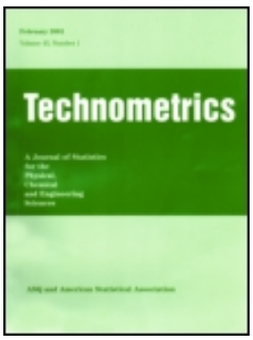

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: http://www.tandfonline.com/loi/utch20

Short-Term Wind Speed Forecast Using Measurements From Multiple Turbines in A Wind Farm

Arash Pourhabib, Jianhua Z. Huang & Yu Ding

To cite this article: Arash Pourhabib, Jianhua Z. Huang & Yu Ding (2016) Short-Term Wind Speed Forecast Using Measurements From Multiple Turbines in A Wind Farm, Technometrics, 58:1, 138-147, DOI: 10.1080/00401706.2014.988291

To link to this article: http://dx.doi.org/10.1080/00401706.2014.988291

Accepted author version posted online: 22 Jan 2015.
Published online: 22 Feb 2016.

Short-Term Wind Speed Forecast Using Measurements From Multiple Turbines in A Wind Farm

Arash POURHABIB

School of Industrial Engineering
and Management
Oklahoma State University
Stillwater, OK 74074
(*arash.pourhabib@okstate.edu*)

Jianhua Z. HUANG

Department of Statistics
Texas A&M University
College Station, TX 77843
School of Statistics
Renmin University of China
Beijing 100872, China
(*jianhua@stat.tamu.edu*)

Yu DING

Department of Industrial and Systems Engineering
Texas A&M University
College Station, TX 77843
(*yuding@iemail.tamu.edu*)

Turbine operations in a wind farm benefit from an understanding of the near-ground behavior of wind speeds. This article describes a probabilistic spatial-temporal model for analyzing local wind fields. Our model is constructed based on measurements taken from a large number of turbines in a wind farm, as opposed to aggregating the data into a single time-series. The model incorporates both temporal and spatial characteristics of wind speed data: in addition to using a time epoch mechanism to model temporal nonstationarity, our model identifies an informative neighborhood of turbines that are spatially related, and consequently, constructs an ensemble-like predictor using the data associated with the neighboring turbines. Using actual wind data measured at 200 wind turbines in a wind farm, we found that the two modeling elements benefit short-term wind speed forecasts. We also investigate the use of regime switching to account for the effect of wind direction and the use of geostrophic wind to account for the effects of meteorologic factors other than wind. These at best provide a small performance boost to speed forecast.

KEY WORDS: Multiple time-series; Near-ground wind; Probabilistic modeling; Short-term wind speed forecasts; Spatial-temporal models.

1. INTRODUCTION

This article is concerned with developing short-term wind speed forecasts that could potentially benefit turbine operations and control. Wind speed forecasting is commonly categorized as short-term, medium-term, and long-term. There is no sharp division between medium- and long-term forecasts, which can range from days to years. For medium- and long-term forecasts, physics-based numerical weather prediction (NWP) methods (Cassola and Burlando 2012) are typically used. Short-term forecasts, on the other hand, rely primarily on using past wind speed measurements and purely data-driven approaches (Genton and Hering 2007). The literature appears to agree that short-term forecasts refer to a few hours of prediction horizon for which data-driven, statistical models generally outperform the physics-based NWPs. For instance, 6 hours were mentioned by Giebel et al. (2011) as the separator between short-term and medium/long-term forecasts. Our article uses forecasts of 2 to 5 hr ahead.

There has been a rich body of literature on short-term wind speed forecasts; for comprehensive reviews, refer to Giebel et al. (2011) and Zhu and Genton (2012). This article focuses on turbine-specific wind speed forecasts, namely, that our model is constructed based on measurements coming from individual turbines in a wind farm, as opposed to aggregating the whole farm data into a single time-series. The need for turbine-specific wind speed forecasts arises from the need to control turbine operations. For instance, the concept of damage-mitigating control (Santos 2007) depends on the trade-off between the wind power generation and the level of stress exerting on the turbine structure by wind. As a turbine generates more power under high wind, the resulting high stresses can lead to reliability problems. Turbine operators employ pitch and yaw controls to reduce the maximum and accumulative stresses at sensitive locations, for example, at the root of a blade, although this results in some loss of production. To perform the damage-mitigating control (and many other types of control), operators need to know the types of wind conditions to anticipate in the near future. Once wind speed is predicted, the expected power generation can be estimated using, for instance, a nominal power curve converting wind speed to power (Kusiak, Zheng, and Song 2009). Given

TECHNOMETRICS, FEBRUARY 2016, VOL. 58, NO. 1
DOI: 10.1080/00401706.2014.988291

that such control must be performed on individual turbines, a turbine-specific wind speed forecast model is more helpful than using aggregated farm-level data.

This article presents a probabilistic spatial-temporal model for analyzing local wind fields. We use the spatial information through selecting the informative neighborhood of a turbine and then build an "ensemble learner" using the neighboring turbines. For handling nonstationarity in wind speed measurements, we employ a time epoch mechanism. We conduct numerical analysis using 2 years of wind and power data collected on 200 turbines in an inland wind farm and demonstrate the merit of the proposed model.

The remainder of this article is organized as follows. Section 2 reviews some of the widely used models for short-term wind speed forecasting. Section 3 presents three proposed models and discusses the selection of appropriate loss functions. Section 4 applies them and compares their performance with four commonly used models. Section 5 summarizes this article and offers suggestions for future research.

2. COMMONLY USED SHORT-TERM WIND SPEED FORECAST APPROACHES

In this section, we review some existing models for short-term wind speed forecasting. Let $Y_i(t) = Y(\mathbf{s}_i; t)$ denote the wind speed at time t measured at location $\mathbf{s}_i$ for $i = 1, 2, \ldots, I$. Let $\mathbf{Y}(t) = [Y_1(t), Y_2(t), \ldots, Y_I(t)]^T$. Suppose we observe the wind speed at locations $\mathbf{s}_i$, $i = 1, 2, \ldots, I$, and at times $t = 1, 2, \ldots, T$. We want to make an h-step ahead prediction, denoted as $\widehat{Y}_i(t + h)$. Assume the temporal resolution is in hours, so that h-step ahead means h-hour ahead. For the short-term forecasts made in this article, h is up to 5 hr.

We note that some studies directly forecast the wind power, that is, they define $Y_i(t)$ as the power output from a turbine or a wind farm, rather than wind speed. Since methods used for power forecasts do not differ fundamentally from those for wind speed forecasts, we focus on different mechanisms of using the temporal and spatial information for forecasts without differentiating whether it is for wind speed forecasts or power forecasts.

Statistical models developed for short-term wind speed forecasts can be categorized into temporal models and spatial-temporal models. The basic idea behind temporal models is that wind speed at each time point is partially predicted by the wind speed in its near past, and thus time-series models can be built for producing short-term forecasts. There is a rich body of literature on time-series models for short-term forecasts, particularly autoregressive models (Brown, Katz, and Murphy 1984; Huang and Chalabi 1995; Schlink and Tetzlaff 1998), autoregressive moving average models (Torres et al. 2005; Erdem and Shi 2011), and autoregressive integrated moving average models (Palomares-Salas et al. 2009).

Specifically, temporal models use an independent model for each respective time series $\{Y_i(t) : t = 1, 2, \ldots, T\}$, for $i = 1, 2, \ldots, I$. The simplest case, known as persistence forecasting, assumes

$$\widehat{Y}_i(t + h) = Y_i(t), \qquad \text{for} \quad i = 1, 2, \ldots, I, \tag{1}$$

that is, the wind speed "persists" over time for the following h hours. Despite its simplicity, the persistence model has been considered as a reference model in the literature (Giebel et al. 2011).

More sophisticated methodologies include the autoregressive moving average model of order (p, q), denoted by ARMA(p, q), and expressed as

$$Y_i(t) = c + \sum_{\ell=1}^{p} \phi_\ell Y_i(t - \ell) + \sum_{\ell=1}^{q} \theta_\ell \epsilon(t - \ell) + \epsilon(t), \quad \text{for} \quad i = 1, 2, \ldots, I, \tag{2}$$

where c is a constant, ϕ_ℓ and θ_ℓ are the autoregressive and moving average parameters, respectively, and $\epsilon(t) \sim (0, \sigma^2)$ for $t = 1, 2, \ldots, T$. A special case is when $q = 0$; this results in an autoregressive model of order p, denoted by AR(p). Brown, Katz, and Murphy (1984), by explicitly considering non-Gaussian distribution and diurnal nonstationarity, found that AR(1) and AR(2) models outperform the persistence model for hourly data. In general, low-order AR models are considered most suitable for short-term forecasts (Katz and Skaggs 1981; Huang and Chalabi 1995). However, successful implementation of higher-order AR models for short-term forecasts has also been reported, for instance, by Schlink and Tetzlaff (1998). Sophisticated AR models, for example, those built upon generalized logit-normal distributions (Pinson 2012), or Markov-switching autoregressive models that handle seasonal fluctuations (Ailliot and Monbet 2012), have been proposed to model wind power/speed time series (Ailliot and Monbet 2012). In addition, ARMA models, that is, $q \neq 0$, have been used for short-term forecasts (Daniel and Chen 1991; Kamal and Jafri 1997; Torres et al. 2005). Other temporal models for wind power/speed forecasts include the Kalman filter (KF; Crochet 2004; Louka et al. 2008), KF in combinations with NWP (Cassola and Burlando 2012) as well as hybrid time-series KF (Liu, Tian, and Li 2012), hybrid intelligent algorithms (Wan et al. 2014), and kernel density estimation (Qin, Li, and Xiong 2011; Bessa et al. 2012).

The basic idea of spatial-temporal models is the wind speed characteristics of a region resemble the characteristics of neighboring regions. This idea has encouraged researchers to incorporate the spatial dependency of wind speed into their models (Alexiadis, Dokopoulos, and Sahsamanoglou 1999; Gneiting et al. 2006; Hering and Genton 2010; Kusiak and Li 2010; He et al. 2014; Tastu et al. 2014). For example, a straightforward extension of the AR(p) model yields the Vector AR(p) model (Johansen 1995), or simply VAR(p), which considers wind speed at multiple locations. VAR(p) is defined as

$$\mathbf{Y}(t) = \mathbf{c} + \sum_{\ell=1}^{p} \mathbf{\Psi}_\ell \mathbf{Y}(t - \ell) + \boldsymbol{\epsilon}(t), \tag{3}$$

where $\mathbf{c}$ is an $I \times 1$ constant vector, $\mathbf{\Psi}_\ell$ is an $I \times I$ matrix of autoregressive coefficients for $\ell = 1, 2, \ldots, p$, and $\boldsymbol{\epsilon}(t)$ is the $I \times 1$ error vector such that $\mathbb{E}\{\boldsymbol{\epsilon}(t)\} = \mathbf{0}$, $\mathbb{E}\{\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}(t)^T\} = \mathbf{\Omega}$, where $\mathbf{\Omega}$ is a diagonal matrix with nonnegative entries, and $\mathbb{E}\{\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}(t - k)^T\} = \mathbf{0}$, for $k \neq 0$. See de Luna and Genton (2005) for an application of VAR(p) for wind speed forecasts.

It is important to understand how the spatial information helps with wind speed forecasts. When multiple wind farms are geographically dispersed and far away from each other (e.g., 100 miles in distance), wind measurements at an upstream location can inform wind speed forecasts at a downstream location.

Under a typical traveling speed of the weather system, using multiple wind farm measurements can improve wind prediction for a few hours ahead. When the wind measurement locations are closer, say a few tens of miles apart, the look-ahead time is naturally shortened to a few or tens of minutes. Most spatio-temporal models for wind speed forecasts, including Alexiadis, Dokopoulos, and Sahsamanoglou (1999), Gneiting et al. (2006), Hering and Genton (2010), and Tastu et al. (2014), employ weather system information. This approach, which we call between-farm spatial models, is most effective when multiple wind measurements come from relatively large regions for a small number of locations.

Unlike the between-farm setting, our interest is to make turbine-specific forecasts for up to a few hours using wind speed data measured at all turbine sites within a wind farm. In other words, we consider a set of measurements in a relatively small region for a large number of locations within a wind farm. Given the close vicinity between turbines within a farm, the information propagating through the weather system is not the primary reason for prediction enhancement. But the wind measurements taken at other turbine sites do bear relevance. For example, Kusiak and Li (2010) and He et al. (2014) used either an artificial neural network or a linear regression to take wind measurements at multiple turbine sites as inputs and predicted the wind power aggregated at the farm level.

One important question in modeling spatial dependency in local wind fields for individual turbines is selecting the subset of relevant turbine sites. He et al. (2014) allowed their linear regression model to assign weights to different turbine sites based on historical observations, whereas Kusiak and Li (2010) set a threshold on the correlation between a turbine site and the target site. The two approaches are similar as they both depend on correlations calculated by using the original wind speed measurements. Our proposed model selects the informative neighborhood of turbine sites based on similarity in the rate of change in wind speed, rather than that in the wind speed itself. The proposed model also uses the time epoch mechanism to model nonstationarity. The spatio-temporal treatments collectively are an improvement over the use of the persistent model and other time series models.

3. SPATIAL-TEMPORAL AUTOREGRESSIVE MODELS

We treat wind speed at each location as a random variable. Different probability distributions have been proposed in the literature to model wind speed, including Weibull (Yu and Tuzuner 2008) and truncated normal (Gneiting et al. 2006). We use the truncated normal distribution, because it can model nonnegativity of the wind speed and its quantiles can be easily computed. Assume the wind speed $Y_i(t) = Y(\mathbf{s}_i; t)$ follows a truncated normal distribution $\mathcal{N}^+(\mu_i(e_t), \sigma^2(e_t))$ at time t and location $\mathbf{s}_i$ for $i = 1, 2, \ldots, I$ (location $\mathbf{s}_i$ will be often shortened as location i), where e_t denotes the "epoch" at time t, that is, a section of days in a period of time in which the wind speed can be assumed stationary (He et al. 2014). For example, 6 a.m. to 12 p.m. in the month of January is selected as one epoch in our analysis.

Our objective is to develop a model for the parameters of the truncated normal distribution by considering both temporal and spatial dependency in the field. The use of epochs as just mentioned is a way to accommodate nonstationarity in wind speed. Below, we explain how to handle the spatial dependency through a regularized learning formulation and then determine how best to evaluate forecasts out of sample through loss functions.

3.1 Gaussian Spatial-Temporal Autoregression

To begin, we note that single time-series models can be derived from the assumption that wind speed is some (parametric or nonparametric) function of the past wind speed values. To account for the spatial dependency of the wind speed, we assume that the conditional mean of the wind speed is a function of the wind speeds at not only the target site but also other locations in the region. Specifically, the conditional mean of the wind speed at location i can be written as

$$\mu_i(e_t) = c + \sum_{\ell=1}^{p} \sum_{j \in J_i} a_{ij\ell} Y_j(t-\ell), \quad \text{for } i = 1, 2, \ldots, I, \tag{4}$$

where c is a constant, p represents the history of time that can be informative to model the mean of the distribution, $a_{ij\ell}$ are the parameters that show the spatial-temporal dependency, and $J_i \subset 1, 2, \ldots, I$ is a set of locations whose wind speeds have predictive power of the wind speed at the target location i. Here, we fix the order of temporal part p; Section 4.3 discusses its determination. Section 3.2 discusses the data-driven selection of the informative neighborhood J_i, and Section 4 discusses the selection of time epochs, e_t.

Dealing with large-scale datasets that influence the model through J_i, we proceed by imposing a natural structure on the spatial-temporal coefficients through parameterization. Subsequently, we adaptively select the neighborhood size such that most important information is captured via employing a smaller number of locations. These tasks, while maintaining the model interpretability, facilitate the solution procedure enormously as we explain below.

Assume the spatial-temporal parameters $a_{ij\ell}$ can be decomposed into the respective spatial and temporal parts,

$$a_{ij\ell} = a^s_{ij} a^t_{i\ell} \qquad \text{for} \quad i = 1, 2, \ldots, I, \quad j \in J_i, \quad \ell = 1, 2, \ldots, T. \tag{5}$$

A key observation in modeling the spatial parameter a^s_{ij} is that wind speeds at closer geographic proximity contribute more in explaining the change in the wind speed at the target site, whereas the speeds at distant locations give less information. One way to model this type of dependency is through a Gaussian kernel. Specifically,

$$a^s_{ij} = \exp\left[-\left(\mathbf{s}_i - \mathbf{s}_j\right)^T \mathbf{\Lambda}_i \left(\mathbf{s}_i - \mathbf{s}_j\right)\right], \text{ for } i = 1, 2, \ldots, I, \tag{6}$$

where $\mathbf{\Lambda}_i = \text{diag}\{\lambda_{i1}, \lambda_{i2}\}$, and λ_{i1} and λ_{i2} are the parameters modeling the spatial decay in the longitudinal and latitudinal directions, respectively. In other words, this Gaussian kernel assigns "weights" to different locations and the weights continuously diminish as the distance increases. Therefore, this strategy is based on a continuous modeling of the spatial dependency, that is, we replace the spatial part of the coefficients in Equation (4) by new location-specific parameters $\mathbf{\Lambda}_i$.

For the temporal part $a^t_{i\ell}$, we can make a similar argument, that is, an exponential delay in weighting but in terms of time distance. This leads to the following equation,

$$a^t_{i\ell} = \exp\left[-\lambda_{i3}\ell\right], \qquad \text{for} \quad i = 1, 2, \ldots, I, \tag{7}$$

where λ_{i3} is a parameter modeling the temporal decay. Using Equations (5)–(7), we reduce the large number of spatial-temporal parameters for location i to the three parameters λ_1, λ_2, and λ_3.

Let $\mathbf{A}_i$ denote an $I \times p$ matrix of spatial dependency for location i. This means if $j \in J_i$, the (k, ℓ) entry of the matrix is a^s_{ij}, otherwise it is zero. Define matrix $\mathbf{D}_i$ as a $p \times p$ diagonal matrix whose (ℓ, ℓ)th entry is $a^t_{i\ell}$. Then write Equation (4) as

$$\mu_i(e_t) = c + \operatorname{tr}\left(\mathbf{A}_i\mathbf{D}_i\mathcal{Y}^T(t)\right), \qquad \text{for} \quad i = 1, 2, \ldots, I, \tag{8}$$

where $\mathcal{Y}$ is an $I \times p$ matrix whose ℓth column is $\mathbf{Y}(t-\ell)$, and the superscript T denotes the transpose. We call model (8) the Gaussian spatial-temporal autoregression of order p, or, simply GSTAR(p).

To estimate the parameters in Equation (8), we follow a regularized least-square estimation procedure. Specifically, we consider the optimization problem,

$$\min U(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}) = \sum_{\ell=1}^{T} \mathcal{L}\left\{Y_i(\ell+h) - \bar{Y}_i, \operatorname{tr}\left(\mathbf{A}_i\mathbf{D}_i\mathcal{Y}^T(\ell)\right)\right\} + \gamma \operatorname{Pen}(\mathbf{A}_i), \tag{9}$$

where $\bar{Y}_i = \frac{1}{T}\sum_{\ell=1}^{T} Y_i$, $\mathcal{L}\{.,.\}$ is a loss function (see Section 3.5 for details), γ is a penalty coefficient, and $\operatorname{Pen}(\mathbf{A}_i)$ is a penalty term that controls the size of the neighborhood, as discussed in more detail in Section 3.2. We solve optimization problem (9) using numerical steepest descent methods.

We model the variance of wind speed as a linear combination of volatility, which measures the magnitude of recent changes in wind speed (Gneiting et al. 2006). Specifically,

$$\sigma_i^2(e_t) = b_0 + b_1 v_i(t), \qquad \text{for} \quad i = 1, 2, \ldots, I, \tag{10}$$

where

$$v_i(t) = \left[\frac{1}{2|J_i|}\sum_{j\in J_i}\sum_{\ell=0}^{1}\left\{\left(Y_j(t-\ell) - Y_j(t-\ell-1)\right)^2\right\}\right]^{\frac{1}{2}}, \tag{11}$$

$|J_i|$ is the number of elements in J_i, and we estimate b_0 and b_1 through least-square estimation: we use the sample variance on the left-hand side of Equation (10) and regress it on $v_i(t)$ to estimate b_0 and b_1. Based on our wind data, which have a large variability (generally true for almost all wind data), this equation tends to yield positive estimates of b_0 and b_1, guaranteeing the positivity of the estimated $\sigma_i^2(e_t)$. However, if for some data we observe negative values for the estimated $\sigma_i^2(e_t)$, we can replace the negative estimate with a small positive number.

The predicted value for location i at h-step ahead will be the α-quantile of the truncated normal distribution,

$$\widehat{Y}(t+h) = \widehat{\mu}_i(t+h) + \alpha\Phi^{-1}\left[\alpha + (1-\alpha)\Phi\left(-\frac{\widehat{\mu}_i(t+h)}{\widehat{\sigma}_i(t+h)}\right)\right], \tag{12}$$

where $\widehat{\mu}_i(\cdot)$ is the estimated mean found through Equation (8) in which $t+h$ denotes a prediction time that falls in the epoch e_t (likewise, $\widehat{\sigma}_i(\cdot)$ can be found through Equation (10)), and the value of α should be decided based on the loss function $\mathcal{L}(\cdot,\cdot)$. Section 4.2 discusses the optimal choice for α.

3.2 Selecting the Neighborhood Through Regularization

Despite the fact that a Gaussian kernel, as in Equation (6), has already been used to weigh locations based on their relative distance from the target site, our analysis reveals that a pure distance-based determination of informative sites is insufficient and ineffective. It is important to identify the informative neighborhood for location i, that is, J_i. Based on our study, we find that two locations are informative to each other if they have similar rates of change in wind speed for a given period. In other words, using the correlation among the rate of change of the wind speeds can help determine the spatial dependency.

Specifically, let $Z_i(t) = \frac{dY_i^s(t)}{dt} \approx Y_i^s(t) - Y_i^s(t-1)$, for $i = 1, 2, \ldots, I$, where $Y_i^s = \frac{Y_i}{m(Y_i)}$, where $m(Y_i) = \max\{Y_i(t); t = 1, 2, \ldots, T; i = 1, 2, \ldots, I\}$. Next, define a sample covariance matrix for Z as

$$\boldsymbol{\rho} = \frac{1}{T}\sum_{\ell=1}^{T}\left(\mathbf{Z}(\ell) - \bar{\mathbf{Z}}\right)\left(\mathbf{Z}(\ell) - \bar{\mathbf{Z}}\right)^T, \tag{13}$$

where $\mathbf{Z}(\ell) = [Z_1(\ell), Z_2(\ell), \ldots, Z_I(\ell)]^T$, for $\ell = 1, 2, \ldots, T$ and $\bar{\mathbf{Z}} = \frac{1}{T}\sum_{\ell=1}^{T}\mathbf{Z}(\ell)$.

As we have mentioned, including penalty $\operatorname{Pen}(\mathbf{A}_i)$ helps us find a sparse representation of the informative neighborhood by using the information embedded in $\boldsymbol{\rho}$. To ensure that we select a small neighborhood having a high correlation in the rate of change with the target site, $\operatorname{Pen}(\mathbf{A}_i)$ performs three steps: (a) it thresholds the entries of $\boldsymbol{\rho}$ with respect to $\beta \in [0, 1]$; (b) it creates a new matrix whose entries are the inverse of the entries of the matrix obtained in step (a) (with the convention that inverse of zero is ∞); and (c) it calculates the Frobenius norm of the product between the matrix obtained after step (b) and $\mathbf{A}_i$ with the convention that $0 \times \infty = 0$. Specifically, let $\boldsymbol{\rho}^\beta$ denote the matrix $\boldsymbol{\rho}$ after thresholding with respect to β,

$$\rho^\beta_{jk} = \rho_{jk} \quad \text{if} \quad \rho_{jk} \geq \beta \quad \text{otherwise} \quad \rho^\beta_{jk} = 0, \tag{14}$$

where ρ^β_{jk} and ρ_{jk} are the (j,k)th entries of $\boldsymbol{\rho}^\beta$ and $\boldsymbol{\rho}$, respectively, for $j, k \in \{1, 2, \ldots, I\}$. Then, let $\boldsymbol{\rho}^\beta_{\text{inv}}$ define the entrywise inverse of matrix $\boldsymbol{\rho}^\beta$,

$$\rho^\beta_{\text{inv,jk}} = \frac{1}{\rho^\beta_{jk}}, \tag{15}$$

where $\rho^\beta_{\text{inv,jk}}$ is the (j,k)th entry of $\boldsymbol{\rho}^\beta_{\text{inv}}$ for $j, k \in \{1, 2, \ldots, I\}$. Assume $\rho^\beta_{jk} = 0$ implies $\rho^\beta_{\text{inv,jk}} = \infty$. Finally, define

$$\operatorname{Pen}(\mathbf{A}_i) = \|\mathbf{A}_i^T\boldsymbol{\rho}^\beta_{\text{inv}}\|_F, \tag{16}$$

where $\|.\|_F$ denotes the Frobenius norm, and again, use the notational convention that $0 \times \infty = 0$.

In other words, $\operatorname{Pen}(\mathbf{A}_i)$ is a scalar obtained by imposing a sparse structure on $\mathbf{A}_i$ in which entries with associated sample correlation of the derivative smaller than β are 0. If the sample correlation of the derivative is small (but larger than β), the associated entry in $\mathbf{A}_i$ is penalized more, whereas if

the sample correlation of the derivative is large, the associated entry in $\mathbf{A}_i$ is slightly penalized. This defines the informative neighborhood $J_i = \{j : \rho_{ij}^{\beta} \neq 0\}$. Section 4.1 uses this approach to select neighborhoods.

3.3 Regime Switching Gaussian Spatial-Temporal Autoregression

Following the regime switching approach developed by Gneiting et al. (2006), we extend model (8) to account for effect of wind direction. Regimes are determined according to the wind direction, denoted by θ. Consider a partition of the interval $[0^\circ, 360^\circ)$, with 0° representing due north, where each segment of the partition defines a regime. For example, represent an east–west two-regime partition as $\boldsymbol{r} = \{[0^\circ, 180^\circ)\}$, meaning that when $0^\circ \leq \theta(t) < 180^\circ$, it is the east regime, whereas when $180^\circ \leq \theta < 360^\circ$, it is the west regime. Fit a separate model for each regime as

$$\mu_i(e_t) = c + \operatorname{tr}\left(\mathbf{A}_i(\theta(t), \boldsymbol{r})\mathbf{D}_i\mathcal{Y}^T(t)\right), \text{ for } i = 1, 2, \ldots, I, \quad (17)$$

where $\boldsymbol{r}$ denotes the forecast regimes, and $\theta(t)$ is the current wind direction at location i. Note that matrix $\mathbf{A}_i(\theta(t), \boldsymbol{r})$ is similar to $\mathbf{A}_i$ as defined in Section 3.1 and that the dependency on $(\theta(t), \boldsymbol{r})$ means considering only those observations that fall in the specific range determined by $(\theta(t), \boldsymbol{r})$ in the training stage for each regime. Based on the regime at time t, predict the wind speed at time $t + h$, according to the specific trained model for that θ. We call model (17) the regime-switching GSTAR of order p and denote it by RGSTAR(p). Specifically, we solve

$$\begin{aligned}\min U(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}) = c + \sum_{\ell=1}^{T} \mathcal{L}\{Y_i(\ell + h) \\ - \bar{Y}_i, \operatorname{tr}\left(\mathbf{A}_i(\theta(t), \boldsymbol{r})\mathbf{D}_i\mathcal{Y}^T(\ell)\right)\} \\ + \gamma \operatorname{Pen}(\mathbf{A}_i), \end{aligned} \quad (18)$$

which, given a regime, is solved similar to optimization problem (9).

Next, find the regimes in each calendar month by using the previous year's data. Select a group of candidate regimes, for example, east–west or north–south. For each of the candidate regimes fit the model (17) and then choose the regime that yields the smallest training error. Specifically,

$$\boldsymbol{r}^* = \arg\min_{\boldsymbol{r}} \mathcal{E}_m(\boldsymbol{r}), \qquad \text{for } m_o = 1, 2, \ldots, 12, \quad (19)$$

where $\mathcal{E}_m(\boldsymbol{r})$ denotes the prediction error based on some loss function for month m_o. In other words, for each month we choose the regime that yields a smaller prediction error, bearing in mind that selecting a regime with too many partitions reduces the number of data points needed for training in each regime and thus reduces prediction accuracy. For the cases used in this article, we choose the following candidate regimes: (a) two-partition: east–west, north–south, northwest–southeast, southwest–northeast, (b) three-partition: $\{[0^\circ, 120^\circ)\} \cup \{[120^\circ, 240^\circ)\} \cup \{[240^\circ, 360^\circ)\}$, (c) four-partition: $\{[0^\circ, 90^\circ)\} \cup \{[90^\circ, 180^\circ)\} \cup \{[180^\circ, 270^\circ)\} \cup \{[270^\circ, 360^\circ)\}$, and (d) five-partition: $\{[0^\circ, 72^\circ)\} \cup \{[72^\circ, 144^\circ)\} \cup \{[144^\circ, 216^\circ)\} \cup \{[216^\circ, 288^\circ)\} \cup \{[288^\circ, 360^\circ)\}$. We note that for some cases, the regimes with higher partition numbers are not feasible because one of the partitions has no data points. We suggest that regimes with two or three partitions are best and that two-partition regimes constitute the majority of the regimes in this study.

3.4 Using Geostrophic Wind in Regime Switching Gaussian Spatial-Temporal Autoregression

Temperature and air pressure measurements are also used to model wind speeds (Xie et al. 2014). Geostrophic wind is a type of theoretical wind obtained by assuming an exact balance between the air pressure gradient force and the Coriolis force (Focken and Lange 2006). Geostrophic wind can be obtained after some simple calculations on temperature and air pressure measurements (Xie et al. 2014). The actual value of geostrophic wind is in general assumed to be in good accordance with the wind speed close to the ground. "Close to the ground" in this context is about 100 m above ground, that is, the hub height of a typical turbine.

Let $\omega_i(t)$ denote the geostrophic wind at location i at time t and extend model (17) to incorporate geostrophic wind. Specifically,

$$\begin{aligned}\mu_i(e_t) = \operatorname{tr}\left(\mathbf{A}_i(\theta(t), \boldsymbol{r})\mathbf{D}_i\mathcal{Y}^T(t)\right) + \sum_{\ell=1}^{w} \psi(t - \ell)\omega_i(t - \ell), \\ \text{for} \quad i = 1, 2, \ldots, I, \end{aligned} \quad (20)$$

where $\psi(\ell)$ denotes the coefficient of the geostrophic wind at time ℓ and w is the order of the model associated with the geostrophic wind. We call model (20) the RGSTAR Geostrophic Wind of order (p, w) and denote it by RGSTARGW(p,w). We use a two-step approach to find the optimal values of the parameters in model (20). First, given p and w, we fit data to the model RGSTAR, for each respective regime, based on the model discussed in Section 3.3, namely, optimization problem (18). Second, we regress the residuals on the geostrophic wind using least-square estimation,

$$\begin{aligned}&\widehat{Y}_i(t + h) - \operatorname{tr}\left(\widehat{\mathbf{A}}_i(\theta(t), \boldsymbol{r})\widehat{\mathbf{D}}_i\mathcal{Y}^T(t)\right) \\ &= c + \sum_{\ell=1}^{w} \psi(t - \ell)w_i(t - \ell), \quad \text{for} \quad i = 1, 2, \ldots, I, \end{aligned} \quad (21)$$

where $\widehat{\mathbf{A}}_i(\theta(t), \boldsymbol{r})$ and $\widehat{\mathbf{D}}_i$ are the estimated values for the matrix $\mathbf{A}_i(\theta(t), \boldsymbol{r})$ and $\mathbf{D}_i$, respectively, and $\widehat{Y}_i(t + h)$ is the predicted value obtained in the first step.

So far, we have discussed how to incorporate the spatial information into our model for a given order of temporal dependency, namely, for fixed p, q, and/or w. Section 4.3 discusses selecting the temporal order.

3.5 Choice of Loss Functions

In practice, prediction evaluation is based on the loss functions used by the wind industry. Thus, if $\widehat{Y}_i(t + h)$ denotes the predicted wind speed for h-step ahead forecasts at location i for $i = 1, 2, \ldots, I$, a common choice is the mean absolute difference,

$$\text{MAD} = \frac{1}{I}\sum_{i=1}^{I} \left|\widehat{Y}_i(t + h) - Y_i(t + h)\right|. \quad (22)$$

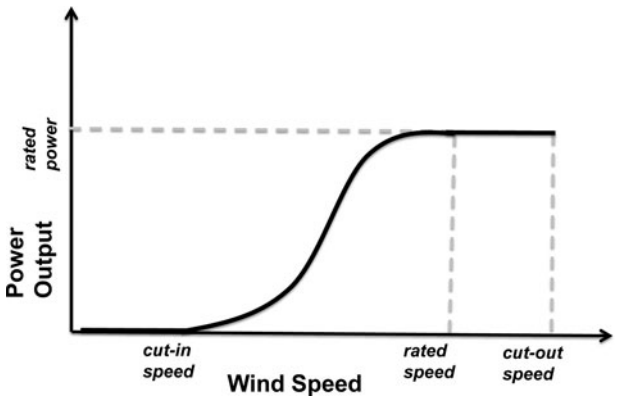

Figure 1. Power curve for a wind turbine.

Hering and Genton (2010), who instead favor measuring the final response affected by the wind speed forecast, used a nonlinear mapping like the power curve associated with a turbine to convert wind speed to a turbine response. As shown in Figure 1, for a wind speed smaller than cut-in speed the energy produced by the turbine is zero. For wind speeds between cut-in speed and rated speed, the energy produced increases monotonically. If the wind speed is beyond the cut-out speed, the operator shuts down the turbine for safety, that is, it produces no energy.

To consider the impact on a turbine's power response, Hering and Genton (2010) proposed the following power curve error (PCE), defining it as

$$\mathrm{PCE}(Y,\hat{Y}) = \begin{cases} \alpha\left(g(Y) - g(\hat{Y})\right) & \text{if } \hat{Y} \le Y, \\ (1-\alpha)\left(g(\hat{Y}) - g(Y)\right) & \text{if } \hat{Y} > Y, \end{cases} \tag{23}$$

where $g(.)$ is the power curve and $\alpha \in (0, 1)$, which is introduced to penalize underestimation and overestimation differently; because in practice, underestimating incurs more cost than overestimating. Therefore, for practical purposes $\alpha > 0.5$. Provided that the PCE is used as the loss function, α in Equation (23) is the quantile of the optimal predictor and is the same as the α shown in Equation (12).

To solve optimization problem (9), determine the loss function $\mathcal{L}$ based on the forecast objective. For example,

$$\min U(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}) = \sum_{\ell=1}^{T} \mathrm{PCE}\left\{Y_i(\ell + h) - \bar{Y}_i, \mathbf{A}_i\mathbf{D}_i\mathcal{Y}^T(\ell)\right\} + \gamma\,\mathrm{Pen}\,(\mathbf{A}_i)\,, \tag{24}$$

where PCE replaces $\mathcal{L}(.,.)$. An alternative criterion is the continuous rank probability score (CRPS), which is akin to PCE, but evaluates the forecast from a distribution perspective. Specifically,

$$\mathrm{CRPS} = \frac{1}{I}\sum_{i=1}^{I}\int_0^1 \left(\hat{F}(X) - \mathbb{I}(X \ge g_i)\right)^2 dX, \tag{25}$$

where $\hat{F}(X)$ is the distributional forecast of power (assuming the power curve is known), g_i is the (normalized) power generated by turbine i, and $\mathbb{I}(\cdot)$ is the indicator function. In this article, we use CRPS for model evaluation, similar to the approach in He et al. (2014), but not for parameter estimation.

4. RESULTS

The dataset comprises information collected between 2008 and 2010 from 200 randomly selected turbines and a few meteorological mast towers located on a wind farm having mostly flat terrain. We note that we use 200 turbines rather than all of them in the wind farm because our industrial partner deems the exact number of turbines confidential. In this wind farm, the elevation differences between the highest and lowest locations are less than 10 m, over a stretch of approximately 160 km^2. Each turbine measures both wind speed every 10 min and the standard deviation during the 10 min period. Each mast tower measures temperature, air pressure, and wind direction as a 10 min average. We note that the use of 10 min blocks is standard industry practice.

We impute the values of the 2%–3% of missing monthly data by using the iterative singular value decomposition (Beckers and Rixen 2003; Maadooliat, Huang, and Hu 2015). For our forecast purpose, we further combine the six 10 min averages in an hour and produce hourly averages. This means that we use a temporal resolution of 1 hr.

We consider the three proposed models: GSTAR(p), RGSTAR(p), and RGSTARGW(p,w). GSTAR(p) is simply the Gaussian spatial-temporal autoregression of order p as described in Section 3.1; RGSTAR(p) is the regime-switching GSTAR in which the regime in each calendar month is decided based on the data in the previous year; and RGSTARGW(p,w) uses the temperature and pressure measurements in the form of geostrophic wind as described in Section 3.4.

Assuming that wind speed in each time epoch is a stationary stochastic process, for each model we define four epochs for each day in a calendar month: (1) 12:00 am to 6:00 am, (2) 6:00 am to 12:00 pm, (3) 12:00 pm to 6:00 pm, and (4) 6:00 pm to 12:00 am. Consequently, we need to fit individual models for each epoch, depending on the epoch to which the forecast horizon belongs. We note that reporting the error average over all turbine cases for each of the four epochs is computationally burdensome. For example, fitting the GSTAR model for one epoch for each month takes about 1 min, but fitting it to 200 turbines for 12 months with four epoch options each takes 6.7 days ($1 \times 4 \times 12 \times 200 = 160$ hr). Since we have three model options and 2 years of data, it would take almost 40 days. Therefore, we randomly assign each of the 200 turbine cases to evaluate one of the epochs for a given month to reduce the computational burden. We report the prediction error for that month, averaged over roughly 50 evaluation cases. We believe that this reported average prediction error is a good approximation of the error that would have been obtained using the entire 200 turbines.

Table 1 summarizes the features of the three models.

Table 1. Features of proposed models

Model	Epochs	Regimes	Geostrophic Wind
GSTAR	Yes	No	No
RGSTAR	Yes	Yes	No
RGSTARGW	Yes	Yes	Yes

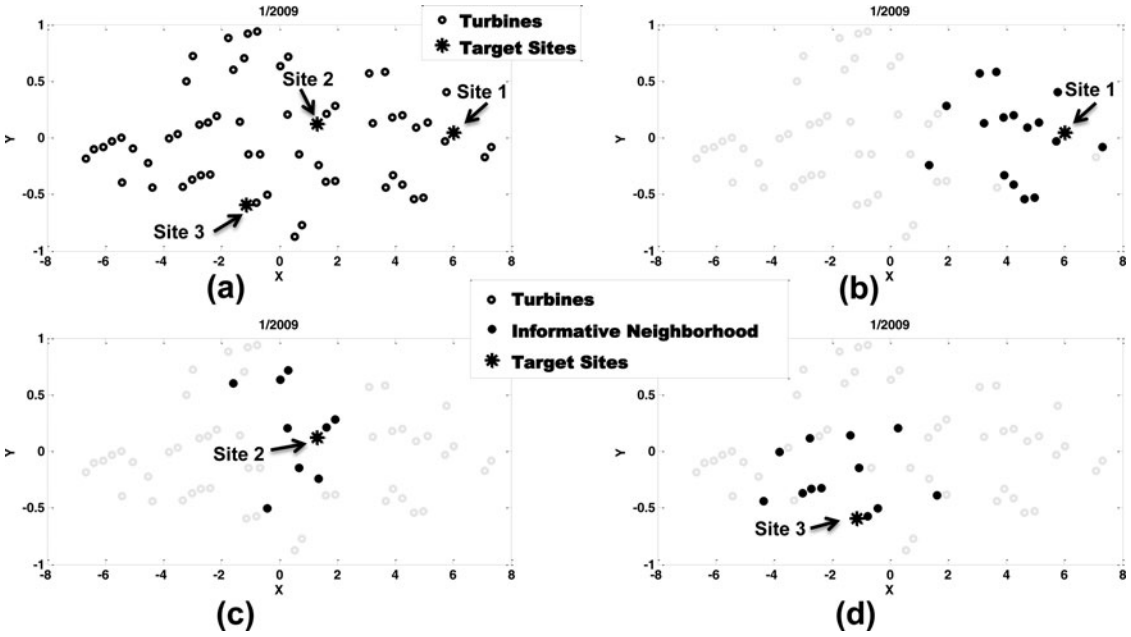

Figure 2. Neighborhood selection in GSTAR for a sample data in January 2009: (a) three different sites and turbines in their neighborhood; (b)–(d) informative neighborhood selection for each site.

The competing models used in this comparison are ARMA(p,q), ARMA*(p,q), vector AR of order p (VAR(p)), and the persistence model (PER). ARMA*(p,q) is the same as ARMA(p,q), but the analysis is performed on the residuals after removing a diurnal trend. For the VAR(p), we select the neighborhood based on the geographical distance smaller than 5 km. For these models, we select $p = 1$, $q = 2$, and $w = 1$. See Section 4.3 for the details of selecting p, q, and w. We compute h-step ahead predictions for $h = 2, 3, \ldots, 5$. Specifically, for each h-step ahead prediction, we train the model using 30 days of hourly data and then make predictions for the next h hours. In doing so, we fit separate models for each prediction horizon. See Section 3.5 for the details of the loss function.

4.1 An Example of an Informative Neighborhood

Figure 2 illustrates how the GSTAR models select informative neighborhoods for the target sites. We note that they select a neighborhood according to the historical similarity in the rate of

Table 2. Prediction results for 2009 and 2010 using PCE. The numbers in parentheses are the standard deviations of the corresponding predictions. Imp. % over PER shows the improvement of the best model over PER. The values in bold indicate the best model for each prediction horizon in each year

Model	2 hr	3 hr	4 hr	5 hr
		2009		
PER	0.054 (0.013)	0.066 (0.17)	0.076 (0.022)	0.089 (0.025)
VAR(1)	0.109 (0.053)	0.115 (0.044)	0.126 (0.042)	0.127 (0.037)
ARMA(1,2)	0.058 (0.016)	0.077 (0.020)	0.088 (0.028)	0.100 (0.028)
ARMA*(1,2)	0.066 (0.018)	0.085 (0.022)	0.094 (0.027)	0.104 (0.028)
GSTAR(1)	**0.050 (0.012)**	**0.058 (0.016)**	**0.066 (0.019)**	**0.080 (0.025)**
RGSTAR(1)	0.055 (0.018)	0.061 (0.020)	0.070 (0.022)	0.085 (0.025)
RGSTARGW(1,1)	0.054 (0.016)	0.059 (0.019)	0.069 (0.021)	0.087 (0.027)
Imp. % over PER	7.4	12.1	13.2	10.1
		2010		
PER	0.047 (0.012)	0.053 (0.015)	0.069 (0.019)	0.074 (0.023)
VAR(1)	0.110 (0.068)	0.129 (0.075)	0.138 (0.058)	0.148 (0.057)
ARMA(1,2)	0.055 (0.014)	0.068 (0.017)	0.085 (0.023)	0.094 (0.026)
ARMA*(1,2)	0.096 (0.015)	0.104 (0.017)	0.118 (0.020)	0.121 (0.022)
GSTAR(1)	**0.042 (0.013)**	0.052 (0.016)	**0.063 (0.019)**	**0.071 (0.020)**
RGSTAR(1)	0.043 (0.017)	**0.051 (0.017)**	0.067 (0.022)	0.073 (0.021)
RGSTARGW(1,1)	**0.042 (0.014)**	0.054 (0.016)	0.066 (0.021)	**0.071 (0.021)**
Imp. % over PER	10.6	3.8	8.7	4.1

change in wind speed (see Section 3.2) instead of geographical proximity.

In Section 3.2, the parameter β in Equation (14) affects the choice of the neighborhood: a smaller β leads to a larger neighborhood for a target site, because it causes $\boldsymbol{\rho}^{\beta}$ to have fewer zero entries, whereas a large β creates a smaller informative neighborhood, because it causes $\boldsymbol{\rho}^{\beta}$ to have more zero entries. Recall that for forecasts ranging from 2 hr to 5 hr ahead, the weather-related information, which propagates through wind, contributes little to the model's performance. For this reason, we choose similar values for each horizon but with a slight ascending trend. For the 2 hr ahead forecasts, we choose $\beta = 0.85$ and for the 5 hr ahead forecasts, $\beta = 0.92$, and then we linearly interpolate the values of β for the 3 hr and 4 hr ahead forecasts. The rationale for this ascending trend is because we observe that as the prediction horizon lengthens, the impact of the informative neighborhood mechanism weakens, which suggests there is no need to aggregate a large number of models from the neighboring turbines.

4.2 Forecasts and Comparisons

All models use the wind speed data adjusted according to a standard industry practice that converts the after-wake wind data to free stream wind data. We train each model using 1 month of data and then make h-step ahead forecasts for $h = 2, \ldots, 5$ hr. Doing so for turbine $i = 1, 2, \ldots, 200$, for each year gives us 200×12 prediction results for each model, and we report the average values and the standard deviations of the prediction errors obtained based on using the PCE as the loss function, where we normalize the power curve so the rated power is 1. We choose $\alpha = 0.73$ as suggested by Hering and Genton (2010). The point forecast for the GSTAR models is the α-quantile of the predictive normal distribution.

Prior to our comparison, we assess the goodness-of-fit of the GSTAR model. One measure for the goodness-of-fit is the coefficient of determination R^2, which is the ratio between the variation explained by the model to the total variation. For the three GSTAR models, we calculate this measure for a given turbine in each month. For most cases, this measure is greater than 0.93, suggesting a good fit for the model. We also took a look at the residual plots (not shown) and did not detect any obvious patterns.

Table 2 shows the results of our comparison. Generally, the three GSTAR models consistently outperform the VAR and ARMA models, sometimes by more than 10% reduction in terms of PCE. The fact that ARMA performs better than VAR testifies to the importance of choosing an appropriate neighborhood. A poor selection of neighborhood, for example, selecting based on geographical distance, can mislead a model into making worse predictions.

Table 3 shows the average improvement in prediction made on individual turbines for the three GSTAR models. The first number is the percentage of turbines for which the models work better than the persistence model and the second number in parentheses is the average percentage of improvement over the persistence model for those turbines. The table indicates that the GSTAR model that is identified as the best model in Table 2 outperforms the persistent model for over 50% of the turbines, and in some cases, the GSTAR model preforms better on close to 75% of the turbines. Moreover, on those turbines, the improvements are more than 15% and even as much as 29%.

Table 3. Percentage of improvement over PER in 2009 and 2010. The data pairs in parentheses are the percentage of turbines that see an improvement and the average improvement percentage for those turbines. The values in bold correspond to the best model in Table 2

Model	2 hr	3 hr	4 hr	5 hr
GSTAR(1) (2009)	(**66**, **18.0**)	(**75**, **20.1**)	(**77**, **20.3**)	(**70**, **19.8**)
RGSTAR(1) (2009)	(53, 23.9)	(60, 27.6)	(58, 30.1)	(57, 27.3)
RGSTARGW(1,1) (2009)	(49, 24.9)	(60, 28.9)	(60, 29.1)	(49, 30.8)
GSTAR(1) (2010)	(**63**, **22.1**)	(54, 18.6)	(**63**, **19.5**)	(**59**, **15.9**)
RGSTAR(1) (2010)	(61, 31.9)	(**56**, **27.1**)	(54, 28.7)	(53, 26.5)
RGSTARGW(1,1) (2010)	(**66**, **29.2**)	(46, 26.3)	(54, 28.1)	(**53**, **26.8**)

However, the persistent model outperforms the three GSTAR models in some cases of volatility, for example, sudden changes in wind speeds. This finding suggests the need for a model to capture more complex temporal patterns and trends.

Table 4 shows some of the results using CRPS to give a sense of the quality of predictive distribution. We randomly select 20 turbines, apply the GSTAR model, and compare the results with ARMA(1,2). Table 4 suggests that the GSTAR model does better than the ARMA(1,2) model for capturing the predictive distribution.

Finally, we investigate whether the GSTAR model is sensitive to the value of α used in PCE. In the model building step, we choose $\alpha = 0.73$ following the suggestion made by Hering and Genton (2010). In practice, the optimal value of α may change over time, and a variation of α around 0.73 is expected. Hence, we conduct a sensitivity analysis by changing the value of α between 0.6 and 0.8, and then we average the PCE over this range. We randomly select 100 turbines and use 2009 data. Table 5 shows that the performance of the model is not sensitive to the specific value of 0.73 chosen for the model fitting.

4.3 Role of Informative Neighborhood and Temporal Dependency

We speculate that the performance of the GSTAR models depends upon an effective aggregation of a group of predictive models initiated from individual turbines in its informative neighborhood. Therefore, we compare the change in the neigh-

Table 4. CRPS values using 20 randomly selected turbines and 2009 data

Model	2 hr	3 hr	4 hr	5 hr
ARMA(1,2)	1.166	1.480	1.845	2.353
GSTAR	0.997	1.198	1.478	1.532

Table 5. Average PCE with α varying in [0.6, 0.8] for 100 turbines in the year of 2009. The values in parentheses are the average of standard deviations for each case

Model	2 hr	3 hr	4 hr	5 hr
PER	0.071 (0.007)	0.073 (0.008)	0.064 (0.009)	0.077 (0.010)
GSTAR	0.057 (0.007)	0.066 (0.007)	0.053 (0.009)	0.074 (0.008)

borhood selected by the GSTAR when it performs better than the persistence model. Intuitively, the selected (informative) neighborhood should change less to give the GSTAR model an edge.

We define a metric of quantifying the changes in a neighborhood $\mathrm{ch}(i, m)$, $i = 1, 2, \ldots, 200$ and $m = 1, 2, \ldots, 12$:

$$\mathrm{ch}(i, m) = 1 - \frac{|J_i(m) \cap J_i(m+1)|}{|J_i(m)|}, \tag{26}$$

where $J_i(m)$ denotes the informative neighborhood selected by the GSTAR at month m for turbine i, $|J_i(m)|$ is the number of elements in $J_i(m)$, and $\mathrm{ch}(i, m)$ is a number between 0 and 1. To obtain a robust statistic, we set a threshold, $t^{\mathrm{ch}} \in (0, 1)$, and find how often the change in the informative neighborhood is smaller than t^{ch}. More precisely,

$$S_c = \frac{1}{12I} \sum_{i,m} \left\{ \mathbb{I}\left(\mathrm{ch}(i, m) \le t^{\mathrm{ch}}\right) \right\}. \tag{27}$$

Simply speaking, when t^{ch} is chosen as a small value, S_c indicates the likelihood of a consistent neighborhood; the bigger S_c is, the more consistent the neighborhood. Based on the intuition expressed above, if GSTAR outperforms the persistence model, S_c should have a value more than 0.5, implying that it is more likely for the neighborhood to have a small change (smaller than t^{ch}) than otherwise.

To test this, we randomly select 60 turbines (therefore $I = 60$ in Equation (27)), use the data of 2010, and both GSTAR and the persistence models to produce an h-hour ahead forecast, for $h = 2, \ldots, 5$. We choose $t^{\mathrm{ch}} = 0.2$ and estimate the value of S_c by counting the proportion of times $\mathrm{ch}(i, m)$ is smaller than 0.2 when GSTAR performs better. Table 6 shows that when the GSTAR model performs better for smaller horizons, the neighborhood more often stays unchanged, and that the GSTAR model still maintains its edge for longer horizons when the neighborhood changes more often. These results indicate that when predicting for shorter horizons, the informative neighborhood plays a more important role than when predicting for longer horizons.

We also believe that the GSTAR model is superior as a result of using epochs for modeling nonstationarity in wind data. Undertaking an analysis similar to what we presented for neighborhood selection, we observed that the time epoch mechanism has a bigger impact on longer prediction horizons (4 hr or 5 hr) and a smaller impact on shorter prediction horizons (2 hr or 3 hr). The combined effect of the informative neighborhood mechanism and the time epoch mechanism does not appear to be monotonic in time, which explains why the GSTAR model's improvement over the persistent model as shown in Table 2 does not show any monotonic trend.

We explained earlier that the improvement mechanism of the GSTAR models is different from that of the between-farm

Table 6. Proportion of the times that the informative neighborhood stays largely unchanged

	2 hr	3 hr	4 hr	5 hr
S_c	0.68	0.59	0.53	0.53

spatial models. The comparisons presented in Table 2 reaffirm this understanding. We note that RGSTAR and RGSTARGW, the GSTAR variants using regime-switching and geostrophic wind, perform no better than the GSTAR model for most of the cases. The use of regime-switching and geostrophic wind is believed to benefit the between-farm spatial models when winds traveling through weather systems play a larger role. Due to different combination of spatial and temporal scales in our problem, inclusion of these modeling elements appears to produce only a marginal benefit.

One final issue regarding model specification is the orders of the model, p, q, and w. One approach to find the order p is the partial autocorrelation function (Brockwell and Davis 2009). Partial autocorrelation of lag $k > 0$ is the correlation between the terms with indices t and $t + k$, with the linear dependence of the terms $t + 1$ to $t - k + 1$ removed. For the datasets used in this article, the partial autocorrelation of lag 1, for both geostrophic winds and wind speeds is dominant, which suggests $p = 1$ in VAR, GSTAR, RGSTAR, and RGSTARGW, and $w = 1$ in RGSTARGW. This finding is consistent with the claim that for meteorological data, a low order for AR models, often $p = 1$, would be sufficient (Katz and Skaggs 1981). Similarly, to determine q in ARMA, we use Bayesian information criterion (BIC; Torres et al. 2005; Erdem and Shi 2011). BIC suggests $p = 1$ and $q = 2$ for most of the cases, so we use ARMA(1,2) as a competing algorithm.

5. CONCLUSION

This article presented a model for making short-term forecasts of wind speed in local wind fields. It focused on turbine-specific forecasts that can benefit turbine operations and control. Our proposed model and two variants employed both spatial and temporal information to enhance the forecast quality. Testing the three models using 2 years of real data demonstrated that incorporating spatial information into the proposed models assisted in obtaining more accurate wind speed forecasts and that the use of the spatial information depended on selecting the most informative neighborhood. We also found that using the rate of change in wind speeds as a criterion for selecting the informative neighborhood was better than basing the selection on distance only.

The results reported in Table 3 indicate that our proposed models did not always outperform the persistent model on all 200 turbines. This is in contrast to the large body of literature using a single time series dataset, claiming to make improvement over the persistent model. Thus, we maintain that it is unlikely that any temporal-only model can consistently outperform the persistent model in a comprehensive comparative study. Our experience appears to suggest that for the highly volatile near-ground wind field, any attempts to model the temporal dependency as a linear function will be unlikely to succeed. We

suggest that future research should examine the use of other nonlinear mechanisms to fully identify the benefits of capturing the nonstationarity in wind data. Nonlinear models in the mode of Giannakis and Majda (2012) may be more capable of handling the nonlinear dependency for meteorological data, but determining how such models could be applied to wind field modeling is an ongoing pursuit.

ACKNOWLEDGMENTS

Pourhabib and Ding were partially supported by the grants from NSF (CMMI-0926803 and CMMI-1300236). Huang's research was partially supported by NSF grant no. DMS-1208952.

[Received December 2013. Revised October 2014.]

REFERENCES

Ailliot, P., and Monbet, V. (2012), "Markov-Switching Autoregressive Models for Wind Time Series," *Environmental Modelling & Software*, 30, 92–101. [139]

Alexiadis, M. C., Dokopoulos, P. S., and Sahsamanoglou, H. S. (1999), "Wind Speed and Power Forecasting Based on Spatial Correlation Models," *IEEE Transactions on Energy Conversion*, 14, 836–842. [139]

Beckers, J.-M., and Rixen, M. (2003), "EOF Calculations and Data Filling From Incomplete Oceanographic Datasets," *Journal of Atmospheric and Oceanic Technology*, 20, 1839–1856. [143]

Bessa, R. J., Miranda, V., Botterud, A., Wang, J., and Constantinescu, E. M. (2012), "Time Adaptive Conditional Kernel Density Estimation for Wind Power Forecasting," *IEEE Transactions on Sustainable Energy*, 3, 660–669. [139]

Brockwell, P. J., and Davis, R. A. (2009), *Time Series: Theory and Methods*, New York: Springer. [146]

Brown, B. G., Katz, R. W., and Murphy, A. H. (1984), "Time Series Models to Simulate and Forecast Wind Speed and Wind Power," *Journal of Climate and Applied Meteorology*, 23, 1184–1195. [139]

Cassola, F., and Burlando, M. (2012), "Wind Speed and Wind Energy Forecast Through Kalman Filtering of Numerical Weather Prediction Model Output," *Applied Energy*, 99, 154–166. [138,139]

Crochet, P. (2004), "Adaptive Kalman Filtering of 2-Metre Temperature and 10-Metre Wind-Speed Forecasts in Iceland," *Meteorological Applications*, 11, 173–187. [139]

Daniel, A., and Chen, A. (1991), "Stochastic Simulation and Forecasting of Hourly Average Wind Speed Sequences in Jamaica," *Solar Energy*, 46, 1–11. [139]

de Luna, X. and Genton, M. G. (2005), "Predictive Spatio-Temporal Models for Spatially Sparse Environmental Data," *Statistica Sinica*, 15, 547–568. [139]

Erdem, E., and Shi, J. (2011), "ARMA Based Approaches for Forecasting the Tuple of Wind Speed and Direction," *Applied Energy*, 88, 1405–1414. [139,146]

Focken, U., and Lange, M. (2006), *Physical Approach to Short-Term Wind Power Prediction*, New York: Springer. [142]

Genton, M. G., and Hering, A. S. (2007), "Blowing in the Wind," *Significance*, 4, 11–14. [138]

Giannakis, D., and Majda, A. J. (2012), "Nonlinear Laplacian Spectral Analysis for Time Series With Intermittency and Low-Frequency Variability," *Proceedings of the National Academy of Sciences*, 109, 2222–2227. [147]

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., and Draxl, C. (2011), *The State-of-the-Art in Short-Term Prediction of Wind Power: A Literature Overview* (2nd ed.), ANEMOS.plus. Available at *http://orbit.dtu.dk/fedora/objects/orbit:83397/datastreams/file_5277161/content*. [138,139]

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006), "Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center," *Journal of the American Statistical Association*, 101, 968–979. [139,140,141,142]

He, M., Yang, L., Zhang, J., and Vittal, V. (2014), "A Spatio-Temporal Analysis Approach for Short-Term Forecast of Wind Farm Generation," *IEEE Transactions on Power Systems*, 29, 1611–1622. [139,140,143]

Hering, A. S., and Genton, M. G. (2010), "Powering up With Space-Time Wind Forecasting," *Journal of the American Statistical Association*, 105, 92–104. [139,143,145]

Huang, Z., and Chalabi, Z. S. (1995), "Use of Time-Series Analysis to Model and Forecast Wind Speed," *Journal of Wind Engineering and Industrial Aerodynamics*, 56, 311–322. [139]

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Cambridge: Cambridge University Press. [139]

Kamal, L., and Jafri, Y. Z. (1997), "Time Series Models to Simulate and Forecast Hourly Averaged Wind Speed in Quetta, Pakistan," *Solar Energy*, 61, 23–32. [139]

Katz, R. W., and Skaggs, R. H. (1981), "On the Use of Autoregressive-Moving Average Processes to Model Meteorological Time Series," *Monthly Weather Review*, 109, 479–484. [139,146]

Kusiak, A., and Li, W. (2010), "Estimation of Wind Speed: A Data-Driven Approach," *Journal of Wind Engineering and Industrial Aerodynamics*, 98, 559–567. [139,140]

Kusiak, A., Zheng, H., and Song, Z. (2009), "On-Line Monitoring of Power Curves," *Renewable Energy*, 34, 1487–1493. [138]

Liu, H., Tian, H.-Q., and Li, Y.-F. (2012), "Comparison of Two New ARIMA-ANN and ARIMA-Kalman Hybrid Methods for Wind Speed Prediction," *Applied Energy*, 98, 415–424. [139]

Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., and Kallos, G. (2008), "Improvements in Wind Speed Forecasts for Wind Power Prediction Purposes Using Kalman Filtering," *Journal of Wind Engineering and Industrial Aerodynamics*, 96, 2348–2362. [139]

Maadooliat, M., Huang, J. Z., and Hu, J. (2015), "Integrating Data Transformation in Principal Components Analysis," *Journal of Computational and Graphical Statistics*, 24, 84–103. [143]

Palomares-Salas, J. C., de la Rosa, J. J. G., Ramiro, J. G., Melgar, J., Agera, A., and Moreno, A. (2009), "ARIMA vs. Neural Networks for Wind Speed Forecasting," in *Proceedings of the 2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, Hong Kong, pp. 129–133. [139]

Pinson, P. (2012), "Very-Short-Term Probabilistic Forecasting of Wind Power With Generalized Logit-Normal Distributions," *Journal of the Royal Statistical Society*, Series C, 61, 555–576. [139]

Qin, Z., Li, W., and Xiong, X. (2011), "Estimating Wind Speed Probability Distribution Using Kernel Density Method," *Electric Power Systems Research*, 81, 2139–2146. [139]

Santos, R. A. (2007), "Damage Mitigation Control for Wind Turbines," Ph.D. dissertation, University of Colorado, Boulder, CO. [138]

Schlink, U., and Tetzlaff, G. (1998), "Wind Speed Forecasting From 1 to 30 Minutes," *Theoretical and Applied Climatology*, 60, 191–198. [139]

Tastu, J., Pinson, P., Trombe, P. J., and Madsen, H. (2014), "Probabilistic Forecasts of Wind Power Generation Accounting for Geographically Dispersed Information," *IEEE Transactions on Smart Grid*, 5, 480–489. [139]

Torres, J., Garca, A., De Blas, M., and De Francisco, A. (2005), "Forecast of Hourly Average Wind Speed With ARMA Models in Navarre Spain," *Solar Energy*, 79, 65–77. [139,146]

Wan, C., Xu, Z., Pinson, P., Dong, Z. Y., and Wong, K. P. (2014), "Optimal Prediction Intervals of Wind Power Generation," *IEEE Transactions on Power Systems*, 29, 1166–1174. [139]

Xie, L., Gu, Y., Zhu, X., and Genton, M. G. (2014), "Short-Term Spatio-Temporal Wind Power Forecast in Robust Look-Ahead Power System Dispatch," *IEEE Transactions on Smart Grid*, 5, 511–520. [142]

Yu, Z., and Tuzuner, A. (2008), "Wind Speed Modeling and Energy Production Simulation With Weibull Sampling," in *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pp. 16. [140]

Zhu, X., and Genton, M. G. (2012), "Short-Term Wind Speed Forecasting for Power System Operations," *International Statistical Review*, 80, 2–23. [138]