

Quantifying sources of uncertainty in reanalysis derived wind speed



Stephen Rose ^{a,*}, Jay Apt ^{a,b}

^a *Tepper School of Business, Carnegie Mellon University, Pittsburgh, USA*

^b *Department of Engineering & Public Policy, Carnegie Mellon University, Pittsburgh, USA*

ARTICLE INFO

Article history:

Received 23 July 2015

Received in revised form

17 December 2015

Accepted 4 March 2016

Keywords:

CFS reanalysis

Wind integration

Linear mixed-effect model

ABSTRACT

Reanalysis data are attractive for wind-power studies because they can offer wind speed data for large areas and long time periods and in locations where historical data are not available. However, reanalysis-predicted wind speeds can have significant uncertainties and biases relative to measured wind speeds. In this work we develop a model of the bias and uncertainty of CFS reanalysis wind speed that can be used to correct the data and identify sources of error. We find the CFS reanalysis data underestimate wind speeds at high elevations, at high measurement heights, and in unstable atmospheric conditions. For example, at a site with an elevation of 500 m and hub height of 80 m, a CFS reanalysis wind speed of 8 m/s is 0.2 m/s higher to 1.3 m/s lower than the measured wind speed. We also find a seasonal bias that correlates with surface roughness length used by the reanalysis model during the spring season. The corrections we propose reduce the average bias of reanalysis wind speed extrapolated to hub height to nearly zero, an improvement of 0.3–0.9 m/s. These corrections also reduce the RMS error by 0.1–0.4 m/s, a small improvement compared to the uncorrected RMS errors of 1.5–2.4 m/s.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Reanalysis models interpolate in time and space between historical meteorological measurements and estimate the values of unmeasured quantities. Reanalysis data are attractive for wind-power studies because they can offer wind speed data for large areas (sometimes the entire world) and long time periods and in locations where historical data are not available. The current generation of reanalysis models, the Modern-Era Retrospective Analysis for Research and Applications (MERRA), Climate Forecast System (CFS), and the ECMWF Reanalysis (ERA-Interim) estimate meteorological variables with spatial resolutions of 0.313° at the equator (0.7° for ERA) [1–3]. They estimate wind speeds at various heights: all offer 10 m above ground level and pressure levels; MERRA offers 50 m above ground as well. These models also estimate variables useful for modeling the atmospheric stability.

Previous research has used reanalysis data for large-scale analyses of wind power, such as wind resource assessment [4,5], long-term trends in wind speed [6,7], daily to yearly variability [8–10], and, extreme wind events [11,12]. However, the spatial resolutions

of current reanalysis models limit their accuracy in areas with complex terrain [13] and the temporal resolutions limit their accuracy for periods less than one day. Also, the reanalyses are likely to be less accurate in areas with few meteorological stations, such as high latitudes and parts of Africa and the Southern Hemisphere.

Many studies have addressed some of the limitations of reanalysis data by dynamically downscaling the data using physics-based models with better spatial and temporal resolution. For example, the U.S. National Renewable Energy Laboratory commissioned two large studies based on dynamically-downscaled wind [14,15]. However, dynamic downscaling requires significant computing time. Also, dynamically-downscaled models may be only slightly more accurate than reanalysis models for applications such as wind resource assessments and long-term (month to year) average wind speeds. Sharp et al. say “In some respects, raw CFS reanalysis data is also very close to the results obtained when using downscaled data, although accuracy appears to vary dependent on location.” [16].

In this paper, we characterize differences between instantaneous wind speeds from CFS reanalysis and 1-h average historical wind speeds greater than 2 m/s at heights of 10–100 m from sites in the U.S. Great Plains. Our goal is to quantify these differences and develop corrections to improve the accuracy and precision of reanalysis data without dynamic downscaling. We develop a model of the bias and uncertainty of CFS reanalysis wind speed

* Corresponding author. Current address: Humphrey School of Public Affairs, University of Minnesota, Minneapolis, MN, USA.

E-mail address: roses@umn.edu (S. Rose).

than can be used to correct the data. The model we propose also quantifies the contributions of various sources of error, which may suggest improvements to the reanalysis model. These results are specific to the CFS reanalysis and the U.S. Great Plains, but we expect the underlying methods will be applicable to other reanalyses.

2. Method

We use data from the CFS reanalysis [2] to calculate horizontal wind speeds at typical wind turbine heights using two different extrapolation methods. We then fit linear mixed-effects (LME) models to the data. Those LME models both allow us to correct reanalysis wind speeds for systematic biases and uncertainties, and to quantify the sources of both. Linear mixed-effects models, which are also called hierarchical, multilevel, or random-effects models estimate a model that allows for both between-site and within-site variation [17]. We fit the LMEs to the data described in Sections 2.1 and 2.2 using the “nlme” package, version 3.1–120 for R, release 3.1.3 [18,19].

LME models have not previously been applied to wind speeds, although one study develops a hierarchical model (a synonym for LME) of the distribution of extreme wind speeds [20]. Several studies statistically downscale wind speeds measured or modeled at lower spatial resolutions using Bayesian hierarchical models, which are closely related to LME models [21,22]. Finally, several studies have used LME models to study the effects of policies on wind energy [23,24] or the effects of wind farms on wildlife [25].

2.1. Empirical data

We validate the extrapolated reanalysis wind speeds against historical measurements from the U.S. Great Plains. These historical data are 1-h average horizontal wind speeds measured at heights between 10 m and 100 m above ground level (AGL). These measurements are taken at 162 sites shown in Fig. 1, with base elevations of 133–1463 m above mean sea level (AMSL). The data were collected by various government agencies in eleven U.S. states (MN, ND, OK, IA, WI, NE, MO, SD, CO, KS, IL) and compiled into a single database by the University of North Dakota Energy & Environmental Research Center, which performed quality control on the data [26]. Many of the measurements were collected by anemometers attached to existing telecommunications towers, others were collected by temporary 20-m towers from state and federal anemometer loan programs. We randomly divide the measurement sites into a “training” group of 109 sites and a “validation” group of 53 sites listed in the online Supporting Information (see Table 1). We exclude certain periods of bad data from some sites listed in the online Supporting Information (see Fig. 2).

Some towers have two anemometers at the same measurement height; in those cases we select the maximum of the two measured wind speeds at each time step on the assumption that the lower measurement is partially shadowed by the tower. We exclude data from sites within approximately 3 km of mountains or significant terrain features, sites with inhomogeneous fetch within 1 km (as judged from satellite photos), and sites within 1 km of trees or buildings.

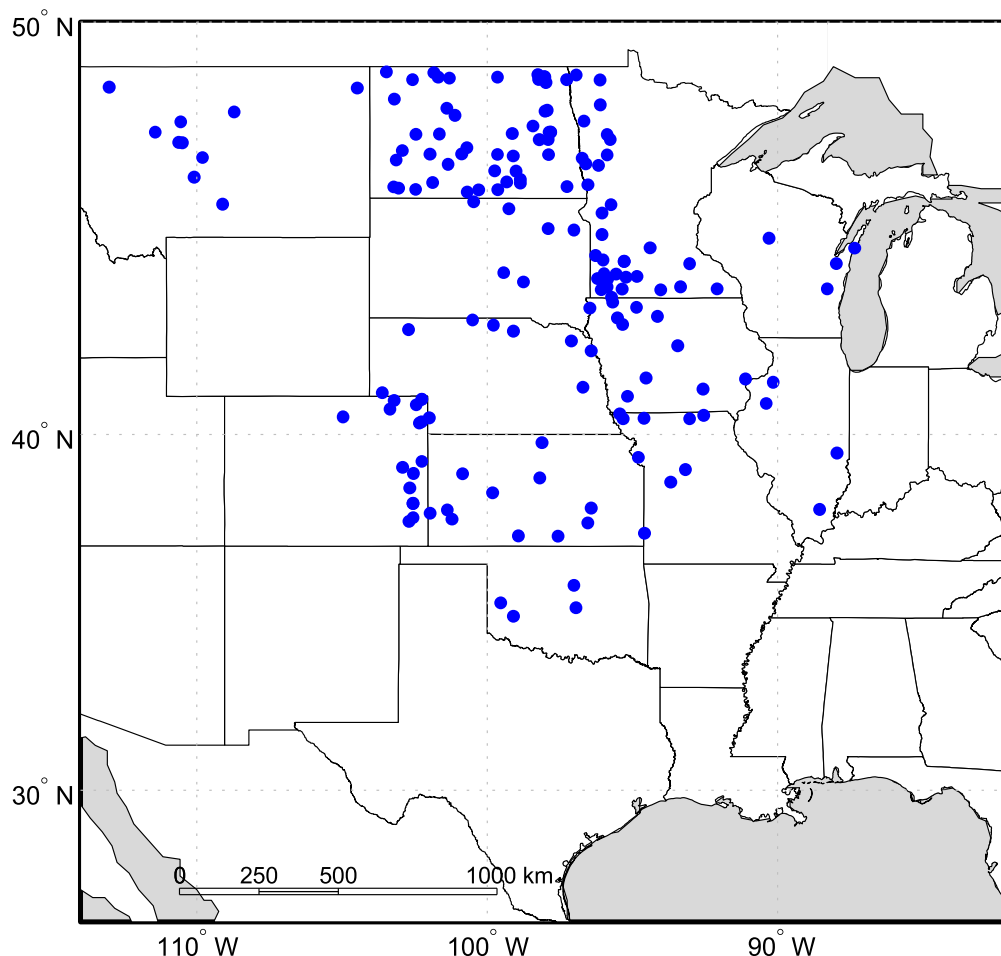


Fig. 1. Locations of the measurement sites in the U.S. Great Plains.

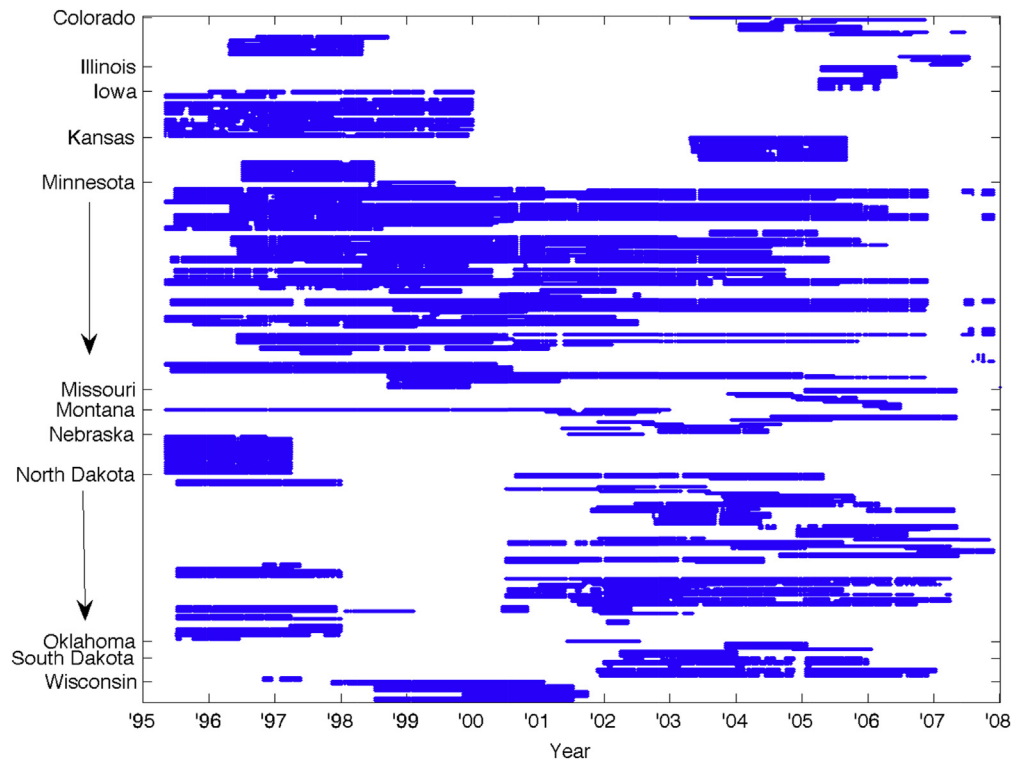


Fig. 2. Temporal coverage of measured wind speed data, grouped by state. In this paper we are concerned with comparing these measured data with corresponding periods of reanalysis data.

Table 1

Summary statistics for measured data used in this analysis. Each sampling site may have multiple time series collected at different measurement heights.

	Sampling sites	Time series	Base elevation (AMSL)	Measurement height (AGL)	Mean wind speed
Training	109	227	133–1463 m	10–100 m	3.4–9.9 m/s
Validation	53	110	254–1390 m	10–80 m	4.4–8.8 m/s

2.2. Reanalysis data

The reanalysis data we use in this work come from the CFS reanalysis, which uses numerical weather prediction models to interpolate meteorological measurements to a grid of locations [2]. Specifically, we analyze 10-m horizontal wind speed with a temporal resolution of 1 h (analysis and 1–5 h forecasts) for the period 1995–2008 and a spatial resolution of ~ 35 km at the equator (T382 grid). These wind speeds are state variables, which represent instantaneous samples of wind speed [27]. We discuss the consequences of comparing instantaneous wind speeds to 1-h averages of measured wind speed in Section 3.2. When the locations of historical measurement sites (the towers) do not coincide with the reanalysis data grid points we bilinearly interpolate the reanalysis variables to the historical site location using the 4 nearest grid points. The CFS raw data are well-fitted by Weibull distributions with scale parameters in the range 4.1–5.9 and shape parameters in the range 2.0–2.4. These shape parameters are within the commonly-observed range of 1.6–2.4 [28].

The CFS reanalysis model we analyze is similar to the MERRA model, which is commonly used in wind power studies because one of its standard outputs is wind speed at 50-m height. However, we expect MERRA would yield results similar to those we find in this work because the two models are closely-related; Sharp et al. say, “CFSR and MERRA are based on the same set of observations

and use similar models to extrapolate these data over space and time, to the same temporal scope.” [16].

2.3. Extrapolating reanalysis wind speed to hub height

We analyze reanalysis wind speed extrapolated from 10-m height using two commonly used methods: a logarithmic vertical wind speed profile with surface roughness length z_0 taken directly from the reanalysis model [29] and a power-law vertical profile with an exponent inferred from reanalysis wind speeds above and below the desired height [12,30,31]. We attempted to interpolate reanalysis wind speeds from the three lowest model layers assuming a logarithmic profile, but did not use that approach because it could interpolate only speeds that increased monotonically with height. The reanalysis data contain periods of speeds that do not increase monotonically with height, which is physically realistic in stable atmospheric conditions [32]. However, we were able to associate only some of those periods with stable conditions.

2.3.1. Neutral logarithmic profile with reanalysis surface roughness

The neutral logarithmic vertical wind profile is a function of surface roughness z_0 and wind speed u at some reference height z_{ref} , typically 10 m [33]. This vertical profile gives wind speed at an arbitrary height $u(z)$ as:

$$\bar{u}(z) = \bar{u}(z_{\text{ref}}) \frac{\ln(z/z_0)}{\ln(z_{\text{ref}}/z_0)} \quad (1)$$

We use the location- and time-specific values of the surface roughness length z_0 from the reanalysis data (see, for example, Fig. 6), as used by Huang et al. [29].

2.3.2. Power law profile

The power-law vertical wind profile is a function of wind speed u at some reference height z_{ref} , typically 10 m, and a power-law exponent α [33]. This vertical profile gives wind speed at an arbitrary height $u(z)$ as:

$$\bar{u}(z) = \bar{u}(z_{\text{ref}}) \left(\frac{z}{z_{\text{ref}}} \right)^\alpha \quad (2)$$

Some previous work has estimated varying values of α from wind speeds at different heights [30,31]. We estimate α at each hour and each site similarly, using the closest reanalysis wind speeds above u_{hi} and below u_{lo} the desired height according to the following formula:

$$\alpha = \frac{\ln\left(\frac{u_{\text{hi}}}{u_{\text{lo}}}\right)}{\ln\left(\frac{z_{\text{hi}}}{z_{\text{lo}}}\right)} \quad (3)$$

In the cases that z_{lo} is lower than 10 m above ground level, we substitute the reanalysis wind speed at 10 m u_{10} . We exclude extrapolated wind speeds less than 0 m/s and greater than 30 m/s because they are physically-unrealistic artifacts from the process of inferring the exponent α .

3. Results

We present two related results: First, we fit LME models to reanalysis wind speed extrapolated to hub height and use those models to correct for biases and uncertainties in the reanalysis data. Second, we fit similar models to raw reanalysis wind speed outputs in order to analyze the sources of bias and uncertainty in the reanalysis model.

3.1. Bias and uncertainty correction for reanalysis wind speed at hub height

We propose a LME model correct the biases and uncertainties in the reanalysis-predicted wind speed. The general model, given in (4), estimates measured horizontal wind speed at a given site as a function of reanalysis-predicted wind speed, month of year, altitude of the site, and measurement height. We fit this model to data from the “training” data set described in Section 2.1 and then test the fitted model by using it to correct data from the “validation” data set (see Table 1).

$$\begin{aligned} y_{ij} &= b_1 x_{ij} + b_2 h_j + b_3 z_j + \alpha_j + \epsilon_{ij} \\ a_j &\sim \mathcal{N}(0, \sigma_a) \\ \epsilon_{ij} &\sim \mathcal{N}(0, \sigma_\epsilon) \end{aligned} \quad (4)$$

with the following fixed effects:

y_{ij} measured wind speed at site j and time i [m/s]
 x_{ij} reanalysis-predicted wind speed at site j and time i [m/s]
 Separate groups for months Mar., Apr., May, June, and one group for July–Feb.

h_j altitude of site j [m above sea level]
 z_j measurement height of site j [m above ground level]

and the following random effects:

α_j between-site error, drawn from normal dist. with mean 0 and std. dev. σ_α
 ϵ_{ij} Residual error, drawn from normal dist. with mean 0 and std. dev. σ_ϵ

Previous studies have used several methods for extrapolating reanalysis wind speed to hub height and have investigated wind speeds averaged over different periods, so we fit families of models to data with different extrapolation methods and averaging periods. We fit models to wind speeds averaged over 1, 3, 6, 9, 12, 18, and 24 h and reanalysis wind speeds extrapolated vertically using two methods (described in Section 2.3), for a total of 14 models. We fit these models to all measurements higher than 10 m above ground level and wind speeds greater than 2 m/s. The unextrapolated 10-m wind speeds are qualitatively different: the ratio of measured to reanalysis wind speed is consistently greater than 1 for the 10-m data and consistently less than 1 for the extrapolated data. Fitted models for 10-m data are included in Section 3.2. We exclude reanalysis-predicted wind speeds less than 2 m/s in order to fit the model using Ordinary Least Squares (OLS) regression. The wind speed data are censored at 0 (i.e. wind speeds cannot be lower than zero), which biases OLS estimates. Censored LME regression gives unbiased estimates but is orders of magnitude more computationally intensive than OLS regression. We validate the approach of using OLS excluding low wind speeds by using both methods to fit models to a random subset of the data and confirming that they estimate the same parameter values.

In Table 2 we present coefficients of the model fitted to 6 h-average wind speeds. These models have a conditional R^2 value of approximately 0.66, which means the model explains 66% of the variance of the data [34]. This value is similar to R^2 values estimated in previous studies for regression models fit to hourly reanalysis data from individual sites. Liléo et al. calculate individual-site R^2 values of 0.40–0.80 for CFS reanalysis data and 0.56–0.78 for MERRA reanalysis data at 0.995 sigma level (~42 m above ground). Carvalho et al. estimated an average R^2 value of 0.61 for CFS reanalysis data 60–80 m height above ground [13]. The Crown Estate estimates individual-site R^2 values of 0.69–0.93 for MERRA reanalysis data at 50-m height [35]. Our results compare well to these in spite of the fact that we estimate a more general model for many sites that explains some of the variation between sites.

The fitted coefficients in Table 2 that relate reanalysis to measured wind speed ($b_1 - b_5$) are larger than the slope of 0.80 estimated previously by Rose and Apt for quarterly wind energy [10]. There are two differences between the model in (4) and the model proposed by Rose et al. that explain the difference in estimated slopes. First, wind energy is proportional to wind speed squared for the typical range of wind speeds and the slopes coefficients $b_1 - b_5$ are less than 1, so we expect the previously-estimated slope for wind energy to be smaller than the slopes estimated for wind speed. Second, the model in (4) includes terms for site elevation and measurement height that explain more of the bias than the single slope term in the previously-proposed model.

In Fig. 3 we plot the same model overlaid on the corresponding data for the months July–February: the left figure plots measured vs. reanalysis wind speed for all sites with overlaid with the model calculated for average values and the right figure plots data for a single site overlaid with models calculated for average and site-specific conditions. These results are representative of the results fitted to other averaging periods; the detailed results are given the

Table 2

Coefficients for the model in (4), fitted to data averaged over 6 h. All coefficients are statistically significant at a < 1% level.

		Log, model roughness (z_0)	Power-law, inferred alpha (α)
Slope	July–Feb (b_1)	0.92	0.89
	March (b_2)	0.96	0.94
	April (b_3)	0.97	0.97
	May (b_4)	0.99	0.98
	June (b_5)	0.95	0.93
	Elevation (b_6)	1.40e-3 (m/s)/m	1.61e-3 (m/s)/m
	Meas. height (b_7)	0.011 (m/s)/m	0.017 (m/s)/m

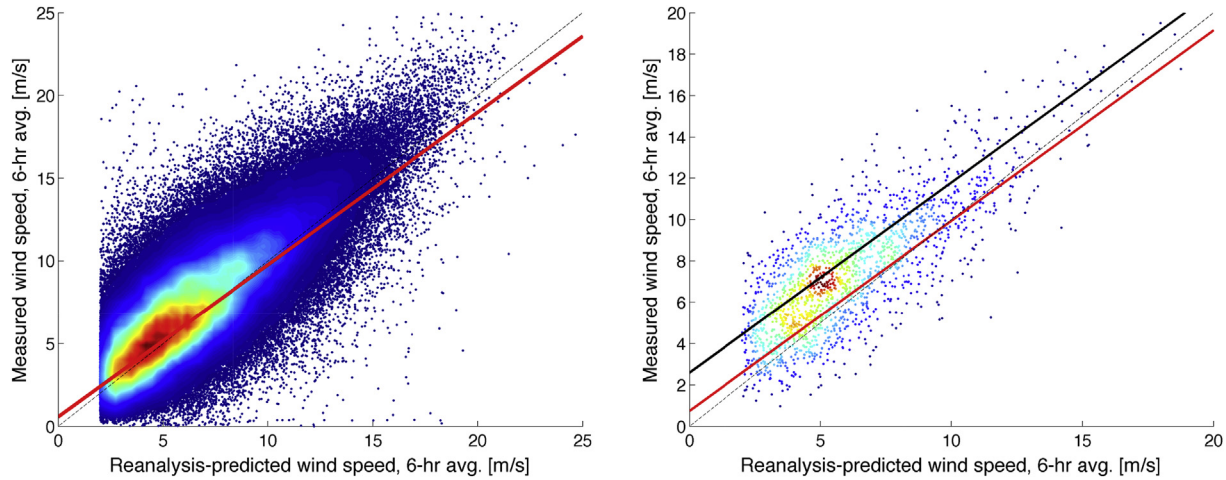


Fig. 3. Measured vs. reanalysis-predicted 6-h average wind speeds from July–February, extrapolated with log profile and modeled roughness. Left: data from all sites at measurement heights greater than 10 m. Right: data from site 158, with a 1463-m elevation and 50-m measurement height. Both are overlaid with a solid red line representing the best-fit model for average site characteristics and a dashed line showing a 1:1 relationship. The single-site data on the right is also overlaid with a solid black line representing the model that includes adjustments for the elevation and measurement height of that specific site. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

online [Supporting Information](#).

The fitted models in Table 2 reveal that the reanalysis wind speeds generally under-predict measured wind speeds in the typical operating range and typical locations of wind turbines in the U.S. Great Plains. The under-prediction increases with site altitude and turbine hub height and during the Spring (March–June) but decreases with wind speed. For example, at a site with an elevation of 500 m and a hub height of 80 m, a reanalysis wind speed of 4 m/s under-predicts by an average of 0.3–1.4 m/s, depending on the season and vertical profile used for extrapolation. However, a reanalysis wind speed of 14 m/s can range from an average over-prediction of 0.8 m/s to an average under-prediction of 1.2 m/s. Finally, Fig. 3 shows that the model fits the data better at moderate speeds (e.g. 3–8 m/s) where the data density is high than at high speeds (e.g. > 14 m/s) where the data density is low. However, a poor fit at wind speeds higher than approximately 14 m/s is not important for wind power applications because typical turbines reach their maximum rated power output in that speed range.

The model given in Eq. (4) can be used to correct for the biases and reduce the uncertainties in reanalysis-predicted wind speed. For example, the correction for log-extrapolated 6-h average wind speed in April at 1000 m above sea level and 80 m above ground is $0.97x + (1000 \times 1.40 \times 10^{-3}) + (80 \times 0.011)$, where x is the reanalysis wind speed extrapolated from 10-m to hub height. To quantify the effect of this correction, we use the model in (4) to calculate “corrected” wind speed from raw reanalysis data extrapolated to the desired measurement height and then calculate the bias (reanalysis minus measured speed) and root-mean squared error (RMSE) relative to actual measured wind speed at the same height. The models are fitted to the “training” data set but

the bias and RMSE are calculated with the “validation” data set. Fig. 4 and Fig. 5 show the effects of this correction: it reduces the mean bias to nearly zero, but reduces the RMSE only modestly.

Fig. 4 shows that the LME model we propose is quite effective in reducing the average wind speed bias to nearly zero: the uncorrected biases are -0.9 to -0.4 m/s, but the biases after applying our model are -0.1 – 0.2 m/s. Fig. 5 shows that the RMSE decreases with longer averaging periods, but applying the correction we propose decreases RMSE for all averaging periods. The RMSE is higher for higher ranges of wind speed, but this is not surprising because the RMSE units are absolute, rather than relative. Together, these results show that our proposed model significantly improves the accuracy of reanalysis-predicted wind speed.

3.2. Sources of bias and uncertainty in 1-hr reanalysis data at 10-m height

The model we propose in (4) offers a way to correct for bias and uncertainties in the reanalysis wind speeds, but it also quantifies the sources of bias and uncertainty. Identifying those sources is may offer insights into the way the CFS reanalysis models wind speed in the boundary layer. In order to analyze the sources of bias and uncertainty independent of the vertical extrapolation method, we fit the slightly different LME model given in (5) to unextrapolated 10-m wind speeds greater than 2 m/s, which are the raw output of the reanalysis model.

$$\begin{aligned}
 y_{ij} &= b_{1...5}x_{ij} + b_6h_j + b_{7...9}\Psi + \alpha_j + \epsilon_{ij} \\
 \alpha_j &\sim \mathcal{N}(0, \sigma_\alpha) \\
 \epsilon_{ij} &\sim \mathcal{N}(0, \sigma_\epsilon)
 \end{aligned}
 \quad (5)$$

This model given in (5) removes the term for measurement height z because all measurements have the same height and because it is not statistically significant in the model. However, it adds an additional term Ψ related to atmospheric stability, to assess whether stability contributes significantly to wind speed error. This term Ψ , defined in (6), is a standard correction for atmospheric stability in the logarithmic vertical wind speed profile:

$$\Psi = \begin{cases} 2 \ln\left(\frac{1+a}{2}\right) + \ln\left(\frac{1+a^2}{2}\right) - 2 \arctan(a) + \pi/2 & \text{if } -10 \leq \frac{z}{L} < 0 (\text{unstable}) \\ -5\frac{z}{L} & \text{if } 0 \leq \frac{z}{L} < 0.5 (\text{stable}) \\ -\left(\frac{z}{L} + 0.66\left(\frac{z}{L} - 14.3\right)\exp\left(-0.35\frac{z}{L}\right) + 9.52\right) & \text{if } 0.5 \leq \frac{z}{L} < 7 (\text{very stable}) \end{cases} \quad (6)$$

where $a = (1 - 16z/L)^{1/4}$ [32]. It consists of three different sub-terms for unstable, stable, and very stable atmospheric conditions. There is no explicitly defined sub-term for neutral stability, where the correction is zero, because neutral stability corresponds to $z/L = 0$. The “stable” term yields a correction of 0 when $z/L = 0$. The online [Supporting Information](#) gives details of the calculation of the Obukhov length L from reanalysis data.

We fit the model given in (5) to 1-h average wind speeds measured at 10-m height and summarize the fitted coefficients in Table 3.

The most important terms in the model in (5) are the “slope” parameters $b_1 - b_5$, which represent the ratio of measured wind speed to reanalysis wind speed. The fitted values in Table 3 show that the reanalysis model consistently underestimates measured 1-h average wind speeds at 10-m height and that the underestimate is larger in Spring months (March–June). The slopes range from 1.17 for non-Spring months (b_1) to 1.35 for the month of May (b_4). Although no previous studies we are aware of have explicitly reported the slope of fitted regression models, several studies have shown results consistent with the positive slope like we find for 10-m data. Carvalho et al. write, “There seems to be a somewhat linear variation of the bias [where bias is reanalysis minus measured speed] with the measured wind speed: for low wind speeds the bias tends to be positive and high, gradually diminishing with increasing measured wind speed and for strong wind speeds the biases are now negative and again high in value.” [13] Cannon et al. plot a linear least square fit that shows the same trend we find, but do not report the fitted model parameters [12].

We test several hypotheses to explain why 10-m reanalysis wind speeds underestimate measured wind speeds. First, the measured wind speeds are averaged over 1 h, but the reanalysis wind speeds to which we compare them are “state variables” (i.e. instantaneous samples) [27]. If this were a source of error, it should decrease as both data sets are averaged over longer periods, but we fit similar models to data averaged over much longer periods (days – months) and found consistent underestimation. Second, Carvalho et al. suggest that the simplified terrain used in reanalysis models can cause underestimation of wind speed. To test this hypothesis, we added two predictor variables to the model in (5): one for site elevation relative to the average elevation within a 3-km radius and one for standard deviation of terrain elevations within 3 km, but the coefficients for those variables were not statistically significant. Third, the underestimation of measured wind speeds has a seasonal pattern that may be caused by the surface roughness length models used in the reanalysis data. Fig. 6 shows that the increase in ratio of

measured to reanalysis wind speed coincides with the springtime increase in surface roughness length used in the reanalysis model. We estimated surface roughness lengths from measured wind shear (speeds at various measurement heights) assuming a logarithmic vertical profile, but the values we estimated were too noisy to meaningfully compare to the roughness lengths from the reanalysis model.

The “elevation” parameter b_6 is marginally statistically significant ($p = 0.047$) and the magnitude is small, which suggests that bias in reanalysis wind speed is not strongly influenced by effects correlated with elevation. For example, the fitted coefficient of 2.53×10^{-4} (m/s)/m means that the reanalysis data underestimate wind speed by an average of 0.38 m/s at the highest-elevation site we analyze (1494 m). However, the coefficient for the elevation parameter estimated from vertically-extrapolated wind speeds in Section 3.1 is statistically significant ($p \approx 0$) and the magnitude is almost an order of magnitude larger (1.42×10^{-3}). To understand why, we re-fit the model in (5) to 10-m data without the terms for stability ($b_7 - b_9$) in order to make the estimated coefficients more comparable to the model in (4). Fitting this reduced model to the 10-m data gives an estimate of 6.0×10^{-4} (m/s)/m for the elevation coefficient b_6 , which is closer to the estimates for extrapolated data in Table 2. Sharp et al. find that reanalysis errors are larger for sites above 600 m because average wind speed increases with altitude [16]. We also find that reanalysis errors are larger for sites with higher average measured wind speeds, but we do not find a clear relationship between average reanalysis wind speed and error.

We fit separate coefficients for each of the three sub-terms of Ψ : b_7 for unstable conditions, b_8 for stable conditions, and b_9 for very stable conditions. The coefficients for the unstable (b_7) sub-term is statistically significant and the magnitude is relatively large. For example, the fitted value of b_7 (0.666) means that the reanalysis wind speed underestimates measured wind speed by approximately 1.6 m/s in very unstable ($z/L = -10$) conditions. This suggests that the reanalysis model does not model wind speeds accurately in unstable atmospheric conditions. This is consistent with our finding that the solar zenith angle [36] lagged 3 h behind local time is a statistically-significant predictor variable. The coefficient for the stable (b_8) sub-term is statistically significant (see Table 3), but the magnitude is small so we do not think the reanalysis model is significantly inaccurate in stable (and neutral) conditions.

4. Conclusions

This paper characterizes the differences between CFS reanalysis and measured wind speeds to correct for biases and uncertainties in the reanalysis data. Previous studies calculate measures of error (e.g. RMSE) and correlation between reanalysis and measured wind speed, but do not offer a way to correct the errors. Previous research by Rose and Apt [10] proposes a simple linear model to correct biases in quarterly wind power and a novel correction for errors

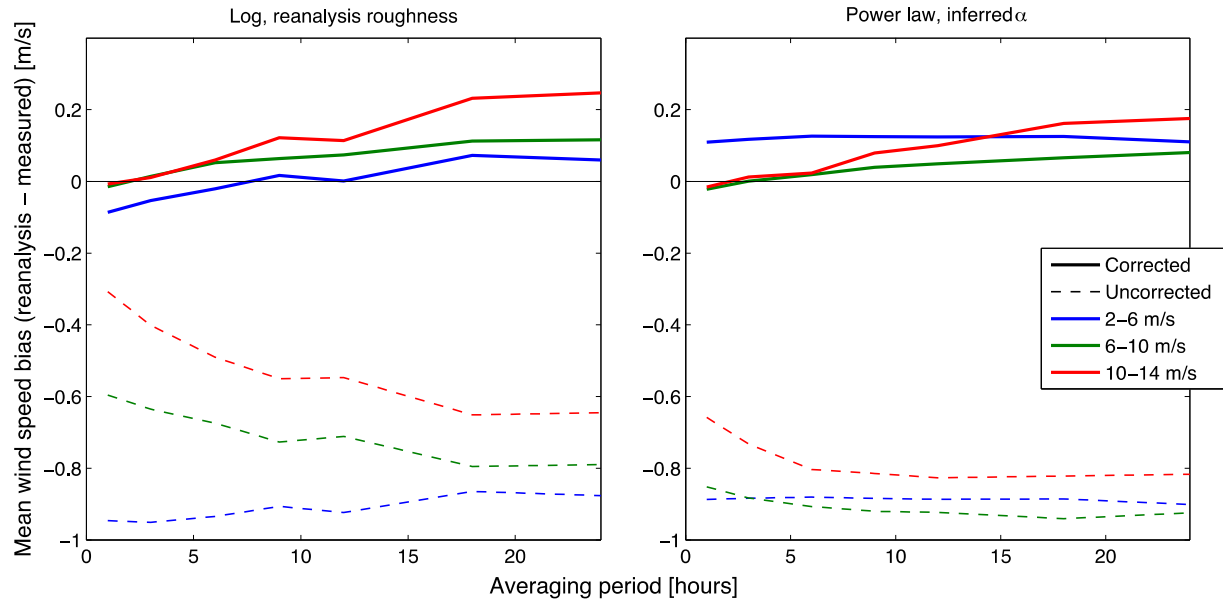


Fig. 4. Mean bias (defined as reanalysis minus measured) of reanalysis-predicted wind speed, with and without correction from the model we propose. Dashed lines plot bias without correction and solid lines plot bias after correction in Eq. (4) has been applied.

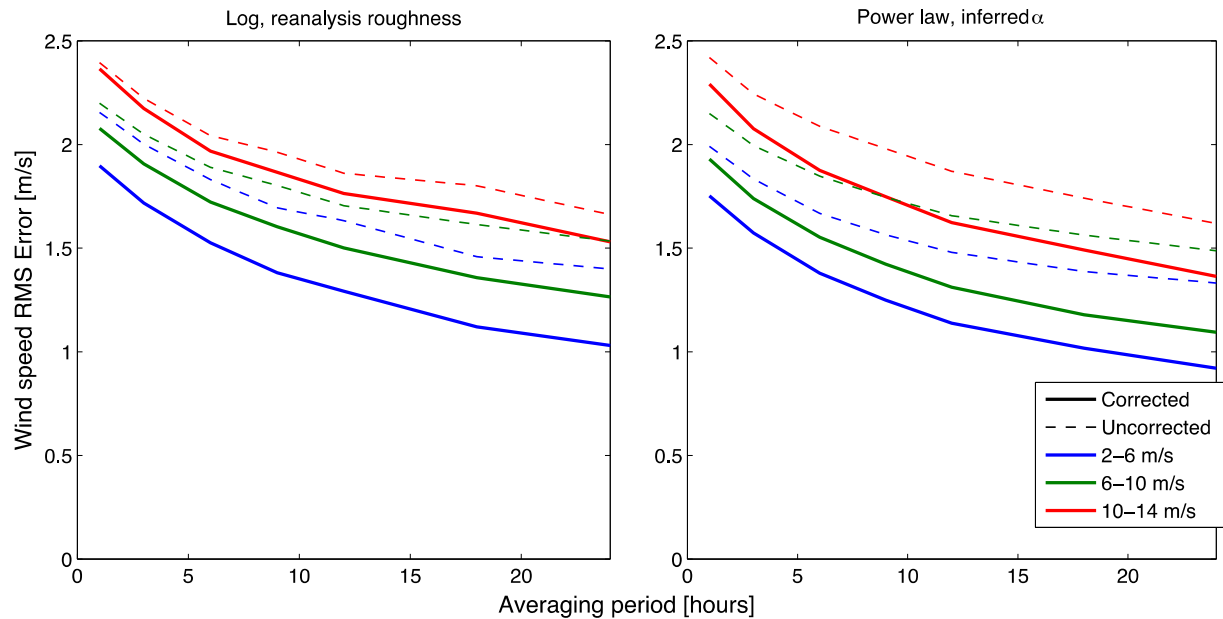


Fig. 5. RMS error of reanalysis-predicted wind speed, with and without correction from the model we propose. Dashed lines plot bias without correction and solid lines plot bias after correction in Eq. (4) has been applied.

Table 3

Coefficients for the model in (5), fitted to 1-h average data measured at 10-m height.

		Coefficient	Std. err.	DF	t-value	p-value
Slope	July–Feb (b_1)	1.17	1.00e-3	1.22e6	1173	0
	March (b_2)	1.26	1.31e-3	1.22e6	961	0
	April (b_3)	1.30	1.40e-3	1.22e6	925	0
	May (b_4)	1.35	1.48e-3	1.22e6	915	0
	June (b_5)	1.30	1.65e-3	1.22e6	786	0
	Elevation (b_6)	2.53e-4 (m/s)/m	1.25e-4	56	2.03	0.047
Ψ	Unstable (b_7)	0.666	8.98e-3	1.22e6	74.2	0
	Stable (b_8)	−0.0616	4.10e-3	1.22e6	−15.0	0
	Very stable (b_9)	−2.08e-3	3.94e-3	1.22e6	−0.53	0.60

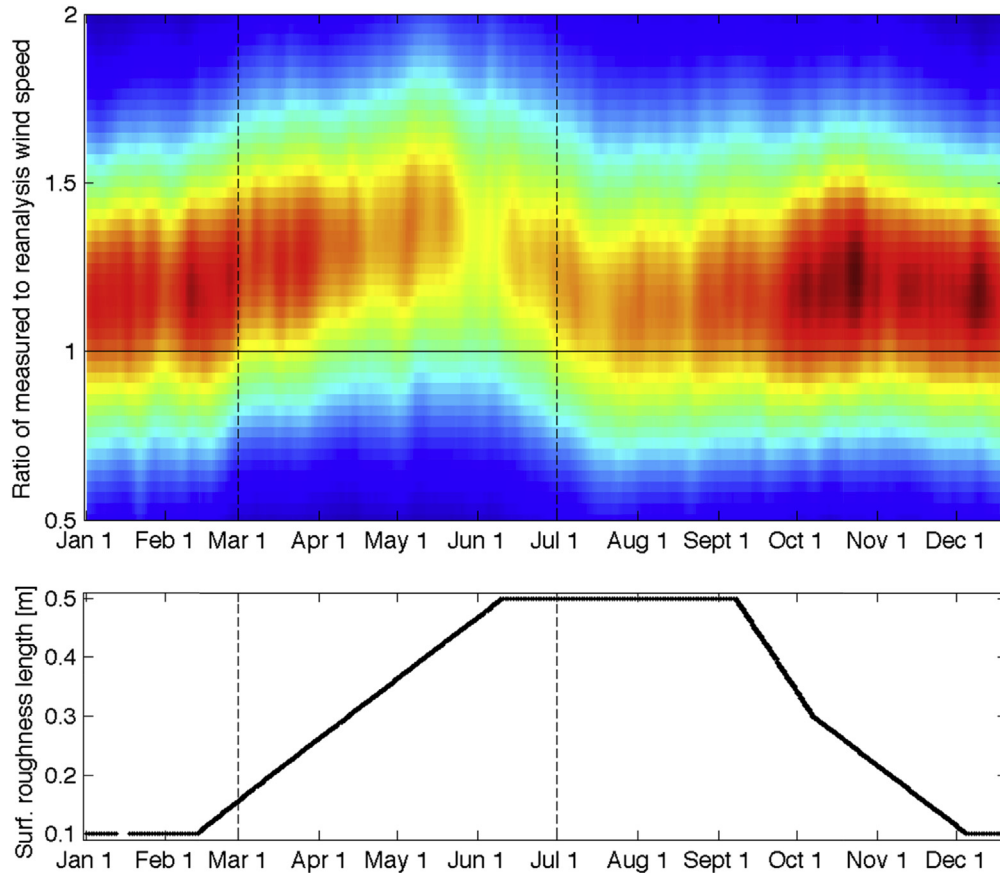


Fig. 6. The ratio of measured to reanalysis-predicted wind speeds at 10-m height as a function of day of year (top) and typical surface roughness length profile used in the reanalysis model (bottom). The vertical dashed lines show the start and end dates of the month-specific coefficients $b_2 - b_5$.

introduced by vertically extrapolating wind speeds in unstable and stable atmospheric conditions. In the current paper we propose more sophisticated linear models that take into account altitude, measurement height, and seasonal variations in the relationship between reanalysis and measured wind speeds.

The corrections we propose reduce the average bias of reanalysis wind speed extrapolated to hub height by 0.3–0.9 m/s, a large improvement that makes the average bias of the corrected wind speeds near zero. The corrections we propose also reduce the RMS error by 0.1–0.4 m/s, which is a small improvement. Averaging reanalysis wind speeds over several hours further reduces the RMS error, but the errors are still large. For example, the RMS error for 24-h average wind speeds after our corrections have been applied is still 1–1.5 m/s, a significant fraction of the mean wind speed.

The fitted coefficients of our model show that the reanalysis model under-predicts wind speeds more at higher elevations (e.g. 1.4–1.6 m/s at 1000 m above sea level) and more at higher measurement heights (e.g. 1.1–1.7 m/s at 100 m above ground level). For example, the correction for log-extrapolated 6-h average wind speed in April at 1000 m above sea level and 80 m above ground is $0.97x + (1000 \times 1.40 \times 10^{-3}) + (80 \times 0.011)$, where x is the reanalysis wind speed extrapolated from 10-m to hub height. These results are relatively consistent for two different methods of extrapolating reanalysis wind speeds 10 m to higher heights above ground level, though we estimate a smaller magnitude of under-prediction using the raw 10-m wind speeds output by the reanalysis.

The coefficients of our fitted models show that the ratio of measured to reanalysis wind speeds is higher during the Spring

season than the rest of the year. For the reanalysis data extrapolated to hub height, where the ratio of measured to reanalysis wind speeds is less than 1, the ratio becomes nearer to 1 during the Spring. For the un-extrapolated reanalysis data where the ratio is greater than 1, the ratio grows farther from 1 in the Spring. We show this seasonal pattern is correlated with the springtime increase in surface roughness lengths used by the reanalysis model. However, surface roughness lengths we inferred from the measured wind speed data were too noisy to fit a function relating them to the modeled roughness lengths.

Finally, the coefficients of our fitted models also show that the reanalysis wind speeds under-predict measured wind speeds in unstable atmospheric conditions. The magnitude of this under-prediction as a function of the z/L stability criterion is consistent with the stability correction for unstable conditions proposed by Paulson [37]. In cases where z/L cannot be estimated, we find that the solar zenith angle, lagged 3 h behind local time, can be substituted as a predictor variable in the model.

4.1. Extensions of this work

This work compares wind speeds measured over relatively flat, homogeneous terrain in the U.S. Great Plains to data from the CFS reanalysis. The Great Plains region we analyze is an important one, because it contains the majority of wind power development in the United States. However, a similar analysis of other important wind-power regions, such as the plains of northern Europe, would help validate the model. A similar analysis comparing wind speeds predicted by a different reanalysis model, such as MERRA, would be

useful to determine whether different reanalyses use better sub-models of the phenomena that we find affect the bias and uncertainty.

Acknowledgements

This work was supported by grants to the RenewElec project by the Richard King Mellon Foundation (5562) and the Doris Duke Charitable Trust. This research was also supported in part by the Climate and Energy Decision Making (CEDM) center, created through a cooperative agreement between the National Science Foundation (SES-0949710) and Carnegie Mellon University. We are grateful to George Young and John Zack for their meteorological insights and advice, and Fallaw Sowell and Stephen Karolyi for their help with mixed-effects models, Bradley Stevens for providing the measured data, and Bob Dattore for providing the reanalysis data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.renene.2016.03.028>.

References

- [1] M.M. Rienecker, M.J. Suarez, R. Gelaro, R. Todling, Julio Bacmeister, E. Liu, et al., MERRA: NASA's modern-era retrospective analysis for research and applications, *J. Clim.* 24 (2011) 3624–3648, <http://dx.doi.org/10.1175/JCLI-D-11-00015.1>.
- [2] S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, et al., The NCEP climate forecast system reanalysis, *Bull. Amer Meteor Soc.* 91 (2010) 1015–1057, <http://dx.doi.org/10.1175/2010BAMS3001.2>.
- [3] D.P. Dee, S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, et al., The ERA-interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor Soc.* 137 (2011) 553–597, <http://dx.doi.org/10.1002/qj.828>.
- [4] X. Lu, M.B. McElroy, J. Kiviluoma, Global potential for wind-generated electricity, *Proc. Natl. Acad. Sci.* 106 (2009) 10933–10938, <http://dx.doi.org/10.1073/pnas.0904101106>.
- [5] S. Liléo, O. Petrik, Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis, in: *Ewea 2011, Brussels, 2011*, pp. 1–18.
- [6] S.C. Pryor, R.J. Barthelmie, Assessing climate change impacts on the near-term stability of the wind energy resource over the United States, *Proc. Natl. Acad. Sci.* 108 (2011) 8167–8171.
- [7] S.C. Pryor, R.J. Barthelmie, D.T. Young, E.S. Takle, R.W. Arritt, D. Flory, et al., Wind speed trends over the contiguous United States, *J. Geophys. Res. Atmos.* 114 (2009), <http://dx.doi.org/10.1029/2008JD011416>.
- [8] S.C. Pryor, R.J. Barthelmie, J.T. Schoof, Inter-annual variability of wind indices across Europe, *Wind Energy* 9 (2006) 27–38, <http://dx.doi.org/10.1002/we.178>.
- [9] M.C. Brower, M.S. Barton, L. Lledó, J. Dubois, A Study of wind speed variability using global reanalysis data, *AWSTruepower* (2013).
- [10] S. Rose, J. Apt, What can reanalysis data tell us about wind power? *Renew. Energy* 83 (2015) 963–969, <http://dx.doi.org/10.1016/j.renene.2015.05.027>.
- [11] X.G. Larsén, J. Mann, Extreme winds from the NCEP/NCAR reanalysis data, *Wind Energy* 12 (2009) 556–573, <http://dx.doi.org/10.1002/we.318>.
- [12] D.J. Cannon, D.J. Brayshaw, J. Methven, P.J. Coker, D. Lenaghan, Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain, *Renew. Energy* 75 (2015) 767–778, <http://dx.doi.org/10.1016/j.renene.2014.10.024>.
- [13] D. Carvalho, A. Rocha, M. Gómez-Gesteira, C. Silva Santos, WRF wind simulation and wind energy production estimates forced by different reanalyses: comparison with observed data for Portugal, *Appl. Energy* 117 (2014) 116–126, <http://dx.doi.org/10.1016/j.apenergy.2013.12.001>.
- [14] M.C. Brower, Development of Eastern Regional Wind Resource and Wind Plant Output Datasets, National Renewable Energy Laboratory, Golden, CO, 2009.
- [15] C.W. Potter, D. Lew, J. McCaa, S. Cheng, S. Eichelberger, E. Gritmit, Creating the dataset for the western wind and solar integration study (U.S.A.), *Wind Eng.* 32 (2009) 325–338, <http://dx.doi.org/10.1260/0309-524X.32.4.325>.
- [16] E. Sharp, P. Dodds, M. Barrett, C. Spataru, Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information, *Renew. Energy* 77 (2015) 527–538, <http://dx.doi.org/10.1016/j.renene.2014.12.025>.
- [17] J.C. Pinheiro, D.M. Bates, Mixed-effects Models in S and S-PLUS, Springer-Verlag, New York, 2000, <http://dx.doi.org/10.1007/b98882>.
- [18] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R Core Team, nlme: Linear and Nonlinear Mixed Effects Models, (n.d.), <http://CRAN.R-project.org/package=nlme>.
- [19] R Core Team, R: a language and environment for statistical computing, <http://www.R-project.org/>, 2015.
- [20] L. Fawcett, D. Walshaw, A hierarchical model for extreme wind speeds, *Appl. Stat.* 55 (2006) 631–646, <http://dx.doi.org/10.2307/3879116>.
- [21] C.K. Wikle, R.F. Milliff, D. Nychka, L.M. Berliner, Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds, 96 (2011) 382–397, <http://dx.doi.org/10.1198/016214501753168109>.
- [22] T.J. Hoar, R.F. Milliff, D. Nychka, C.K. Wikle, L.M. Berliner, Winds from a Bayesian hierarchical model: computation for atmosphere-ocean research, *J. Comput. Graph. Statistics* 12 (2003) 781–807, <http://dx.doi.org/10.1198/1061860032616>.
- [23] C. Hitaj, Wind power development in the United States, *J. Environ. Econ. Manag.* 65 (2013) 394–410, <http://dx.doi.org/10.1016/j.jeem.2012.10.003>.
- [24] S. Jenner, F. Groba, J. Indvik, Assessing the strength and effectiveness of renewable electricity feed-in tariffs in European Union countries, *Energy Policy* 52 (2013) 385–401, <http://dx.doi.org/10.1016/j.enpol.2012.09.046>.
- [25] G.B. Stewart, A.S. Pullin, C.F. Coles, Poor evidence-base for assessment of windfarm impacts on birds, *Environ. Conserv.* 34 (2007) 1, <http://dx.doi.org/10.1017/S0376892907003554>.
- [26] T.K. Simonsen, B.G. Stevens, Regional wind energy analysis for the Central United States, in: *Global WINDPOWER 2004*, Chicago, 2004, pp. 1–16.
- [27] M. Decker, M.A. Brunke, Z. Wang, K. Sakaguchi, X. Zeng, M.G. Bosilovich, Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations, *J. Clim.* 25 (2012) 1916–1944, <http://dx.doi.org/10.1175/JCLI-D-11-00004.1>.
- [28] (Chapter 10) M.C. Brower (Ed.), *Wind Resource Assessment*, John Wiley & Sons, Inc., Hoboken, 2012.
- [29] J. Huang, X. Lu, M.B. McElroy, Meteorologically defined limits to reduction in the variability of outputs from a coupled wind farm system in the Central US, *Renew. Energy* 62 (2014) 331–340, <http://dx.doi.org/10.1016/j.renene.2013.07.022>.
- [30] E. Holt, J. Wang, Trends in wind speed at wind turbine height of 80 m over the contiguous United States using the North American Regional Reanalysis (NARR), *J. Appl. Meteorol. Clim.* 51 (2012) 2188–2202, <http://dx.doi.org/10.1175/JAMC-D-11-0205.1>.
- [31] S.M. Fisher, J.T. Schoof, C.L. Lant, M.D. Therrell, The effects of geographical distribution on the reliability of wind energy, *Appl. Geogr.* 40 (2013) 83–89, <http://dx.doi.org/10.1016/j.apgeog.2013.01.010>.
- [32] S. Emis, *Wind Energy Meteorology*, Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 2013, <http://dx.doi.org/10.1007/978-3-642-30523-8>.
- [33] M.C. Brower, *Wind Resource Assessment*, John Wiley & Sons, Inc., Hoboken, 2012.
- [34] S. Nakagawa, H. Schielzeth, A general and simple method for obtaining R^2 from generalized linear mixed-effects models, *Methods Ecol. Evol.* 4 (2013) 133–142, <http://dx.doi.org/10.1111/j.2041-210x.2012.00261.x>.
- [35] The Crown Estate, UK MERRA Validation with Offshore Meteorological Data, The Crown Estate, London, 2014.
- [36] I. Reda, A. Andreas, *Solar Position Algorithm for Solar Radiation Applications*, National Renewable Energy Laboratory, Golden, CO, 2008.
- [37] C.A. Paulson, The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer, *J. Appl. Meteorol. Clim.* 9 (1970) 857–861, [http://dx.doi.org/10.1175/1520-0450\(1970\)009<0857:TMROWS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2).