z500) anomaly obtained from ERA5 to cluster weather patterns in the cold (October to March) and warm (April to September) seasons from 1981 through 2020. In cold seasons, four WRs are identified. They are Alaskan ridge (AkR), Pacific trough (PT), Arctic high (ArH), and Arctic low (ArL), consistent with the findings in earlier studies[10,15,23]. The AkR, PT, and ArL consist of meridionally oriented ridge and trough anomalies that resemble Rossby wave trains, which have a similar structure to the intermediate 10–30-day timescale wave identified in ref. [30]. The AkR regime accounts for 21.3% of the days, featuring an anomalous ridge centered near Alaska (Fig. 1a). The PT regime (Fig. 1b) features an anomalous trough centered near the Aleutians. The PT regime occurs more frequently (36.9%) than the AkR regime. The ArH regime (Fig. 1c) occurs on 21.2% of the days and is associated with a strong anomalous high centered over Canada and near Greenland, coinciding with an anomalous low over the western North Atlantic Ocean, which leads to an enhanced meridional pressure gradient. The circulation anomaly of the ArH regime is similar to the negative phase of the NAO (NAO–). The Arctic low (ArL, Fig. 1d) regime, accounting for 20.6% of the days, is associated with the negative phase of PNA (PNA–). The PT and ArH regimes are associated with a trough-ridge pattern that extends from the Pacific to the continent, resembling the midlatitude atmospheric pattern during El Niño episodes[31], while the Pacific components of the ArL regime bear similarities to the atmospheric pattern during La Niña episodes. The four WRs (AkR, PT, ArH, and ArL) have the pattern correlations of 0.10, 0.54,

−0.72, 0.18 with the NAO and −0.40, 0.57, 0.52, −0.71 with the PNA (computed using the monthly mean of Z500).

The warm-season WRs are less studied as the PNA- and NAO-induced anomalies are not as pronounced as those in boreal winter[32]. The warm-season WRs generally resemble the four regimes identified in the cold season but with smaller magnitudes (Fig. 1e–h). We find that 18.6% and 35.9% of the days in warm seasons are governed by the AR and PT regimes, respectively, while AH and AL regimes occur at 20.0% and 25.4%, respectively.

Persistence and transitioning of WRs

The daily weather can be viewed as synoptical-scale perturbations on the quasi-stationary weather regimes that persist longer than individual weather disturbances[33]. On the other hand, the transitioning of a WR is frequently responsible for abrupt changes in wind speed (i.e., wind ramps) in addition to local factors. We investigate the evolution of a WR by selecting the days associated with the WR and counting the frequency of each WR that appears on the following day. Such frequencies reflect the likelihood, measured as a percentage of days of a WR either persists within the same WR or transitioning into a different WR. As a result, over 80% of the days in each regime in cold seasons persist to the following day, while the remaining days change to other regimes (Fig. 2). Because of the west-to-east propagation of the wave trains, 10.8% of AkR days transition to PT. The PT and ArH share a similar trough-ridge pattern over the eastern Pacific. 11.7% of ArH days and 7.3% of PT days evolve into each other on the following day. The AkR is least likely to transit to ArH due to their distinct patterns (only 1.8%). The warm-season WRs show 72.6–76.3% of persistency, weaker than in cold seasons. More frequent transitioning occurs between the PT regimes and the others. Less than 5% of the days transit from ArH to ArH or to AkR, and vice versa.

Another property of the persistence of a WR is the duration. We quantify this property by counting the consecutive days that are identified as a WR before transitioning to another. In cold seasons, the mean durations of AkR and PT regimes are 9.3 and 11.8 days, respectively, while the durations of ArH and ArL regimes are 10.9 and 15.3 days, respectively (Fig. 3a). The most prolonged duration is found in PT regime (78 days), which relates to the strong El Niño in 1982–1983 winter. Mean durations in warm seasons are generally shorter than in cold seasons (Fig. 3b) due to weaker Z500 anomaly and stronger interaction with local thermal factors. AkR and PT
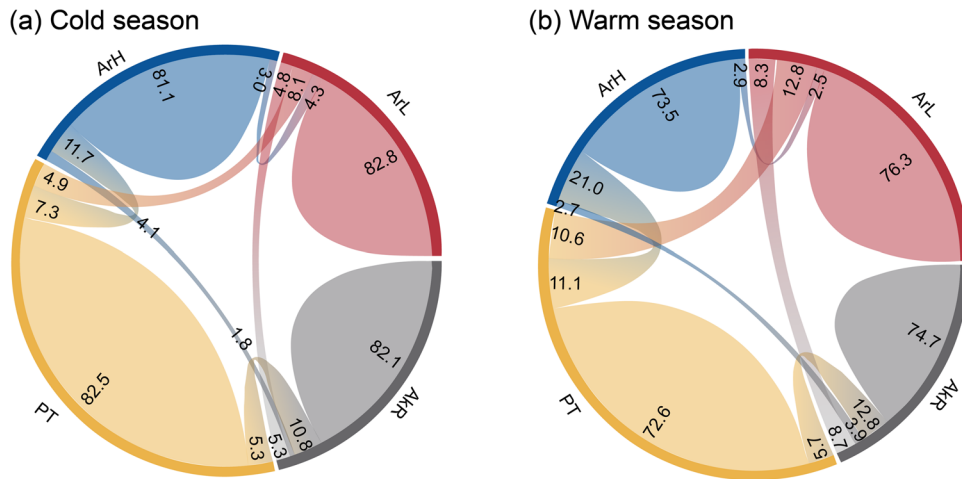
Fig. 2 Progressing between WRs. a Cold seasons and **b** warm seasons. The numbers indicate the percentage of a WR persisting or progressing to another on the following day.
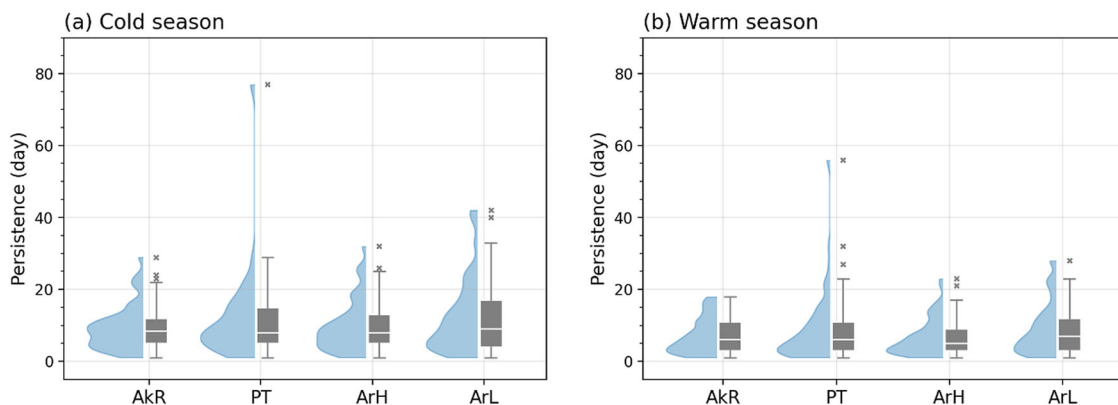
Fig. 3 Duration of each regime. a Cold seasons (October–March) and **b** warm seasons (April–September). The shaded curve shows the probability distribution function. The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The central line indicates the median. The whiskers extend from the box by 1.5× the interquartile range. The outliers are plotted individually as "×".

regimes persist for 7.1 and 8.6 days on average, while the ArH and ArL regimes persist for 6.7 and 9.0 days, respectively.

Impact of WRs on hub-height wind speed

The climatology of the hub-height wind speed over the CONUS is obtained by averaging the ERA5 100-m wind over the 40-year period. Climatologically, larger wind speeds (annual mean wind speed $>7.5\,\mathrm{m\,s^{-1}}$) mainly occur over the Great Plains and the Midwestern United States and become weaker near the west and southeast coasts (Supplementary Fig. 2).

The WRs' impact on wind speed is represented by the mean of wind speed anomaly associated with each WR with respect to the climatology of the daily mean. In general, the WR-induced wind anomalies are stronger in cold seasons than in warm seasons (Fig. 4), corresponding to more intense circulation anomalies (Fig. 1). The cold-season AkR regime is associated with a blocking pattern over the northeastern Pacific that significantly slows down the large-scale wind speed west of the Rocky Mountains and the Great Plains but increases wind speed over the eastern slope of the Rocky Mountains and Midwestern U.S. (Fig. 4a). In contrast, the PT regime is associated with the decreases in wind speeds across the CONUS (except for the north Great Plains and the east coast), which is associated with a high-pressure center dominating North American (Fig. 1b). The ArH regime is associated with an anomalous high-pressure center over the central North American land, surrounded by anomalous lows

over the ocean (Fig. 1c), which leads to smaller wind anomalies over $0.4\,\mathrm{m\,s^{-1}}$ (Fig. 4c). The ArL regime is associated with dipole anomalies of Z500 over land with a low in the west and a high in the east, resulting in an increase of wind speed over most of the CONUS, except for the East Coast (Fig. 4d). The WRs lead to over $\pm1\,\mathrm{m\,s^{-1}}$ wind speed anomalies over the eastern slope of the Rocky Mountains, Great Plains, and Columbia River Basin which hosts many wind farms[34]. The significant changes in the wind anomalies introduce changes in power production.

The influence of WRs on hub-height wind speed decreases considerably in warm seasons due to small Z500 anomalies (Fig. 1) and the weak linkage between large-scale circulations and local processes[35]. The wind speed associated with AkR increases east of the Rocky Mountains, southern Great Plains, and the northeastern U.S. (Fig. 4e). The PT induces a reduction of wind speeds over the Midwest U.S., but insignificant changes are found over the rest of the CONUS (Fig. 4f). The ArH regimes is associated with a decrease of wind speed over the CONUS except for the east coast (Fig. 4g). In contrast, the ArL regime leads to an increase of wind speed over the Great Plains and western U.S., while decreasing wind speed along the east coast (Fig. 4h).

Relationship between WRs and wind speed

To quantify the impact of WRs on local wind resources and further investigate to what degree the WRs can explain the variability of
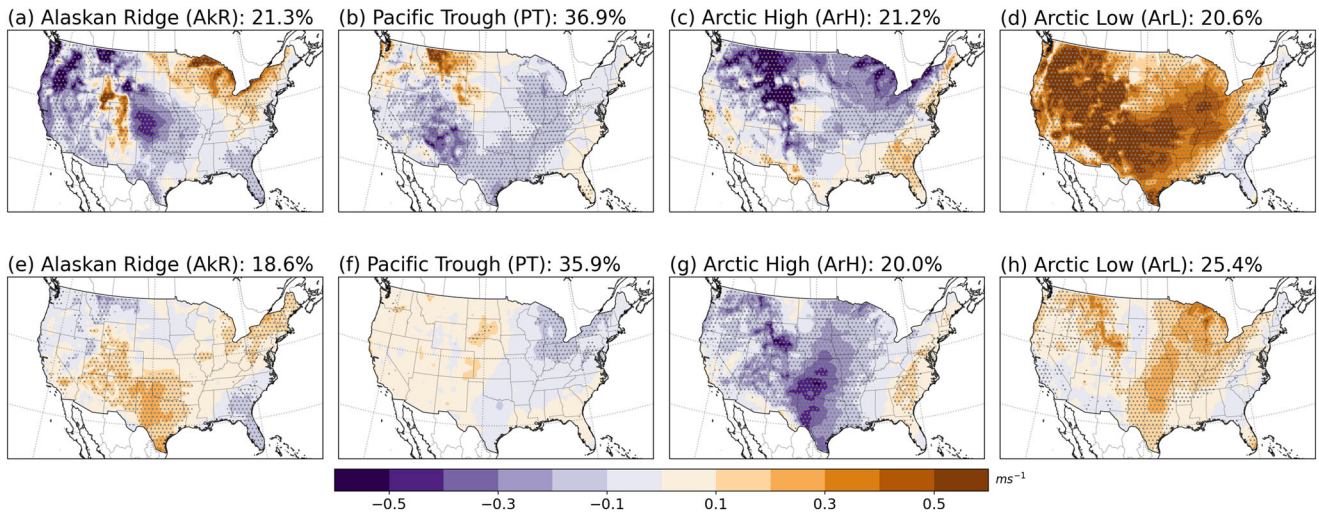
Fig. 4 The same as Fig. 1, but for daily 100-m wind speed mean anomalies (ms^{-1}). a–d Cold (October–March) and **e–h** warm (April–September) seasons. Dots indicate the statistical significance of the wind speed anomalies at the 5% significant level according to a two-side bootstrap sampling test.
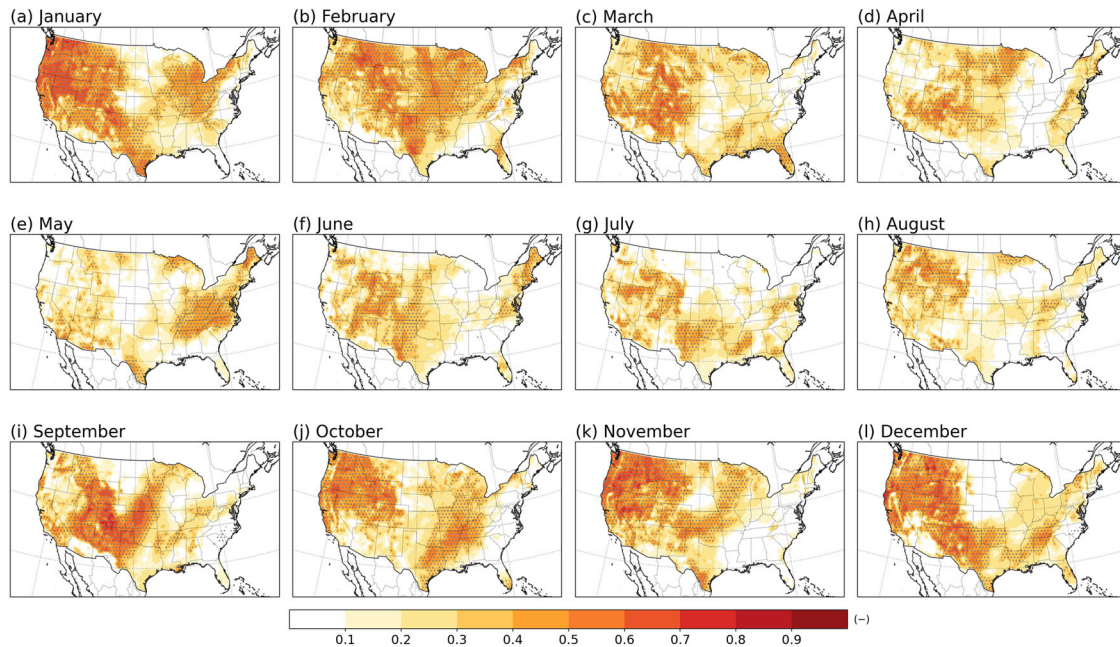

Fig. 5 Correlation coefficient between the WR-related and ERA5 100-m wind speed for each month. a–l Correlations for each month. Dots indicate the statistical significance of the correlation at the 5% significance level according to the Spearman significance test.

winds and power production, we reconstruct the WR-related hub-height wind speed over different timescales using the prototypes from the first-step SOM training (see "Methods"). Note that as S2S timescales span from 2 weeks to one season, here we choose to present the results at the monthly scale as an example to demonstrate the spatial distributions of the properties of WR-related wind speed (the aggregated results across multiple time-scales are organized in next section). These properties include the temporal correlation (Fig. 5) and standard deviation ratio (Fig. 6) of reconstructed wind speed as opposed to the original ERA5 monthly mean, representing the timing and magnitude of the variabilities, respectively.

Besides the large-scale weather patterns, wind speed is affected by topography, small-scale atmospheric and surface perturbations, and other factors. We hypothesize that WRs have higher explanatory power for regions and periods in which large-scale circulations play

critical roles in local weather, namely holding significant correlation and/or standard deviation ratio (dotted areas in Figs. 5 and 6). Here we define the explanatory power of WRs as the percentage of the variance of WR-related wind speed/power production in the total variance of actual wind/power production, namely R^2, and will be fully discussed in the next section.

In general, the correlations between the WR-related wind and actual wind are larger in cold seasons than in warm seasons, as the local and small-scale perturbations (mostly due to thermal effects) are more influential in warm seasons. From November to January, the western U.S. shows a correlation coefficient over 0.4 and a standard deviation ratio above 0.7 between WR-related and actual winds, suggesting that the large-scale circulation is the main driver of local wind variability. Low correlations and small standard deviation ratios over the eastern United States in December are likely due to the weakening impact of NAO on local wind[36]. The NAO-associated
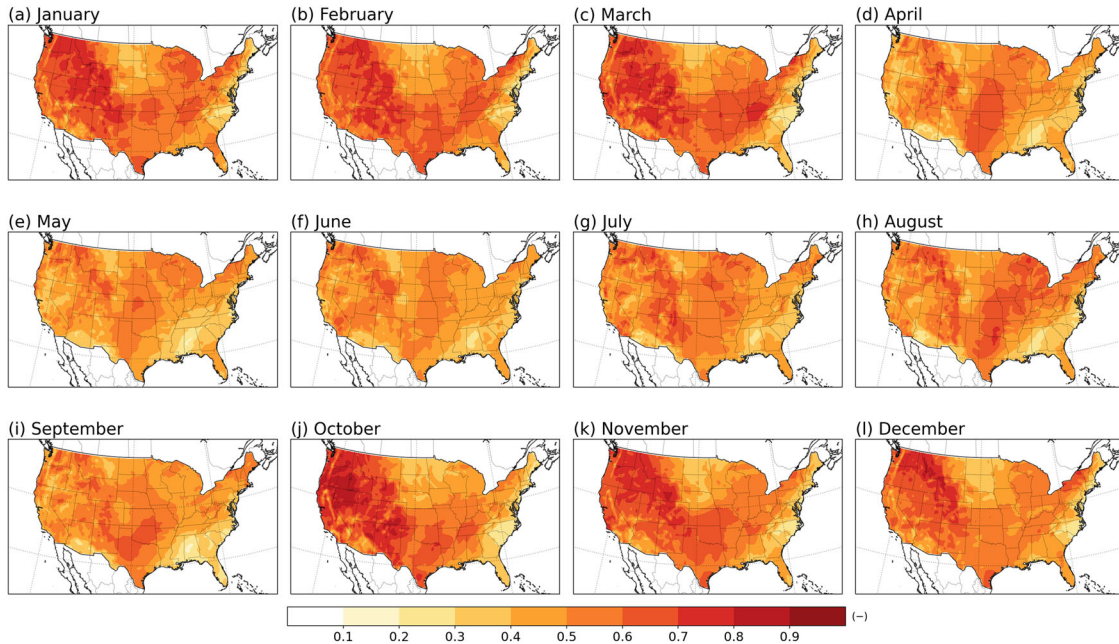
Fig. 6 Standard deviation ratio between the WR-related and ERA5 100-m wind speed. a–l Standard deviation ratio for each month.
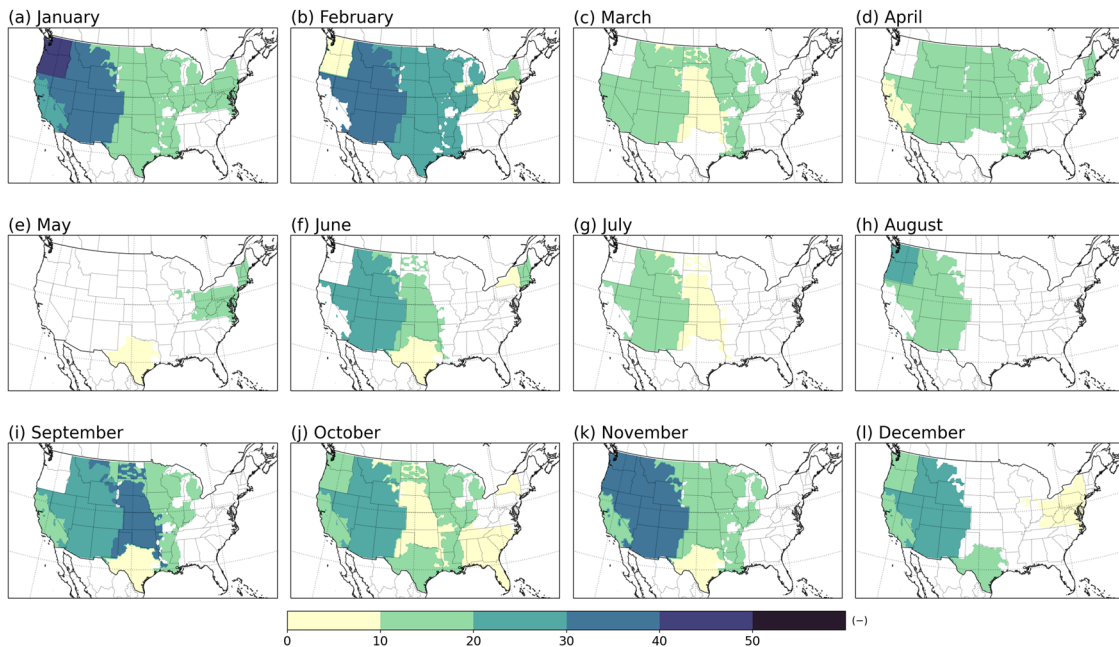
Fig. 7 Explanatory power (%) of WRs to regional power production. a–l Explanatory power for each month.

surface pressure anomalies located over the Atlantic Ocean in December (not shown) and far off the U.S. east coast result in a weak impact on local winds. Significant correlations and high standard deviation ratios are also found in the western U.S., Great Plains, and Midwestern U.S. in many months of warm seasons.

Impact of WRs on power production

We obtain WR-related power production by scaling WR-related wind speed using the International Electrotechnical Commission (ICE) class 2 power curve (see "Methods") and calculate the explanatory power of WRs to the total power production to assess how much the variation of power production can be explained by

WRs. Note that the explanatory power is defined as R^2 between WR-related and ERA5 wind-translated power productions. We aggregate the explanatory power into the subregions, including seven ISOs, two non-ISOs, and the Department of Energy's Wind Forecast Improvement Project II (WFIP2[37], also part of Bonneville Power Administration service area) area (Supplementary Fig. 1)[38].

On a monthly basis, the explanatory power of WRs for power production (Fig. 7) is distributed similarly to correlation patterns of wind speed (Fig. 5), as regional power production mainly depends on local wind speed. The differences between the explanatory power for wind speed and power production are mainly due to the non-linear power curve and the impact of the cut-in and cut-off wind speed of turbines. As we see in Fig. 7, the largest
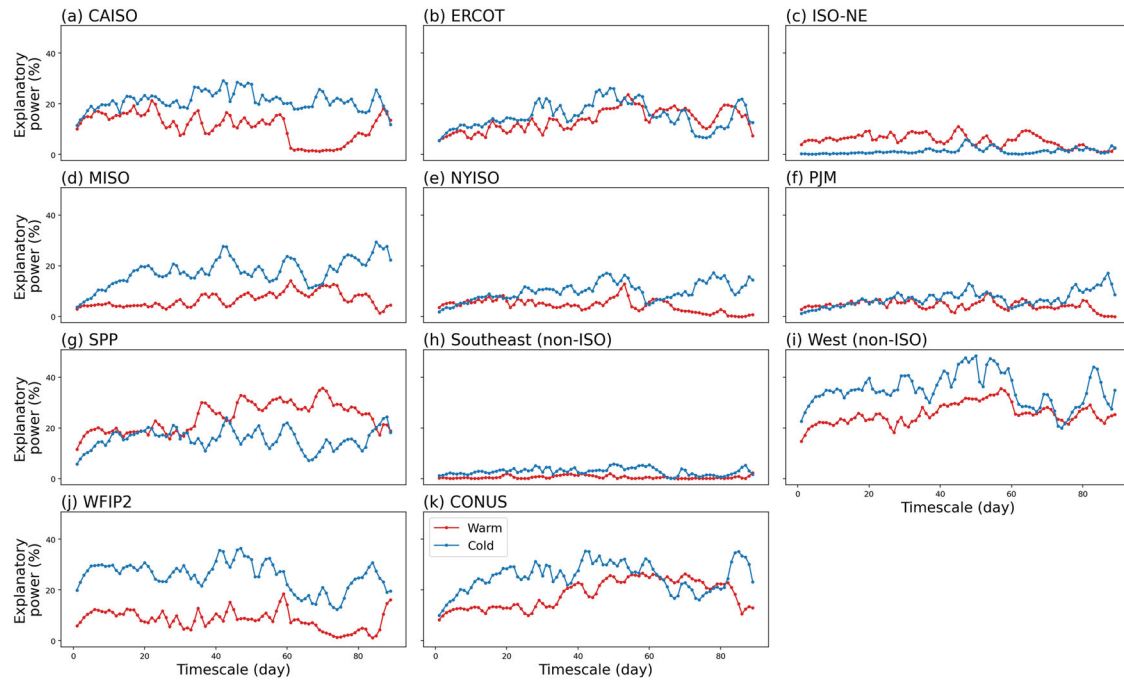
Fig. 8 Explanatory power (%) of WRs to regional power production as a function of averaging window. a–k Explanatory power for each sub-region.

explanatory power is found in January, more than 40% of the variability in the WFIP2 study region and followed by 30% in the Northwest and Southwest (Non-ISO) regions. WRs have explanatory power ranging over 10–40% in months except for May over the western U.S., including CAISO, Northwest, and Southwest, while ranging from 10 to 30% in January, February, April, June, September, and November over the central U.S., including SPP and ERCOT. Relatively low explanatory power is found in the Midwestern U.S. and the east coast.

We further examine how the explanatory power of WRs to the regional power production changes with the timescales and organize the results in Fig. 8. In general, the explanatory power increases with the increase of timescales (averaging time window). Over the CONUS, the explanatory power of cold-season WRs increases rapidly from one day to 2 weeks and then becomes stable from 2 weeks to one month, suggesting that the WRs are one of the main factors driving power production variability at these timescales (Fig. 8k). The explanatory power shows more fluctuations beyond one month, which is likely caused by other low-frequency atmospheric phenomena, such as ENSO. Similar patterns are found for CAISO, ERCOT, MISO, West (non-ISO), and WFIP2 regions with one exception of the WIFP2 region where the explanatory power of WRs decreases from monthly to seasonal timescales. In contrast, the warm-season explanatory power remains low from one to 50 days timescales (Fig. 8k), while it rapidly increases from 50 to 90 days, indicating that WRs have a stronger impact at the seasonal timescales for warm seasons. This pattern is mainly contributed by CAISO, ERCOT, SPP, Southeast, and West non-ISOs (Fig. 8a, b, g, i).

DISCUSSION

The relationship of the WR and other well-known low-frequency oscillations and the application and limitation of the WR-wind relationship are discussed in this section. The WRs, resulting from the interaction between planetary and synoptical-scale atmospheric waves, have been found tele-connected to tropical sea surface temperature forcing at weeks to months lead time[15,38,39], providing additional predictability. We further test the relationship

between North American WRs and Madden-Julian Oscillation (MJO) and find an increasing frequency of ArH's occurrence 1 week after MJO phase 6, ArL 5–15 days after MJO phase 3, and AkR 4 weeks after MJO phase 3 (Supplementary Fig. 3), which is consistent with the findings of ref. [15]. Phase 3 and phase 6 of MJO feature strong dipole anomalies in tropical diabatic heating with convection anomalous centers of opposite signs in the eastern Indian Ocean and western Pacific Ocean, which propagates to the North Pacific and North America through Rossby wave trains. The physical connection between the MJO and WRs with weeks of lead time indicates the source of predictability. However, since each MJO phase persists for about 1 week, it is difficult to distinguish the impact of a MJO phase from the previous one[39]. Meanwhile, the influence of MJO can interfering constructively with other low-frequency atmospheric oscillations such as ENSO, therefore posing challenges in terms of isolating the true MJO effect. We recognize the additional predictability of MJO and will investigate its impact on regional wind forecasts in future work.

The dependence of local wind resources on WRs indicates the regions where power production is sensitive to future climate change. A strengthened PNA pattern[40] and increasing frequency and persistence of NAO+[41] are projected by CMIP5 models in future warming scenarios, suggesting a growing impact of WRs on power production in the future. Co-deployment of other renewables could balance the WRs impact on national electricity[6], and further study is encouraged.

The WR-wind relationship obtained in this study can help improve the forecasting skill of dynamical and statistical models[42]. Climate models are expected to have better performance in predicting large-scale atmospheric circulations than local wind. Unrealistically representing the WR–local wind relationship can lower the skill of hub-height wind forecast even with a skillful forecast of weather patterns. Therefore, the relationship identified in this study is a benchmark for evaluating the climate model performance, although the statistical relationship cannot replace dynamic model predictions. On the other hand, the S2S models use very coarse resolutions, which can introduce large biases in modeled winds as well as other variables. Therefore, statistical post-processing such as calibration, downscaling, or bridging are

(a) Flowchart of SOM algorithm
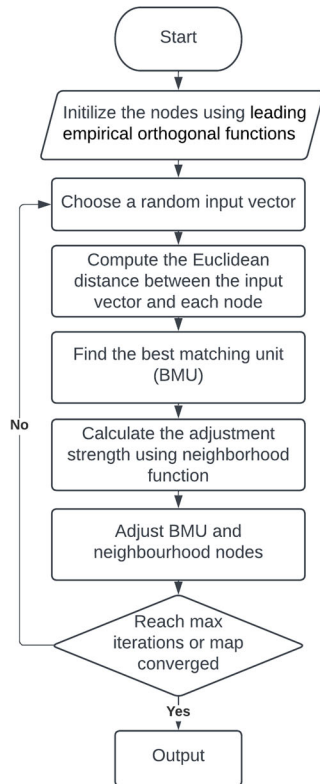
(b) Concept diagram of SOM

Fig. 9 Schematic diagram of SOM analysis. a Flowchart of SOM algorithm and **b** concept diagram of SOM.

needed to translate atmospheric variables to energy properties[8]. Particularly, bridging methods allow transforming forecasts of the state of the climate into another variable of interest over well-defined large areas, if a robust linkage between them is found in observational records.

In practice, the application of the WR-wind relationship to S2S forecasts relies on skillful WR forecasts in climate models. Current state-of-art climate models still have limited forecast skills for weather patterns at the S2S timescales, although better than for surface variables. The current models underestimate the pre-dictable signals (the predictable fraction of the total variability) of the climate variability by an order of magnitude[17]. To improve the S2S forecasting skill, many programs have been initiated, such as the subseasonal to seasonal project by the World Weather Research Program (WWRP) and the World Climate Research Program (WCRP)[42] and the S2S for clean energy (S2S4E) funded by the European Union H2020 Framework Programme for Research and Innovation (https://s2s4e.eu/). The horizon of forecasting skill for large-scale weather patterns has been significantly extended in the past decades. However, translating the predictability to the renewable energy sector is still at an early stage. The robust WR-wind linkage suggests a potential way to improve the S2S forecasting skill for wind resources and to better understand forecast uncertainty. Besides the limitation due to the performance of current climate models, the WR-wind relationship varies with location and season. For example, WRs are closely connected to local wind over the western and central U.S. but hardly explain any of the variability of wind speed along the U.S. east coast.

We provide the assessment of WR-related variability of hub-height wind speed and power production over CONUS across multiple timescales. The WRs are determined by daily ERA5 500-hPa geopotential height (Z500) anomalies over a period of 40

years (1981–2020) using a two-stage clustering procedure method that combines SOM and K-means. Four WRs are identified over North America: The Alaskan ridge (AkR), Pacific trough (PT), Arctic high (ArH), and Arctic low (ArL). These WRs resemble the North American atmospheric circulations representing the PNA, NAO, and ENSO teleconnections. These WRs are found to persist 1–2 weeks before transitioning to another and have a larger impact on wind resources in cold seasons than in warm seasons and over the western and central U.S. than over the eastern U.S. Accordingly, the explanatory power of WRs to power production is significantly larger in the western and central U.S. The driving effect of WRs on local wind resources is evidenced by the significant correlation between WR-related and local wind speed/power production. WRs have larger explanatory power in the western and central U.S., including West (non-ISO), CAISO, SPP, ERCOT, and MISO than other subregions. These regions together supply over 80% of the total capacity by the end of 2020[1]. The linkage between WRs and local wind resources at the S2S timescales suggests a potential source of predictability that benefits wind power management and planning, especially in the western United States. This study focuses on land-based wind resources as the ISOs and other subregions provide geographic areas that serve to aggregate the wind power production regionally. The knowledge gained in this study is transferable to offshore wind resource assessment.

METHODS

Two-stage clustering procedure

We apply a universal two-stage procedure[24] to ERA5 Z500 anomalies to determine the WRs that influence North America. In the first stage, we train SOM to generate a low-dimensional discretized representation of the data (16 prototypes) in the
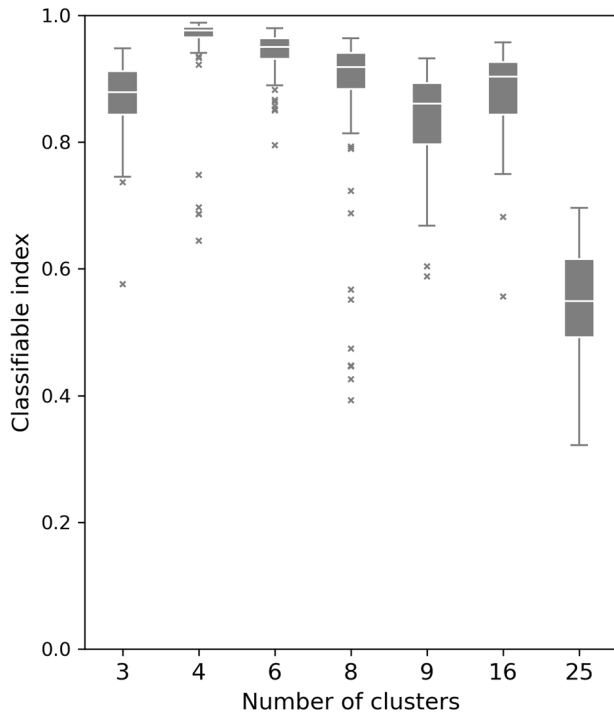
Fig. 10 Classifiable index as a function of the number of regimes. The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The central line indicates the median. The whiskers extend from the box by 1.5× the interquartile range. The outliers are plotted individually as "×".

original feature space while preserving the topological properties of the data. SOM is an artificial neural network commonly employed for clustering analyses, which projects the high-dimensional data to a visually comprehensible two-dimensional map. In the second stage, we use the 16 SOM-generated prototypes as the input of K-means for clustering. The second stage is applied to reduce the number of named regimes and keep them consistent with those found in the previous studies[12,15,18]. Three benefits are associated with this two-stage procedure: first, using the SOM-generated prototypes substantially reduces the number of dimensions of the variable that is taken by K-means clustering so reduces the computational time; second, it largely eliminates outliers as the prototypes constructed by SOM are considered as local averages of the data so improves the efficiency and accuracy of K-means clustering; third, the topological preserving prototypes can be directly used to reconstruct WRs-related wind resources to serve the later analysis. Directly using K-means for clustering is not recommended, as K-means is highly sensitive to the initial positions of the centroids and outliers[43] and not suitable for high-dimensional dataset[44].

Before applying the two-stage procedure, we calculate the Z500 anomaly by subtracting the climatological daily mean from the daily Z500 at each grid point over the 1981–2020 annual cycle. Some previous studies also filtered the Z500 anomalies by performing a principal component analysis (PCA) to reduce the dimension of the input vector[15] for K-means clustering. Then the Z500 anomalies are weighted by the cosine of the latitude to account for reprehensive areas. With the two-stage procedure, the first-step SOM training completes the dimension reduction and outlier elimination for the second step K-means clustering with an additional benefit that preserves the topology of the data points in contrast to simply linear approaches (e.g., PCA).

The workflow of SOM is summarized in Fig. 9a. During the SOM training stage, the initial nodes for SOM are determined by the leading empirical orthogonal functions (EOFs). Then we add input

vector data (daily Z500 anomaly; red dot in the example shown in Fig. 9) to the map of SOM-generated prototypes and find the best matching unit (BMU). The BMU is the prototype that has the smallest Euclidean distance to the input vector. The BMU and neighbor nodes are adjusted towards the input vector to better represent the data distribution (Fig. 9b). A neighborhood function is applied to determine the number of neighborhood nodes to be adjusted and the strength of adaption, depending on the order number of the current iteration and distance between the neighborhood node and BMU.

In this study, the SOM analysis is performed over 10°–70°N, 150°–40°W, which is chosen to include the Pacific jet exit region and North Atlantic variability[12,15]. The Z500 anomalies in October–March and April–September derived from ERA5[45] (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5) are used to identify the dominant WRs in the cold and warm seasons, respectively. The impact of a WR on the wind speed is measured by averaging the wind speed anomalies associated with the given WR with respect to the climatology of daily mean over the 1981–2020 annual cycle over all days.

The choice of the number of prototypes (k) is arbitrary, so we test the values of k that range in 3, 4, 6, 8, 9, 16, and 25. For each k, the robustness of the regime clusters is measured by a classifiability index (CI) following previous studies[15,46,47]. The maximum spatial correlation coefficient between the clusters obtained from the full dataset and many random halves (100 halves in this study) of the data is calculated to construct the CI[47]. Therefore, the CI measured the reproductivity of the k partitioning[33], with perfect partitioning leading to a CI equals to 1. Figure 10 shows the CI values calculated from cold seasons in 1981–2020, which exhibit two local peaks for k = 4 and k = 16. The 4-cluster partition is the most compact representation for cold-season WRs, however, leads to paired two patterns (not shown) in the results. Another local peak of CI found at k = 16 generally displays the variants of four regimes while reflecting disturbances associated with synoptical-scale atmospheric waves. Therefore, the 16-cluster partition is selected to generate SOM prototypes.

Estimation of power production

We use generic power curves to estimate the power production that might be available at different locations across the CONUS following the approach described in the WIND ToolKit[48]. The turbines are categorized into three classes based on the rated power, following the classification criteria defined by International Electrotechnical Commission (ICE) 61400-1 (ICE 2005). We calculate the normalized power production from hourly 100-m wind speed using the ICE Class 2 power curve shown as an example. Note that we test the three ICE-defined classes, and all generate similar results.

WR-related wind speed and power production

To quantify the impact of WRs on the variation of hub-height wind speed and power production, we reconstruct the WR-related wind and power production over different time intervals using the SOM prototypes. The impact (I) of a prototype on wind speed or power production during month m of year y is defined as:

$$I(p, m, y) = \frac{1}{N}\sum_{d=1}^{N} W(p) \tag{1}$$

where N is the number of days categorized as a prototype p in month m during 1981–2020 except for the year y, and W is the mean anomaly of wind speed or power production. This leave-one-out approach is applied to avoid overfitting.

The WR-related winds or power production are derived based on the concept of inverse distance weighting average in the SOM input data space. We calculate the impact based on the 16

prototypes identified by SOM to reconstruct the timeseries of WR-related wind resources (wind speed and power production) over different timescales. The Euclidean distance (D) between an input vector (Z500 anomaly averaged over a specific time interval) and a prototype is used to calculate the weight (w) of the impact of a given weather pattern (namely a SOM prototype). Then the WR-related variable (W_{WR}) is obtained by weighted-averaging the impact of the nearest four prototypes to reduce the computational cost. Specifically, we calculate W_{WR} using Z500 anomaly averaged over different time intervals (n, i.e., from 1-day to 90-day) to examine the WR's impact at different timescales.

$$w(p, n, m, y) = \frac{1}{D(p, n, m, y)} \tag{2}$$

$$W_{WR}(k, m, y) = \frac{\sum_{p=1}^{1} w(p, n, m, y) \cdot I(p, m, y)}{\sum_{p=1}^{4} w(p, n, m, y)} \tag{3}$$

The WR-related power production is aggregated to subregions. Seven isolated system operators (ISOs), including the Southwest Power Pool (SPP), Electric Reliability Council of Texas (ERCOT), Midcontinent Independent System Operator (MISO), California Independent System Operator (CAISO), ISO New England (ISO-NE), PJM Interconnection (PJM), and New York Independent System Operator (NYISO), two non-ISO regions, i.e., West and Southeast, and the Wind Forecast Improvement Project (WFIP2[37]) region (approximately covering Washington and Oregon States and also part of Bonneville Power Administration service area), are used to estimate regional power production (Supplementary Fig. 1). The temporal spearman correlation and standard deviation ratio (defined as $rstd = \frac{std(W_{WR})}{std(W_{ERA5})}$) between the monthly reconstructed and full wind speed are used to evaluate the skill of reconstruction. The explanatory power (calculated as R^2) is obtained to quantify the percentage of the variability of power production explained by WRs.

Significance test
We calculate the statistical significance of the composite maps of Z500 and wind speed anomalies associated with each WR by a bootstrap sampling with replacement. We construct the 5% significance level using 50,000 resamples per regime over all days in the period of 1981–2020, following ref. [12]. Random days are selected in the blocks corresponding to the observed regime "events" to test the null hypothesis that the composites are the result of random subsampling of days. Specifically, the following steps are taken: (1) for each regime occurrence, the number of consecutive days in each WR is computed to produce "blocks" of the regime days; (2) each set of consecutive days is randomized to produce random sets of numbers with the same structure as observed and selected days of data from all days in the period of 1981–2020; (3) step (2) is repeated 50,000 times; (4) At each grid point, if the observed value lies beyond the 2.5/97.5 percentiles of the resampled distribution, then it is classified as statistically significant.

Finally, the statistical significance of the Spearman correlation, standard deviation ratio, and explanatory power are determined by Spearman significance test. We construct the 5% significance level to test the null hypothesis that the two timeseries of WR-related and full wind speed/power production are independent.

DATA AVAILABILITY

CODE AVAILABILITY

REFERENCES
1. Wiser, R. et al. Land-based wind market report: 2022 Edn. https://doi.org/10.2172/1893263 (2022).
2. Biden Administration Jumpstarts Offshore Wind Energy Projects to Create Jobs. https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/29/fact-sheet-biden-administration-jumpstarts-offshore-wind-energy-projects-to-create-jobs/ (2021).
3. Orlov, A., Sillmann, J. & Vigo, I. Better seasonal forecasts for the renewable energy industry. Nat. Energy 5, 108–110 (2020).
4. Soret, A. et al. Sub-seasonal to seasonal climate predictions for wind energy forecasting. J. Phys. Conf. Ser. 1222, 012009 (2019).
5. Alessandrini, S. & Sperati, S. Characterization of forecast errors and benchmarking of renewable energy forecasts. in Renewable Energy Forecasting (ed. Kariniotakis, G.). 235–256 (Woodhead Publishing, 2017).
6. Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I. & Wernli, H. Balancing Europe's wind-power output through spatial deployment informed by weather regimes. Nat. Clim. Chang 7, 557–562 (2017).
7. Das, A. & Roy, S. B. Evaluation of subseasonal to seasonal forecasts over India for renewable energy applications. Adv. Geosci. 56, 89–96 (2021).
8. Lledó, L., Ramon, J., Soret, A. & Doblas-Reyes, F.-J. Seasonal prediction of renewable energy generation in Europe based on four teleconnection indices. Renew. Energy 186, 420–430 (2022).
9. Buizza, R. & Leutbecher, M. The forecast skill horizon. Q J. R. Meteor Soc. 141, 3366–3382 (2015).
10. Stan, C. & Straus, D. M. Is blocking a circulation regime? Mon. Weather Rev. 135, 2406–2413 (2007).
11. Bao, M. & Wallace, J. M. Cluster analysis of Northern Hemisphere wintertime 500-hPa flow regimes during 1920–2014*. J. Atmos. Sci. 72, 3597–3608 (2015).
12. Lee, S. H., Furtado, J. C. & Charlton-Perez, A. J. Wintertime North American weather regimes and the Arctic stratospheric polar vortex. Geophys. Res. Lett. 46, 14892–14900 (2019).
13. Ghil, M. & Robertson, A. W. "Waves" vs. "particles" in the atmosphere's phase space: a pathway to long-range forecasting? Proc. Natl Acad. Sci. USA 99, 2493–2500 (2002).
14. Ghil, M., Groth, A., Kondrashov, D. & Robertson, A. W. The Gap Between Weather and Climate Forecasting. in Sub-seasonal to seasonal prediction (ed. Robertson, A. W.). 119–142 (Elsevier, 2019).
15. Vigaud, N., Robertson, A. W. & Tippett, M. K. Predictability of recurrent weather regimes over North America during winter from submonthly reforecasts predictability of recurrent weather regimes over North America during winter from submonthly reforecasts. Mon. Weather Rev. 146, 2559–2577 (2018).
16. Roller, C. D., Qian, J.-H., Agel, L., Barlow, M. & Moron, V. Winter weather regimes in the Northeast United States. J. Clim. 29, 2963–2980 (2016).
17. Smith, D. M. et al. North Atlantic climate far more predictable than models imply. Nature 583, 796–800 (2020).
18. Robertson, A. W., Vigaud, N., Yuan, J. & Tippett, M. K. Towards identifying sub-seasonal forecasts of opportunity using North American weather regimes. Mon. Weather Rev. 148, 1861–1875 (2020).
19. Lee, J. C. K., Lee, R. W., Woolnough, S. J. & Boxall, L. J. The links between the Madden-Julian Oscillation and European weather regimes. Theor. Appl. Climatol. 141, 567–586 (2020).
20. Beck, C., Philipp, A. & Streicher, F. The effect of domain size on the relationship between circulation type classifications and surface climate. Int. J. Climatol. 36, 2692–2709 (2016).
21. Cortesi, N., Torralba, V., González-Reviriego, N., Soret, A. & Doblas-Reyes, F. J. Characterization of European wind speed variability using weather regimes. Clim. Dyn. 53, 4961–4976 (2019).
22. Bloomfield, H. C., Brayshaw, D. J. & Charlton-Perez, A. J. Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. Meteorol. Appl. 27, e1858 (2020).
23. Straus, D. M., Corti, S. & Molteni, F. Circulation regimes: chaotic variability versus SST-forced predictability. J. Clim. 20, 2251–2272 (2007).
24. Vesanto, J. & Alhoniemi, E. Clustering of the self-organizing map. IEEE Trans. Neural Netw. 11, 586–600 (2000).

25. Liu, Y., Qian, Y. & Berg, L. K. Local-thermal-gradient and large-scale-circulation impacts on turbine-height wind speed forecasting over the Columbia River Basin. *Wind Energy Sci.* **7**, 37–51 (2022).

26. Díaz-Esteban, Y. & Raga, G. B. Weather regimes associated with summer rainfall variability over southern Mexico. *Int. J. Climatol.* **38**, 169–186 (2017).

27. Guèye, A. K. et al. Weather regimes over Senegal during the summer monsoon season using self-organizing maps and hierarchical ascendant classification. Part I: synoptic time scale. *Clim. Dyn.* **36**, 1–18 (2011).

28. Haupt, S. E. et al. A method to assess the wind and solar resource and to quantify interannual variability over the United States under current and projected future climate. *J. Appl. Meteorol. Clim.* **55**, 345–363 (2016).

29. Huang, H. et al. Where does the dust deposited over the Sierra Nevada snow come from? *Atmos. Chem. Phys.* **22**, 15469–15488 (2022).

30. Blackmon, M. L., Lee, Y.-H. & Wallace, J. M. Horizontal structure of 500 mb height fluctuations with long, intermediate and short time scales. *J. Atmos. Sci.* **41**, 961–980 (1984).

31. Weinberger, I., Garfinkel, C. I., White, I. P. & Oman, L. D. The salience of non-linearities in the boreal winter response to ENSO: Arctic stratosphere and Europe. *Clim. Dyn.* **53**, 4591–4610 (2019).

32. Visbeck, M. H., Hurrell, J. W., Polvani, L. & Cullen, H. M. The North Atlantic Oscillation: past, present, and future. *Proc. Natl Acad. Sci. USA* **98**, 12876–12877 (2001).

33. Michelangeli, P.-A., Vautard, R. & Legras, B. Weather regimes: recurrence and quasi stationarity. *J. Atmos. Sci.* **52**, 1237–1256 (1995).

34. Wilczak, J. M. et al. The second wind forecast improvement project (WFIP2): observational field campaign the second wind forecast improvement project (WFIP2): observational field campaign. *Bull. Am. Meteorol. Soc.* **100**, 1701–1723 (2019).

35. Efthymiadis, D., Jones, P. D., Briffa, K. R., Böhm, R. & Maugeri, M. Influence of large-scale atmospheric circulation on climate variability in the Greater Alpine Region of Europe. *J. Geophys. Res. Atmos.* **112**, D12104 https://doi.org/10.1029/2006JD008021 (2007).

36. Portis, D. H., Walsh, J. E., Hamly, M. E. & Lamb, P. J. Seasonality of the North Atlantic oscillation. *J. Clim.* **14**, 2069–2078 (2001).

37. Shaw, W. J. et al. The second wind forecast improvement project (WFIP2): general overview. *Bull. Am. Meteorol. Soc.* **100**, 1687–1699 (2019).

38. Younas, W. & Tang, Y. PNA predictability at various time scales. *J. Clim.* **26**, 130715122904005 (2013).

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

40. Chen, Z., Gan, B., Wu, L. & Jia, F. Pacific-North American teleconnection and North Pacific Oscillation: historical simulation and future projection in CMIP5 models. *Clim. Dyn.* **50**, 4379–4403 (2018).

41. Fabiano, F., Meccia, V. L., Davini, P., Ghinassi, P. & Corti, S. A regime view of future atmospheric circulation changes in northern mid-latitudes. *Weather Clim. Dyn.* **2**, 163–180 (2021).

42. Vitart, F. & Robertson, A. W. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *NPJ Clim. Atmos. Sci.* **1**, 3 (2018).

43. Mingoti, S. A. & Lima, J. O. Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *Eur. J. Oper. Res.* **174**, 1742–1759 (2006).

44. Misra, S., Li, H. & He, J. Robust geomechanical characterization by analyzing the performance of shallow-learning regression method using unsupervised clustering methods. in *Machine learning for subsurface characterization (ed. Hammon K.)*. 129–155 (Gulf Professional Publishing, 2020).

45. Hersbach, H. et al. The ERA5 global reanalysis. *Q J. R. Meteor. Soc.* **146**, 1999–2049 (2020).

46. Vigaud, N. & Robertson, A. W. Convection regimes and tropical-midlatitude interactions over the Intra-American Seas from May to November. *Int. J. Climatol.* **37**, 987–1000 (2017).

47. Hannachi, A. On the origin of planetary-scale extratropical winter circulation regimes. *J. Atmos. Sci.* **67**, 1382–1401 (2010).

48. King, J., Clifton, A. & Hodge, B. Validation of power output for the WIND toolkit. *United States, Web* https://doi.org/10.2172/1159354 (2014).

AUTHOR CONTRIBUTIONS

L.K.B., Y.Q., Y.L., and S.F. designed the research. Y.L. performed the analysis with contributions from all authors. Y.L. and S.F. led the writing and editing of the manuscript. All authors contributed to the interpretation of the results.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00403-5.

Correspondence and requests for materials should be addressed to Ye Liu or Sha Feng.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.