IOP Conf. Series: Journal of Physics: Conf. Series **1037** (2018) 072038 doi:10.1088/1742-6596/1037/7/072038

Determining variabilities of non-Gaussian wind-speed distributions using different metrics and timescales

J C-Y Lee[1,2], M J Fields[2], J K Lundquist[1,2] and M Lunacek[2]

[1]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, CO 80309, USA
[2]National Wind Technology Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA

Email: Joseph.Lee@nrel.gov

Abstract. Quantification of long-term wind-speed variability is a critical component in wind resource assessment, and effective wind-farm operations require proper assessment of this variability. Yet, wind-speed variations differ across averaging temporal scales because hourly mean wind speeds fluctuate more than yearly averages. In this study, we quantify the influence of averaging timescale to the resultant variability. We assess three spread metrics (standard deviation, coefficient of variation, and robust coefficient of variation) and two distribution measures (skewness and kurtosis) based on 38 years of wind speeds from the National Aeronautics and Space Administration's MERRA-2 reanalysis data set over the contiguous United States. The spatial distributions of wind-speed variability differ with metrics and timescales: wind speeds of fine temporal resolution generate strong variabilities that dilute spatial contrasts; small sample size becomes a constraint in calculating interannual variabilities via annual means and leads to inaccurate results. Overall, we find that metrics based on monthly data portray the largest spatial differences of wind-speed variability. Although standard deviation yields consistent geographical projections, none of the wind-speed data of any time frame are perfectly Gaussian. Therefore, the robust coefficient of variation, a statistically robust and resistant approach, appears to be the ideal metric for quantifying wind-speed variabilities based on monthly mean data.

1. Introduction

Part of the long-term investment decisions in the wind energy industry rest on the accurate assessment of wind-speed variations at a location. Moreover, skillful long-term wind-resource predictions allow the investors to ensure that they have adequately sized the minimum debt obligations to coincide with expected minimum productions—and therefore, revenue. Hence, quantifying the uncertainty in winds is critical for the wind resource assessment process.

Wind speed varies across timescales: gusts alter wind speeds in seconds and climate oscillations change wind patterns year to year. To efficiently quantify the variation of a specific wind-speed distribution, a spread statistic is used to summarize the information into one number. Of all the spread metrics, standard deviation (σ) is the most commonly used tool in the industry to assess wind-speed variability, especially interannual variability (IAV) [1]. Although the industry uses the Weibull distribution to quantify wind-speed distributions, its two parameters are not spread metrics by definition; hence, σ becomes the default tool to assess IAV. However, σ is subjected to assumptions of the nature of the data, including the distribution shape, and it is influenced by outliers; so, σ is not statistically robust or resistant [2]. The metric uses sample mean, which represents the center of a sample when the data follow a Gaussian distribution. When the data are not Gaussian distributed, characterizing the data using the mean and the σ is not completely valid. As a result, using σ alone to quantify variability without addressing the characteristics of the data can lead to misleading results and interpretations. For example, distribution parameters such as skewness and kurtosis can determine the degree of deviation of a certain distribution from the perfect Gaussian distribution. Thus, when the data sample is skewed with a sharp peak and deviates from Gaussian, σ alone does not offer the most accurate perspective in determining wind-speed variability.

The goal of the study is to improve the wind-resource assessment process by assessing the inadequacy of the Gaussian assumption in calculating wind-speed variabilities of various averaging timescales. Herein, we evaluate and compare wind-speed statistics of timescales from hourly means to yearly averages. We expect that wind speeds of various temporal resolutions possess distinct shapes, spreads, and distribution attributes. We also challenge the validity of the Gaussian assumption, particularly in annual mean wind speeds, because the calculation of IAV requires annual mean wind speeds. Overall, we demonstrate that perfect Gaussian distributions of wind speeds are rare regardless of the choice of timescales; hence, quantifying variabilities with a robust and resistant metric is ideal.

2. Methods

2.1. Data set

We use the hourly horizontal wind components in the National Aeronautics and Space Administration's Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) reanalysis data set [3] from 1980 to 2017. Using the wind speeds at 10 m, 50 m, 850 hPa, and 500 hPa, we calculate the wind speeds at 80 m above the surface, a presumed turbine hub height, via the power law and the derived shear components. Herein, we compute the mean hub-height wind speeds at various time resolutions: hourly, daily, weekly, monthly, seasonal (winter: December, January, and February; spring: March, April, and May; summer: June, July, and August; fall: September, October, and November), and annual. Our samples include 333,120 hours, 13,880 days, 1,982 weeks, 456 months, 152 weeks, and 38 years of wind speed at each model grid point. In this study, we focus on the wind resources in the contiguous United States (CONUS). Note that spatial averaging and temporal smoothing are applied to the MERRA-2 data set, and hence, the modeled wind speeds are only close approximations of the actual wind resource. Even though the industry typically uses 10-minute data, the highest temporal resolution of the MERRA-2 data set is hourly data.

2.2. Metrics

We compute several metrics to represent the spread as well as the characteristics of the wind-speed distribution at each MERRA-2 grid point. The statistical measures include a simple spread metric, σ, and two normalized spread metrics that are divided by an average metric, coefficient of variation (CoV) in equation (1), and robust coefficient of variation (RCoV) in equation (2):

$$\text{Coefficient of variation (CoV)} = \frac{\sigma}{\text{mean}} \tag{1}$$

$$\text{Robust coefficient of variation (RCoV)} = \frac{\text{median}|\text{x-median}(\text{x})|}{\text{median}} \tag{2}$$

Like σ, large values of CoV and RCoV represent strong variations in the data. Of the three spread metrics evaluated, only RCoV is statistically robust and resistant [2]. Additionally, RCoV is the ideal metric to assess and connect the long-term variabilities of wind speeds and actual wind energy productions [4]. Moreover, to contrast variabilities geographically, we use spatially normalized metrics:

$$\text{Spatially normalized RCoV}_{\text{i}} = \frac{\text{RCoV}_{\text{i}}}{\text{CONUS median RCoV}} . \tag{3}$$

We calculate and normalize the variabilities for each grid cell i. We derive the spatially normalized σ and CoV in the same fashion.

The two distribution parameters chosen are skewness, (equation (4), which indicates symmetry, and kurtosis, equation (5), which represents tailedness:

$$\text{Skewness} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{\frac{3}{2}}} \tag{4}$$

$$\text{Kurtosis} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{2}} - 3 \tag{5}$$

A perfect Gaussian distribution has zero skewness and zero kurtosis. Positive skewness indicates the distribution tends toward low values, and positive kurtosis signifies the distribution tends to cluster near the center [2].

3. Results

An ideal location possesses strong wind energy content and little variability. Based on the hourly mean wind speeds, the region with the best wind resources in the CONUS resides in central United States (figure 1). On average, the Plains, Upper Midwest, and eastern parts of some mountain states record some of the highest hourly mean hub-height wind speeds in the country between 1980 and 2017 (figure 1).

The absolute values of RCoV do not offer fair and easy geographical comparisons among averaging timescales. Accounting for the variability in fine temporal resolution, RCoVs of hourly wind speeds in the central United States are moderate compared to the other regions of the CONUS (figure 2a). However, expanding the averaging time frames from hourly to yearly mean data reduces the absolute values of RCoV: the variations decrease with fewer data points (figure 2). For example, the RCoVs of monthly mean wind speeds (figure 2a) are larger than the RCoVs of annual mean wind speeds (figure 2f) by nearly an order of magnitude. Furthermore, RCoV demonstrates weak spatial contrasts. For instance, wind speeds in Wyoming fluctuate more relative to other states in hourly means (figure 2a), but not in daily or annual means (figure 2b and 2f). Because one of our goals is to find metrics that highlight geographical variations—and because the variations within one map in each panel of figure 2 are small relative to inter-timescale comparisons—the rest of the paper focuses on the results from spatially normalized metrics using the CONUS median within that map.

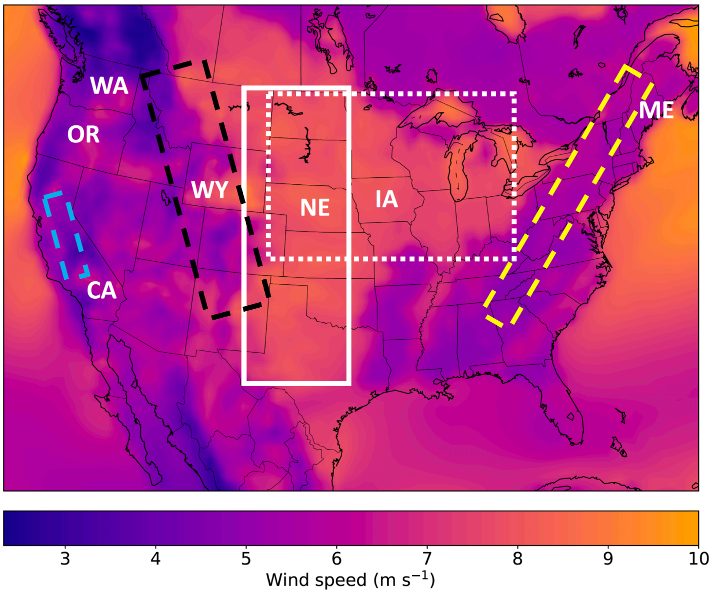

Figure 1. Hourly mean wind speed at 80 m above surface of the CONUS over 38 years, overlaying with specific locations that are mentioned in the paper, including states of California (CA), Iowa (IA), Maine (ME), Nebraska (NE), Oregon (OR), Washington (WA), and Wyoming (WY). The white boxes with solid and dotted lines in central United States identify the Great Plains and the Upper Midwest, respectively. The yellow, black, and blue boxes bounded by dashed lines, respectively, mark the Appalachian Mountains along the East Coast, Rocky Mountains, and Sierra Nevada in California.

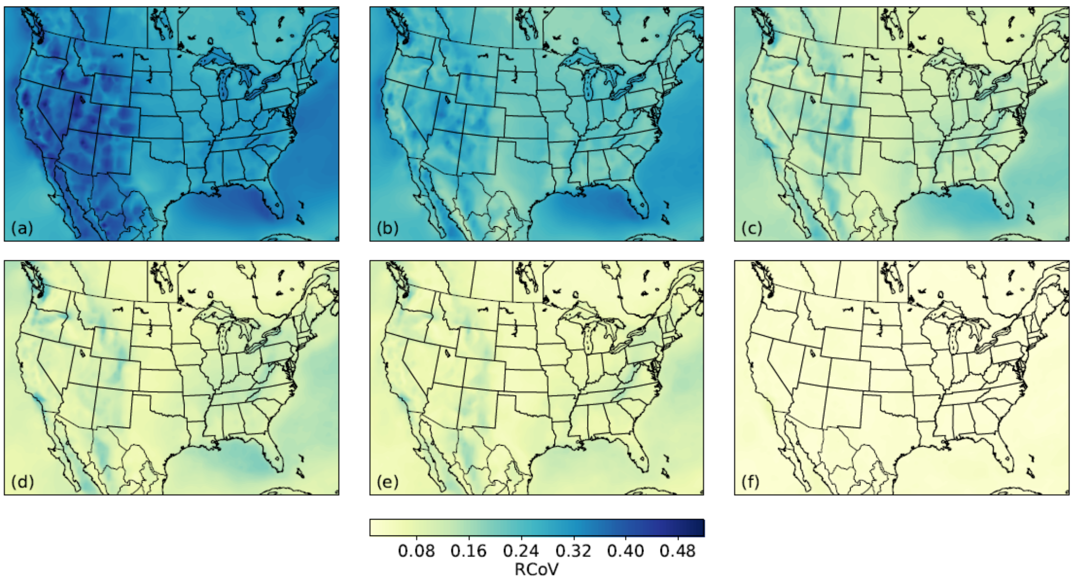

Figure 2. Spatial distribution of RCoV calculated using (a) hourly, (b) daily, (c) weekly, (d) monthly, (e) seasonal, and (f) annual-mean wind speeds of the CONUS over 38 years.

The distributions of spatially normalized RCoV display geographical distinctions in intramap and intermap comparisons (figure 3), in contrast to the results of RCoV in figure 2. The spatial pattern in each panel of figure 3 remains the same as those in figure 2, and the normalization provides fair comparisons between timescales. For instance, the high wind-speed variabilities in the Rocky Mountains become obvious in the weekly (figure 3c), monthly (figure 3d), and seasonal data (figure 3e) data. In particular, the wind-speed variabilities in Wyoming are generally above the national average in each averaging time frame (figure 3), which are different from the results in figure 2. Moreover, the wind speeds in the central part of the United States remain moderately variable across timescales. In general, the spatially normalized RCoV and CoV demonstrate analogous spatial results across averaging time frames (figure 3 and 4), and the results from monthly means and seasonal means illuminate the largest geographical differences, both on land and offshore.

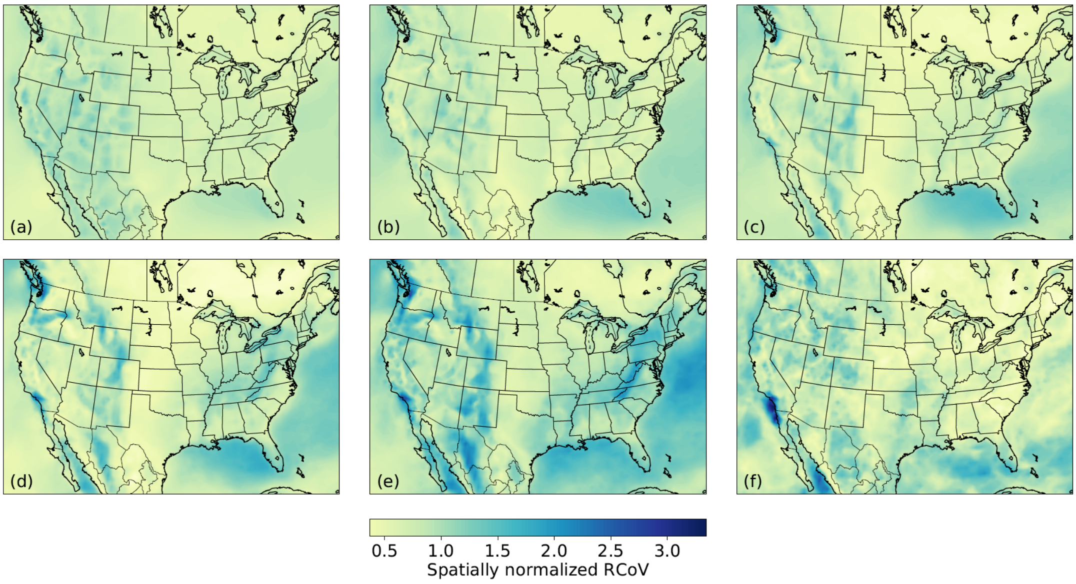

Figure 3. As in figure 2, but for each timescale, the RCoV values are spatially normalized with the CONUS median of RCoV for that map.

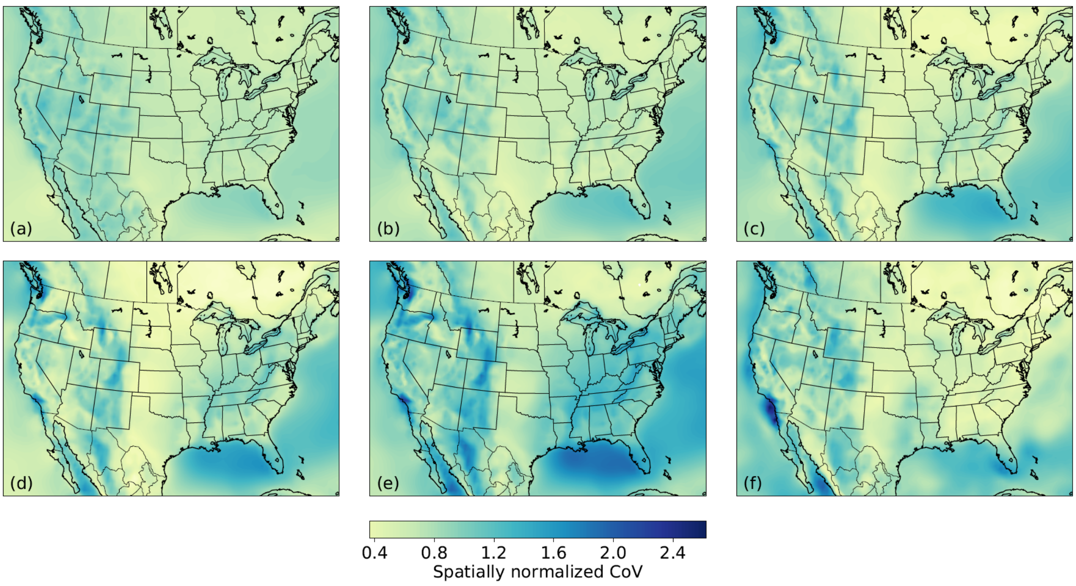

Figure 4. As in figure 5, but for spatially normalized CoV.

Similarly, the geographical discrepancies from the spatially normalized σ also emerge only with timeframes longer than a week (figure 5). This trend is the most noticeable along the Rocky Mountains and the western states (figure 5). The pattern in each panel of figure 5 is similar to those in figure 3 and 4, and the largest differences among the panels are located in southern California using seasonal and annual averages (figure 3e and 3f, figure 4e and 4f, figure 5e and 5f).

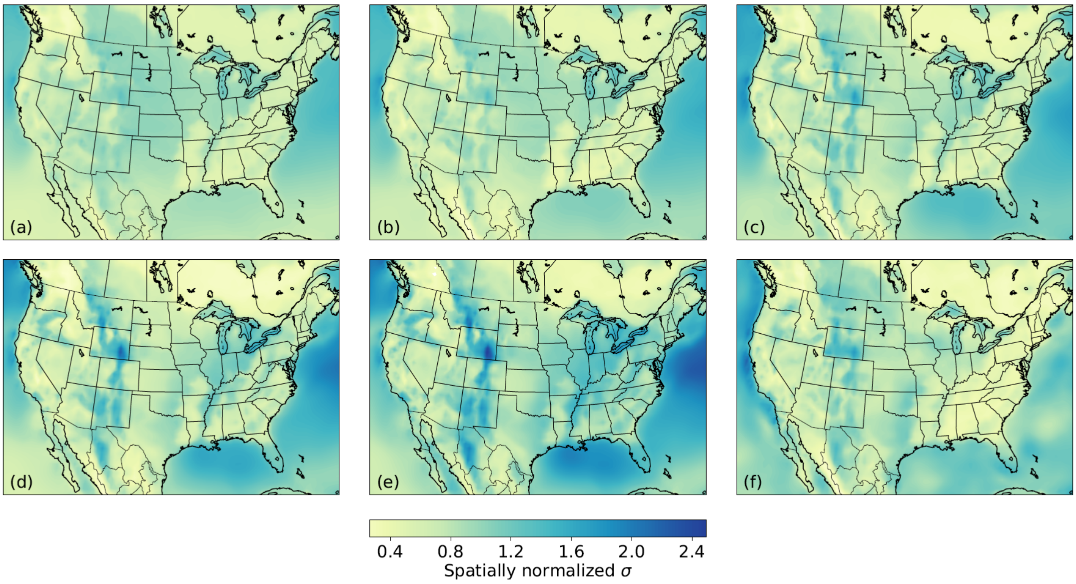

Figure 5. As in figure 5, but for spatially normalized σ.

Moreover, among the three spatially normalized spread metrics, normalized σ demonstrates the highest consistency in depicting geographical variabilities among grid points across timescales. For each MERRA-2 grid point in the CONUS, we calculate the range of the results over different timescales, which is the difference between the maximum and minimum (figure 6). In other words, we compute the largest changes in the values among all the panels in each of the figure 3, 4, and 5. The relatively low and uniform ranges of the normalized σ values indicate that the relative variabilities within the CONUS are consistent among different timeframes (figure 6c). For instance, Maine generally possesses low variabilities compared to other regions (figure 5), and the close-to-zero ranges in the state in figure 6a show that the variabilities in Maine are low, from hourly to yearly mean wind speeds.

In contrast, the normalized spread metrics, especially the RCoV, display temporal inconsistencies in wind-speed variability quantification (figure 6a and 6b). For example, in southern coastal California and the Appalachians, the normalized CoVs and RCoVs increase with longer averaging time periods (figure 3 and 4). Although the ranges of the two normalized spread measures are large in those locations (figure 6a and 6b), the ranges of the three spread metrics generally demonstrate good agreements (figure 6). The ranges of all spatially normalized metrics are near zero in the central United States (figure 6), where the wind resources are consistent and abundant (figure 1).

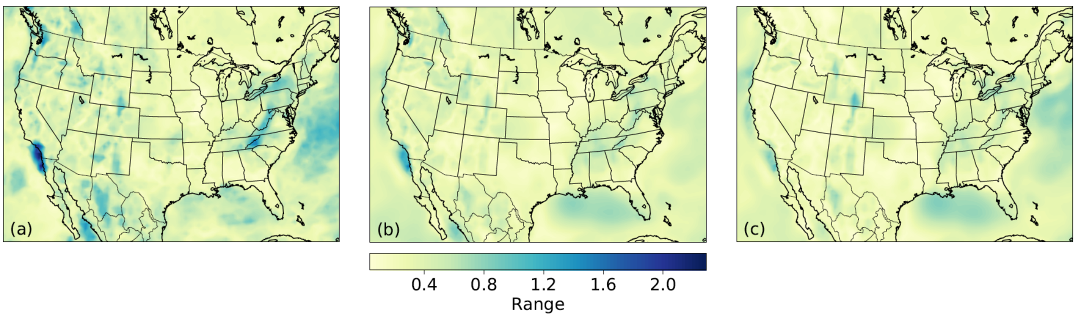

Figure 6. Spatial distribution of range of (a) spatially normalized RCoV, (b) spatially normalized CoV, and (c) spatially normalized σ over the six chosen timescales. Each panel illustrates the range at each grid point of six temporal resolutions in figure 3, 4, and 5, respectively.

In contrast to the spatially normalized spread metrics, distribution parameters, such as skewness and kurtosis, drastically change with averaging timeframes. Specifically, for the hourly, daily, and weekly mean wind-speed data, the skewness values in most of the CONUS tend to be positively skewed, or leaning toward lower wind speeds (figure 7a, 7b, and 7c). The spatial pattern is distinct, especially in the western states. For longer timescales, other parts of the CONUS become more negatively skewed (figure 7d, 7e, and 7f). For example, the seasonal-mean wind-speed distributions in the states east of the Rocky Mountains are largely negatively skewed (figure 7e).

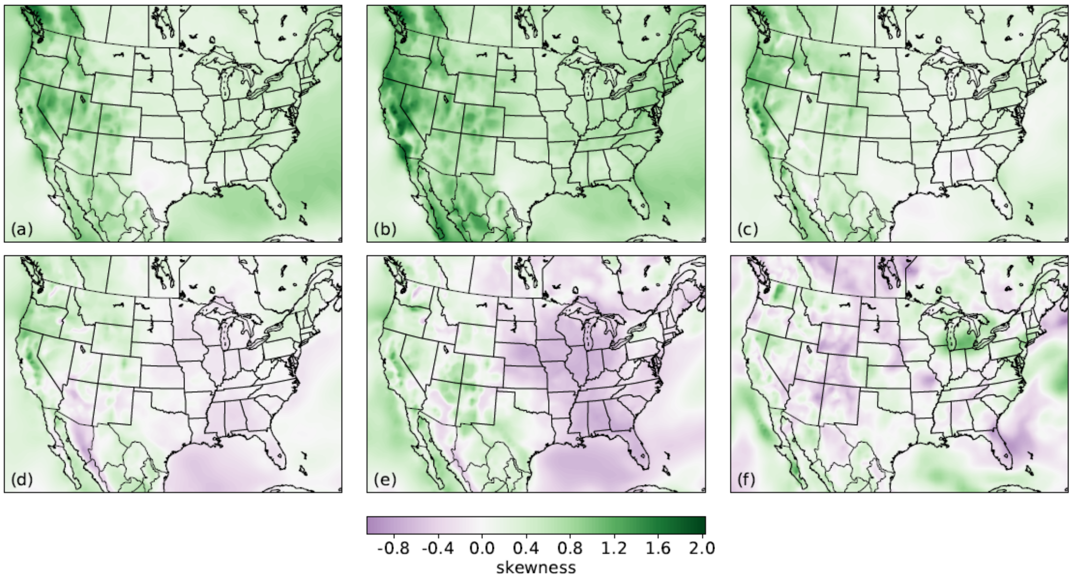

Figure 7. Spatial distribution of skewness of (a) hourly, (b) daily, (c) weekly, (d) monthly, (e) seasonal, and (f) annual-mean wind speed of the CONUS over 38 years.

The changes in kurtosis across averaging times are relatively gradual compared to skewness. The values of kurtosis in most of the CONUS transition from positive in hourly mean data to negative in yearly mean data (figure 8). Although extreme values emerge—for instance, the strongly positive kurtosis values of daily mean wind speeds in the Sierra Nevada (figure 8b)—kurtosis values in coarser timescales generally become more neutral in the CONUS (figure 8d, 8e, and 8f). Moreover, the maximum positive kurtosis values are over 7, whereas the minimum of all kurtosis values in the CONUS across all timescales is -1.35.

The consistent variances of the spatially normalized RCoV across timescales offer fair spatial comparison in wind-speed variability. The interquartile range (IQR) represents the span of a box in figure 9. The IQR of the absolute RCoV in the CONUS decreases with coarser temporal resolution, and the IQR becomes notably small in the yearly mean data (figure 9a). Hence, the spatial differences in absolute RCoV of wind speeds appear trivial in the seasonal-mean and annual-mean data (figure 2e and 2f). Meanwhile, for spatially normalized RCoV, the IQR generally increases with longer timeframes, and more outliers also emerge (figure 9b). Accordingly, the geographical pattern of normalized RCoV is distinct in each timescale (figure 3).

Variances in the distribution metrics also differ with time frames, as in figure 7 and 8. A large portion of the monthly, seasonal, and annual mean wind speeds in the CONUS yields skewness of near-zero (figure 9c), whereas most data of the finer temporal resolutions tend to be positively skewed, echoing figure 7a, 7b and 7c. Of the three coarser averaging timescales, the monthly mean skewness results in the lowest IQR (figure 9c). For kurtosis, only weekly mean wind speeds concentrate around zero, among the data of all timeframes in the CONUS (figure 9d). All temporal distributions have outliers of positive kurtosis values (figure 9d), especially the daily means (figure 8b).

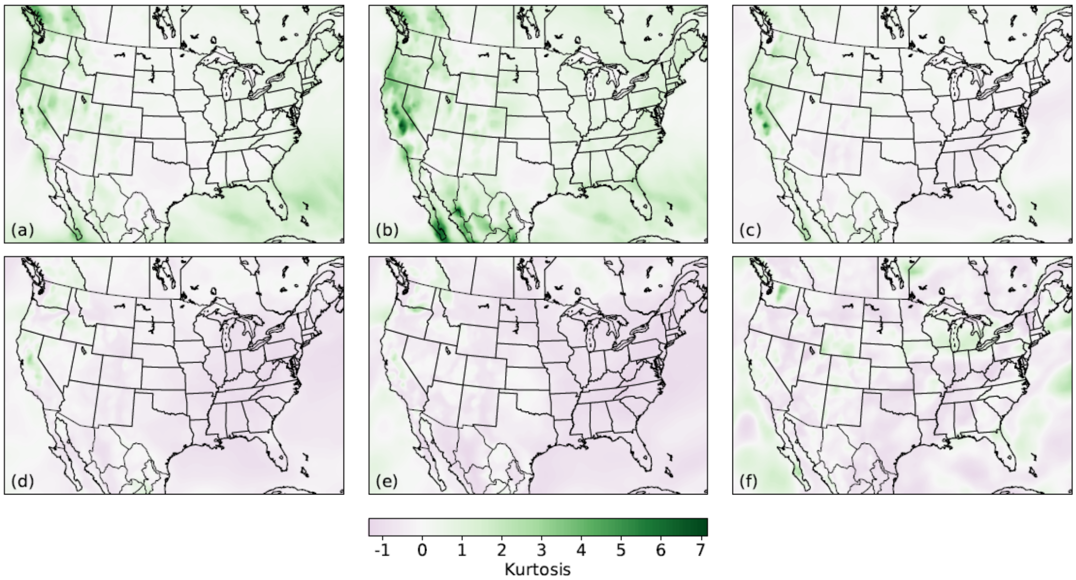

Figure 8. As in figure 7, but for kurtosis.

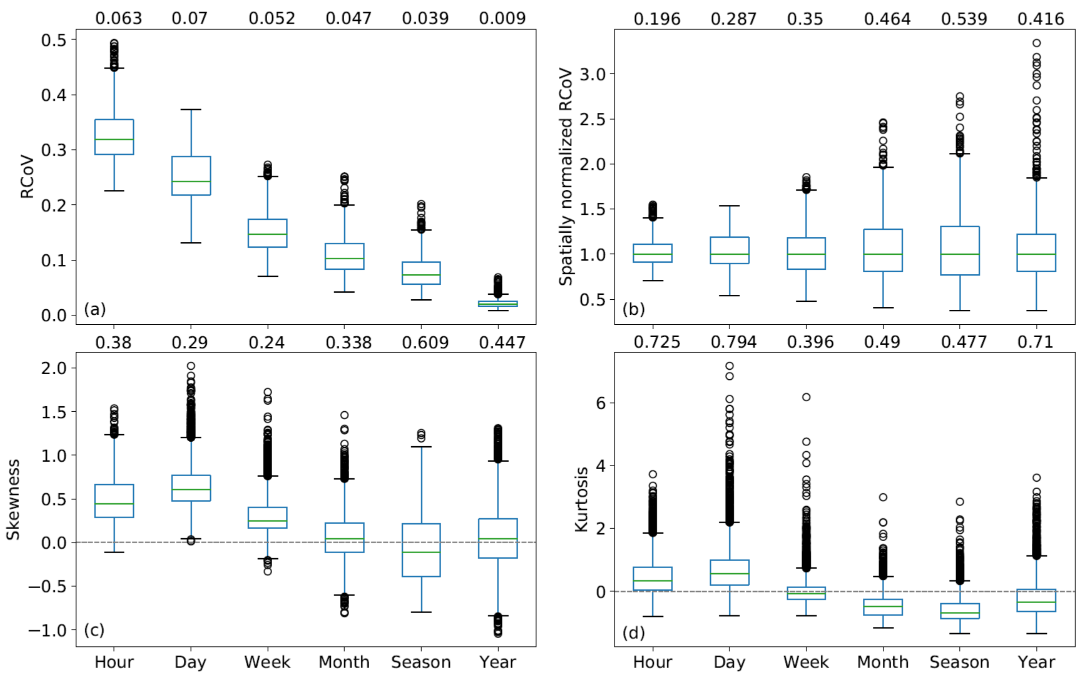

Figure 9. Box plots of (a) RCoV, (b) spatially normalized RCoV, (c) skewness, and (d) kurtosis of all the CONUS grid points across timescales, corresponding to figure 2, 3, 7, and 8, respectively. The numbers on the top of each panel indicate the IQR of the CONUS data in each time resolution, which also illustrate the span of the boxes. The grey horizontal lines in (c) and (d) represent skewness and kurtosis of zero.

IOP Conf. Series: Journal of Physics: Conf. Series **1037** (2018) 072038 doi:10.1088/1742-6596/1037/7/072038

4. Discussion

We illustrate the impact of time resolution of wind speeds in variability calculation and distribution characterization. We calculate and compare the spatial projections of three spread metrics (σ, CoV, and RCoV) and two distribution parameters (skewness and kurtosis) using the MERRA-2 wind speeds at 80 m above the surface. We contrast the statistics using the mean wind speeds of every hour, day, week, month, season, and year from 1980 to 2017.

Using monthly or seasonal mean wind speeds can accurately contrast variabilities spatially, whereas data of too fine or too coarse temporal resolutions lead to geographically indistinct results. On one hand, hourly data contain highly fluctuating winds from wind gusts and mesoscale weather events. Hence, the spread metrics tend to yield large values of variability (figure 2a and 9a) and result in ambiguous spatial distinctions (figure 3a, 4a, and 5a). For example, using hourly data, spatially normalized σ indicates comparatively high variabilities in most of the central United States (figure 5a), and geographically normalized CoV and RCoV display high variabilities in all the states west of the Rocky Mountains (figure 3a and 4a). Hence, in general, the spread of hourly mean wind speeds does not spatially differentiate the variabilities among regions with precision (figure 9a). On the other hand, IAV is a commonly used measure in the industry and requires annual mean data. However, variability quantification using annual mean wind speeds cannot sufficiently distinguish the relative long-term variabilities between states because of the small sample size of yearly data and the weak geographical discrepancies in IAV (figure 3f, 4f, and 5f). Moreover, annual mean wind speeds erode irregular signals of finer temporal timescales that can deviate distribution characteristics (figure 7f and 8f), leading to inaccurate representation of the spread of yearly data. Overall, we recommend using monthly or seasonal mean wind speeds to adequately differentiate regions in relative wind-speed variabilities (figure 3, 4, and 5). Note that geographically indistinct patterns from nonspatially normalized spread metrics are still meaningful for existing commercial wind farms because variabilities at all timescales affect wind power productions and profitability.

Depending on which spread metric is used, contradicting results in variability can emerge. For instance, in the monthly data, the results of spatially normalized CoV and RCoV identify those two states with the lowest variabilities in the CONUS (figure 3d and 4d), whereas using normalized σ indicates moderately high variabilities in Iowa and Nebraska (figure 5d). Nevertheless, in mountainous regions, such as the Rocky Mountains and Appalachians, the spread metrics exhibit agreements in indicating the mountains with high wind-speed variabilities, regardless of the averaging timescales (figure 3, 4, and 5).

Furthermore, results of the distribution metrics—skewness and kurtosis—differ greatly between timescales. Although the kurtosis values of wind speeds demonstrate nearly perfect Gaussian distributions in most of the CONUS (figure 8 and 9d), the skewness values suggest otherwise (figure 7 and 9c). Of all the time resolutions, monthly mean wind speeds generally yield skewness and kurtosis closest to zero (figure 9c and 9d), yet the patterns are not uniform across the CONUS (figure 7d and 8d). Moreover, readers should be mindful of the rapidly changing spatial patterns of skewness across timescales (figure 7). Hence, except for monthly data, the Gaussian assumption is principally inadequate in most of the CONUS for all the averaging time frames of wind-speed data.

Considering that wind-speed distributions are not perfectly Gaussian, using RCoV to quantify long-term variability is advantageous given its statistical robustness and resistance. Even though the spatially normalized σ yields consistent relative variabilities in the CONUS across timescales (figure 6a), readers also need to account for the nonzero skewness and kurtosis values (figure 7, 8, and 9). In longer time frames of seasonal and annual means, the Gaussian assumption can be reasonable with large sample sizes (figure 9c and 9d); for fine time resolutions—namely, the hourly, daily, and weekly data—most results of skewness and kurtosis in the CONUS deviate from zero (figure 9c and 9d). Hence, representing wind-speed variabilities via nonrobust and nonresistant metrics, such as σ and CoV, becomes inadequate, whereas RCoV works for any type of wind-speed distribution. In particular, the geographical distinctions of monthly averages using the spatially normalized RCoV are clear (figure 3d and 9b). Moreover, the results of geographically normalized RCoV in figure 3d also resemble those illustrated in past research [5,6]. Therefore, we recommend using RCoV to evaluate long-term variabilities with monthly mean wind speeds.

Acknowledgements
The Alliance for Sustainable Energy, LLC (Alliance) is the manager and operator of the National Renewable Energy Laboratory (NREL). NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy. This work was authored by the Alliance and supported by the U. S. Department of Energy under Contract No. DE-AC36-08GO28308. Funding was provided by the U.S. Department of Energy Wind Energy Technologies Office.

References
[1] Brower M 2012 *Wind resource assessment: A practical guide to developing a wind project* (Wiley)
[2] Wilks D S 2011 *Statistical methods in the atmospheric sciences* (Academic Press)
[3] Gelaro R, McCarty W, Suárez M J, Todling R, Molod A, Takacs L, Randles C A, Darmenov A, Bosilovich M G, Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, da Silva A M, Gu W, Kim G-K, Koster R, Lucchesi R, Merkova D, Nielsen J E, Partyka G, Pawson S, Putman W, Rienecker M, Schubert S D, Sienkiewicz M, Zhao B, Gelaro R, McCarty W, Suárez M J, Todling R, Molod A, Takacs L, Randles C A, Darmenov A, Bosilovich M G, Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, Silva A M da, Gu W, Kim G-K, Koster R, Lucchesi R, Merkova D, Nielsen J E, Partyka G, Pawson S, Putman W, Rienecker M, Schubert S D, Sienkiewicz M and Zhao B 2017 The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) *J. Clim.* **30** 5419–54
[4] Lee J C Y, Fields M J and Lundquist J K 2017 Assessing variability of wind speed: Comparison and validation of 26 methodologies *Wind Energy Sci.*
[5] Hamlington B D, Hamlington P E, Collins S G, Alexander S R and Kim K-Y 2015 Effects of climate oscillations on wind resource variability in the United States *Geophys. Res. Lett.* **42** 145–152
[6] Gunturu U B and Schlosser C A 2012 Characterization of wind power resource in the United States *Atmos. Chem. Phys.* **12** 9687–702