*Submitted to the Annals of Applied Statistics*

# SPATIO-TEMPORAL SHORT-TERM WIND FORECAST: A CALIBRATED REGIME-SWITCHING METHOD

By Ahmed Aziz Ezzat, Mikyoung Jun and Yu Ding

*Texas A&M University*

Accurate short-term forecasts are indispensable for the integration of wind energy in power grids. On a wind farm, local wind conditions exhibit sizeable variations at a fine temporal resolution. Existing statistical models may capture the in-sample variations in wind behavior, but are often shortsighted to those occurring in the near future, i.e. in the forecast horizon. The calibrated regime-switching method proposed in this paper introduces an action of regime dependent calibration on the predictand (here the wind speed variable), which helps correct the bias resulting from out-of-sample variations in wind behavior. This is achieved by modeling the calibration as a function of two elements: the wind regime at the time of the forecast (and the calibration is therefore regime dependent), and the runlength, which is the time elapsed since the last observed regime change. In addition to regime-switching dynamics, the proposed model also accounts for other features of wind fields: spatio-temporal dependencies, transport effect of wind, and nonstationarity. Using one year of turbine-specific wind data, we show that the calibrated regime-switching method can offer a wide margin of improvement over existing forecasting methods in terms of both wind speed and power.

**1. Introduction.** With the global aspiration towards a more sustainable environment, wind power presents itself as one of the most appealing sources of clean energy [DOE, 2015]. Despite the promising potential, serious challenges still hinder its large scale exploitation, including the intermittency and limited predictability of the wind resource. Hence, improving the accuracy of short-term wind forecasts is essential for integrating wind energy in power grid systems [Pinson, 2013].

There is a rich body of literature on short-term wind forecasting [Giebel et al., 2011], from time series methods [Brown, Katz and Murphy, 1984; Erdem and Shi, 2011], machine learning techniques [Mohandes et al., 2004; Sideratos and Hatziargyriou, 2012], to spatio-temporal models [Dowell and Pinson, 2016; Pourhabib, Huang and Ding, 2016]. In the past decade, there has been a growing recognition for *regime-switching* methods for short-term forecast [Gneiting et al., 2006]. The essence of these regime-switching approaches is to fit statistical models that are conditioned on a finite set of

*Keywords and phrases:* regime-switching, spatio-temporal, wind energy, wind forecast

wind regimes or states, and produce regime-dependent forecasts.

Our review of the literature suggests that a wide array of regime-switching methods assume that the wind regime observed at the time of forecast (or shortly prior to that) will persist during the forecast horizon, and hence, use the model whose parameters are fit specifically to the observed regime to make forecasts. One such approach is the regime-switching autoregressive method of Zwiers and Von Storch [1990]. Several regime-switching models have been proposed in the past decade, mostly sharing the same concept, but with key differences in how regimes are defined, whether using lagged values of wind speed, power, direction, precipitation, temperature, or other exogenous variables [Gneiting et al., 2006; Reikard, 2008; Hering and Genton, 2010; Zhu et al., 2014; Browell, Drew and Philippopoulos, 2018].

Let us refer to the observed regime as the *in-sample* regime, and the regime in the forecast horizon as the *out-of-sample* regime. The aforementioned methods can be perceived as "reacting" to the in-sample regime without anticipating potential changes between the in- and out-of-sample regimes. We hereinafter label them collectively as "reactive" regime-switching approaches. A drawback of the reactive approaches is that they are shortsighted to the changes in wind behavior that are yet to occur in the forecast horizon. If the wind behavior in the near future departs from the observed in-sample attributes, then a considerable discrepancy between the forecasts and the true underlying process can be expected.

In light of that, this paper proposes a calibrated regime-switching (CRS) method. The CRS approach starts by constructing a reactive regime-switching model, which by design, suffers from an inherent bias while extrapolating the in-sample attributes in the form of forecasts. The CRS approach then corrects this bias by introducing a regime-specific calibration to capture potential out-of-sample variations in the predictand (here the wind speed). This is achieved by modeling the forecast calibration as a parametric function of two elements that are shown to be able indicators of out-of-sample changes: the observed regime at the time of the forecast, and the runlength, which is the time elapsed since the most recent regime change.

A parallel line of research in the regime-switching literature concerning the in-sample/out-of-sample regime difference is the Markov-switching (MS) models [Ailliot, Monbet and Prevosto, 2006; Pinson et al., 2008; Ailliot et al., 2015; Hering, Kazor and Kleiber, 2015; Bessac et al., 2016], which use a transition matrix to connect, probabilistically, the in-sample regimes with possible regimes in the forecast horizon. The forecast under an MS model is either the one made by the most probable regime (the hard thresholding approach) or a weighted average of all possible regime-specific forecasts (the

soft thresholding approach).

We perceive both the CRS approach and MS approach as a transition step from the "reactive" regime-switching approaches to the next-generation "proactive" regime-switching models; the latter would ideally predict the out-of-sample regimes directly and then issue forecasts based on the regime predictions. CRS and MS do not involve a direct prediction of the out-of-sample regimes but both nonetheless account for in-sample/out-of-sample changes in wind behavior. In the MS approach, one estimates the transition probabilities from the current regime to other regimes based on the regime change data historically observed, and assumes that such change patterns remain the same in the next forecasting period. On the other hand, CRS employs the runlength variable to sense how likely the base reactive model is going to be biased, due to possible changes in wind behavior, and then uses the training data to estimate the amount of bias to be corrected as a function of both the runlength and the current regime. We show in our case study that the CRS approach has an advantage over MS methods in terms of short-term forecasting, primarily because of the use of runlength, which provides a more specific anticipation of out-of-sample variations in the predictand. The comparison between CRS and MS conveys the message that an improvement in change anticipation can lead to appreciable gains in wind forecasting accuracy. Further discussion is provided in Section 4.

Finally, it is important to highlight that we in this study are specifically interested in turbine-specific, rather than the farm-level aggregated, short-term wind forecasts. Recently, several research studies hinted to the need for the former to replace the latter [Kusiak and Li, 2010; He et al., 2014; Pourhabib, Huang and Ding, 2016]. This is due to the fact that many wind farm operations are carried out at the turbine level and would naturally require turbine-specific forecasts, such as predictive turbine control [Santos, 2007], turbine-specific power estimation [Bessa et al., 2012; Lee et al., 2015], wake propagation prediction [Hwangbo, Johnson and Ding, 2018; You et al., 2017] and repair decisions for wind turbines [Byon and Ding, 2010; Byon, Ntaimo and Ding, 2010]. Understandably, turbine-specific forecasts can be easily transformed to the farm-level estimates either by spatial averaging (for wind speed) or by aggregation (for wind power), while the converse is generally not true.

The remainder of this paper is organized as follows. Section 2 comprises the data description and preliminary data analysis. Section 3 elucidates the building blocks of the CRS approach and outlines its implementation procedure. Section 4 presents a case study using actual wind farm data and the corresponding results. Lastly, we conclude this article in Section 5 and

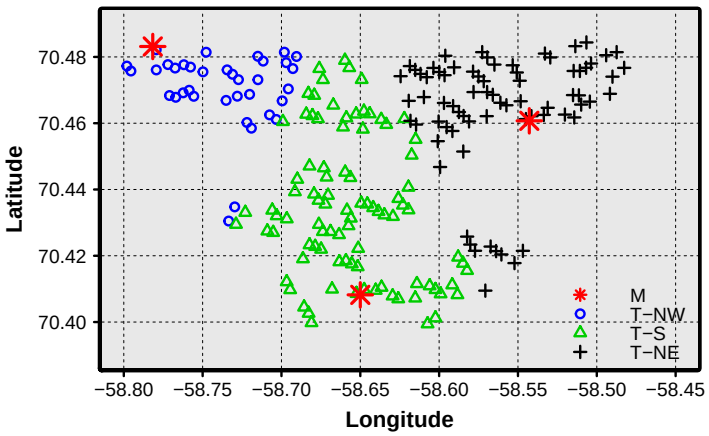


Fig 1: Spatial map of the wind farm (coordinates shifted to maintain confidentiality). M: meteorological masts located in the Northwest (NW), South (S), and Northeast (NE). T: Turbines grouped according to their proximity to the nearest mast, whether NW, S, or NE.

highlight future research directions beyond this work.

**2. Data structure and preliminary analysis.** The data in this study were collected at an onshore wind farm in the United States, located on a relatively flat terrain over approximately $25 \times 17$ square kilometers ($km^2$). Several treatments were carried out to preserve the data provider's confidentiality, such as shifting the geographic coordinates of the turbines, randomly selecting 200 turbines for analysis, and normalizing wind power outputs to the range of $[0, 1]$. Starting from September 2010, one year of turbine-specific hourly wind speed and power values are measured at the hub height of 80 meters on each of the 200 turbines. Additionally, hourly wind speed and direction values are measured at three spatially distant meteorological masts. The spatial map of the wind farm is presented in Figure 1.

Preliminary analysis suggests that the wind farm data exhibit dependencies across space and time. The first five panels of Figure 2 show the partial autocorrelation functions (PACF) of the wind speeds at five different wind turbines for lags ranging from one hour up to 12 hours. Short lags (less than four hours) appear to be of high relevance to the current observed wind speeds at each of the five turbines. Similar trends were observed in multiple research studies [Erdem and Shi, 2011; Kazor and Hering, 2015a; Pourhabib,

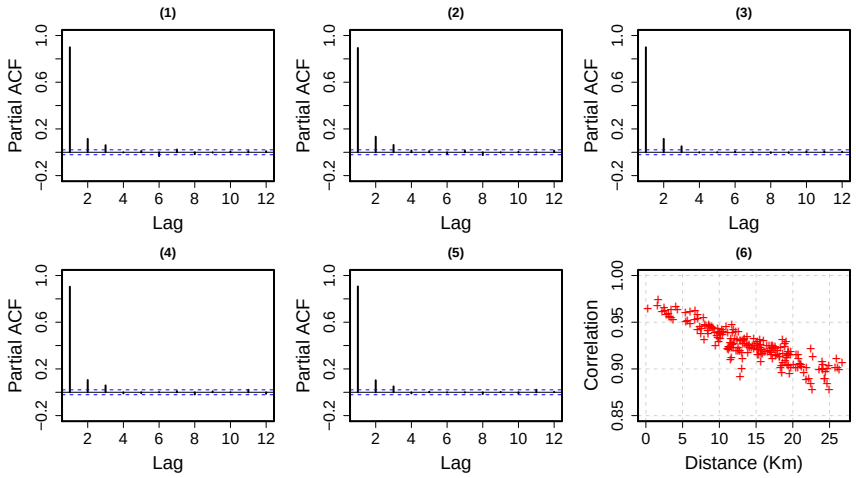


Fig 2: Panels (1)-(5): PACFs of the wind speeds at five wind turbines. The $x$-axis is the time lag in hours. Panel (6): Correlations between wind speeds at an arbitrarily chosen turbine and those at the remaining 199 turbines at zero-hour lag against the separating distances.

Huang and Ding, 2016]. The quick decay of temporal correlation hints to the irrelevance of longer-memory effects to short-term prediction.

The last panel of Figure 2 plots the correlation coefficients between the wind speeds at an arbitrarily chosen turbine and those at the remaining 199 turbines at zero-hour lag against the separating distances between them. The plot suggests that strong spatial correlations exist in the wind farm, ranging from 0.88 to 0.97 for this example. A decreasing trend in the coefficients is notable as the separating distance increases. Even at the near maximal distances, spatial correlations appear to sustain a relatively strong effect and do not vanish. The strong spatial dependence can be explained by wind propagation across a dense grid of spatial locations in the relatively small area of a wind farm. The observations made above advocate modeling both temporal and spatial dependencies when performing short-term wind forecasting.

Similar to previously reported works [Alexiadis et al., 1998; Reikard, 2008; Ezzat, Jun and Ding, 2018], we note that local wind conditions (i.e. wind speed and direction) exhibit changes of sizeable magnitude in short periods of time, suggesting that the wind field under study is highly dynamic. For exploratory purposes, we conduct two univariate change point tests on the first 30 days of spatially averaged data of wind speed and wind direction.

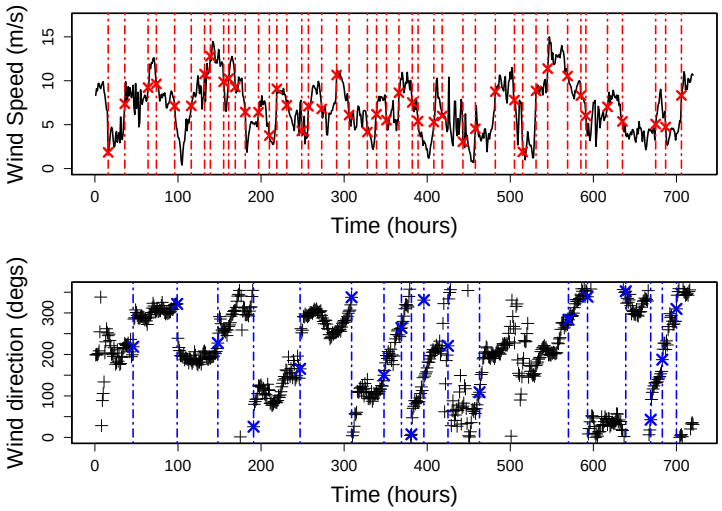


Fig 3: Top panel: change points in one month of spatially averaged wind speed data. Bottom panel: change points in one month of spatially averaged wind direction data. The span of the $x$-axis is a month or 720 hours.

For the wind speed variable, we implement a change point detection with binary segmentation using the package `changepoint` in `R` [Killick and Eckley, 2014]. For the wind direction, which is a circular variable, we implement a binary segmentation version of the circular change point detection [Jammalamadaka and SenGupta, 2001], using the package `circular` [Agostinelli and Lund, 2013]. Figure 3 illustrates the results of the change point detections. The minimum-time-to-change and the median-time-to-change in wind speed are 5 hours and 15 hours, respectively, whereas those in wind direction are 11 hours and 37 hours, respectively. On average, a change in wind speed or wind direction takes place every 10 hours.

From a physical interpretation standpoint, Pinson [2013] explains that the variation in local wind conditions can be observed on multiple frequency ranges, among which slow fluctuations (i.e. days) are driven by synoptic-scale weather variables, whereas higher frequency variations (i.e. minutes to hours) occur due to a combination of interacting physical processes that are difficult to individually pinpoint, yet their collective effect is often notable. The system's short memory noted in the PACFs of Figure 2 is perhaps an implication of this dynamic behavior.

**3. The calibrated regime-switching method.** Let $Y_i(j)$ represent the wind speed recorded at turbine $i$ at time $j$, where $i = 1, ..., N$ turbines and $j = 1, ..., T$ hours. Furthermore, let $\bar{Y}(j)$ and $d(j)$ be the spatially averaged wind speed and direction at time $j$, in m/s and degrees, respectively. Hereinafter, the time index, $t$, is reserved to denote the present time, whereas $j$ denotes an arbitrary time index. A wind speed forecast is to be made at $t+h$ for $h = 1, 2, \ldots, H$, i.e., the forecast horizon could be as far as $H$ hours ahead of the present time.

3.1. *Regime identification.* There are several ways to define regimes in the wind simulation and forecasting literature [Kazor and Hering, 2015a,b]. Clustering-based methods can be run on pre-specified meteorologic variables to identify regimes corresponding to physically interpretable weather patterns [Browell, Drew and Philippopoulos, 2018]. In other instances, regimes can be defined implicitly through latent variables such as in hidden Markov models [Ailliot, Monbet and Prevosto, 2006]. An alternative approach is to divide the space of some prescribed variables into a finite number of partitions by imposing a set of thresholds that are often selected through expert knowledge or data-driven measures. For example, Gneiting et al. [2006] postulated two regimes based on wind direction to reflect the alternation of westerly and easterly winds in the U.S. Pacific Northwest.

In this paper, we follow an approach similar to Gneiting et al. [2006], but we use both the wind speed and direction variables as regime identifiers. This is motivated by our explicit interest in capturing fluctuations of local wind conditions. The wind speed variable is the predictand of interest, and capturing its fluctuations would greatly benefit any data-driven method to forecast it, whereas wind direction is also known to affect wind flow patterns, power production, and spatial correlations of wind speed between wind turbines. So both factors are used in defining our wind regimes.

For our regime definition scheme, we assume there exists a finite number, $R$, of wind regimes, such that the observed regime at time $j$, denoted by $r(j)$, takes on values in $\{1, \ldots, R\}$ depending on the observed values of $\bar{Y}(j)$ and $d(j)$. Suppose that there are $R_1$ disjoint wind speed partitions and $R_2$ disjoint wind direction partitions, then $R = R_1 \times R_2$. Collectively, the regimes are defined by a set of bivariate thresholds, $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{R-1}\}$, where $\mathbf{r}_k = \{r_k^v, r_k^d\}$ comprises the wind speed and wind direction thresholds that separate two adjacent wind speed or wind direction partitions.

Our regime identification starts by selecting the number of regimes $R$ and a set of tentative thresholds $\{\mathbf{r}_k\}_{k=1}^{R-1}$. Our approach for deciding the tentative thresholds is based on the understanding of wind power production.

Later, in Section 4.2, we use a small subset of training data to fine-tune the tentatively selected thresholds in order to boost the forecast performance.

We guide the selection of wind speed thresholds, $r_1^v, \ldots, r_{R-1}^v$, in light of the regions associated with a wind power curve. Figure 4 plots one year of wind speed versus the normalized wind power values recorded at one of the turbines, as well as the histograms of wind speed and power shown, respectively, in the above and side sub-figures. The power curve in Figure 4 is estimated by using the binning method [IEC, 2005], a common industrial practice. The binning estimates are shown in Figure 4 as the red triangles.

Four physically meaningful values of wind speed are critical to defining a power curve, namely: the cut-in-speed, $V_{ci}$, the inflection point, $V_{in}$, the rated speed, $V_r$, and the cut-out speed, $V_{co}$. The cut-in speed $V_{ci}$ is the minimal speed at which turbines start to generate power. The rated speed $V_r$ defines the minimal speed at which the maximum permissible power is produced. This maximum level is maintained until the speed reaches $V_{co}$, beyond which the turbine is halted for its safety. Between $V_{ci}$ and $V_r$, the power curve follows a nonlinear relationship, with an inflection point separating the convex and concave regions. This inflection point, corresponding to $V_{in}$, marks the point at which the turbine control mechanism is used to regulate the power production. For our data, $V_{ci}$, $V_r$ and $V_{co}$, as provided by the manufacturer, are around 3.5, 13.5 and 20 m/s, respectively, whereas $V_{in}$ is estimated by Hwangbo, Johnson and Ding [2017] for modern turbines to be around 9.5 m/s.

These physically meaningful values can be used to set the wind speed thresholds at $V_{ci}$, $V_{in}$, and $V_r$, which partitions wind speed into four ranges. However, we notice that only around 3% of wind speed data in our dataset are higher than $V_r$. We then decide to have a total of three speed partitions after merging the last two partitions and eliminating the threshold at $V_r$. Please also note that while $V_{co}$ is at 20 m/s for our data, $V_{co}$ most commonly takes the value of 25 m/s in commercial turbines. Extending $V_{co}$ from 20 m/s to 25 m/s, however, does not affect our wind regime definition.

With respect to wind direction, Figure 5a shows the roseplot of the spatially averaged wind speed and direction data. We note that wind direction is dominantly westerly, and the year-long spatio-temporal average wind direction is 257° (270° is due west). Figure 5b shows the histograms of the turbine-specific wind speeds for westerly versus easterly partitions, which shows a notable distinction between the behavior of the wind associated with each direction partition. This suggests two wind direction partitions with thresholds set at 0° and 180°, resulting in a westerly regime and an easterly regime. The differences in wind speed histograms are not nearly as

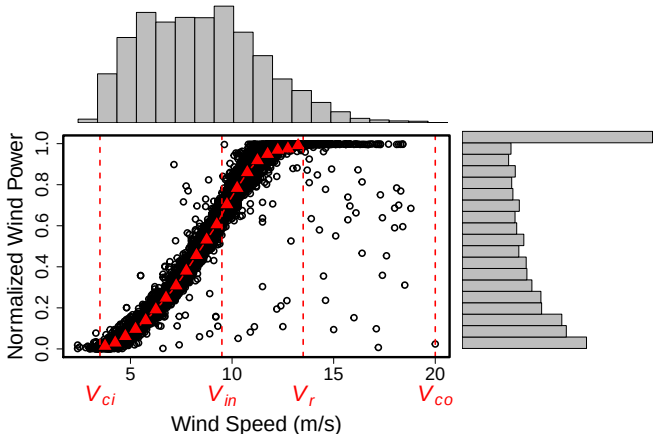


Fig 4: Normalized power versus speed values, histograms of wind speed and power, for one of the turbines on the wind farm. $V_{ci}$: cut-in speed, $V_{in}$: inflection point, $V_r$: rated speed and $V_{co}$: cut-out speed.

striking for northerly versus southerly partitions.

The above analysis motivates us to define a total of $R = 6$ wind regimes corresponding to the combination of three wind speed partitions and two wind direction partitions, as shown in (3.1). Figure 6 shows the time series of the first 30 days of $\bar{Y}(j)$, $d(j)$, and $r(j)$. In Section 4, the values of these tentative thresholds will be refined. That fine-tuning step is intended to further enhance the regime identification procedure in light of the forecasting performance. Our regime identification procedure is not purely data-driven nor solely dictated by physics; rather it borrows the strength of both. While the physical understanding of how wind dynamics affects wind power generation guides our high level decision, like the choice of the number of regimes and the tentative regime thresholds, the farm-specific data informs the subsequent fine-tuning of the regime thresholds.

(3.1)

$$r(j) = \begin{cases} 1, & \bar{Y}(j) \in [0, 3.5) & \& \; d(j) \in [0°, 180°) & \Rightarrow \text{Low speed, easterly wind} \\ 2, & \bar{Y}(j) \in [0, 3.5) & \& \; d(j) \in [180°, 360°) & \Rightarrow \text{Low speed, westerly wind} \\ 3, & \bar{Y}(j) \in [3.5, 9.5) & \& \; d(j) \in [0°, 180°) & \Rightarrow \text{Moderate speed, easterly wind} \\ 4, & \bar{Y}(j) \in [3.5, 9.5) & \& \; d(j) \in [180°, 360°) & \Rightarrow \text{Moderate speed, westerly wind} \\ 5, & \bar{Y}(j) \in [9.5, 20 \text{ or } 25) & \& \; d(j) \in [0°, 180°) & \Rightarrow \text{High speed, easterly wind} \\ 6, & \bar{Y}(j) \in [9.5, 20 \text{ or } 25) & \& \; d(j) \in [180°, 360°) & \Rightarrow \text{High speed, westerly wind} \end{cases}$$

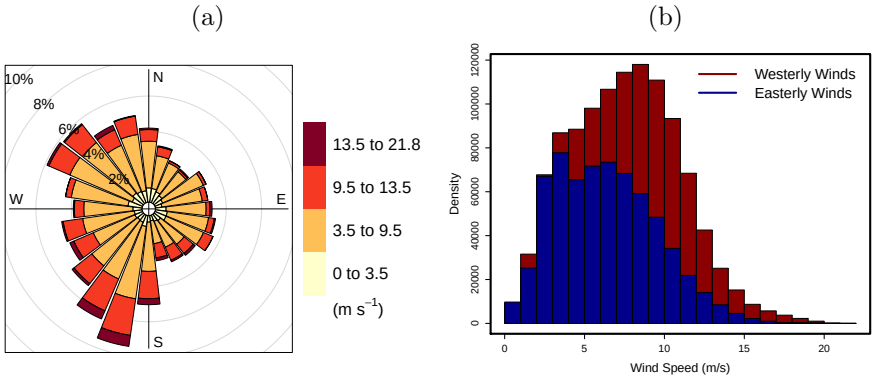


Fig 5: Left panel: Rose plot of spatially averaged wind speeds and directions. Right panel: Histogram of wind speeds for westerly versus easterly winds.

3.2. *The CRS method.* We assume that we have at hand a statistical model, denoted by $\mathcal{M}$. In this paper, $\mathcal{M}$ is a reactive regime-switching model, which, conditioned on $r(t)$, can produce regime-dependent spatio-temporal forecasts, where $f_i^{r(t)}(t+h)$ represents a forecast at location $i$ and time $t+h$, given the current regime of $r(t)$. Its specific choice in the context of our application is to be discussed in Section 4, but our method is generic to different selections of $\mathcal{M}$.

As a reactive regime-switching model, $\mathcal{M}$ suffers from an inherent bias while extrapolating the observed in-sample attributes in the form of short-term forecasts. The aim of the CRS approach is to correct this inherent bias through adding a regime-dependent term, $c_i(t+h|r(t)) \in \mathbb{R}$, to the regime-specific forecast $f_i^{r(t)}(t+h)$. Hereinafter, we refer to $c_i(t+h|r(t))$ as the regime-dependent *forecast calibration*, and the quantity $\hat{f}_i(t+h) = f_i^{r(t)}(t+h) + c_i(t+h|r(t))$ as the *calibrated forecast*.

We assume that the forecast calibration, $c_i(t+h|r(t))$, can be informed by the observed data up to time $t$, denoted by $\mathcal{D}_t$. The dependence on $\mathcal{D}_t$ is signified by the notation $c_i(t+h|\mathcal{D}_t, r(t))$. For simplicity, we assume that $c_i(t+h|\mathcal{D}_t, r(t))$ only varies over time and is fixed across space, that is, $c_i(t+h|\mathcal{D}_t, r(t)) = c(t+h|\mathcal{D}_t, r(t))$, for $i = 1, \cdots, N$, but future research can look into the benefit of varying that quantity over space as well. A general calibration formulation to infer $c(\cdot|\mathcal{D}_t, r(t))$ can be expressed as

$$\min_{c \in \mathcal{C}} \quad \mathcal{L}\big(f_i^{r(t)}(t+h) + c(t+h|\mathcal{D}_t, r(t)), Y_i(t+h)\big), \tag{3.2}$$

where $\mathcal{C}$ is some class of functions to which $c(\cdot|\mathcal{D}_t, r(t))$ belongs, and $\mathcal{L}(\cdot,\cdot)$

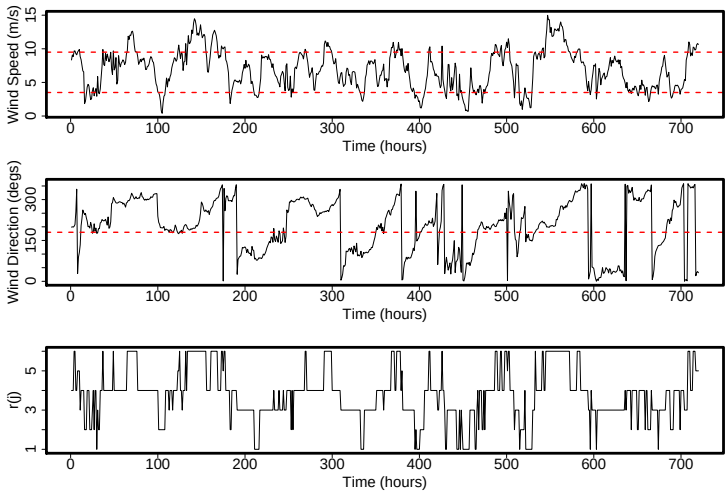


Fig 6: Top panel: Spatially averaged wind speed data (solid line), along with the wind speed regime thresholds at 3.5 and 9.5 m/s (dashed line). Middle panel: spatially averaged wind direction data (solid line), along with the wind direction regime threshold value at 180° (dashed line). Bottom panel: Evolution of the regime variable $r(j) \in \{1, ..., 6\}$. The span of the $x$-axis is a month or 720 hours.

is a loss function that defines a discrepancy measure.

We would like to select a parametric form for $c(\cdot|\mathcal{D}_t, r(t))$ that enables our model to specifically capture the out-of-sample variations in wind speed. We suggest that $c$ can be informed by means of two elements that are shown to be able indicators of out-of-sample change: the observed wind regime at the time of the forecast, $r(t)$, and the runlength, denoted by $x(t+h)$. The runlength, $x(t+h)$, is defined as the time elapsed since the most recent regime change. The value of the runlength at any arbitrary time index $j$ is defined as $x(j) = j - j^*$, where $j^*$ is the time at which the most recent change was observed such that $j^* < \min(j, t)$. For a time point in the forecast horizon, i.e. $j = t + h$, we define the runlength in the forecast horizon as $x(t+h) = t + h - j^*$. The runlength has been recently proposed as an indicator of upcoming changes in the emerging online change detection literature [Adams and MacKay, 2007; Saatçi, Turner and Rasmussen, 2010], but has not been used in the regime-switching modeling or wind forecasting literature in general.

As such, we let the function $c(\cdot)$ depend on two inputs: $\{r(t), x(t+h)\}$. We further propose to use the functional form of a log-normal cumulative distri-

bution function (cdf) to characterize $c(\cdot)$'s relationship with the two inputs. We model $c(\cdot)$ individually in each wind regime, so that $r(t)$ is implicitly incorporated into the relationship with $c(\cdot)$ through a set of regime-specific parameters. Consequently, $c(\cdot)$ is manifested with the runlength, $x(t+h)$, as the single explicit input.

For the $k$-th regime, we denote the regime-specific parameters by $\Psi^k = \{\psi_1^k, \psi_2^k, \psi_3^k\}$, and the calibration function by $c(x(t+h)|r(t)=k)$. The functional form of the log-normal cdf is expressed as in (3.3).

$$c(x(t+h)|r(t)=k) = \psi_1^k \Phi\left(\frac{\ln(x(t+h)) - \psi_2^k}{\psi_3^k}\right), \tag{3.3}$$

where $\Phi(\cdot)$ is the normal cdf. The choice of the lognormal cdf as a calibration function is motivated by its flexibility to model a wide spectrum of change behavior, ranging from abrupt shifts to gradual drifts, depending on the values of parameters in $\Psi^k$. The parameters in $\Psi^k$ are estimated from the data and continuously updated in a rolling forward forecasting scheme. Other selections of calibration functions are discussed in Section 4.

The estimation procedure of parameters in $\Psi^k$ goes as follows. We assume that in a training dataset, we have at hand a sequence of forecasts and their respective true observations. These forecasts were obtained via the model $\mathcal{M}$ in a rolling fashion, such that for the $\ell$-th roll, the data observed up to time $t^\ell$ are used to obtain forecasts from $t^\ell + 1$ till $t^\ell + H$. Suppose that we have $L$ forecasting rolls in the training set. For the $\ell$-th forecasting roll, $\ell = 1, \ldots, L$, we store the following information: the observed wind regime at the time of forecasting, $r^\ell(t)$, the associated runlengths, $x^\ell(t+h)$, the raw forecast via $\mathcal{M}$ at $t+h$, $f_i^{r^\ell(t)}(t+h)$, and the actual observation at $t+h$, $Y_i^\ell(t+h)$, for $h = 1, \ldots, H$ and $i = 1, \ldots, N$. By employing a squared error loss function, (3.2) can be re-written as in (3.4),

$$\min_{\Psi^k} \frac{1}{L^k \times N \times H} \sum_{\ell=1}^{L^k} \sum_{i=1}^{N} \sum_{h=1}^{H} \left[ f_i^{r^\ell(t)}(t+h) + c(x^\ell(t+h)|r^\ell(t)=k) - Y_i^\ell(t+h) \right]^2 \tag{3.4}$$

where $L^k$ denotes the number of forecasting rolls relevant to regime $k$. Solving (3.4) for each regime individually, i.e., for $k = 1, \ldots, R$, gives the least-squares estimate of the parameters in $\{\Psi^k\}_{k=1}^R$.

3.3. *Proposed implementation procedure.* We next propose an implementation procedure for the CRS approach, which comprises three sequential phases: (1) Phase I: generating the raw forecasts (via the model $\mathcal{M}$) in the initialization period, (2) Phase II: estimating the forecast calibration function based on the raw forecasts and the actual observations solicited during

Phase I, (3) Phase III: making continuously rolling-forward forecasts and updates. Phases I and II use a subset of the data, say, a one-month initialization period to set up the CRS machinery. In Phase III, the actual forecasting and testing are carried out for the remaining 11 months.

We assume that the model, $\mathcal{M}$, used for issuing the raw forecasts, can be parameterized by a set of parameters in $\Theta^k$, and is thus denoted as $\mathcal{M}(\Theta^k)$, where the superscript $k$ refers to the dependence of the model parameters on the current observed regime, $r(t) = k$. At each forecasting roll, we estimate $\Theta^k$ using a subset of historical data and obtain raw forecasts from $t+1$ till $t+H$. As such, the parameters in $\Theta^k$ are regime-specific and time-varying.

Using a sliding interval of six hours, this rolling mechanism is continued until we exhaust the initialization period, resulting in $L$ rolls. Once Phase I is finished, the goal of Phase II is to estimate the function $c$ using the information attained in Phase I, where (3.4) is solved for each regime independently in order to estimate the regime-specific parameters $\Psi^k$.

Afterwards, we proceed to Phase III, where the same rolling mechanism is performed. Specifically, at the present time $t$, we estimate $\Theta^k$ and obtain short-term forecasts from $t+1$ till $t+H$. We then compute the runlength values $x(t+h)$ for $h = 1, \ldots, H$. Using $r(t)$, $x(t+h)$, and $\Psi^k$, we calculate the values of $c(x(t+h)|r(t) = k)$ for $h = \{1, ..., H\}$ and use the resulting $c(x(t+h)|r(t) = k)$ to calibrate the raw forecasts. The window is then slid by six hours. At each forecasting roll, the last 30-day of data are used to update $\Psi^k$ by re-solving (3.4), given the newly revealed observations. As such, similar to $\Theta^k$, the parameters $\Psi^k$ are also regime-specific and time-varying. The cycle is repeated until the forecasts for the remaining 11 months are produced.

**4. Case Study: application to data from an onshore wind farm.** In this section, we conduct a case study of the CRS method using the wind farm data explained in Section 2. Our interest is to make one-hour ahead to 12-hour ahead wind speed and power forecasts, i.e., $H = 12$ and $h = 1, ..., 12$. We note that in the literature, there are no precise thresholds between short, medium or long-term forecasting, but the convention is that short-term forecasting is concerned with the prediction for a few hours ahead, a horizon that is critical to subsequent power system operations such as economic dispatch, electricity market operation, and reserve quantification [Pinson, 2013; Xie et al., 2014].

4.1. *Choice of $\mathcal{M}$.* We choose $\mathcal{M}$ as a reactive regime-switching model, for which the parameters are regime-specific, leading to regime-dependent wind speed forecasts. Specifically, we use a Gaussian random field (GRF)

in a spatio-temporal domain which makes use of information from recent wind conditions observed at the $i$-th turbine as well as those from its neighborhood. The model has the following form: $\mathbf{Y} = \mathbf{M}\mathbf{q} + \mathbf{e}$, where $\mathbf{M}$ is a pre-specified $NT \times p$ design matrix, $\mathbf{q} = [q_1, ..., q_p]^T$ is a vector of regression coefficients and $\mathbf{Y}$ is the $NT \times 1$ vector of spatio-temporal data points such that

$$\mathbf{Y} = [Y_1(1), \cdots, Y_N(1), Y_1(2), \cdots, Y_N(2), \cdots, Y_N(T), \cdots, Y_N(T)]^T.$$

Moreover, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a positive-definite covariance matrix, for which the entries are modeled by a stationary covariance function, denoted by $K(\mathbf{u}, w)$, such that $\mathbf{u} = (u_1, u_2)^T$ comprises the longitudinal and latitudinal lags, respectively, and $w$ is the temporal lag.

Defining $K(\cdot, \cdot)$ is a key aspect in a GRF since it dictates the spatio-temporal dependence structure. One particular aspect to consider when defining $K(\cdot, \cdot)$ is the transport effect of dominant winds, related to what is known in the geostatistical literature as the spatio-temporal "asymmetry" [Gneiting, 2002; Stein, 2005; Gneiting, Genton and Guttorp, 2007; Jun and Stein, 2007]. Asymmetry implies that a dominant wind direction causes a discrepancy between a stronger, along-wind spatio-temporal correlation and a weaker, span-wind correlation. Ezzat, Jun and Ding [2018] numerically show that spatio-temporal asymmetry in local wind fields exist and is flow dependent. To account for possible flow dependent asymmetries and following Gneiting, Genton and Guttorp [2007] and Ezzat, Jun and Ding [2018], we model $K(\cdot, \cdot)$ as a convex combination of two components: a fully symmetric non-separable model, denoted by $C_1$, and an asymmetric model, denoted by $C_2$, as in (4.1).

$$K(\mathbf{u}, w) = \sigma^2\big[(1-\lambda)C_1(\mathbf{u}, w) + \lambda C_2(\mathbf{u}, w)\big] + \eta \mathbb{1}_{\{||\mathbf{u}||=|w|=0\}}, \tag{4.1}$$

where $||\cdot||$ is the Euclidean norm, $\mathbb{1}_{\{\cdot\}}$ is an indicator function, $\eta \geq 0$ is the spatio-temporal nugget effect, and $\sigma^2 > 0$ is the spatio-temporal variance. The convex combination coefficient, $\lambda \in [0, 1]$, assigns the weight given to the asymmetric model $C_2$. For $C_1$, we use the non-separable model proposed by Gneiting [2002]:

$$C_1(\mathbf{u}, w) = \frac{1-\delta}{1+\alpha|w|^2}\left[\exp\left(-\frac{c||\mathbf{u}||}{(1+\alpha|w|^2)^{\frac{\beta}{2}}}\right) + \frac{\delta}{1-\delta}\mathbb{1}_{\{||\mathbf{u}||=0\}}\right], \tag{4.2}$$

where $0 \leq \delta < 1$. The parameters $\alpha$, $c \geq 0$ determine the inverse of the temporal and spatial ranges, respectively. The parameter, $\beta \in [0, 1]$, is the

non-separability parameter indicating the strength of interaction between the spatial and temporal components.

The asymmetric model $C_2$ characterizes asymmetry possibly existing in a local wind field. Conceptually, one way to define an asymmetric spatio-temporal covariance function is through the following form:

$$C_2(\mathbf{u}, w) = \mathbb{E}_{\mathbf{V}}\zeta(||\mathbf{u} - \mathbf{V}w||), \tag{4.3}$$

where $\mathbf{V}$ is a random vector in $\mathbb{R}^2$ and $\zeta(\cdot)$ is a valid spatial covariance function [Gneiting, 2002; Gneiting, Genton and Guttorp, 2007].

Appropriate specifications of $\mathbf{V}$ and $\zeta(\cdot)$ can yield different explicit representations of $C_2$. The stochastic nature of local wind fields motivate us to use the asymmetric model proposed by Schlather [2010], which lets $\zeta(x) = \exp(-x^2)$ and $\mathbf{V} \sim \mathcal{N}(\boldsymbol{\mu}, \frac{\mathbf{D}}{2})$, rather than defining $\mathbf{V}$ as constant representing a fixed prevailing flow as suggested by Gneiting, Genton and Guttorp [2007]. As such, the model in (4.3) is re-written as in (4.4):

$$C_2(\mathbf{u}, w) = \frac{1}{\sqrt{|\mathbf{1}_{2\times 2} + w^2\mathbf{D}|}} \exp\big[-(\mathbf{u} - w\boldsymbol{\mu})^T(\mathbf{1}_{2\times 2} + w^2\mathbf{D})^{-1}(\mathbf{u} - w\boldsymbol{\mu})\big], \tag{4.4}$$

where $|\cdot|$ in (4.4) denotes the matrix determinant.

To produce regime-dependent forecasts, we estimate the parameters in (4.1) using only the spatio-temporal data since the most recent regime change. In other words, conditional on $r(t)$, we only use training data that pertains to the most recently observed wind regime. By continuously updating these regime-specific parameters through the rolling mechanism described in Section 3.3, we naturally overcome temporal nonstationarity [Fuentes, 2001; De Luna and Genton, 2005; Pourhabib, Huang and Ding, 2016], which is expected to exist in local wind fields due to atmospheric boundary layer effects resulting in turbulence and wake effect constantly perturbing the wind propagation across the farm. As such, the parameters in (4.1) are both regime-specific, and time-varying. In practice, we only use temporal lags that are smaller than or equal to 4 hours for model training, as dictated by the PACFs of Figure 2. We also impose a minimum of 2 time lags in history to ensure a reliable estimation of the parameters in (4.1). With 200 turbines, this truncation gives between $400 = (200 \times 2)$ to $800 = (200 \times 4)$ data points, which are sufficient for parameter estimation.

We further account for nonstationarity across space by assuming local stationarity within sub-regions of the spatial domain [Fuentes, 2001]. We define three sub-regions of wind turbines based on their proximity to the three masts, as shown in Figure 1. Within each sub-region, we fit the stationary spatio-temporal model of (4.1) and obtain region-specific model parameters.

Maximum Likelihood Estimation (MLE) is used to estimate all model parameters, and is implemented using the routine `nlm` in `R`, except for $\boldsymbol{\mu}$ and $\mathbf{D}$, which are specified based on the region-specific wind velocity information. For each sub-region, we use the most recent history of wind speed and direction data as recorded at the mast to compute a time series of two-dimensional wind velocity vectors. Our estimate for $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ is the sample average of the longitudinal and latitudinal wind velocities of the time series vectors, whereas our estimate for $\frac{\mathbf{D}}{2}$ is the sample covariance matrix. Using the model of (4.1), we can obtain spatio-temporal ordinary kriging-based predictions.

4.2. *Practical considerations.* Before we present the case study results, we would like to stress a few practical considerations. Regarding the choice of the calibration function $c$, we have tried various choices including the exponential function, polynomial function, and log-normal cdf. Our analysis indicates that the log-normal cdf achieves the best performance, i.e., the lowest mean squared discrepancy between calibrated forecasts and actual observations, as measured by (3.4). We note, however, that differences in performance between different calibration functions are not that pronounced, suggesting that other appropriate selections of $c$ are equally acceptable.

Regarding the identification of wind regimes, we were interested in refining the tentative thresholds of (3.1) to boost the performance of the CRS approach. Using the first month of data, we tried 112 different combinations of regime thresholds, chosen as follows: we vary $r_1^v$ from $V_{ci}$ to $V_{ci} + 1.5$ with increments of 0.5 m/s, $r_2^v$ from $V_{in} - 1.5$ to $V_{in}$ with increments of 0.5 m/s, $r_2^d$ from $180° - 45°$ to $180° + 45°$ with $15°$ increment, and set $r_1^d = 360° - r_2^d$. Our analysis indicates that, in terms of calibration performance as measured by (3.4), the best regime thresholds are at $\{0, 4.5, 9.0, 20.0\}$ m/s for wind speed, and $\{45°, 225°\}$ for wind direction, resulting in $R = 6$ regimes.

Figure 7 illustrates the estimated calibration functions for the six regimes. It appears that the wind speed variable is the main determinant of the calibration sign and magnitude. For instance, the first two regimes (top row), which share the same wind speed profile (low wind speeds), both transit to regimes with higher wind speeds. Same finding applies to regimes with moderate wind speeds (regimes 3 and 4), and regimes with high wind speeds (regimes 5 and 6). The wind direction, however, appears to have a secondary, yet still important effect on the magnitude of the calibration. For instance, it appears that the magnitude of change is larger in regime 2 (westerly) than in regime 1 (easterly), and larger in regime 4 (westerly) than in regime 3 (easterly). The opposite happens in regimes 5 (easterly) and 6 (westerly).

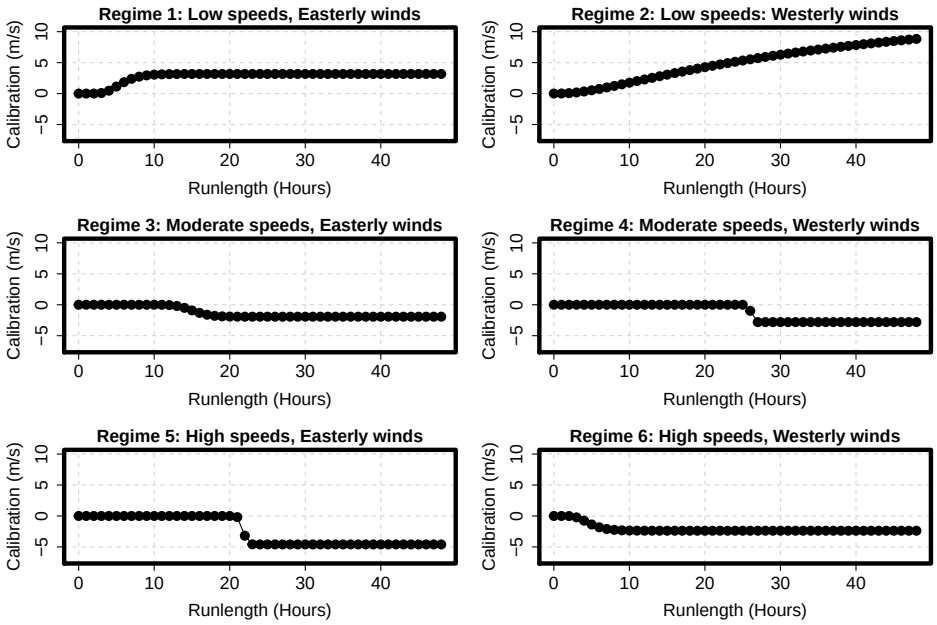


Fig 7: Estimated calibration functions using Phase I data for the six regimes.

The switching behavior difference between gradual shifts like in regimes 1, 2, 3, and 6 and abrupt shifts like in regimes 4 and 5 also implies certain degree of interaction between the two factors. As mentioned in Section 3.3, we allow these functions to change with time by continuously re-estimating $\Psi^k$ at every roll in Phase III.

With respect to the implementation of the CRS approach in Phase III, we decide to impose bounds on the forecast calibration to avoid over-calibrating the forecasts when extrapolating. Our experiments indicate that bounding the calibration quantities in the range $[-3, 3]$ m/s yielded satisfactory performance. Our analysis also suggests that, on average, calibrating forecasts does not offer much benefit in the very short-term horizons, so we decide to only calibrate the forecasts for more than two hours ahead forecasting, which means that CRS reduces to the model $\mathcal{M}$ for $h = 1, 2$.

4.3. *Forecasting results.* The rolling mechanism described in Section 3.3 is implemented on the remaining 11-month data, resulting in a total of 200 turbines $\times$ 12 hours $\times$ 1,339 rolls $=$ 3,213,600 forecasts. Using this massive test set, we compare the performance of the CRS method to the following approaches: asymmetric model (ASYM), separable model (SEP),

persistence forecast (PER), regime-switching autoregressive model (RSAR), Markov-switch autoregressive model (MSAR) and Markov-switch vector autoregressive model (MSVAR).

ASYM is in fact the reactive regime-switching base model, $\mathcal{M}$, used in the CRS approach, with its covariance function as specified in (4.1). Similar to ASYM is SEP, except that for that model, we set $\beta = \lambda = 0$ in (4.1) and freely estimate the rest of the parameters, yielding a separable spatio-temporal model. The persistence model is commonly used as a benchmark in forecasting studies and assumes that the current wind speeds will persist for the entire forecasting horizon.

RSAR is a reactive regime-switching approach, where the model parameters are dependent on $r(t)$, and $r(t)$ is assumed to persist in the forecast horizon. The autoregressive models used in RSAR are low-order autoregressive models like AR(1). Low-order AR models are common choices in the wind forecasting literature [Huang and Chalabi, 1995; Pourhabib, Huang and Ding, 2016]. To produce turbine-specific forecasts, we fit an RSAR model for each turbine.

For the MSAR model, we use Phase I data to estimate a transition probability matrix, $\Pi_{R\times R}$, of which each entry is $\pi_{kk'} = Pr(r(t+1) = k'|r(t) = k)$. Then, we fit six AR(1) models using the historical data classified to each regime. The final forecast at $t+h$ will be the convex combination of the forecasts from the six models, where the combination coefficients correspond to the probability of reaching each regime at $t + h$, that is: $\hat{f}_i(t + h) = \sum_{k=1}^{R} f_i^k(t + h)Pr(r(t + h) = k)$, where $f_i^k(t + h)$ is the forecast at the $i$-th turbine at $t + h$ obtained by fitting an AR(1) model to the data belonging to regime $k$, and $Pr(r(t + h) = k)$, $k = 1, ..., R$, denotes the probability of reaching regime $k$ at $t + h$. At each forecasting roll, the transition matrix is re-estimated using the newly revealed observations.

Generalizing on MSAR is the MSVAR model which further accounts for the spatial dependence. Attempting to use all 200 turbines in a VAR model would require the estimation of a large number of parameters. Therefore, we follow an approach similar to Pourhabib, Huang and Ding [2016], where for each turbine, we fit six VAR(1) models corresponding to each regime using the historical observations from the turbine itself and its nearest five neighbors, and obtain a final weighted prediction. We tested increasing the informative neighborhood up to ten turbines and the change in prediction performance was almost negligible.

We compare all the aforementioned competing models in both wind speed and power domains. For wind power forecasting, we first make a wind speed forecast and then convert it to the corresponding wind power forecast using

the turbine-specific power curve. Figure 8 shows, at one of the randomly selected turbines, the 6-hour ahead (i.e., one sliding interval) forecasts of wind speed and power, made by using CRS, versus the actual observations for five days starting on November 23, 2010.

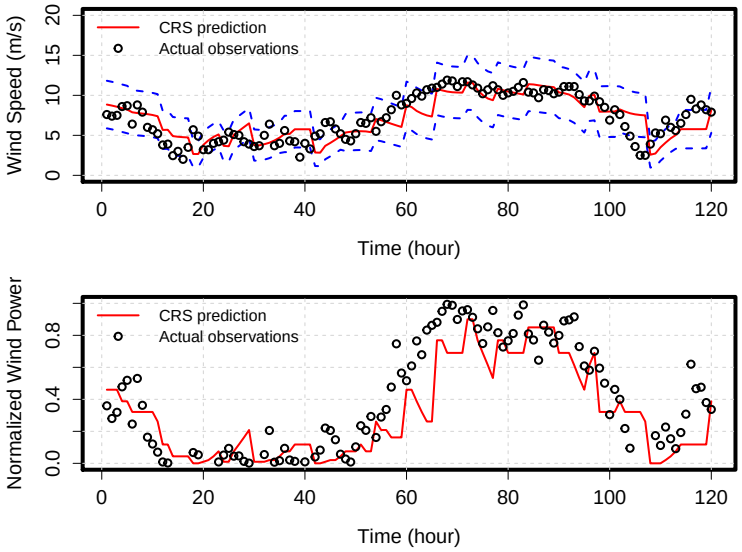


Fig 8: Top panel: 6-hour ahead forecasts of wind speeds at the chosen turbine for five days starting on November 23, 2010. Point forecasts shown by the solid red line, along with $\pm 1$ standard deviation as the dashed blue lines. Actual observations shown as black circles. Similarly, the bottom panel shows the corresponding power forecasts and the actual observations.

We evaluate the overall forecast accuracy over all testing instances using the Mean Absolute Error (MAE), as expressed in (4.5).

$$\text{MAE}(h) = \frac{1}{L \times N} \sum_{\ell=1}^{L} \sum_{i=1}^{N} |\hat{f}_i^{\ell}(t+h) - Y_i^{\ell}(t+h)|, \tag{4.5}$$

where $Y_i^{\ell}(t+h)$ and $\hat{f}_i^{\ell}(t+h)$ are, respectively, the observation and point forecast (from any of the competing models) at the $i$-th location and the $h$-th horizon during the $\ell$-th forecasting roll. For each forecasting horizon $h = 1, ..., 12$, the associated MAE is computed over all turbines and forecasting rolls using the 11-month of test data. The resulting MAE values are presented in Table 1.

We note that MAE as a loss measure suggests the use of the medians of the predictive distributions as optimal point forecasts [Gneiting, 2011a]. Since the mean and median coincide under a Gaussian predictive distribution, the point forecasts used in (4.5) for ASYM and CRS are the raw and calibrated kriging-based predictions, respectively. Similarly, RSAR, MSAR and MSVAR are all based on Gaussian (vector) autoregressive models, and as such, the means of the resulting predictive distributions (in case of RSAR), or the convex combination of the means (in case of MSAR and MSVAR), are used as the point forecasts in (4.5).

In addition to MAE, Pinson, Chevallier and Kariniotakis [2007] note that, in deregulated electricity markets, under-estimating power is often a more costly situation than over-estimating it. To reflect this cost trade-off, Hering and Genton [2010] propose the power curve error (PCE), as expressed in (4.6) to evaluate the power forecast.

(4.6)
$$PCE(P,\hat{P}) = \begin{cases} g\big[P_i(t+h) - \hat{P}_i(t+h)\big], & \text{if } f_i(t+h) \le Y_i(t+h) \\ (1-g)\big[\hat{P}_i(t+h) - P_i(t+h)\big], & \text{if } f_i(t+h) > Y_i(t+h), \end{cases}$$

where $P_i(t+h)$ and $\hat{P}_i(t+h)$ are the normalized power observations and forecasts at $t+h$ and the $i$-th location, and $g$ is the weight given to under-estimates, which is usually set at values higher than 0.5 to penalize under-estimates more than over-estimates. Pinson, Chevallier and Kariniotakis [2007] suggest a value of $g = 0.73$. Under the loss function of (4.6), Gneiting [2011b] shows that the $g$-th quantile of the predictive distribution is an optimal point forecast. For ASYM, SEP, CRS and RSAR, the $g$-th quantile is directly computed from the resulting Gaussian predictive distribution and is used as the input point forecast to the PCE loss function in (4.6). For MSAR and MSVAR, however, the predictive distribution is a mixture of Gaussians, for which the quantiles do not have closed form expressions. Therefore, we numerically compute the $g$-th quantile of the predictive distribution and use it as a point forecast for the PCE loss function in (4.6). In Table 2, we present the average PCE values across all horizons for values of $g$ ranging between 0.5 and 0.8 with 0.1 increment, as well as $g = 0.73$. We stress that when computing MAE in (4.5) and PCE in (4.6) for the CRS approach, $\hat{f}_i(t+h)$ is the calibrated forecast.

Table 1 demonstrates that the CRS approach outperforms the competing models in terms of wind speed forecasting for $h > 1$. We believe that this is mainly the result of better capturing the out-of-sample variations in the wind speed variable using $c(x(t+h)|r(t))$. Additional benefits over temporal-only

TABLE 1
*MAE for wind speed and power forecasting for h-hour ahead, $h = 1, 2, \ldots, 12$. Bold-faced values indicate best performance.*

| MAE for wind speed forecasts issued at $h = 1, 2, \ldots, 12$ | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| ASYM | 1.12 | **1.45** | 1.72 | 1.96 | 2.15 | 2.27 | 2.39 | 2.51 | 2.68 | 2.77 | 2.83 | 2.87 |
| SEP | 1.15 | 1.47 | 1.74 | 1.97 | 2.15 | 2.27 | 2.40 | 2.52 | 2.68 | 2.77 | 2.84 | 2.87 |
| PER | **1.11** | 1.46 | 1.73 | 1.97 | 2.16 | 2.31 | 2.44 | 2.57 | 2.74 | 2.84 | 2.92 | 2.96 |
| RSAR | 1.16 | 1.53 | 1.79 | 2.03 | 2.21 | 2.36 | 2.46 | 2.56 | 2.73 | 2.82 | 2.89 | 2.93 |
| MSAR | 1.23 | 1.64 | 1.92 | 2.14 | 2.28 | 2.38 | 2.45 | 2.48 | 2.54 | 2.59 | 2.62 | 2.63 |
| MSVAR | 1.21 | 1.60 | 1.87 | 2.09 | 2.23 | 2.33 | 2.40 | 2.45 | 2.52 | 2.57 | 2.60 | 2.61 |
| CRS | 1.12 | **1.45** | **1.71** | **1.89** | **2.06** | **2.15** | **2.25** | **2.29** | **2.37** | **2.44** | **2.52** | **2.56** |
| MAE for wind power forecasts issued at $h = 1, 2, \ldots, 12$ | | | | | | | | | | | | |
| Method | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| ASYM | **.121** | **.156** | **.184** | .212 | .227 | .236 | .247 | .261 | .280 | .291 | .294 | .296 |
| SEP | .123 | .158 | .185 | .212 | .227 | .236 | .247 | .261 | .280 | .292 | .295 | .296 |
| PER | .125 | .161 | .189 | .215 | .230 | .241 | .253 | .268 | .286 | .299 | .303 | .304 |
| RSAR | .129 | .169 | .199 | .226 | .241 | .253 | .264 | .278 | .297 | .309 | .314 | .314 |
| MSAR | .132 | .171 | .200 | .220 | .233 | .242 | .249 | .258 | .263 | .267 | .268 | .269 |
| MSVAR | .131 | .170 | .198 | .217 | .228 | .238 | .245 | .256 | .262 | .266 | **.267** | **.267** |
| CRS | **.121** | **.156** | .186 | **.207** | **.220** | **.229** | **.239** | **.244** | **.254** | **.263** | .268 | .271 |

and separable spatio-temporal models come from the incorporation of comprehensive spatio-temporal correlations and flow dependent asymmetries. For the very short-term horizon, $h = 1$, PER offers the best performance, with CRS slightly behind, but still enjoying a competitive performance.

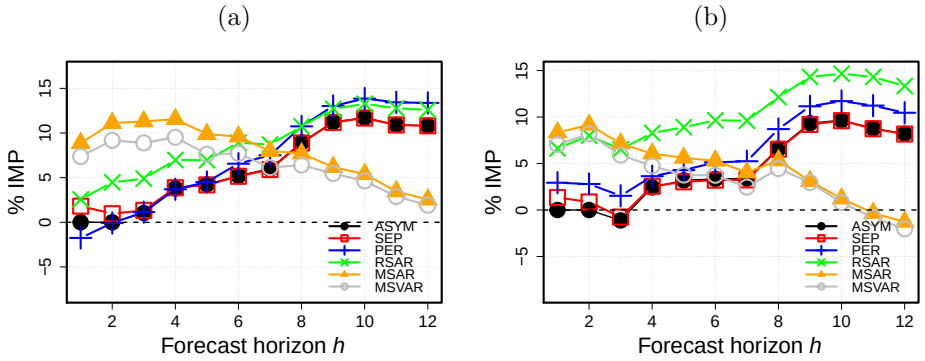


Fig 9: Percentage improvements in MAE of CRS over competing approaches in wind speed (left panel) and wind power (right panel).

Figure 9a presents the percentage improvements, in terms of MAE of wind speed forecast, that CRS achieves over the competing models at different forecast horizons. The percentage improvement over reactive methods such as ASYM, SEP, RSAR and PER is more substantial as the look-ahead

horizon increases. This does not come as a surprise since the farther the look-ahead horizon is, the more likely a change to take place in wind speed, and hence, the benefit of capturing the out-of-sample variations by means of the runlength variable becomes more appreciated.

The trend of the improvement of CRS over the Markov-switching approaches, i.e., MSAR and MSVAR, is opposite. For short-term horizons, the performance of CRS is remarkably better than the MS approaches. As the look-ahead horizon increases, the advantage of CRS over MS models reaches a peak around $h = 4$ hours, and after that, the performance of the MS approaches gradually catches up with that of CRS. This trend suggests that CRS anticipates out-of-sample wind speed variations in the early portion of the forecast horizon better than the MS approaches. We believe that this is rooted in the mechanisms each approach relies on: CRS uses the runlength, whereas MSAR and MSVAR use the transition probabilities.

The analysis presented in Figure 10 helps explain the advantage of CRS over MSVAR. We evaluate how each method (CRS versus MSVAR) handles out-of-sample variations in wind speed. We define an out-of-sample change in wind speed as crossing a wind speed regime threshold set at either 4.5 or 9.0 in the forecast horizon. For a given $h$ between 3-hour ahead and 12-hour ahead, if both the actual observation $Y_i(t+h)$ and its corresponding forecast $\hat{f}_i(t+h)$ cross the same speed threshold, we label that as a true positive; while on the other hand, when neither $Y_i(t+h)$ nor $\hat{f}_i(t+h)$ crosses any speed threshold, we label it as a true negative. In the left panel of Figure 10, we plot the true positive rate (TPR) of CRS versus MSVAR for $h = 3, 4, \ldots, 12$. Similarly, the right panel of Figure 10 plots the true negative rate (TNR). This analysis is performed on all $1,339$ forecasting rolls in the 11 months of test data and provides an empirical estimation of the TPR and TNR. Apparently, CRS does better than MSVAR in terms of both measures in the middle range of the forecast horizon, between 4-hour ahead and 10-hour ahead. This is consistent with the difference between the two methods observed in Figure 9. In terms of the true positive rate, which is the proportion of correct anticipation of a change, CRS performs better than MSVAR for smaller $h$'s, while MSVAR is more conservative in signaling changes under those circumstances, suggesting that the use of runlength offers a higher degree of change anticipation in the shorter horizons.

Similar findings can be extended to the power forecasting results in Table 1; the CRS approach outperforms the competing models for power prediction for most forecasting horizons. Its improvement over the reactive methods is higher as the look-ahead horizon increases, whereas its improvement over the MS approaches is better in the shorter forecast horizons. In the case of

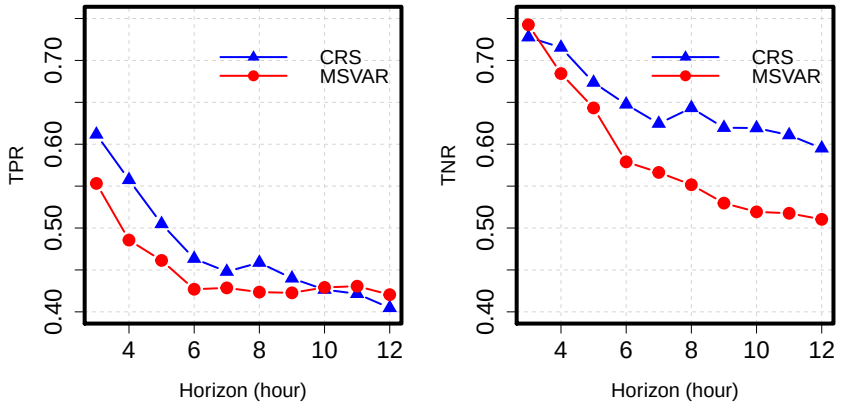


Fig 10: Left panel: the true positive rate (TPR). Right panel: the true negative rate (TNR). Comparisons are between CRS (blue triangles) and MSVAR (red circles) for $h = 3, ..., 12$.

TABLE 2

*Average PCE values across all horizons. Bold-faced values indicate best performance. *indicates the value suggested by Pinson, Chevallier and Kariniotakis [2007].*

| Method | $g = 0.5$ | $g = 0.6$ | $g = 0.7$ | $g = 0.73^*$ | $g = 0.8$ |
|---|---|---|---|---|---|
| ASYM | .116 | .117 | .114 | .111 | .104 |
| SEP | .116 | .118 | .114 | .112 | .105 |
| PER | .118 | .121 | .124 | .125 | .127 |
| RSAR | .123 | .123 | .120 | .117 | .110 |
| MSAR | .113 | .123 | .127 | .124 | .126 |
| MSVAR | .112 | .118 | .122 | .118 | .119 |
| CRS | **.109** | **.110** | **.107** | **.105** | **.097** |

wind power forecast, the performance of MS approaches in the end surpass that of CRS at $h = 11$. The percentage improvements shown in Figure 9b are somewhat different from their counterparts in Figure 9a; the difference is mainly due to the nonlinear speed-to-power conversion.

In Table 2, it appears that the improvement of CRS over the competing models is also realizable in terms of PCE. The CRS approach performs well compared to the competing approaches, especially when the under-estimation is penalized more severely than over-estimation (namely $g > 0.5$), which describes the more realistic cost trade-off in power systems.

The improvements presented above are indeed significant from a practical point of view. With sometimes double digit percentage improvements in wind speed forecasts, using the proposed method can lead to major benefits in a wide set of operational analytics on the wind farm such as predictive turbine control, power estimation and economic dispatch, among others.

It is still important, however, to test if these improvements are significant from a statistical point of view. Similar to Hering and Genton [2010], we implement the large sample test introduced by Diebold and Mariano [1995] for comparing the forecasting accuracy of two models at a specific forecasting horizon. For $h = 1, \ldots, 12$, and $i = 1, \ldots, 200$, we implement a one-sided version of the test, corresponding to a sample size of $1,339$ forecasts per turbine per horizon. Figure 11 shows the boxplots of the 200 $p$-values of the pairwise comparisons for the CRS approach against the competing models at each horizon. Again, the improvements from the CRS approach are mostly significant against the reactive methods in larger forecast horizons ($h > 3$), while the difference between the MS approaches and CRS is significant at the small and moderate time lags. Recall that CRS adds the forecast calibration only when $h \geq 3$, which means that CRS is the same as ASYM for $h = 1$ and $h = 2$. For this reason, the first two p-values, corresponding to $h = 1$ and $h = 2$ in the top-left panel of Figure 11, are supposed to be one; these p-values are trivial and thus not shown.

4.4. *Adaptation to and impact on probabilistic forecasting.* Probabilistic forecasting stems from the importance of characterizing distributions associated with point forecasts to subsequent optimal decision makings [Pinson, 2013; Gneiting and Katzfuss, 2014]. While our discussion so far primarily focuses on the $h$-hour ahead point forecasting, the resulting method can be placed in the framework of a probabilistic forecasting, simply because the essence of the CRS approach is to make an adjustment to the mean prediction of a statistical model that can be used for making probabilistic forecasts.

To elaborate, our choice of $\mathcal{M}$ is ASYM, which is one type of Gaussian random field model, for which the point forecast is the mean of a predictive normal distribution. In other instances, the point forecast can be the mean of a truncated normal distribution as in Gneiting et al. [2006] and Pourhabib, Huang and Ding [2016]. In all cases, the CRS approach inherits the inherent uncertainty from $\mathcal{M}$, but offers an enhancement to reduce bias in the point forecasting. In other words, having started with a probabilistic forecasting using the base model, CRS aims at reducing the bias of the original forecasting without affecting its predictive variance.

To illustrate this point, we show in Figure 12 the forecast at one of the turbines from an arbitrarily selected forecasting roll. In the left panel, it shows that the calibrated forecasts (in blue) by CRS are able to pick up an out-of-sample change in wind speed, in contrast to the reactive ASYM model (in green), which fails to do so, as it solely relies on extrapolating

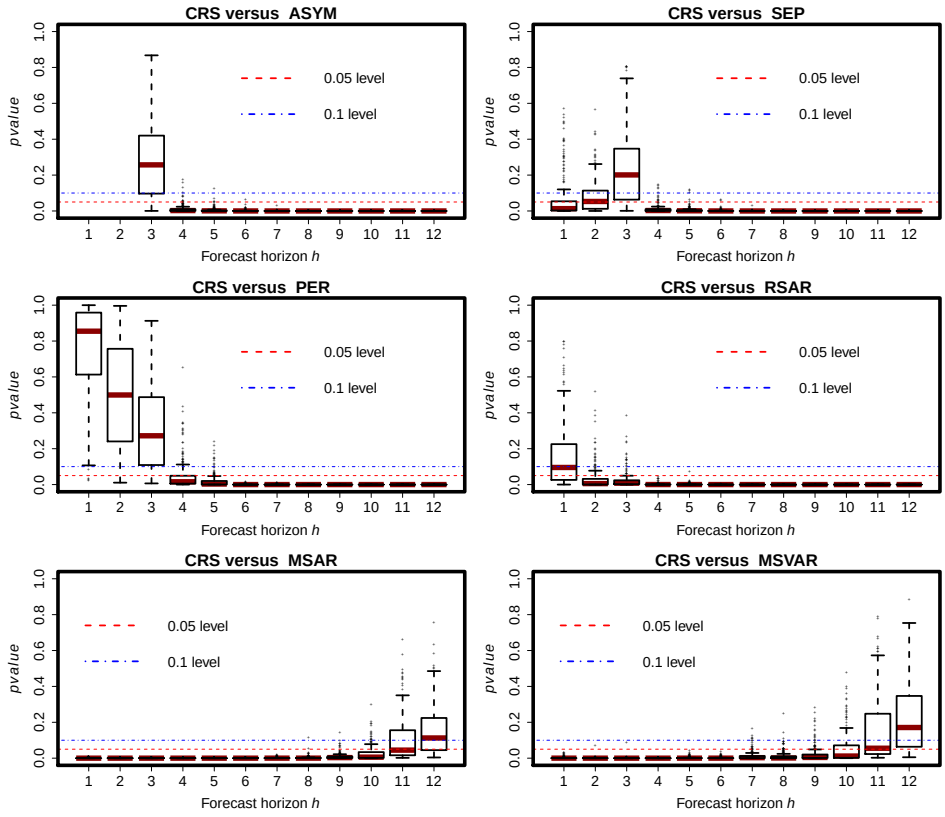


Fig 11: Boxplots of $p$-values generated from conducting the one-sided test of Diebold and Mariano [1995] to compare the turbine-specific CRS forecasts against those from the competing models at different horizons.

the observed in-sample wind regime. The right panel depicts the predictive distributions associated with the point forecasts of CRS and ASYM at $h = 4$. The benefit of the CRS approach is to correct the inherent bias in the forecasts of the reactive ASYM model by shifting the mean of the resulting distribution towards the true value, while preserving the variance of the original model.

To summarize, the CRS method is a bias correction approach, rather than a variance reduction endeavor. The uncertainty in the resulting forecasts depends largely on the selection of $\mathcal{M}$. Since the CRS approach is generic to different selections of $\mathcal{M}$, the decision-maker then has the luxury to select or tune $\mathcal{M}$ accordingly to achieve a bias-variance trade-off as dictated by the application under study.

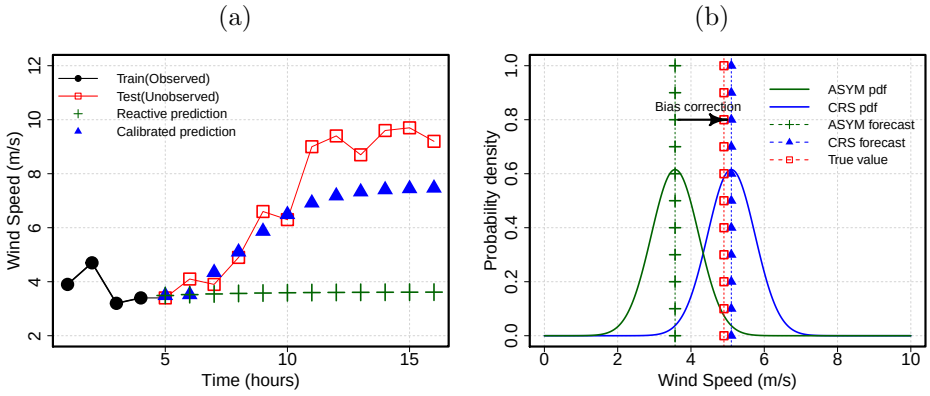


Fig 12: Left panel: CRS versus ASYM predictions for one turbine from an arbitrarily selected forecasting roll. Calibrated predictions shown without imposing the bounds discussed in Section 4.2. Right panel: Predictive distributions for the point forecast at $h = 4$. Absolute errors of point predictions at $h = 4$ are, respectively, 0.20 and 1.33 for CRS and ASYM.

TABLE 3
*CRPS values of wind speed prediction for h-hour ahead, $h = 3, \ldots, 12$. Bold-faced values indicate best performance.*

| Method | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|
| ASYM | 1.34 | 1.55 | 1.71 | 1.82 | 1.94 | 2.05 | 2.20 | 2.28 | 2.35 | 2.38 |
| CRS | **1.32** | **1.47** | **1.62** | **1.70** | **1.79** | **1.83** | **1.91** | **1.97** | **2.05** | **2.08** |

A standard evaluation measure in probabilistic forecasting is the continuous ranked probability score (CRPS). In Table 3, we compare the calibrated forecasts (CRS) versus the uncalibrated, reactive forecasts (ASYM) in terms of CRPS using the 11-month worth of test data. Note that the essential difference between CRS and ASYM is the calibration part. For all $h \geq 3$ cases, CRS outperforms ASYM, as much as 13.6% for some $h$'s. This comparison empirically supports the benefit of bias reduction rendered by CRS and highlights the impact of CRS under a probabilistic forecasting framework.

**5. Conclusions and future directions.** This paper proposes the calibrated regime-switching method for short-term wind forecasting. The essence of the CRS approach is to calibrate raw forecasts from a reactive regime-switching statistical model to capture the out-of-sample variations in the wind speed variable taking place in the forecast horizon. Extensive testing using one year, 200-turbine wind farm data suggests that the CRS approach can offer substantial benefit that enhances the forecast accuracy over a wide

spectrum of existing methods in both wind speed and wind power domains.

One important message conveyed in this paper is that an improvement in change anticipation can lead to appreciable improvements in forecasting quality. The CRS approach is an important step towards steering the focus of the literature and practice from the reactive "regime detection" models towards the next-generation proactive "regime anticipation" models. A fully proactive regime-switching model would of course not confine itself to solely calibrating reactive regime-switching forecasts, but would, instead, address the more general and challenging problem of predicting directly the out-of-sample wind regimes, thus naturally producing predictions that are adjusted to future regime changes. Albeit not fully proactive yet, we hope that the CRS approach paves one of the pathways and makes a solid step forward towards attaining that goal.

Some other ongoing and future research directions related to the CRS approach include, but are not limited to, looking into nonparametric modeling of the calibration function, using different indicators to inform the calibration action, and reducing both bias and variance in a probabilistic forecasting framework. Varying the calibration function over space, in addition to time, as mentioned in Section 3.2, is also a matter of promising future research.

**Acknowledgements.** This research was partially supported by NSF grants no. CMMI-1300560, IIS-1741173, DMS-1613003, and NIH grant no. P42ES027704.

## References.

Adams, R. P. and MacKay, D. J. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*.

Agostinelli, C. and Lund, U. (2013). R package circular: circular statistics (version 0.4-7). *URL https://r-forge.r-project.org/projects/circular*.

Ailliot, P., Monbet, V. and Prevosto, M. (2006). An autoregressive model with time-varying coefficients for wind fields. *Environmetrics* **17** 107–117.

Ailliot, P., Bessac, J., Monbet, V. and Pene, F. (2015). Non-homogeneous hidden Markov-switching models for wind time series. *Journal of Statistical Planning and Inference* **160** 75–88.

Alexiadis, M., Dokopoulos, P., Sahsamanoglou, H. and Manousaridis, I. (1998). Short-term forecasting of wind speed and related electrical power. *Solar Energy* **63** 61–68.

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.

Bessac, J., Ailliot, P., Cattiaux, J. and Monbet, V. (2016). Comparison of hidden and observed regime-switching autoregressive models for (u, v)-components of wind

fields in the Northeast Atlantic. *Advances in Statistical Climatology, Meteorology and Oceanography* **2** 1–16.

Browell, J., Drew, D. R. and Philippopoulos, K. (2018). Improved very short-term spatio-temporal wind forecasting using atmospheric regimes. *Wind Energy* **on-line** 1–12.

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.

Byon, E. and Ding, Y. (2010). Season-dependent condition-based maintenance for a wind turbine using a partially observed Markov decision process. *IEEE Transactions on Power Systems* **25** 1823–1834.

Byon, E., Ntaimo, L. and Ding, Y. (2010). Optimal maintenance strategies for wind turbine systems under stochastic weather conditions. *IEEE Transactions on Reliability* **59** 393–404.

De Luna, X. and Genton, M. G. (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica* **15** 547–568.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13** 253–263.

DOE (2015). Wind Vision: A New Era for Wind Power in the United States. US Department of Energy, Washington D.C.

Dowell, J. and Pinson, P. (2016). Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid* **7** 763–770.

Erdem, E. and Shi, J. (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy* **88** 1405–1414.

Ezzat, A. A., Jun, M. and Ding, Y. (2018). Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting. *IEEE Transactions on Sustainable Energy* **9** 1437–1447.

Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics* **12** 469–483.

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 Edition Technical Report, ANEMOS.plus.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97** 590–600.

Gneiting, T. (2011a). Making and evaluating point forecasts. *Journal of the American Statistical Association* **106** 746–762.

Gneiting, T. (2011b). Quantiles as optimal point forecasts. *International Journal of Forecasting* **27** 197–207.

Gneiting, T., Genton, M. G. and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistics of Spatio-Temporal Systems* (B. Finkenstaedt, L. Held and V. Isham, eds.) 4, 151–175. Monographs in Statistics and Applied Probability. Chapman & Hall/CRC Press.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *The Annual Review of Statistics and its Application* **1** 125–151.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G. and Aldrich, E. (2006). Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method. *Journal of the American Statistical Association* **101** 968–979.

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.

Hering, A. S. and Genton, M. G. (2010). Powering up with space-time wind forecasting. *Journal of the American Statistical Association* **105** 92–104.

Hering, A. S., Kazor, K. and Kleiber, W. (2015). A Markov-switching vector autoregressive stochastic wind generator for multiple spatial and temporal scales. *Resources* **4** 70–92.

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.

Hwangbo, H., Johnson, A. L. and Ding, Y. (2017). A production economics analysis for quantifying the efficiency of wind turbines. *Wind Energy* **20** 1501–1513.

Hwangbo, H., Johnson, A. L. and Ding, Y. (2018). Spline model for wake effect analysis: Characteristics of a single wake and its impacts on wind turbine power generation. *IISE Transactions* **50** 112–125.

IEC (2005). Wind Turbines-Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines. IEC-International Electrotechnical Commission, IEC-61400-12, Geneva, Switzerland.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific.

Jun, M. and Stein, M. (2007). An approach to producing space-time covariance functions on spheres. *Technometrics* **49** 468–479.

Kazor, K. and Hering, A. S. (2015a). The role of regimes in short-term wind speed forecasting at multiple wind farms. *Stat* **4** 271–290.

Kazor, K. and Hering, A. S. (2015b). Assessing the performance of model-based clustering methods in multivariate time series with application to identifying regional wind regimes. *Journal of Agricultural, Biological, and Environmental Statistics* **20** 192–217.

Killick, R. and Eckley, I. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software* **58** 1–19.

Kusiak, A. and Li, W. (2010). Estimation of wind speed: A data-driven approach. *Journal of Wind Engineering and Industrial Areodynamics* **98** 559–567.

Lee, G., Ding, Y., Genton, M. G. and Xie, L. (2015). Power curve estimation with multivariate environmental factors for inland and offshore wind farms. *Journal of the American Statistical Association* **110** 56–67.

Mohandes, M. A., Halawani, T. O., Rehman, S. and Hussain, A. A. (2004). Support vector machines for wind speed prediction. *Renewable Energy* **29** 939–947.

Pinson, P. (2013). Wind energy: forecasting challenges for its operational management. *Statistical Science, Special Issue on Mathematics of Planet Earth* **28** 564–585.

Pinson, P., Chevallier, C. and Kariniotakis, G. N. (2007). Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems* **22** 1148-1156.

Pinson, P., Christensen, L., Madsen, H., Sørensen, P. E., Donovan, M. H. and Jensen, L. E. (2008). Regime-switching modelling of the fluctuations of offshore wind generation. *Journal of Wind Engineering and Industrial Aerodynamics* **96** 2327–2347.

Pourhabib, A., Huang, J. Z. and Ding, Y. (2016). Short-term wind speed forecast using measurements from multiple turbines in a wind farm. *Technometrics* **58** 138–147.

Reikard, G. (2008). Using temperature and state transitions to forecast wind speed. *Wind Energy* **11** 431–443.

Saatçi, Y., Turner, R. and Rasmussen, C. E. (2010). Gaussian process change point models. In *Proceedings of the 27th International Conference on Machine Learning, (ICML-10)* 927–934.

Santos, R. A. (2007). Damage Mitigating Control for Wind Turbines, PhD thesis, Uni-

versity of Colorado at Boulder.

Schlather, M. (2010). Some covariance models based on normal scale mixtures. *Bernoulli* **16** 780–797.

Sideratos, G. and Hatziargyriou, N. D. (2012). Probabilistic wind power forecasting using radial basis function neural networks. *IEEE Transactions on Power Systems* **27** 1788–1796.

Stein, M. (2005). Space-time covariance functions. *Journal of the American Statistical Association* **100** 310–321.

Xie, L., Gu, Y., Zhu, X. and Genton, M. (2014). Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Transactions on Smart Grid* **5** 511–520.

You, M., Byon, E., Jin, J. and Lee, G. (2017). When wind travels through turbines: A new statistical approach for characterizing heterogeneous wake effects in multi-turbine wind farms. *IISE Transactions* **49** 84–95.

Zhu, X., Genton, M. G., Gu, Y. and Xie, L. (2014). Space-time wind speed forecasting for improved power system dispatch. *Test* **23** 1–25.

Zwiers, F. and Von Storch, H. (1990). Regime-dependent autoregressive time series modeling of the Southern Oscillation. *Journal of Climate* **3** 1347–1363.