SPATIO-TEMPORAL MODELING AND ANALYSIS FOR WIND ENERGY APPLICATIONS

A Dissertation

by

AHMED AZIZ EZZAT

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Yu Ding
Committee Members,	Chanan Singh
	Mikyoung Jun
	Natarajan Gautam
Head of Department,	Mark Lawley

August 2019

Major Subject: Industrial & Systems Engineering

ABSTRACT

The promising potential of wind energy as a source for carbon-free electricity is still hampered by the uncertainty and limited predictability of the wind resource. The overarching theme of this dissertation is to leverage the advancements in statistical learning for developing a set of physics-informed statistical methods that can enrich our understanding of local wind dynamics, enhance our predictions of the wind resource and associated power, and ultimately assist in making better operational decisions. At the heart of the methods proposed in this dissertation, the wind field is modeled as a stochastic spatio-temporal process. Specifically, two sets of methods are presented.

The first set of methods is concerned with the statistical modeling and analysis of the transport effect of wind—a physical property related to the prevailing flow of wind in a certain dominant direction. To unearth the influence of the transport effect, a statistical tool called *the spatio-temporal lens* is proposed for understanding the complex spatio-temporal correlations and interactions in local wind fields. Motivated by the findings of the spatio-temporal lens, a statistical model is proposed, which takes into account the transport effect in local wind fields by characterizing the spatial and temporal dependence in tandem. Substantial improvements in the accuracy of wind speed and power forecasts are achieved relative to several existing data-driven approaches.

The second part of this dissertation comprises the development of an advanced spatio-temporal statistical model, called *the calibrated regime-switching model*. The proposed model captures the regime-switching dynamics in wind behavior, which are often reflected in sudden power generation ramps. Tested on 11 months of data, double-digit improvements in the accuracy of wind speed and power forecasts are achieved relative to six approaches in the wind forecasting literature.

This dissertation contributes to both methodology development and wind energy applications. From a methodological point of view, the contributions are relevant to the literatures on spatio-temporal statistical learning and regime-switching modeling. On the application front, these methodological innovations can minimize the uncertainty associated with the large-scale integration of wind energy in power systems, thus, ultimately boosting the economic outlook of wind energy.

To My Dear Parents

Prof. Aziz Ezzat and Mrs. Iman Metawi

To My Beloved Wife

Sherine Elgamal

ACKNOWLEDGMENTS

Five years ago, I took the decision to travel thousands of miles away from my home at the beautiful city of Alexandria, Egypt and start my Ph.D. at Texas A&M University. Little did I know about the rich experience I am going to have along the way, the multitude of friendships I am going to make, and the amount of knowledge I was about to acquire during my doctoral journey. Because endings are nothing but new beginnings, it is important, every while and then, to look back, reflect on the past, and acknowledge those people who have helped make it a wonderful experience. Indeed, a written section in a dissertation is never enough to express my gratitude to those people, but nonetheless, a Ph.D. in Industrial Engineering has taught me that exact optimality is often difficult to attain, and as such, at this exceptional moment in time, I will do my best to approximate my feelings through the following few lines.

I was very fortunate to have Prof. Yu Ding as my Ph.D. advisor. I still remember my first semester at Texas A&M University. Overwhelmed by the transition to graduate school, my course performance wasn't at its best, yet, being the smart and visionary person he is, Prof. Ding believed in my capabilities and knew that the best is yet to come. The following semester, his support started to pay off—straight A's in coursework, and a draft of our first paper was almost ready, which eventually made its way to one of the top-tier applied statistics journals. Since day one, he always used to tell me: "*You have to treat yourself as a faculty. Think, act and speak like one!*" Few years down the road, this piece of advice helped me secure a tenure track position at Rutgers University. What Prof. Ding has taught me is way beyond writing an academic paper, or conducting scholarly research—he has always been an example of a successful and caring academic and mentor.

I would like to express my gratitude to Prof. Mikyoung Jun from the Department of Statistics. In fact, this dissertation started with a project in her spatial statistics course offered in my second year of Ph.D. Since then, her suggestions and guidance were instrumental to enhance the quality of my research. I would like to also thank my committee members: Professors Natarajan Gautam and Chanan Singh for their continuous support and recommendations in all phases of this work.

I owe a lot of gratitude to former and current members of our research group: Dr. Hoon Hwangbo (now assistant professor at UTK), Dr. Yanjun Qian (now assistant professor at VCU), Erika Sy (now at Dow Chemical Company), Imtiaz Ahmed, Abhinav Prakash, Shilan Jin, Jiaxi Xu, David Perez and Jason Lawley. Our random, yet frequent conversations about various intriguing topics constantly provided me with a stimulating and enjoyable working environment.

On a more personal level, no words can express how grateful I am to my dear parents: Prof. Aziz Ezzat and Mrs. Iman Metawi. Tracing back to my early childhood, I can easily see how the many little things they were keen to do during the past 25+ years, have now added up and led to this remarkable moment in time where I conclude my Ph.D. journey. I still remember my father leaving his work aside to read my first poem that I've written in elementary school, enthusiastically cheering for me and my team on the sidelines during my soccer matches at Alexandria Sporting Club, explaining to me intriguing concepts and stories in mathematics, engineering and history, and lastly, how he consistently encouraged me to pursue my graduate studies in the United States. As for my mother, much of what I am and what I have achieved up to this day, is attributed in first place to her sincerity and dedication to our family. The amount of love and gratitude I owe to her is inexpressible. I am also forever indebted to my siblings, Dr. Noha Aziz Ezzat and Mrs. Lamia Aziz Ezzat for their continuous support. Despite life separating us apart in three different countries, our bond never stopped growing stronger and more vivid.

Last but not least, I would like to express my love, appreciation, and sincere gratitude to my beloved wife, Sherine Elgamal. At the time I started my Ph.D., we were engaged, yet, thousands of miles apart. Our love has endured two years of long distance relationship, until we tied the knot in my third year of Ph.D. Since then, my life has been constantly changing to the better. The emotional support that she provided me during this journey was invaluable. Sherine is a sincere aggie—she successfully earned a Masters degree at Texas A&M Mays Business School. I slid the majestic Aggie ring onto her finger back in 2017, and now, it's her turn!

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Yu Ding and Natarajan Gautam of the Department of Industrial & Systems Engineering, Professor Mikyoung Jun of the Department of Statistics, and Professor Chanan Singh of the Department of Electrical & Computer Engineering.

All work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was partially supported by the U.S. National Science Foundation under grant no. IIS-1741173.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND LITERATURE REVIEW

1.1 Motivation

In its aspiration towards a cleaner, more sustainable future, the U.S. Department of Energy (DOE) has proposed an ambitious plan to fulfill 35% of the total U.S. electricity demand through wind energy by the year 2050 [22]. Despite the abundant wind resource in the United States, the promising potential of wind energy is still hampered by serious technical and scientific challenges that remain to be resolved in order to reach the 35%-by-2050 goal. Among which, the variable and intermittent nature of the wind resource stands as a major challenge that hinders the large-scale integration of wind energy in modern-day power grids. Unlike conventional generators, wind power cannot be produced at will. In contrast, it is almost fully dependent on the state of the wind resource at the generation site, and exhibits sizeable variations at a multitude of spatial and temporal resolutions.

This stochastic characteristic of the wind resource makes the operational activities in a power grid with sizeable wind energy penetration heavily reliant on accurate predictions of the wind resource and associated power generation. Examples of these operational activities are numerous at different spatial and temporal scales. At the turbine level, for instance, the wind farm operator is tasked to find the optimal turbine-specific pitch and yaw settings, in a near real-time fashion, to maximize the turbine's production while preserving its safety against extreme weather conditions; a problem known in the wind industry as predictive turbine control [74]. At the farm level, the optimal hourly market participation needs to be pre-determined for up to a day-ahead with the ultimate objective of maximizing operational profits [66]. At the grid level, power system operators have to quantify the optimal hourly reserve capacities to ensure cost-efficient supply-demand planning [82]. The quality of these operational activities obviously rely on accurate predictions of wind speed, wind direction or wind power, issued at the proper spatial and temporal scales at which actions need to be taken.

It therefore comes as no suprise that improvements in predicting the wind resource and associated power production, translate to better, more cost-efficient operational decisions. A 2011 study suggested that a 20% improvement in the accuracy of wind generation forecasts, at a wind energy penetration level of 24%—which is still lower than the DOE's 35%-by-2050 goal—can generate up to almost $\$975$M of annual operating cost savings in the U.S. [56]. This economic impact calls upon capable predictive models that can accurately describe and forecast the wind resource and wind power production at a multitude of spatial and temporal scales, and this is, in fact, the overarching theme of this dissertation.

1.2 Statistical versus physics-based approaches for wind forecasting

Generally speaking, wind forecasting models can be classified into "statistical" and "physics-based" models. The statistical approach uses the data collected in wind farms to construct a data-driven model that can extrapolate the historical trends, patterns and correlations to the near or distant future. On the other hand, physics-based approaches, also referred to as Numerical Weather Prediction (NWP), numerically solve a system of discretized differential conservation equations of mass, momentum, and energy in the atmosphere, and forecast wind speed or other weather conditions based on the solution of the physical equations. Oftentimes, a finer resolution "meso-scale" model is nested within a large-scale NWP in order to capture the wind dynamics within a finer spatial resolution. The meso-scale model takes the outputs of an NWP as input, and makes a forecast of the wind velocity vector, or equivalently the wind speed and wind direction, at a time horizon in the future [54].

The dominating consensus in the wind forecasting practice is that statistical models are often preferred over their physics-based counterparts for shorter temporal horizons and finer spatial resolutions. In the literature, a precise threshold between what constitutes a "short" temporal horizon does not appear to exist, but in one of the widely cited survey papers on wind forecasting, a 6-hour temporal range was identified as the horizon after which physics-based models tend to outperform statistical models [31]. In other studies, the envelope was pushed to as much as a day ahead [86]. There is also a "gray" area in between where a combination of statistical and physics-based models

may be favorable [80].

The reason why physics-based models at finer spatial and temporal resolutions are less favorable compared to the statistical models is primarily caused by their high computational cost and complexity. Physics-based models depend on a discretization of the spatial domain. Theoretically, the grid size can be made finer but at the expense of slower computation, resulting in a coarser temporal resolution. NWPs are typically deterministic, and hence, multiple runs have to be conducted with different initial conditions, in order to produce an ensembled forecast, for better forecast accuracy as well as uncertainty quantification; doing so undoubtedly slows down NWP even further. Running NWPs also needs many physical measurements to identify boundary conditions for solving the differential equations, but those measurements may not be available at the finer spatial resolution. The limitations listed above raise the need for statistical, data-driven models and methods to complement NWP models.

This dissertation falls under the umbrella of statistical approaches described above. But the set of methods developed herein are not purely "data-driven." Instead, the construction of the models in this dissertation is guided by the physical features unique to wind fields, and should thus be better described as "physics-informed" statistical methods. One main conclusion of this dissertation is that a physically guided approach in the statistical modeling of wind dynamics yields substantial improvements in the predictive and explanatory power relative to the purely data-driven approaches which are indifferent to the physical properties of wind fields.

The proposed statistical models in this dissertation are intended to describe wind dynamics and power generation at relatively fine spatial and temporal resolutions. Specifically, the models developed herein are concerned with a temporal resolution that is as short as one hour but up to 12 hours in the future, and a spatial resolution designated by specific turbine locations on a wind farm. It is worth noting that time horizons, as short as few hours ahead do play an important role in wind farm operations such as turbine control, scheduling transmission and dispatch [40].

In addition, the term "local" is used in this dissertation, following the terminology in [69], to differentiate the within-farm wind field, studied here, from large-scale wind fields that have

long been studied in the geostatistical and atmospheric science literature in which the wind travels across substantial distances [39].

It is also important to highlight that the methods developed in this dissertation issue turbine-specific, rather than farm-level forecasts. Recently, several research studies highlighted the need for the former to replace the latter [53, 61, 69]. 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, including predictive turbine control [74], turbine-specific power estimation [9, 55], wake propagation prediction [85, 45] and repair decisions for wind turbines [13, 14]. Moreover, 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.

To conclude this Section, I re-iterate the scope and contributions of this dissertation in light of the definitions and clarifications introduced above. The goal of this dissertation is to propose a set of physics-informed statistical methods that will be used to make accurate short-term, turbine-specific forecasts of the wind speed and wind power. The value of these forecasts stems from their potential contribution towards better, more cost-efficient operational activities at the turbine, farm, and grid levels, thus eventually boosting the economic outlook of wind energy. The scope of this dissertation is closely related to the field of spatio-temporal modeling, since the wind field, in all of the proposed methods, is modeled as a stochastic spatio-temporal process. Therefore, a detailed overview on spatio-temporal modeling is presented in Section 1.3 to highlight the research gaps and motivations of the methods proposed in Sections 2 and 3 of this dissertation.

This dissertation also contributes towards the methodology of wind forecasting by constructing statistical models that can account for salient physical properties of wind fields such as the transport effect of wind and the regime-switching dynamics. For the transport effect, a *nonseparable* spatio-temporal model is proposed, which—as the name implies—characterizes the spatial and temporal dependence structure in tandem, rather than separately as traditionally performed in the literature. For the regime-switching aspect, a calibration formulation is proposed, which makes use of a runlength variable as an informative proxy to anticipate, rather than retrospectively detect, the out-

of-sample changes in wind behavior. In light of that, Sections 1.3 and 1.4 present an overview on nonseparable spatio-temporal models and regime-switching methods, respectively, in an attempt to pave the way for the methodological contributions of this dissertation, for which the details will be unraveled in Sections 2 and 3.

1.3 Spatio-temporal modeling and covariance functions

Many real-world processes evolve over both space and time. Environmental and geophysical processes such as atmospheric flow, pollutants' concentrations, large-scale wind and water fields have long been regarded in the geostatistical literature as typical examples of spatio-temporal processes [18, 29, 49]. More recently, several engineering applications, outside these typical fields, have also been studied in a spatio-temporal context [58, 59]. Here, the general background related to spatio-temporal modeling and analysis, and its relevance to wind energy applications, is reviewed.

At the heart of spatio-temporal modeling is the concept of spatio-temporal dependence. Take the wind speed on a wind farm as an example. It is reasonable to expect that the wind speeds observed at two neighboring turbines would be similar to each other, since spatially close turbines are often subject to similar wind characteristics. This is referred to as *spatial correlation*. Figure 1.1 presents the evolution of hourly wind speeds as recorded at two neighboring wind turbines on an onshore wind farm in the United States for a span of 24 hours. The spatial correlation is manifested in the similarity between the two time series shown in Figure 1.1. On the other hand, the wind speeds recorded at the present time are expected to look more similar to those observed in temporal vicinity, too, say, to what was observed one hour ago, than to those observed several hours ago, and this is referred to as *temporal correlation*, and can be characterized by the degree of autocorrelation of different time lags of a single time series. Generally speaking, observations that are close to one another in space and/or time are more likely to be similar to each other than to those that are spatially or temporally apart. In Section 3 of this dissertation, we present an empirical analysis of the spatio-temporal dependencies in local wind fields.

To start, let us define some important notations. Let Y_i refer to the wind speed as observed at a

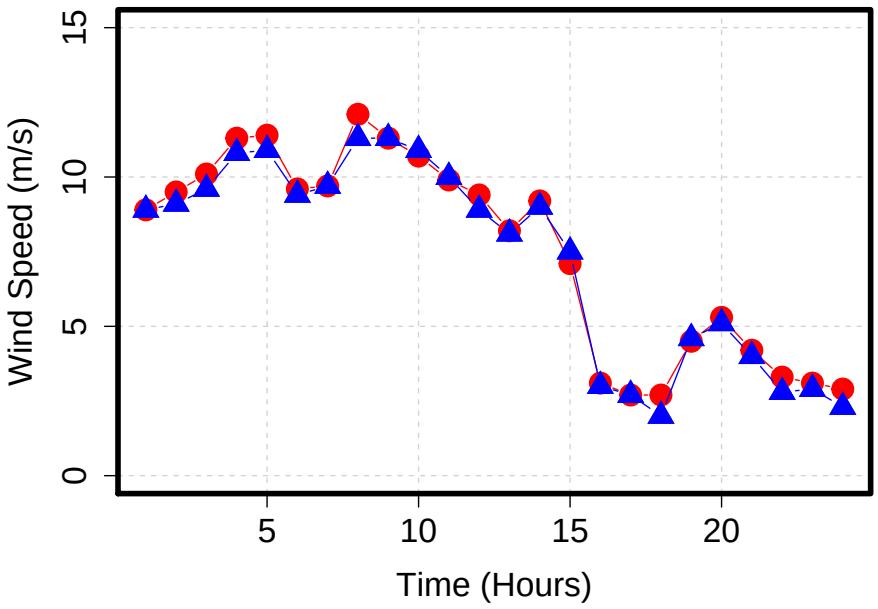


Figure 1.1: Hourly wind speeds as recorded at two neighboring wind turbines on an onshore wind farm in the United States over the span of 24 hours.

particular wind turbine, with $i = 1, ..., n$, where n is the total number of wind turbines on the wind farm. The turbine locations are denoted by $[\mathbf{s}_1, \ldots, \mathbf{s}_n]^T$ such that $\mathbf{s}_i \in \mathbb{R}^2$, $i = 1, ..., n$, is a pair of longitude and latitude coordinates over the landscape. The goal of spatial modeling is to build a statistical model that can take as input an n-dimensional vector of wind speed measurements corresponding to the n turbine locations, denoted by $\mathbf{Y} = [Y_1, \ldots, Y_n]^T$, and find a best-fitting conditional distribution for the wind speed at any arbitrary location $\mathbf{s}^*$. A point prediction can thus be made out of the conditional distribution, and is denoted by $\hat{Y}_{i^*}$.

A popular approach in the geostatistical literature for spatial modeling is the Gaussian Random Field (GRF), also known as a Gaussian Process (GP). GRFs or GPs are one of the most commonly used methods in the geostatistical community to model a wide variety of geophysical and environmental applications. The pillar assumption in GRF models is that the joint distribution of any finite set of inputs follows a multivariate Gaussian distribution. Under this assumption, a GRF model is

defined by a mean structure $\mu(\cdot)$, a stochastic process term $\eta(\cdot)$, and a noise term ϵ, as in (1.1).

$$Y_i = \mu_i + \eta_i + \epsilon_i \quad i = 1, \ldots, n. \tag{1.1}$$

The mean structure can be either expressed as a constant, that is $\mu_i = \mu \ \forall i$, or a parametric function of the spatial coordinates stored in $\mathbf{s}_i$ and/or other pre-specified exogenous covariates stored in $\mathbf{x}_i$ that are thought to capture the overall trend of the response. If μ_i is chosen to have a functional form of $\mathbf{s}_i$ and $\mathbf{x}_i$, then we would have $\mu_i = \mathbf{f}^T(\mathbf{s}_i, \mathbf{x}_i)\boldsymbol{\beta}$, where $\mathbf{f}(\mathbf{s}_i, \mathbf{x}_i) = [f_1(\mathbf{s}_i, \mathbf{x}_i), \ldots, f_\ell(\mathbf{s}_i, \mathbf{x}_i)]^T$ is an $\ell \times 1$ vector of known regression functions, and $\boldsymbol{\beta} = [\beta_1, ..., \beta_\ell]^T$ is the $\ell \times 1$ vector of corresponding unknown regression coefficients.

The term $\eta(\cdot)$ is assumed to be a zero mean Gaussian process with its pairwise covariance denoted by σ_{ij}, where $\sigma_{ij} = \text{Cov}\{\eta_i, \eta_j\}$, $i = 1, \ldots, n$, and $j = 1, \ldots, n$. The term $\epsilon(\cdot)$ is the zero mean, i.i.d Gaussian noise, such that $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \delta\mathbf{I})$, where $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]^T$, $\mathbf{I}$ is the $n \times 1$ identity matrix and $\delta > 0$ denotes the noise variance, also known in the spatial statistics literature as the nugget effect. Let us use $\boldsymbol{\Sigma}$ to denote the resulting $n \times n$ covariance matrix whose (i, j)-th entry is defined as $\sigma_{ij} + \delta \, \mathbb{1}_{\{i=j\}}$. Under this setting and for the set of input locations $[\mathbf{s}_1, \ldots, \mathbf{s}_n]^T$, and the set of input covariates $[\mathbf{x}_1, \ldots, \mathbf{x}_n]^T$, the corresponding output is multivariate normal, that is, $\mathbf{Y} \sim \mathcal{N}(\mathbf{F}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\mathbf{F}$ is the $n \times \ell$ matrix of covariate values such that its (i, j)-th element is $f_j(\mathbf{s}_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, and $j = 1, \ldots, \ell$.

A key issue in fitting GRFs is to determine the entries of the covariance matrix $\boldsymbol{\Sigma}$, which can be in turn dictated through the choice of a suitable parametric covariance function denoted by $K(\cdot, \cdot; \Phi)$, where Φ represents a set of covariance parameters, also known as hyperparameters. By assuming stationarity, the covariance function, only depends on the spatial lag, $\mathbf{u} = \mathbf{s}_i - \mathbf{s}_{i'}$, and as such, the expression of the original covariance function can be simplified as a single input function, $K(\mathbf{u}; \Phi)$, which is used to map $\mathbf{u} \in \mathbb{R}^2$ to $\mathbb{R}$. Intuitively speaking, the stationary covariance structure $K(\cdot)$ defines a measure of similarity between a pair of spatial data points Y_i and $Y_{i'}$ located at $\mathbf{s}_i$ and $\mathbf{s}_{i'}$, respectively, and the similarity solely depends on their spatial lag $\mathbf{u} = \mathbf{s}_i - \mathbf{s}_{i'}$,

rather than the specific location of $\mathbf{s}_i$ or $\mathbf{s}_{i'}$. The farther $\mathbf{s}_i$ is from $\mathbf{s}_{i'}$, the smaller the value of $K(\cdot)$ should be. If the covariance function $K(\cdot)$ is further assumed to be isotropic, it only depends on the Euclidean distance between a pair of locations, i.e. $||\mathbf{u}||$, and is thus invariant to both translation and rotation. Obviously, the class of isotropic covariance functions is a subset of the class of stationary covariance functions.

In the past two decades, geostatisticians have derived numerous stationary covariance functions [70]. Examples of stationary parametric covariance functions include but are not limited to the squared exponential and the Matérn covariance functions [70]. We here present the Matérn covariance function due to its popularity and because many covariance functions are special cases of it. For more details about covariance functions, please refer to the classical texts [70, 19].

The spatial Matérn covariance function is defined as in (1.2).

$$K(\mathbf{u}) = \alpha \, \frac{2^{1-\nu}}{\Gamma(\nu)} \big(\frac{\sqrt{2\nu}||\mathbf{u}||}{\ell} \big)^{\nu} \mathcal{K}_{\nu} \big(\frac{\sqrt{2\nu}||\mathbf{u}||}{\ell} \big), \tag{1.2}$$

where $||\cdot||$ is the Euclidean norm, $\mathcal{K}_{\nu}$ is the modified Bessel function, $\alpha > 0$ denotes the sill or marginal variance, $\nu > 0$ denotes the smoothness, and ℓ represents the scale parameter which is the spatial range until which the spatial correlation has non-trivial effect. Combining the parameters in (1.2) with the noise variance δ, the set of hyperparameters needed to compute the entries of $\mathbf{\Sigma}$ can be denoted by $\Phi = \{\alpha, \nu, \ell, \delta\}$.

In GRFs, the vector of hyperparameters Φ is often estimated, in conjunction with the vector of regression coefficients $\boldsymbol{\beta}$, in a data-driven way through a maximum likelihood estimation (MLE) by solving the optimization problem in (1.3).

$$(\hat{\Phi}, \hat{\beta}) = \underset{\Phi, \beta}{\operatorname{argmax}} \left\{ - \ln(|\mathbf{\Sigma}|) - (\mathbf{Y} - \mathbf{F}\boldsymbol{\beta})^T \mathbf{\Sigma}^{-1} (\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}) \right\}, \tag{1.3}$$

where $\hat{\Phi}$ and $\hat{\boldsymbol{\beta}}$ contain the MLE estimates for Φ, and $\boldsymbol{\beta}$, respectively. The right hand side in (1.3) is proportional to the logarithm of the likelihood for a GRF. In practice, the optimization problem in (1.3) is numerically solved using gradient descent-based methods, commonly implemented by

using the routine `nlm` in `R`. One possible approach is to perform the parameter estimation iteratively, where, at every step of the estimation procedure, new values for the hyperparameters are suggested, and then $\hat{\boldsymbol{\beta}}$ are computed using the generalized least squares estimator (GLS) for $\boldsymbol{\beta}$, which is expressed in (1.4). Then, the log likelihood is computed, up to a constant, by plugging in the values of the hyperparameters and GLS coefficients in (1.3). The procedure is repeated until convergence to a local minima.

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Y}. \tag{1.4}$$

Given the estimated values for the hyperparameters and regression coefficients, one can obtain a point prediction, also known as a kriging-based prediction, at any arbitrary location $\mathbf{s}^*$ as in (1.5). Such kriging-based estimate is known to be a Best Linear Unbiased Predictor (BLUP) [70].

$$\hat{Y}_{i^*} = \mathbb{E}\{Y_{i^*} | \mathbf{Y}\} = \mathbf{f}^T(\mathbf{s}^*, \mathbf{x}^*)\hat{\boldsymbol{\beta}} + \mathbf{k}^T \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y} - \mathbf{F}\hat{\boldsymbol{\beta}}), \tag{1.5}$$

where the $n \times 1$ vector $\mathbf{k}$ contains the pairwise covariances between $\mathbf{Y}$ and Y_{i^*} computed using $K(\cdot; \hat{\Phi})$. In addition to the point prediction from (1.5), one can also obtain a measure of uncertainty around the point prediction of (1.5), i.e. the point-wise predictive variance, as expressed in (1.6).

$$\hat{\sigma}^2_{i^*} = \text{Var}\{Y_{i^*} | \mathbf{Y}\} = K(\mathbf{0}; \hat{\Phi}) - \mathbf{k}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{k} + (\mathbf{f}^T(\mathbf{s}^*, \mathbf{x}^*) - \mathbf{k}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{F})(\mathbf{F}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{F})^{-1} (\mathbf{f}^T(\mathbf{s}^*, \mathbf{x}^*) - \mathbf{k}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{F})^T. \tag{1.6}$$

With the predictive mean and predictive variance in place, a full characterization of the conditional distribution of the forecast at $\mathbf{s}^*$ is possible using the first two moments in (1.5) and (1.6), respectively. This is attributed to the fact that the predictive distribution from a GRF is also Gaussian. Such conditional distribution is important for subsequent decision-making under uncertainty, like in several power grid operations, related to what is known in the forecasting literature as "probabilistic" forecasting [66, 37].

Now, we are ready to make the transition from spatial modeling to spatio-temporal mod-

eling. Considering a wind field that evolves over both space and time, spatio-temporal models are natural extensions of spatial models. Similar to the notation established for the spatial models, we let $Y_i(j)$ denote the wind speed at the ith turbine, and jth time index, for $i = 1, \ldots, n$, and $j = 1, \ldots, T$ hours. Hereinafter, we denote by j any arbitrary time index, while by t we specifically mean the present time. The goal of spatio-temporal modeling is then to build a model that takes as input the $nT \times 1$ vector of spatio-temporal observations, corresponding to n wind turbine locations $[\mathbf{s}_1, \ldots, \mathbf{s}_n]^T$ and T time indices $[1, \ldots, T]^T$, denoted by $\mathbf{Y} = [Y_1(1), \cdots, Y_N(1), Y_1(2), \cdots, Y_N(2), \cdots, Y_N(T), \cdots, Y_N(T)]^T$, and find the best-fitting conditional distribution of the wind speed at any arbitrary location $\mathbf{s}^*$, and future time index $t+h$, where $h \in \mathbb{Z}^+$. The corresponding point prediction is denoted by $\hat{Y}_{i^*}(t+h)$.

Similar to what we established for spatial modeling, a parametric spatio-temporal covariance function is denoted by $K(\cdot, \cdot, \cdot, \cdot)$ and is used to compute the entries of the $nT \times nT$ spatio-temporal covariance matrix $\mathbf{\Sigma}$. By assuming stationarity, the covariance function only depends on the spatial and temporal lags, $\mathbf{u}$ and w, respectively, and as such, its expression is simplified as $K(\mathbf{u}, w)$. By further imposing isotropy, the covariance function only depends on the Euclidean norms of the spatial and temporal lags, i.e. $||\mathbf{u}||$ and $||w||$, respectively.

Now, a lingering question to fully characterize the spatio-temporal model is how to define a covariance function $K(\mathbf{u}, w)$ in a spatio-temporal setting. As mentioned before, in the spatial-only setting, there exists several well-established valid covariance structures such as the Matérn function of (1.2). In the spatio-temporal setting, however, defining a covariance function over both space and time is not as straightforward. Oftentimes, a set of additional assumptions is imposed to simplify the construction of the spatio-temporal covariance structure. Here, we discuss two key assumptions that have been studied by the geostatistical research community in the past two decades—the separability and symmetry assumptions.

A covariance structure is said to be *separable* if it factors into the product of purely spatial and temporal components such that $K(\mathbf{u}, w) = K^s(\mathbf{u}) \cdot K^t(w)$. As such, one could independently model the covariance structure across space and time through separate covariance functions. For

instance, one could use a spatial Matérn for $K^s(\mathbf{u})$ and a temporal Matérn for $K^t(w)$, and the product of these two valid positive-definite covariance functions indeed produces a valid positive-definite covariance function [19]. Such approach is easy to implement and has notable computational benefits. The final spatio-temporal covariance matrix can be computed efficiently as the Kronecker product of two smaller spatial and temporal covariance matrices, as expressed in (1.7).

$$\mathbf{\Sigma}_{nT\times nT} = \mathbf{\Sigma}^s_{n\times n} \otimes \mathbf{\Sigma}^t_{T\times T}, \tag{1.7}$$

where $\mathbf{\Sigma}_{nT\times nT}$, $\mathbf{\Sigma}^s_{n\times n}$ and $\mathbf{\Sigma}^t_{T\times T}$ are the spatio-temporal, spatial and temporal covariance matrices, respectively.

Despite the computational benefits and ease of implementation, separable models have limiting modeling capabilities which do not always align with reality. By independently modeling the spatial and temporal covariance structures, separability overlooks the interaction between the spatial and temporal components and implies *symmetry* in the spatio-temporal covariance structure [19]. A covariance structure is said to be *symmetric* if $K(\mathbf{u}, w) = K(-\mathbf{u}, w) = K(\mathbf{u}, -w) = K(-\mathbf{u}, -w)$ [33]. In other words, the covariance between sites $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ at time indices j and j', is the same as that between $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ at time indices j' and j. Mathematically, this implies a time-reversibility, i.e. $\text{Cov}\{Y_i(j), Y_{i'}(j')\} = \text{Cov}\{Y_i(j'), Y_{i'}(j)\}$. Separability implies symmetry, but the converse cannot be guaranteed [36].

The limiting capabilities of separable models in capturing complex correlation patterns in real-world processes motivated the geostatistical community in the past decade to establish classes of nonseparable, but still symmetric, models [18, 33], as well as asymmetric nonseparable models that can further account for the lack of spatio-temporal symmetry in geophysical processes [78, 65, 49, 36]. In fact, an important research area in the geostatistical literature is to devise statistical tests for the hypotheses of separability and symmetry in spatio-temporal datasets [63, 78, 30, 57]. Section 2 of this dissertation explores the validity of the separability and symmetry assumptions in light of the transport effect property in local wind fields on a farm, and consequently establishes a physically justifiable spatio-temporal model.

Once a parametric covariance function $K(\mathbf{u}, w)$ is imposed, be it separable or not, the hyperparameters can be estimated using the MLE approach as in (1.3), and a point prediction $\hat{Y}_{i*}(t+h)$, as well as its predictive variance $\sigma^2_{i*}(t+h)$, at any location $\mathbf{s}^*$ and future time $t+h$, can be obtained in the same way as in (1.5) and (1.6). These two moments can be used in tandem to characterize the conditional distribution of the forecast.

1.4 Regime-switching models for short-term wind forecasting

Statistical wind forecasting models can either be focused on wind speed or wind power prediction. While there are pros and cons for each strategy, advantages of the former are represented in the ability of obtaining forecasts of the wind resource, which are useful for an additional set of operational analytics such as predictive turbine control [74] and predictive maintenance scheduling in wind farms [13]. It is well established in the wind forecasting practice and literature that wind speed forecasts can be readily converted to wind power estimates using power curves constructed either using the industry standard binning method [46] or other advanced kernel-based approaches [48, 55]. In the wind industry, a power curve refers to the relationship between wind speed and wind power. In this dissertation, wind speed is the predictant of interest, but wind power forecasts are also computed and evaluated.

Time-series modeling has been widely used in the short-term wind forecasting literature, including the classical approaches such as the autoregressive (AR), autoregressive moving average (ARMA), and vector autoregressive (VAR) models [12, 62, 24, 23], or more recently, machine learning methods such as support vector machines and artificial neural networks [64, 8, 76, 73]. These methods generally rely on the idea of temporal dependence, where recent lagged values of wind variables are assumed to be informative about incoming values in the near future. Beside temporal dependence, wind fields have been shown to exhibit strong signs of spatial dependence [81, 32, 66]. As such, spatio-temporal models have been proposed for short-term forecasting using the information from the lagged values at the target site as well as from neighboring sites [6, 53, 79, 61, 23, 69]. The relatively newly discovered benefit of spatio-temporal approaches is the reason behind our review of spatio-temporal modeling in Section 1.3.

Recent studies have also shown that wind fields exhibit strong signs of nonstationarity [69, 83]. One common theme about the models presented in Section 1.3 is that they are all stationary, i.e. they inherently assume stationarity in the process being modeled. In reality, however, the wind behavior in local wind fields exhibit sizeable high-frequency changes [71, 25]. According to [66], 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. In [69], the detrimental effect of sudden high-frequency wind speed changes on the forecast accuracy of their models is noted, indicating the need for more capable methods to capture these fluctuations. Also, it is reported in [80] that classical statistical models are not powerful in dynamic environments with fast-changing winds, while in [79], the notable effect of evolving wind dynamics on volatile power generation levels is pointed out.

Such nonstationarity in time, often referred to in the literature as the "regime-switching" nature of wind, refers to the fluctuating nature of wind characteristics over a spectrum of physical states or conditions. Motivated by the need to capture that physical feature, time-series and spatio-temporal models that account for regime-switching dynamics have been recently proposed, yielding the family of regime-switching methods.

The essence of these regime-switching methods is to fit statistical models that are conditioned on a finite set of wind regimes or states, and produce regime-dependent forecasts. First, a number of regimes is pre-defined and indexed from $1, ..., R$, such that at any arbitrary time point j, the wind regime information is denoted by $r(j) \in \{1, ..., R\}$. The model then uses a set of parameters specific to each wind regime to produce regime-dependent forecasts. One such model is the regime-switching autoregressive model of [88], which is presented in (1.8), where, according to $r(j)$, the wind speed at the ith turbine and time j, denoted by $Y_i(j)$, is modeled as an autoregressive model of order p, AR(p), using a set of regime-dependent parameters $[a_0^{r(j)}, a_1^{r(j)}, \ldots, a_p^{r(j)}]^T$. Here, $\epsilon_i(j)$ is a zero-mean normally distributed random variable whose variance can be regime-

dependent as well. Also, the order of the autoregressive model can be regime-dependent. In other words, p can be different across different regimes, and hence, can be replaced in the formulation of (1.8) by $p^{r(j)}$ to reflect its dependency on the observed wind regime.

$$Y_i(j) = a_0^{r(j)} + \sum_{j'=1}^{p} a_{j'}^{r(j)} Y_i(j - j') + \epsilon_i(j). \tag{1.8}$$

Depending on how $r(j)$ is defined, extensions of the model in (1.8) can be constructed leading to more sophisticated regime-switching models. For instance, a spatio-temporal extension to the model in (1.8) has been proposed in [38] and can be expressed as in (1.9). The model uses wind direction measurements at upstream locations to identify regime information in downstream locations. In other words, specification of $r(j)$ can be governed by the observed wind direction at one of the upstream locations, determined using expert knowledge of the geographic region under study.

$$Y_i(j) = a_0^{r(j)} + \sum_{i'=1}^{n} \sum_{j'=1}^{p} a_{i'j'}^{r(j)} Y_{i'}(j - j') + \epsilon_i(j), \tag{1.9}$$

where $Y_i(j)$ is the wind speed at time j at location i, and n is the total number of spatial locations.

The model in (1.9) has been further generalized by introducing the wind direction as a circular variable in the model formulation to relax the model's dependence on arbitrary regime selections [40]. Other variants of the regime-switching model in (1.8) used different environmental variables other than wind speed, direction or power to define the regime variable $r(j)$. For instance, lagged measurements of temperature to determine $r(j)$ have been previously used [71]. The use of lagged values of precipitation as an indicator of regime information has also been recently reported [11].

One common limitation of the above regime-switching models is their inherent assumption that the current regime, $r(t)$, persists in the forecasting horizon, that is, $r(t + h) = r(t)$, $h = 1, \ldots, H$. To address this limitation, Markov-switching autoregressive (MSAR) models, and their generalizations, Markov-switching vector autoregressive (MSVAR) models, have been proposed to allow for regime switches in the forecast horizon. These models use a group of (vector) autoregressive models similar to (1.8), but primarily differ from the above models by assuming that the switch

between the regimes is triggered by a Markov chain, rather than via the lagged values of a specified explanatory variable [68, 4, 66, 77].

For the MSAR model, a transition probability matrix, $\mathbf{\Pi}_{R\times R}$ is constructed, for which the entries $\pi_{kk'} = Pr(r(j) = k'|r(j-1) = k)$ can be estimated from a training dataset. Then, a set of AR(p) models is trained using the historical data belonging to each regime. The final prediction at time $t+h$ can then be the convex combination of the predictions from the set of R models, where the combination coefficients correspond to the probability of reaching each regime at time $t+h$, that is: $\hat{Y}_i(t+h) = \sum_{k=1}^{R} \hat{Y}_i^k(t+h) Pr(r(t+h) = k)$, where $\hat{Y}_i^k(t+h)$ is the prediction at the ith turbine at time $t+h$ obtained by fitting an AR(p) 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 time $t+h$. The MSAR model obviously does not take into account spatial dependencies, and hence, it can be fit to each wind turbine independently. Generalizing on MSAR is the MSVAR model which further accounts for the spatial dependence and can be fit to a set of observations from a group of neighboring turbines.

In Section 3 of this dissertation, a statistical forecasting model titled: "the calibrated regime-switching (CRS) model" is proposed. The essence of the CRS model is to account for both in- and out-of-sample regime switches by introducing a regime-specific calibration to the predictions from a spatio-temporal model. This regime-specific calibration is modeled as a parametric function of two elements that are shown to be able indicators of out-of-sample changes in wind behavior, namely: the observed regime at the time of the forecast $r(t)$, and a runlength variable, which is defined as the time elapsed since the most recent regime change. Tested on 11 months of data in a rolling forward forecasting fashion, the CRS approach yields substantial improvements over a wide set of time-series, spatio-temporal, and regime-switching methods that are widely used in the wind forecasting literature.

1.5 Wind farm data

The data in this study were collected at an onshore wind farm in the United States, located on a relatively flat terrain over approximately 25×17 square kilometers (km^2). Several treatments were

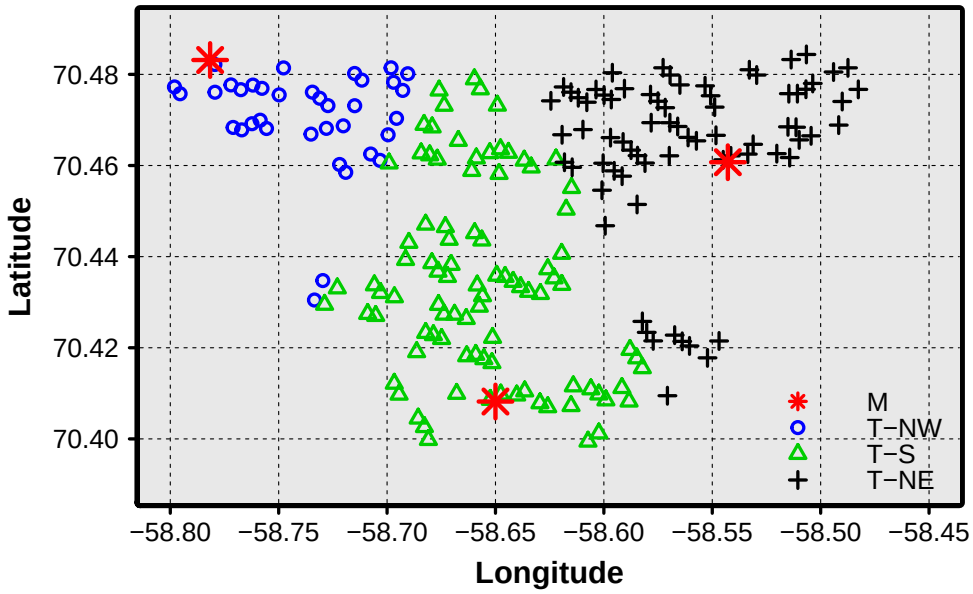


Figure 1.2: Spatial map of the wind farm (coordinates shifted to maintain confidentiality). M: meteorological masts located in the Northwest (NW), South (S), and Northeast (NE). T: Turbines grouped according to their proximity to the nearest mast, whether NW, S, or NE.

carried out to preserve the data provider's confidentiality, such as shifting the geographic coordinates of the turbines by a constant, randomly selecting 200 turbines for analysis, and normalizing wind power outputs to the range of $[0, 1]$. Two datasets are used throughout this dissertation and are referred to as: Datasets 1 and 2, respectively. Dataset 1 comprises turbine-specific hourly measurements of wind speeds recorded at the hub height of 80 meters. The longitudes and latitudes of the wind turbines are also provided.

Dataset 2 contains measurements that come from three spatially distant meteorological masts located in the northwestern, southern, and northeastern regions of the wind farm, respectively. It consists of hourly measurements of wind speed and wind direction. The measurements are taken at a height of 80 meters as well. The total number of observations in Dataset 1 is equal to 200 turbines $\times$ 24 hours $\times$ 365 days $= 1,752,000$ entries, while that in Dataset 2 is 3 masts $\times$ 24 hours $\times$ 365 days $= 26,280$ entries. The spatial map of the wind farm is presented in Figure 1.2.

1.6 Forecast evaluation metrics

In this dissertation, multiple metrics are used to evaluate the quality of the forecasts generated using the proposed methods. Among the widely used metrics in the forecasting literature are the root mean squared error (RMSE) and the mean absolute error (MAE), which are defined as in (1.10) and (1.11), respectively.

$$\text{RMSE}(h) = \sqrt{\frac{1}{L \times n} \sum_{\ell=1}^{L} \sum_{i=1}^{n} \left[\hat{Y}_i^{\ell}(t+h) - Y_i^{\ell}(t+h)\right]^2}, \tag{1.10}$$

$$\text{MAE}(h) = \frac{1}{L \times n} \sum_{\ell=1}^{L} \sum_{i=1}^{n} |\hat{Y}_i^{\ell}(t+h) - Y_i^{\ell}(t+h)|, \tag{1.11}$$

where L is the number of forecasting rolls, $\ell \in \{1, \ldots, L\}$ is the roll index, n is the number of spatial locations or wind turbines, $Y_i^{\ell}(t+h)$ and $\hat{Y}_i^{\ell}(t+h)$ are, respectively, the observation and point forecast (from any of the proposed methods) at the i-th location and the h-th horizon during the ℓ-th forecasting roll. For each forecasting horizon $h = 1, ..., 12$, the associated RMSE and MAE metrics are computed over all turbines and forecasting rolls.

In addition to RMSE and MAE, it has been reported that, in deregulated electricity markets, under-estimating power is often a more costly situation than over-estimating it [67]. To reflect this cost trade-off, the power curve error (PCE), which is an asymmetric error loss, has been proposed, as expressed in (1.12), to evaluate the power forecast [40].

$$PCE(P, \hat{P}) = \begin{cases} g\big[P_i(t+h) - \hat{P}_i(t+h)\big], & \text{if } \hat{Y}_i(t+h) \leq Y_i(t+h) \\ (1-g)\big[\hat{P}_i(t+h) - P_i(t+h)\big], & \text{if } \hat{Y}_i(t+h) > Y_i(t+h), \end{cases} \tag{1.12}$$

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. A value of $g = 0.73$ has been reported

in the literature [67]. Under the asymmetric loss function of (1.12), it can be shown that the g-th quantile of the predictive distribution is an optimal point forecast [35].

1.7 Organization of this dissertation

This dissertation proposes a set of physics-informed statistical models, tools and analytics that are intended to capture salient physical properties of wind fields: stochasticity, spatial and temporal dependence, transport effect, and regime-switching dynamics, and is accordingly organized as follows. Section 2 develops a statistical tool, called *the spatio-temporal lens* that unearths the influence of the transport effect of wind on the spatio-temporal correlation structure in a local wind field. Based on insights generated from the spatio-temporal lens, a physically justifiable asymmetric nonseparable spatio-temporal model is proposed to capture the transport effect of wind. The proposed model achieves promising improvements in predictive accuracy over a set of existing forecasting approaches in the literature.

Section 3 develops an advanced statistical spatio-temporal model, called the "the calibrated regime-switching model" to capture another important physical feature of wind fields: the regime-switching dynamics. The essence of the model is to introduce a calibration formulation that makes use of a runlength variable as an informative proxy for potential out-of-sample variations in the wind behavior. Tested on a massive validation set of more than $3,000,000$ data instances, substantial improvements in wind speed and power forecasting accuracy are demonstrated over a wide set of existing forecasting approaches in the literature including time-series based, spatio-temporal, and regime-switching models.

Finally, Section 4 summarizes the key contributions and highlights potential extensions beyond this dissertation.

2. SPATIO-TEMPORAL ASYMMETRY OF LOCAL WIND FIELDS AND ITS IMPACT ON SHORT-TERM WIND SPEED AND POWER FORECASTING *

The transport effect refers to the physical property pertaining to processes like wind in which a prevailing flow exists over time in a certain dominant direction. In this Section, a statistical tool called *the spatio-temporal lens* is proposed to unearth the influence of the transport effect of wind on the complex spatio-temporal dependence structure in a local wind field. The implementation of the spatio-temporal lens on one year of turbine-specific wind farm measurements suggests that the transport effect of wind generates complex spatio-temporal interactions and correlation patterns that are not captured by the widely used separable spatio-temporal models. Upon this revelation, a question of interest is *how to construct a physically justifiable spatio-temporal model which does not overlook the influence of the transport effect of wind?* Unlike separable models which characterize the dependence structure over space and time independently, a spatio-temporal model is proposed to capture the complex nonseparable spatio-temporal dependence in local wind fields, owing to the transport effect of wind. Using the proposed model, substantial improvements in wind speed and power forecast accuracy are demonstrated over existing data-driven approaches in the wind forecasting literature.

2.1 Spatio-temporal asymmetry and the transport effect of wind

Unlike "static" processes, wind propagates along a dominant flow direction over time—it blows from high- to low-pressure regions. This dynamic behavior is what is called in the atmospheric science and geostatistical literature as the transport effect. One of the main limitations of the widely used separable spatio-temporal models, which—as the name implies—separately characterize the spatial and temporal dependence in a process, is their inability to account for the complex interactions and correlation patterns that are attributed to the transport effect of wind. To elaborate, let us first briefly re-iterate the definitions of separability and symmetry, which were introduced in

Section 1, as these definitions represent the basis for the subsequent analysis in this Section of the dissertation.

A covariance structure is said to be separable if it factors into the product of purely spatial and temporal components such that $K(\mathbf{u}, w) = K^s(\mathbf{u}) \cdot K^t(w)$, where $\mathbf{u} \in \mathbb{R}^2$ and $w \in \mathbb{R}$ are the spatial and temporal lags, respectively. Despite their computational benefits, separable models have limiting modeling capabilities which seldom align with reality. For example, separability overlooks the interaction between the spatial and temporal components and implies symmetry in the spatio-temporal covariance structure [19]. A covariance structure is symmetric if $K(\mathbf{u}, w) = K(-\mathbf{u}, w) = K(\mathbf{u}, -w) = K(-\mathbf{u}, -w)$ [33]. Separability implies symmetry, but the converse cannot be guaranteed [36]. That is, the class of separable spatio-temporal models is a subset of the class of symmetric spatio-temporal models.

Due to their attractive computational properties, separable spatio-temporal models dominated the geostatistical literature up to until the beginning of the 21$^{\text{st}}$ century [19]. In recent years, a number of works have pointed out that symmetry is not a physically justifiable assumption for large-scale atmospheric processes in which there is a transport effect, i.e. there exists a dominant air or water flow over time, making the correlation in one direction often stronger than in other directions [33, 81, 78, 36, 49, 17]. To illustrate, take an atmospheric process as an example, where there is known, due to geographical expertise, that there exists an easterly atmospheric flow between two measurement stations (air blows from east to west). Then, it is not reasonable to assume that the covariance between the wind characteristics at the easterly station at a certain time j denoted by $Y_i(j)$ and that at the westerly station after some time in the future denoted by $Y_{i'}(j + h)$, is exactly the same as the covariance between the wind characteristics at the westerly station at time j denoted by $Y_{i'}(j)$ and that at the easterly station in the future $Y_i(j + h)$, with $h \in \mathbb{Z}^+$. As such, this assumption of symmetry should not be appropriate for a process that has a prevailing flow over time.

The limiting capabilities of separable models in capturing the complex correlation patterns in real-world processes motivated the geostatistical community in the past decade to establish classes

of nonseparable, but still symmetric, models that capture spatio-temporal interactions [18, 33], as well as asymmetric nonseparable models that, in addition to the interactions, account for the lack of spatio-temporal symmetry [78, 65, 49, 36]. By establishing an analogy between the atmospheric processes studied in the aforementioned literature and the wind field on a farm, one expects that the lack of symmetry is extended to the farm-level wind dynamics, and thus, anticipates a successful application of nonseparable asymmetric models in short-term wind forecasting. Our review of the literature, however, has suggested the contrary—to the best of our knowledge, spatio-temporal asymmetry has not been reported in wind energy applications [53, 61, 69].

Triggered by that observation, our goal is to explore the influence of the transport effect of wind on the separability and symmetry of spatio-temporal dependence structure in the context of a local wind field and the reasons behind its disregard in the related literature. Our analysis indicates that a straightforward use of the year-long spatio-temporal data described in Section 1.5 would lead to the conclusion that local wind fields are approximately symmetric; a conclusion that explains the widespread use of separable models. Our later findings suggest that this conclusion is misleading because it overlooks the physical nature of local wind fields, where the local wind dynamics take place on a much more granular scale compared to large-scale processes. In order to see the fine-scale asymmetric patterns, unnoticed previously in the spatio-temporal data, we need some magnifying mechanisms that look deeper into the spatial and temporal scales, analogous to the use of optical lens for seeing tiny specimens and features transparent to naked eyes. The magnifying mechanisms are pertinent to the data in space and time, thereby referred to as a pair of "*spatio-temporal lens*." Using the spatio-temporal lens, we conclude that, contrary to common practice, substantial signs of asymmetry do exist in the wind farm data.

In light of these findings and making use of the devised spatio-temporal lens, a physically justifiable asymmetric nonseparable model is proposed. Substantial improvements in forecast accuracy, in terms of both wind speed and power, are achievable using the proposed model. When combined with certain machine learning methods such as support vector machines, additional improvements can be realized.

2.2 Asymmetry detection and quantification

One active research area in the geostatistical literature is to develop statistical tests for certain properties of covariance functions like separability and symmetry in the context of environmental or geophysical applications [63, 78, 30, 57]. One approach is through an empirical method to quantify asymmetry in terms of spatio-temporal semi-variogram, which is a negative-definite measure of dissimilarity in geostatistics [78]. Specifically, the spatio-temporal empirical semi-variogram between two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ at time lag w is defined as:

$$g(\mathbf{s}_i, \mathbf{s}_{i'}, w) = \frac{1}{2(T-w-1)} \sum_{j=1}^{T-w-1} \{Y_i(j+w) - Y_{i'}(j)\}^2. \tag{2.1}$$

Introduce two semi-variograms between $\mathbf{s}_i$ and $\mathbf{s}_{i'}$: $g(\mathbf{s}_i, \mathbf{s}_{i'}, w)$ and $g(\mathbf{s}_{i'}, \mathbf{s}_i, w)$. Both of them represent the dissimilarity between the two spatial sites, but $g(\mathbf{s}_i, \mathbf{s}_{i'}, w)$ means that measurements taken at $\mathbf{s}_{i'}$ are w time lag behind that at $\mathbf{s}_i$, whereas $g(\mathbf{s}_{i'}, \mathbf{s}_i, w)$ means that measurements at $\mathbf{s}_i$ are behind those at $\mathbf{s}_{i'}$. Assuming stationarity in space-time with a positive-definite covariance function $K(\cdot, \cdot)$, then $\mathbb{E}[g(\mathbf{s}_i, \mathbf{s}_{i'}, w) - g(\mathbf{s}_{i'}, \mathbf{s}_i, w)] = K(\mathbf{s}_{i'} - \mathbf{s}_i, w) - K(\mathbf{s}_i - \mathbf{s}_{i'}, w)$. As such, when the two semi-variogram quantities are the same, the wind field is said to be symmetric because their expected difference, which is equal to the expected difference in covariances, is zero. But when there is a dominant wind blowing from $\mathbf{s}_i$ towards $\mathbf{s}_{i'}$, the propagation of wind velocities from $\mathbf{s}_i$ towards $\mathbf{s}_{i'}$ should generate a positive value for the difference $a(\mathbf{s}_i, \mathbf{s}_{i'}, w) := g(\mathbf{s}_i, \mathbf{s}_{i'}, w) - g(\mathbf{s}_{i'}, \mathbf{s}_i, w)$, indicating a lack of symmetry. In a more general sense, when there is a dominant wind direction, denoted by d (in the above example, from $\mathbf{s}_i$ towards $\mathbf{s}_{i'}$), a non-zero $a(\mathbf{s}_i, \mathbf{s}_{i'}, w)$ exhibits itself as a positive value most of the time. In our research, we signify the dominant wind direction through an extra input, i.e., $a(\mathbf{s}_i, \mathbf{s}_{i'}, w, d)$, and use it to detect the existence of asymmetry in wind data and quantify its strength.

2.2.1 Yearly Scenario

Here, we show the conclusion that would result when attempting to quantify asymmetry in a local wind field by using the procedure outlined above which was designed for a large-scale atmospheric process [78].

First, we fit a parametric diurnal trend for the wind speed data using least squares estimation. Specifically, we use the diurnal trend function expressed in (2.2), originally proposed in [38], to capture the global temporal variation in wind speeds. The fitted trend is then subtracted from the actual wind speed data to get the residuals, which will be used in quantifying asymmetry.

$$f(t) = \beta_0 + \beta_1 \sin\big(\frac{2\pi t}{24}\big) + \beta_2 \cos\big(\frac{2\pi t}{24}\big) + \beta_3 \sin\big(\frac{4\pi t}{24}\big) + \beta_4 \cos\big(\frac{4\pi t}{24}\big). \tag{2.2}$$

For our wind data, the yearly average of wind direction is $\bar{d} = 264.24^\circ$ (where 0° represents north), meaning that the average dominant wind on the wind farm is from west to east. In this analysis, we rely on the wind direction data from one of the 3 masts in Figure 1.2. So, for every pair of turbines i and i' located at $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ such that $\mathbf{s}_i$ is west of $\mathbf{s}_{i'}$, we compute $g(\mathbf{s}_i, \mathbf{s}_{i'}, w) - g(\mathbf{s}_{i'}, \mathbf{s}_i, w)$ using the residuals in place of Y in (2.1). We repeat this computation for every pair of turbines and for different time lags ranging from 0 to 24 hours. All of the computed quantities are then transformed into the correlation scale. For the ℓ-th pair of turbines, the resulting quantities at each temporal lag w are the spatio-temporal asymmetry, denoted by $a^\ell(\mathbf{s}_i, \mathbf{s}_{i'}, w, \bar{d})$.

Denote the collection of asymmetry values at each temporal lag by $A(\mathbf{s}, w) = \{a^\ell(\mathbf{s}_i, \mathbf{s}_j, w)\}_{\ell=1}^{L}$, where L is the total number of turbine pairs, and its 50^{th} percentile as $\bar{A}(\mathbf{s}, w)$. Algorithm 1 summarizes the asymmetry quantification procedure for the yearly scenario, and the results are displayed in Figure 2.1 which shows the 25^{th}, 50^{th} and 75^{th} percentiles of $A(\mathbf{s}, w)$ for $w \in \{0, \ldots, 24\}$ with a 3-hour increment. All median asymmetry values in Figure 2.1 are slightly positive, indicating a potential tendency towards asymmetric behavior. The largest median occurs at a temporal lag, which we denote by $w^* = 12$, and is approximately 0.024 on the correlation scale. To put this value in the context, we note that an asymmetry value of 0.12 for asymmetric large-scale wind

flow over Ireland was previously reported [33]. Similar figures are reported in [36], ranging between 0.04 to 0.14 and averaged at 0.11. As such, an asymmetry of 0.024 appears to be rather weak to justify the existence of asymmetry in the local wind field. From a modeling perspective, one would understandably trade such weak asymmetry for computational efficiency and simplicity gained by the symmetry assumption. This conclusion explains the broad use of separable models in the wind forecasting literature.

Algorithm 1 Asymmetry quantification for the yearly scenario

0. Compute the average wind direction $\bar{d}$ for the whole year.
1. Find the dominant direction $R(d)$, where $R(\bar{d}) \in$ {N, E, S, W}.
2. Fit the duirnal trend of (2.2) and use the residuals for asymmetry quantification.
3. Initialize the time lag $w = 0$.

repeat

4. Initialize the wind turbine index $i = 1$.

repeat

5. Determine the subset $S(i, R(\bar{d}))$ of wind turbines that are considered downstream with respect to i. For instance, if $R(\bar{d}) = W$, then determine the subset of all wind turbines that are east of turbine i.
6. Initialize the index $i' = 1$.

repeat

7. Compute the quantities $g_1 := g(\mathbf{s}_i, \mathbf{s}_{i'}, w)$ and $g_2 := g(\mathbf{s}_{i'}, \mathbf{s}_i, w)$ using the residuals in place of Y in (2.1), where $\mathbf{s}_i$ is the location of the ith wind turbine and $\mathbf{s}_{i'}$ is the location of the i'th turbine of the subset $S(i, R(\bar{d}))$.
8. Compute the asymmetry $a^{\ell}(\mathbf{s}_i, \mathbf{s}_{i'}, w)$ as the difference between g_1 and g_2, transform into correlation scale by dividing by the variance, where ℓ denote the index of the turbine pair.
9. Compute the distance in kilometers $\text{dist}^{\ell}(\mathbf{s}_i, \mathbf{s}_{i'})$ between turbines i and j.
10. Update the index $i' = i' + 1$

until $i' = \mathcal{I} + 1$, such that $\mathcal{I} = |S|$

11. Update the index $i = i + 1$.

until $i = n + 1$

12. Compute the 25^{th}, 50^{th} and 75^{th} percentiles of $A(\mathbf{s}, w) = \{a^{\ell}(\mathbf{s}_i, \mathbf{s}_{i'}, w)\}_{\ell=1,\ldots,L}$ and denote the 50^{th} percentile as $\bar{A}(\mathbf{s}, w)$.
13. Update $w = w + 1$.

until $w > 24$

14. Determine w^* such that $\bar{A}(\mathbf{s}, w^*) = \max_{1 \le w \le 24} \{\bar{A}(\mathbf{s}, w)\}$.
15. Bin $A(\mathbf{s}, w^*)$ versus separating distances $\text{dist}(\mathbf{s}) = \{\text{dist}^{\ell}(\mathbf{s}_i, \mathbf{s}_{i'})\}_{\ell=1,\ldots,L}$ to get the overall estimates of asymmetry level at varying separating distances.

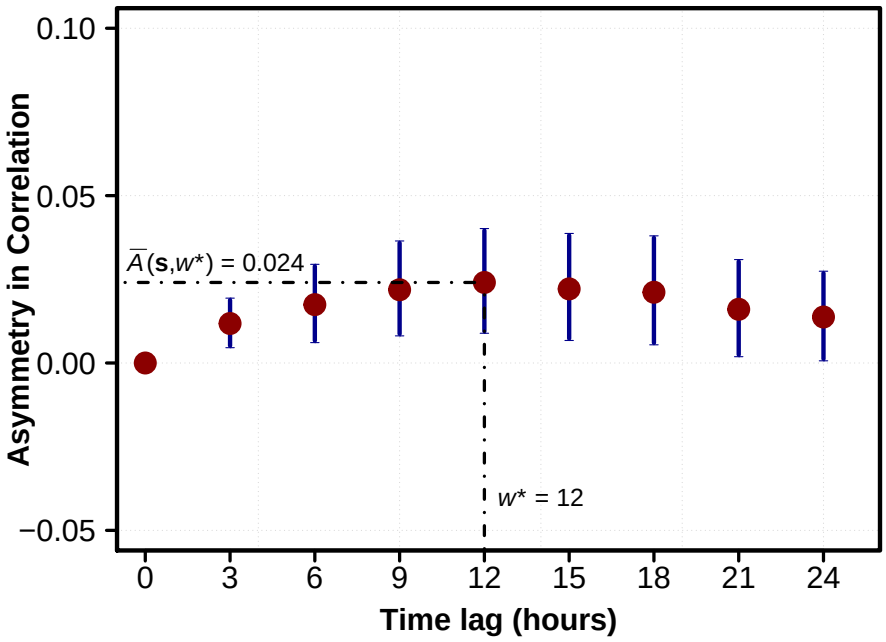


Figure 2.1: The 25^{th}, 50^{th} and 75^{th} percentiles of asymmetry values $A(\mathbf{s}, w)$ versus different time lags w. The time lag at which the maximal median asymmetry is realized is denoted by w^*.

Our hypothesis, however, is that the weak asymmetry in Figure 2.1 is due to the non-optimal handling of wind farm data. In general, the common practice is to decompose data using regular calendar periods, like a week, a month, or a year. The wind data is grouped for the whole year when producing Figure 2.1. In a large-scale atmospheric process, a dominant wind can persist for a sustained period of time and travel a substantial distance; these patterns can be pre-identified through climatological expertise over a region of interest, and as such, regular calendar decompositions appear to be a reasonable choice. For a local wind field, however, observational data suggest that alternations in local winds occur at a relatively high rate, resulting in several distinct wind characteristics at each wind alternation. In such setting, regular calendar periods rarely contain a single dominant wind scenario; rather, they contain various dominant winds that create multiple asymmetries having distinct directions and magnitudes. Consequently, aggregating the heterogeneous, and perhaps opposite, asymmetries leads to an underestimation of the true asymmetry level.

The physical differences between local wind fields and large-scale atmospheric processes require careful adjustments to the spatio-temporal resolution used to analyze wind measurements, in order to reveal the underlying asymmetry pattern. We show next that upon implementing a pair of spatio-temporal "lens" that unearth the fine-scale variations in local wind fields, strong degrees of asymmetry are detected.

2.2.2 Spatio-temporal lens for asymmetry quantification

Here, a statistical tool, called "*the spatio-temporal lens*" is proposed to unearth the underlying asymmetric behavior of local wind fields. The devised lens comprises two components: a temporal adjustment and a spatial adjustment.

A. Temporal adjustment

For exploratory purposes, we start off with performing a series of arbitrary, calendar period-based temporal decompositions. Specifically, we perform seasonal, monthly and weekly decompositions of the wind farm data and compute the asymmetry level in each sub-interval. Take the seasonal decomposition as an example, the average wind direction is computed for the fall season and is denoted by $\bar{d}_1$ and the asymmetry level between all pairs of turbines within that season is computed at the time lag w_1^* that maximizes the median asymmetry level for the fall season, resulting in the vector $\{a^\ell(\mathbf{s}_i, \mathbf{s}_{i'}, w_1^*, \bar{d}_1)\}_{\ell=1}^{n_1}$. We repeat this process for the remaining three seasons, resulting in the wind directions $\bar{\mathbf{d}} = \{\bar{d}_1, \bar{d}_2, \bar{d}_3, \bar{d}_4\}$, the optimal time lags $\mathbf{w}^* = \{w_1^*, w_2^*, w_3^*, w_4^*\}$ and the corresponding asymmetry vectors. We then group the asymmetry values into three subgroups, corresponding to three distance ranges: 0-10 kilometers (km), 10-20 km, and 20-30 km. The $25^{th}, 50^{th}$ and 75^{th} percentiles of asymmetry values in each subgroup are then computed. Algorithm 2 summarizes the procedure for the asymmetry quantification using arbitrary temporal decompositions, and the results are displayed in Figure 2.2 which plots the $25^{th}, 50^{th}$ and 75^{th} percentiles of asymmetry values under each scenario versus the separating distances.

It is apparent that adjusting the temporal lens by which we search for the asymmetry to capture the fine temporal details leads to estimates of stronger asymmetries. However, the adjustment

Algorithm 2 Asymmetry quantification for arbitrary temporal decompositions

0. Decompose the year into P arbitrary intervals. For instance, $P = 4$ for seasonal decompositions.
1. Initialize the interval index $p = 1$.

repeat

2. Compute the average wind direction $\bar{d}_p$ for the p^{th} interval.
3. Define the dominant wind $R(\bar{d}_p)$ correspondent to $\bar{d}_p$ where $R(\bar{d}_p) \in$ {N, E, S, W}.
4. Repeat the Steps 2-14 of Algorithm 1.
5. Update the interval index $p = p + 1$

until $p = P + 1$

6. Aggregate all the turbine pairs from all the sub-intervals into a single two-dimensional dataset $D = \{d_\ell, a_\ell\}_{\ell=1,\ldots,L}$, where d_ℓ and a_ℓ denote the distance in kilometers and the asymmetry in correlation between the ℓ^{th} pair of turbines.
7. Bin the asymmetry levels versus separating distances to get the overall estimate of asymmetry at varying separating distances.

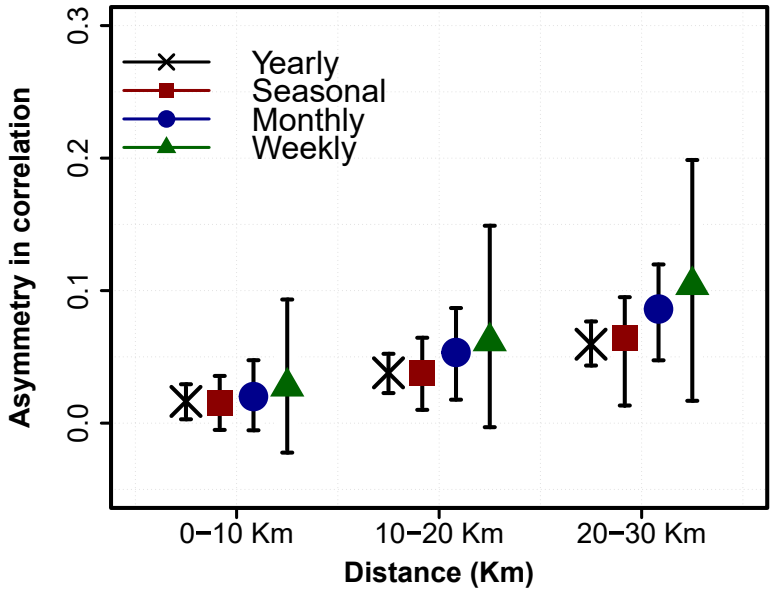


Figure 2.2: The 25^{th}, 50^{th} and 75^{th} percentiles of asymmetry values of temporal decompositions versus separating distance in kilometers.

based purely on calendar periods seems not effective enough, because the decomposition intervals are created arbitrarily. If we can identify the time points when the dominant wind changes its direction, say, from northerly to westerly, we can then isolate the time intervals in which a unique dominant wind persists and consequently detect the underlying asymmetry in such intervals. We call such intervals as the "prevailing periods." This calls for the use of a change point detection procedure.

In order to identify the change points of the wind direction over the year, a rolling binary segmentation of a circular change point detection test is developed, which is a modified version of the method described in [47]. Given hourly wind direction observations for one year, denoted by $\mathbf{d} = \{d_1, d_2, ..., d_T\}$, where T is the total number of wind direction observations, we are interested in detecting the intermediate points at which changes have occurred in the wind direction. We assume that the wind direction variable $d \sim CN(\mu, \kappa)$, i.e. it follows a von Mises, or circular normal distribution with parameters $0 \leq \mu < 2\pi$ and $\kappa \geq 0$ as the mean and concentration parameters, respectively. The von Mises distribution can be expressed as in (2.3).

$$f(d, \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left\{\kappa \cos(d - \mu)\right\}, \qquad 0 \leq d < 2\pi, \tag{2.3}$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero. The choice of the distributional form is motivated by the fact that wind direction is a circular variable and the von Mises distribution can characterize its directionality [60].

The parameter vector for the change detection test would be $\{\omega, \mu_1, \mu_2, \kappa\}$, where ω is the index of the change point to be detected, μ_1 and μ_2 are the means of the before and after subsequences and κ is the concentration parameter. Specifically, the null hypothesis for this change point test is that $H_0 : \omega = T$ indicating no change versus $H_1 : 1 \leq \omega \leq T - 1$ indicating a change occurring at the ω^{th} observation. The generalized likelihood ratio method is used to conduct the test, where H_0 is rejected whenever $\lambda_\omega > c$ such that $\lambda_\omega = \sup_{j \in \{1,...,T\}} (Q_{1j} + Q_{2j}) - Q$ where Q, Q_{1j} and Q_{2j} denote the resultant lengths for the sequence and the resulting subsequences, respectively, and c is a critical cut-off value determined by the prescribed significance level; for more details, please

refer to [47]. Because the test has no simple known distributional form, we resort to Monte Carlo simulation in order to determine the 95^{th} cut-off value which will be used in this test. In the Monte Carlo simulation, we initially tried a run size of 10,000 runs, but it turned out that the results are only trivially different from that obtained at a run size of $1,000$ runs.

Since we are searching for an unknown number of potential change points, a common approach is to implement a binary segmentation algorithm, where the most significant change point in the whole dataset is detected, then the next significant change point is searched for in the resulting subsequences before and after the already-detected change point. For more details about binary segmentation, please refer to [28]. Since we are expecting change points to occur on a fine temporal resolution in local wind fields, we implement a one-month rolling binary segmentation, in which the first month of data is analyzed separately and then the period of interest is shifted by one month from the last change point detected in the first month for the next round of detection. The overall procedure of the rolling binary segmentation is described in Algorithm 3. In doing so, we make use of the `R` packages `circular` and `changepoint` [2, 52].

The top panel of Figure 2.3 shows the detected change points for the first two weeks of data. The bottom panel of Figure 2.3 shows the distribution of the length of the prevailing periods. The results show that, on average, a dominant wind direction lasts for 3.04 days with a standard deviation of 2.46 days. For 50% of the prevailing periods, the wind direction alternated in less than 2.27 days. The maximum interval of time in which a dominant wind direction is found to be persistent is 15.5 days, while the shortest prevailing period's length is found to be 6 hours. A total of 119 change points are detected in the yearlong wind data, leading to 120 prevailing periods identified over the year. These statistics indicate the fine temporal scale at which wind dynamics take place, resulting in a dynamic alternation of dominant winds over time. Quantifying asymmetry for individual prevailing periods would capture the lack of symmetry associated with each distinct dominant wind, and hence, provide reliable asymmetry estimates.

Algorithm 3 Rolling binary segmentation for circular change point detection.

0. Input the wind direction vector $\mathbf{d}$ of size T.
1. Select the first 1-month of data denoted by $\mathbf{d}^{(1)}$ such that $|\mathbf{d}^{(1)}| = 24 \times 30 = 720$ data points.
2. Let $z = 1$, where z is the month index.

repeat

3. Compute the MLE estimate for the concentration parameter associated with $\mathbf{d}^{(z)}$ and denote it by $\hat{\kappa}$.
4. Set $i = 1$.

repeat

5. Simulate a sample $\mathbf{d}_i^s$ of 720 values from a von Mises distribution with mean zero and concentration parameter $\hat{\kappa}$.
6. Determine the resultant length R_i of sample $\mathbf{d}_i^s$.

until $i = 1000$

7. Determine the 95^{th} percentile of the vector $\{R_i\}_{i=1}^{1000}$ and use it as a cut-off value for the hypothesis test.
8. Detect the index of the most significant change point ω and divide the sequence into two subsequences using ω as the separating point.
9. Set $j = 1$

repeat

10. For the j^{th} subsequence, proceed with binary segmentation until no further change points are found.

until $j = 3$

11. Denote the last detected change point within the 1-month interval of data as ω^*.
12. Let $z = z + 1$ and $\mathbf{d}^{(z)}$ would be the 1-month of data starting with d_{ω^*+1}.

until $T - \omega^* < \epsilon$, where $\epsilon \in \mathbb{Z}^+$ is a sufficiently small number .

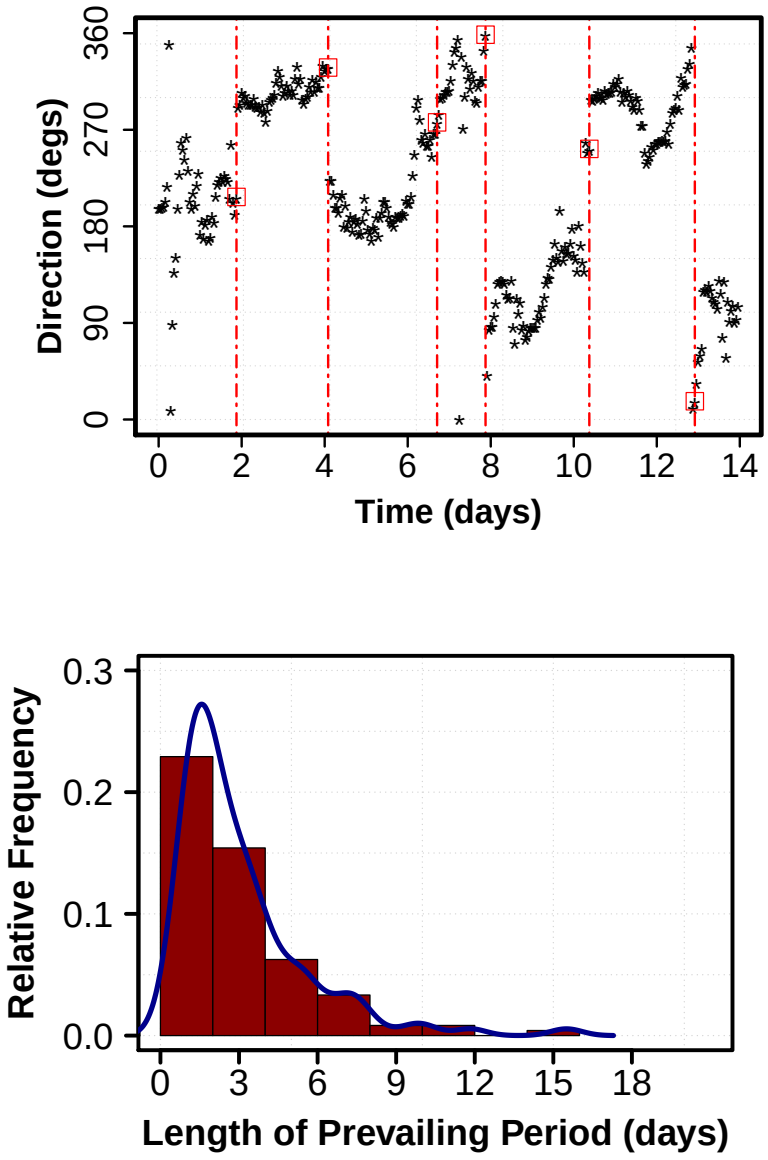


Figure 2.3: Top panel: Change detection for first two weeks of wind direction data. Bottom panel: Distribution of the prevailing periods' length in days.

B. Spatial adjustment

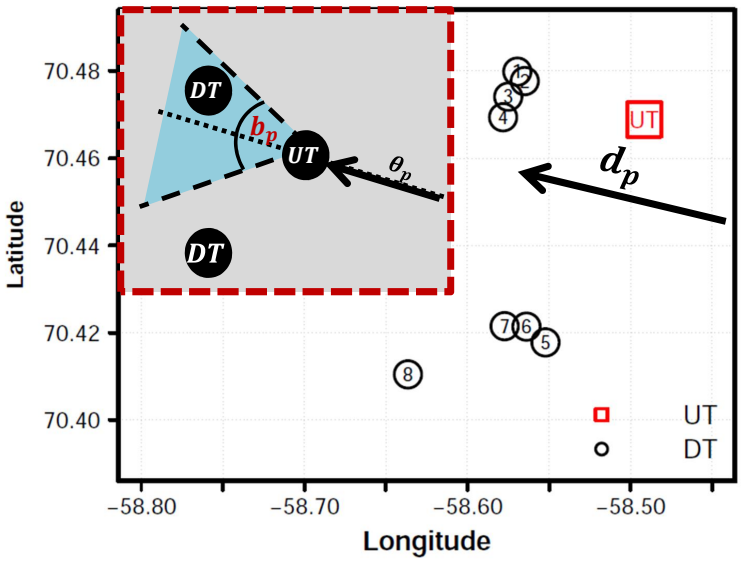
In addition to the temporal adjustment, the spatial configuration of the turbines on a wind farm appears to be a factor that affects the asymmetry level between a pair of turbines in a local wind field. Physically, asymmetry exists when wind propagates from an upstream turbine to a downstream one, implying that the latter is in the along-wind direction with respect to the former.

To show the effect of the spatial configuration of turbines on asymmetry, we pick an 8-hour prevailing period, where the wind is easterly, as illustrated in Figure 2.4. We compute the Pearson correlations between the wind speeds at the upstream turbine (UT) and those in a subset of 8 arbitrarily selected downstream turbines (DTs 1-8) after a time lag of $w = 3$; we denote this correlation by C_1. It is apparent that the correlations in the along-wind direction (between UT and DTs 1-4) are larger than those in the span-wind direction (between UT and DTs 5-8). Also, we compute the pairwise correlations at a time lag of $-w$ (when the DTs are leading the UT) and denote it by C_2. The difference between C_1 and C_2, denoted by ΔC, is largely positive for the along-wind DTs, indicating the existence of a strong asymmetry pattern. These observations prompt the need to select only the along-wind turbines in asymmetry quantification.

For a given prevailing period p, the dominant wind direction is denoted as d_p and a distinct spatial bandwidth b_p will be selected, as shown in Figure 2.4. Our spatial bandwidth selection procedure, as illustrated in the upper left corner of Figure 2.4, executes as follows: we vary the bandwidth in the range $[2.5^\circ, 45^\circ]$ in increments of 2.5° and then select the bandwidth that maximizes the median asymmetry as the optimal bandwidth b_p^*. With the spatial adjustment, the asymmetry metric $a(\cdot)$ takes in one more input and is now denoted as $a(\mathbf{s}_1, \mathbf{s}_2, w_p, d_p, b_p)$.

C. Asymmetry quantification using the spatio-temporal lens

In light of the spatial and temporal adjustments described above, we determine an optimal time lag w_p^* and bandwidth b_p^* that constitute our final pair of spatial and temporal lens, maximizing the median asymmetry level in each prevailing period. Figure 2.5 which shows the 25^{th}, 50^{th} and 75^{th} percentiles of the asymmetry level versus the separating distance subgroups for the different



	C_1	C_2	ΔC
DT1	0.89	0.51	0.38
DT2	0.85	0.17	0.69
DT3	0.79	0.24	0.55
DT4	0.89	0.35	0.54
DT5	0.2	0.26	-0.06
DT6	0.18	0.14	0.04
DT7	0.38	0.38	0.00
DT8	0.52	0.30	0.22

Figure 2.4: Effect of the spatial configuration of the wind turbines on asymmetry quantification. "UT" and "DT" denote an upstream and a downstream turbine, respectively.

scenarios thus considered: yearly, seasonal, monthly, weekly, temporal-only lens scenario and spatio-temporal lens scenario.

It is apparent that applying the spatio-temporal lens detects much higher asymmetry levels. For instance, at separating distances greater than 20 km, all of the turbine pairs exhibit positive asymmetry and 50% of them exhibit an asymmetry level higher than 0.2 on the correlation scale, a level considered significant in the past study [33] and nearly an order of magnitude greater than the median asymmetry of 0.024 detected earlier on the yearly data. The median asymmetry values of all distance subgroups are classified in Table 2.1, where 93% of the prevailing periods exhibit positive median asymmetry, nearly a quarter of them exhibit a greater than 0.2 median asymmetry, and more than 41% of them exhibit a median asymmetry larger than 0.1, the level of asymmetry previously reported in [33, 36] for signaling the existence of appreciable asymmetric behavior in the large-scale atmospheric processes.

The findings made above indicate that not only does strong asymmetry exist in local wind fields, but also the discovered asymmetry appears to fluctuate spatially and temporally, and in

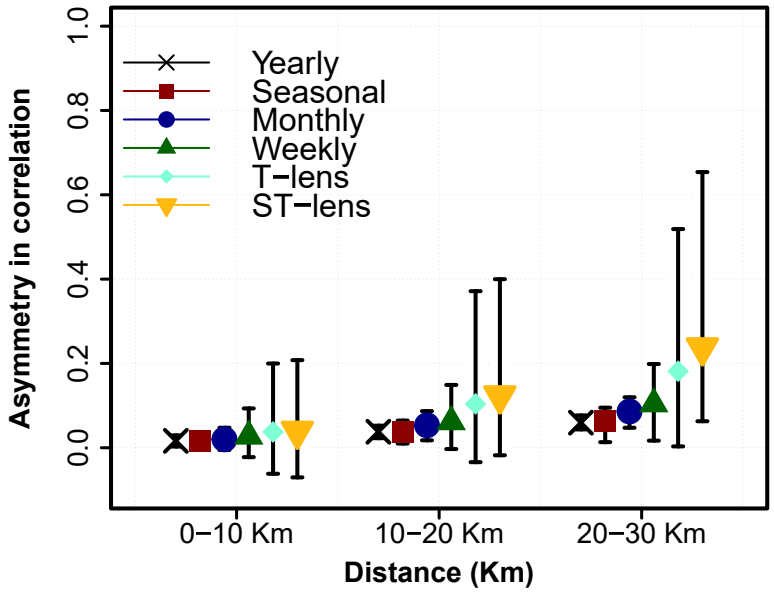


Figure 2.5: The 25^{th}, 50^{th} and 75^{th} percentiles of asymmetry values of different scenarios versus separating distance in kilometers.

Table 2.1: Classification of prevailing periods according to the median asymmetry level.

Group	Range	Percentage
1.	$\bar{A}(\mathbf{s}_1, \mathbf{s}_2, w_p^*, d_p, b_p^*) \leq 0$	7%
2.	$0 < \bar{A}(\mathbf{s}_1, \mathbf{s}_2, w_p^*, d_p, b_p^*) < 0.05$	27%
3.	$0.05 \leq \bar{A}(\mathbf{s}_1, \mathbf{s}_2, w_p^*, d_p, b_p^*) < 0.1$	25%
4.	$0.1 \leq \bar{A}(\mathbf{s}_1, \mathbf{s}_2, w_p^*, d_p, b_p^*) < 0.2$	20%
5.	$0.2 \leq \bar{A}(\mathbf{s}_1, \mathbf{s}_2, w_p^*, d_p, b_p^*)$	21%

both magnitude and direction. Each prevailing period appears to have unique asymmetry pattern, creating a temporal fluctuation of asymmetry throughout the year. The asymmetry appears to exhibit a spatial variation as well, by taking on high values in the along-wind and low values in the span-wind directions. Such spatial and temporal variation of the asymmetry pattern is attributed to the dynamic nature of the transport effect of wind. As such, separable spatio-temporal models appear to be overly simplified in modeling farm-level local wind dynamics, which are shown to

exhibit strong signs of asymmetry owing to the dynamic transport effect of wind.

D. Asymmetry and wake effect

By capturing the asymmetry in a local wind field, we can enrich our understanding of complex physical phenomena on a wind farm such as the wake effect. The spatio-temporal dynamics within a wind farm are affected by the wake effect because the rotating turbine blades cause changes in the speed, direction and turbulence intensity of the propagating wind [16].

In Figure 2.6, we divide the wind farm for each prevailing period based on the wind direction into two regions having approximately the same number of turbines. The first region is considered as the set of "free-stream" wind turbines that receive a relatively wake-free, less turbulent wind and the second region is considered as the set of "wake" wind turbines which are in the wake of other turbines and receive the disturbed, turbulent wind. We plot the medians of the asymmetry for each region. The free-stream region appears to exhibit higher asymmetry, which is consistent with the physical understanding since the less-turbulent wind is the driving force creating the asymmetry. This analysis indicates that asymmetry level spatially varies on a wind farm due to wake effect. Incorporating such pattern in a spatio-temporal model could benefit model fitting and prediction, as well as aid researchers in wake characterization.

2.3 Statistical short-term wind forecasting considering asymmetry of local wind fields

In light of the findings in Section 2.2, a framework for short-term wind forecasting is proposed that takes into account the asymmetry of local wind dynamics attributed to the transport effect of wind. The flowchart in Figure 2.7 presents the steps of the proposed forecasting procedure.

We want to note that to perform h-hour ahead forecasts and following the terminology presented in Section 2.2, only the data in the preceding prevailing period that share similar wind and asymmetry characteristics is used for model training. This implies that a small subset of data relevant to the current prevailing period is used for model training. Doing so, the benefit is twofold. First, it eliminates the computational burden of fitting nonseparable spatio-temporal models. Second, it makes use of a local informative spatio-temporal neighborhood that is most relevant to

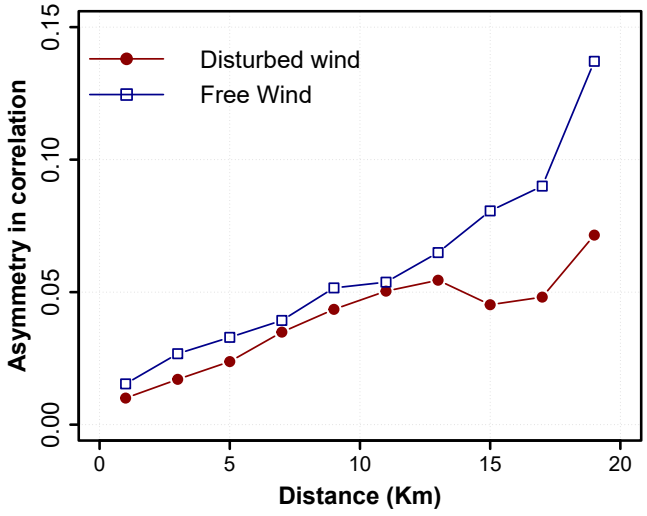


Figure 2.6: Wake effect and its implications on spatio-temporal asymmetry.

short-term horizons. We note informative spatial neighborhoods were used for short-term speed forecast [69], while local temporal windows for short-term time-series power predictions were also employed [83].

2.3.1 Models for short-term wind forecasting

In Figure 2.7, one major component is the fitting of an asymmetric nonseparable spatio-temporal model. Such models have been proposed in the past, although not used for local wind fields. We use a constant mean GRF model as in (1.1). For the covariance function, similar to [36], we define it as a convex combination of two covariance structures which can be expressed as in (2.4).

$$K_a(\mathbf{u}, w) = \sigma^2 \Big\{ (1-\lambda) K_{NS}(\mathbf{u}, w) + \lambda K_T(\mathbf{u}, w) \Big\} + \eta \mathbb{1}_{\{||\mathbf{u}||=|w|=0\}}, \tag{2.4}$$

where K_T is an asymmetric correlation function to be given below and K_{NS} is a nonseparable symmetric correlation function, as in [33], such that:

$$K_{NS}(\mathbf{u}, w) = \frac{1-\delta}{1+\alpha|w|^2} \left(\exp\left[-\frac{c||\mathbf{u}||}{(1+\alpha|w|^2)^{\frac{\beta}{2}}} \right] + \frac{\delta}{1-\delta} \mathbb{1}_{\{||\mathbf{u}||=0\}} \right). \tag{2.5}$$

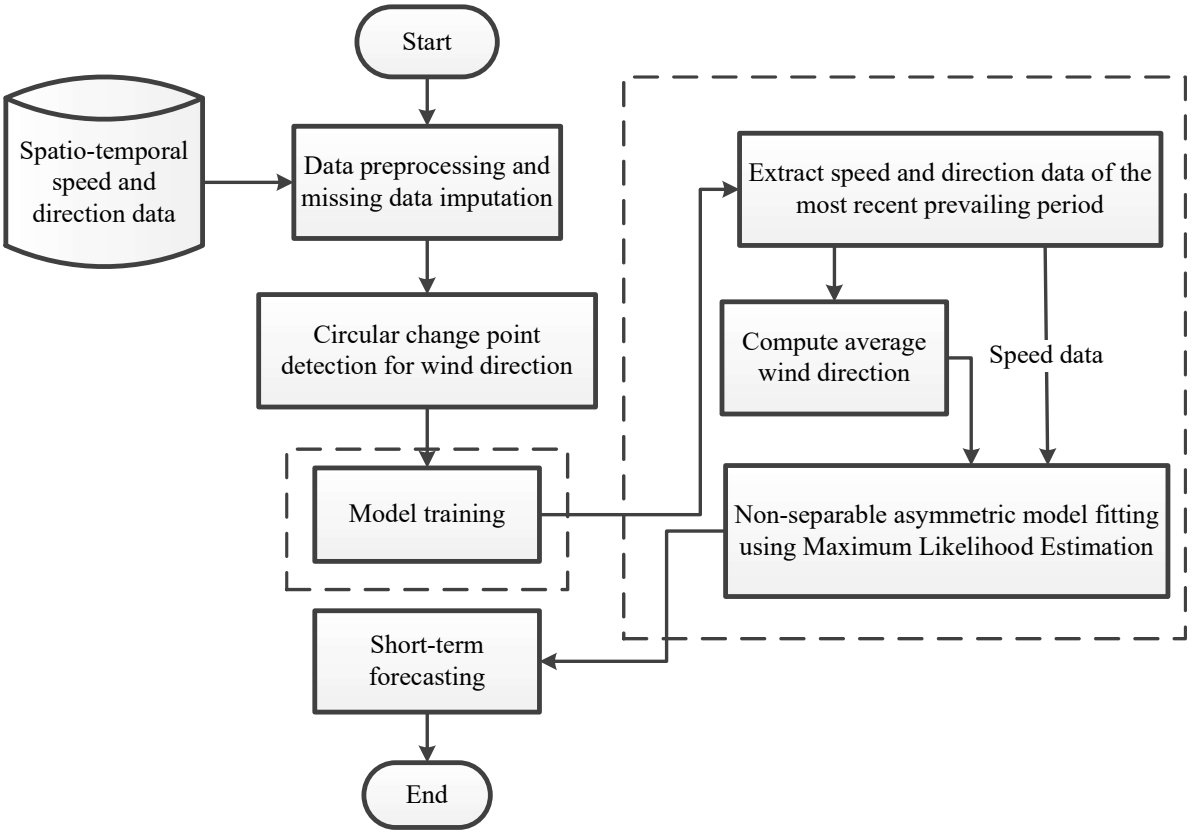


Figure 2.7: Flowchart of short-term wind forecasting using asymmetric modeling.

In (2.4) and (2.5), $\mathbb{1}_{\{\cdot\}}$ is an indicator function, $\mathbf{u} = (u_1, u_2)^T$ is the spatial lag consisting of longitudinal and latitudinal components u_1 and u_2 and $||\cdot||$ is the Euclidean norm. The parameters α and c determine the temporal and spatial ranges, whereas $0 \leq \delta < 1$ and η are the spatial and spatio-temporal nugget effects, respectively, such that α, c and η are all non-negative, and $\sigma^2 > 0$ is the spatio-temporal variance. The nonseparability parameter β represents the strength of the spatio-temporal interaction and the asymmetry parameter λ represents the lack of symmetry; $\beta, \lambda \in [0, 1]$.

There are different approaches to define K_T. A simpler one is defined in [36] as a Lagrangian compactly supported function:

$$K_T(\mathbf{u}, w) = \left(1 - \frac{1}{2||\mathbf{V}||}||\mathbf{u} - \mathbf{V}w||\right)_+, \tag{2.6}$$

where $(\cdot)_+ = \max(\cdot, 0)$ and $\mathbf{V} = (V_1, V_2)^T$ is a two-dimensional vector having a longitudinal and latitudinal component, respectively, and is defined based on the knowledge of the weather system. For example, if the dominant wind is known to be strictly westerly, then $\mathbf{V}$ is chosen to be $(V_1, 0)^T$, i.e., a non-zero longitudinal wind velocity reflecting the traveling of the wind along the longitudinal axis. A generalized version of K_T is proposed in [75], which is, instead of using a constant vector, defines $\mathbf{V}$ as a random variable that follows a multivariate normal distribution, i.e., $\mathbf{V} \sim \mathcal{N}(\boldsymbol{\mu}, \frac{\mathbf{D}}{2})$. As such, K_T is defined as follows:

$$K_T(\mathbf{u}, w) = \frac{1}{\sqrt{|\mathbf{1}_{2\times 2} + w^2\mathbf{D}|}} \exp\big[-(\mathbf{u} - w\boldsymbol{\mu})^T(\mathbf{1}_{2\times 2} + w^2\mathbf{D})^{-1}(\mathbf{u} - w\boldsymbol{\mu})\big], \tag{2.7}$$

where $|\cdot|$ in (2.7) denotes the matrix determinant. Our research shows that using the generalized K_T boosts the forecast quality, and for this reason, our forecast model uses the K_T in (2.7). This does not come as a surprise since local wind dynamics are highly random, and as such, $\mathbf{V}$ is best described as random rather than constant. We refer to this asymmetric nonseparable model hereinafter as ASYM.

We compare ASYM to a set of existing models in the literature and practice, namely: separable

spatio-temporal models (SEP1, SEP2), persistence forecasts (PER), and support vector machines (SVMs). We also define a new model as a hybrid between ASYM and SVM and refer to that as HYB. We begin by explaining the separable spatio-temporal models (SEP1, SEP2). By setting $\beta = \lambda = 0$ in (2.4), we get a symmetric separable model. Two variants of this separable model are used: SEP1 and SEP2. The model SEP1 borrows the estimated parameters of ASYM but sets $\beta = \lambda = 0$. As for SEP2, we first set $\beta = \lambda = 0$ before parameter estimation, and then freely estimate the remaining parameters from the data. For ASYM, SEP1, and SEP2, a Maximum Likelihood Estimation (MLE) is used to estimate the parameters of the model as in (1.3). Once all parameters are estimated, a future wind speed value can be calculated using the wind speed observations collected up to the current time as in (1.5).

Forecasts based on the persistence model (PER) are widely used as a benchmark in the renewable industry and are simply obtained by extending the current wind characteristics for the look-ahead horizons such that $\hat{Y}_i(t+h) = Y_i(t), \forall\, h = 1, \ldots, H$ and $i = 1, \cdots, n$. We also compare ASYM against a classical time series method which does not account for spatial correlation on the wind farm (and thus no spatio-temporal asymmetry). A common approach for time-series modeling and forecasting is the autoregressive moving average (ARMA) model, which is defined by the orders of the autoregressive and the moving average parts, p and q respectively, and a set of parameters $\{a, b_1, .., b_p, c_1, ..., c_q\}$. An ARMA$(p, q)$ model for a spatial location i is expressed as in Equation (2.8).

$$Y_i(j) = a + \sum_{l=1}^{p} b_l Y_i(j-l) + \sum_{k=1}^{q} c_k \epsilon_i(j-k) + \epsilon_i(j). \tag{2.8}$$

We fit an ARMA model using the command `arima` in the `R` package `stats`. Our analysis suggests that a better prediction is achieved by using low-order, rather than high-order ARMA models. This aligns with the recommendations of previous studies which suggest that low-order ARMA models are more suitable for short-term forecasts [43, 69]. Therefore, we set $p = q = 1$ and fit an ARMA$(1, 1)$ model for each turbine out of the 200 locations at the wind farm. As such,

we generate 200 independent forecasts and then compute the overall aggregate forecast error.

Machine learning approaches have also been reported to benefit wind forecasting. Amongst those, support vector machines (SVMs) are a popular choice and have been previously implemented for time-series wind speed and power forecasting [64, 84, 73]. Simply speaking, SVM performs a nonlinear mapping of the data into a high-dimensional feature space, as shown in (2.9):

$$Y_i(j) = \beta_0 + \sum_{j'=1}^{d} \beta_{j'} \phi_{j'}(x_i(j)) + \epsilon_i(j), \tag{2.9}$$

where $x_i(j)$ is an explanatory input to be defined later, $\{\phi_{j'}\}_{j'=1}^{d}$ are the so-called feature basis functions and $\{\beta_{j'}\}_{j'=0}^{d}$ are the SVM parameters to be estimated from the data [64]. For an SVM model, instead of explicitly defining the feature basis functions, a kernel function is specified that defines the inner product in the feature space. This is referred to in the machine learning literature as the "kernel trick" [27]. Following the same logic made above regarding the suitability of most recent lagged values with respect to short-term forecasts, we set $x_i(j) = Y_i(j-1)$. We fit an SVM model for the time series training data of each turbine of the 200 locations at the wind farm. Specifically, we use the command `svm` in the `R` package `e1071` and a radial basis kernel function, which is a common choice in the SVM literature [27]. We refer to this model as SVM.

A growing trend in the wind forecasting literature is to hybridize multiple data-driven methods to achieve better prediction accuracy [15, 42]. We follow an approach similar to [15], which was originally proposed for integrating classical time-series models with neural networks for wind speed forecasting. Specifically, we first fit ASYM to the spatio-temporal training data. The choice of ASYM as the base model is motivated by our goal to capture the spatio-temporal asymmetries on the wind farm. We then fit an SVM to the residuals obtained by ASYM to capture any nonlinearities that are not covered by the base model. We refer to this hybrid model, which is presented in (2.10), as HYB, where $Y_i^A(j)$ is the ASYM model fit, $\mathcal{E}_i^S(j)$ represents the SVM model fit to the spatio-temporal residuals obtained after the ASYM fit, and $e_i(j)$ is the final residual term.

$$Y_i(j) = Y_i^A(j) + \mathcal{E}_i^S(j) + e_i(j). \tag{2.10}$$

2.3.2 Case study

We demonstrate the merit of our procedure on four prevailing periods from different times of the year. For all the periods, we select 6 hours for model training. We will forecast up to 4-hours ahead, i.e., $h = \{1, 2, 3, 4\}$. Our choice for the training period is motivated by observing that the shortest prevailing period length, as shown in Section 2.2, is about 6 hours. For short-term wind prediction, previous studies showed that using a longer history of wind measurements is not necessarily helpful, as evident by the low time lag order used in the time series models [12, 24, 69].

As mentioned previously, in ASYM, $\mathbf{V} \sim \mathcal{N}(\boldsymbol{\mu}, \frac{\mathbf{D}}{2})$. To specify $\boldsymbol{\mu}$ and $\mathbf{D}$, we make use of the average wind speed $\bar{Y}$and average wind direction $\bar{d}$ of the training data as follows. First, we set $||\boldsymbol{\mu}|| = \bar{Y}$. Then, given $\bar{d}$, we compute the wind velocity vector in degrees, by decomposing $||\boldsymbol{\mu}||$ into longitudinal and latitudinal components, denoted by $\hat{v}_1$ and $\hat{v}_2$, respectively. These components will be used as the estimate for $\boldsymbol{\mu}$ and their signs would represent the wind directionality information. In our current study, we found that when the 2×2 matrix $\mathbf{D}$ is chosen as a diagonal matrix with its diagonal entries set to $||\boldsymbol{\mu}||$, the parameter estimation procedure produces higher MLE values, suggesting a better fit.

The MLE is implemented in `R` using the routine `nlm`. For instance, for a prevailing period in January, 2011, $\bar{d} = 122.28°$, $\bar{Y} = 20.15$ km/hr, $(\hat{v}_1, \hat{v}_2) = (-0.20°, 0.10°)$ and the MLE's are $\{\hat{\sigma}^2, \hat{\nu}, \hat{\alpha}, \hat{c}, \hat{\eta}, \hat{\beta}, \hat{\lambda}\}$ = $\{19.22, 0.18, 0.46, 0.35, 3.35, 0.99, 0.47\}$. The estimate for the mean term is 21.98 km/hr. The values for $\hat{\beta}$ and $\hat{\lambda}$ indicate the need to consider spatio-temporal interaction and asymmetry.

The log-likelihoods for ASYM, SEP1 and SEP2 are presented in Table 2.2. The log-likelihood values for ASYM are higher than those of SEP1 and SEP2 for all the periods, suggesting a higher explanatory power in favor of the asymmetric model and the importance of incorporating it in the spatio-temporal modeling of local wind fields.

Table 2.2: Log-likelihoods of the asymmetric versus the separable models. Bold-faced values indicate the best performance.

Period	Month	ASYM	SEP1	SEP2
1.	October, 2010	**-2087.84**	-2091.69	-2089.19
2.	December, 2010	**-1980.29**	-2263.43	-2030.04
3.	January, 2011	**-1796.45**	-1826.90	-1799.82
4.	June, 2011	**-2181.68**	-2463.77	-2185.45

Next, we make forecasts based on the models presented in Section 2.3.1 and evaluate them in terms of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as defined in (1.10) and (1.11), respectively. Since here we do not make a rolling forecast, we can omit the index ℓ and set $L = 1$ in the expressions of (1.10) and (1.11). In Section 3 of this dissertation, the forecasts will be evaluated on a rolling forward forecasting fashion. The RMSE and MAE results are illustrated in Tables 2.3 and 2.4, respectively. Please note that for Period 2, the testing data for $h = 4$ had a large amount of missing values when our industrial collaborator provided the data, and for this reason, we were not able to assess the forecast quality for that particular forecast horizon. We, therefore, compute the aggregate error based on the first 3-hour ahead forecasts for Period 2. For Periods 1, 3 and 4, the aggregate measure reported is the average over all 4-hour ahead forecasts.

An important goal for wind farm operators is to accurately predict power production, in addition to wind speed forecasts. These power predictions are valuable for critical decision-making activities such as determining the optimal hourly market participation ahead of time. In practice, turbine-specific power curves are provided by the manufacturer and are used to assess the prediction accuracy of competing models. Since we do not readily have these power curves at hand, we use the binning method over the available one-year worth of data to estimate the turbine-specific power curves. The binning method is a nonparametric method commonly used in the wind industry [46] and is based on discretizing the wind speed domain into a number of bins and then outputting the average power value for each bin as an estimate of power generation. Using the estimated power curves, we predict the power generated at each turbine given the wind speed forecasts. We compare the competing models in terms of RMSE in Table 2.5.

The results presented in Tables 2.3, 2.4, and 2.5 show that the forecasts based on ASYM model outperform those based on separable (SEP1 and SEP2), persistence (PER), time-series (ARMA), and support vector machine (SVM) models, in terms of both wind speed and wind power forecasting. The improvement of ASYM over the separable models is explained by capturing the strong asymmetries discovered in Section 2.2, whereas the improvement over ARMA and SVM is mostly due to the characterization of spatial correlations, and subsequently the asymmetry, both of which the time-series models fail to capture.

Furthermore, hybridizing ASYM with SVM (the HYB model) appears to achieve a further enhancement in prediction accuracy over the ASYM only approach, demonstrating the additional benefit brought by SVM. The improvements of HYB over ASYM range from 1% to 7%, and on average 3.5%, for wind speed forecast, and up to 8%, but on average 4.0%, for wind power forecast.

Combining the strength of the asymmetric modeling and SVM, in terms of wind speed forecast, HYB improves, on the average of the four periods, 22.5% in RMSE (24.3% in MAE, same below) over SEP1, 8.0% (9.7%) over SEP2, 19.5% (19.0%) over PER, 21.2% (24.5%) over ARMA(1,1) and 22.9% (25.1%) over SVM. In terms of wind power forecast, HYB on average improves in terms of reduction in RMSE 23.5% over SEP1, 9.4% over SEP2, 18.2% over PER, 22.8% over ARMA(1,1), 24.3% over SVM. These results are aligned with the findings made in Section 2.2 that local wind fields are strongly asymmetric at the fine-scale spatio-temporal resolutions and as such, spatio-temporal models that capture such physical phenomenon are expected to perform well in short-term forecasts.

2.4 Conclusion

Spatio-temporal wind farm analytics can provide insights to minimize uncertainty while using the wind resource. We demonstrate that, contrary to common practice, local wind fields are strongly asymmetric. This asymmetry is detected upon implementing a set of spatio-temporal adjustments that unearth the fine-scale spatio-temporal dynamics. The asymmetry pattern changes over space and time, in both magnitude and direction. As such, the traditional separable spatio-temporal models appear to be overly simplified in modeling farm-level wind dynamics. Given

Table 2.3: RMSE of wind speed forecasts. Missed the data for $h = 4$ and Period 2 in the original dataset. Bold-faced values indicate the best performance. The percentage improvements are the error inflation rate relative to HYB.

Period	Model	$h = 1$	$h = 2$	$h = 3$	$h = 4$	Average	% improvement
1	ASYM	**0.95**	1.37	2.70	2.86	2.14	1%
	SEP1	1.23	1.63	2.92	3.03	2.34	9%
	SEP2	1.01	1.50	2.85	2.99	2.26	6%
	PER	1.29	1.72	2.98	3.16	2.42	12%
	ARMA	1.63	2.06	3.48	3.62	2.83	25%
	SVM	1.61	1.91	3.33	3.44	2.70	21%
	HYB	**0.95**	**1.36**	**2.69**	**2.85**	**2.13**	
2	ASYM	1.42	2.53	2.31	–	2.14	4%
	SEP1	2.77	3.25	2.65	–	2.90	29%
	SEP2	1.64	2.62	2.30	–	2.22	7%
	PER	1.83	2.88	2.57	–	2.47	17%
	ARMA	1.99	3.05	2.78	–	2.64	22%
	SVM	2.42	3.68	3.41	–	3.21	36%
	HYB	**1.25**	**2.45**	**2.27**	–	**2.06**	
3	ASYM	**0.87**	0.94	1.02	1.35	1.06	2%
	SEP1	1.21	1.25	1.21	1.52	1.31	20%
	SEP2	0.88	1.09	1.14	1.50	1.17	11%
	PER	1.01	1.07	1.36	1.51	1.25	17%
	ARMA	1.11	1.32	1.30	1.65	1.36	24%
	SVM	1.03	1.16	1.34	1.68	1.33	21%
	HYB	0.89	**0.91**	**1.00**	**1.31**	**1.04**	
4	ASYM	1.27	1.45	**1.87**	3.94	2.37	7%
	SEP1	3.38	2.45	2.66	4.19	3.24	32%
	SEP2	1.41	1.59	1.99	3.80	2.40	8%
	PER	1.88	2.10	2.53	5.28	3.25	32%
	ARMA	2.07	1.77	2.14	3.81	2.58	14%
	SVM	1.71	1.74	2.15	3.89	2.54	13%
	HYB	**1.26**	**1.43**	1.92	**3.50**	**2.21**	

Table 2.4: MAE of wind speed forecasts. Missed the data for $h = 4$ and Period 2 in the original dataset. Bold-faced values indicate the best performance. The percentage improvements are the error inflation rate relative to HYB.

Period	Model	$h = 1$	$h = 2$	$h = 3$	$h = 4$	Average	% Improvement
1	ASYM	0.80	1.18	2.48	2.64	1.77	1%
	SEP1	1.06	1.45	2.72	2.81	2.01	12%
	SEP2	0.86	1.31	2.63	2.77	1.89	7%
	PER	1.05	1.49	2.80	2.90	2.06	15%
	ARMA	1.31	1.83	3.29	3.35	2.47	29%
	SVM	1.40	1.69	3.14	3.18	2.35	25%
	HYB	**0.79**	**1.17**	**2.46**	**2.62**	**1.76**	
2	ASYM	1.04	2.28	2.09	–	1.80	5%
	SEP1	2.19	2.80	2.30	–	2.43	30%
	SEP2	1.29	2.36	2.06	–	1.90	10%
	PER	1.49	2.55	2.27	–	2.10	19%
	ARMA	1.62	2.81	2.52	–	2.31	26%
	SVM	2.19	3.41	3.10	–	2.87	39%
	HYB	**0.89**	**2.19**	**2.04**	–	**1.71**	
3	ASYM	**0.72**	0.77	0.85	1.17	0.88	2%
	SEP1	0.99	1.04	1.01	1.30	1.08	21%
	SEP2	0.73	0.92	0.97	1.31	0.98	13%
	PER	0.81	0.84	1.05	1.20	0.98	13%
	ARMA	0.93	1.15	1.14	1.43	1.16	26%
	SVM	0.84	0.94	1.07	1.40	1.06	19%
	HYB	**0.72**	**0.74**	**0.83**	**1.13**	**0.85**	
4	ASYM	1.00	1.20	**1.48**	3.73	1.85	6%
	SEP1	2.68	1.99	2.13	3.75	2.64	34%
	SEP2	1.13	1.33	1.61	3.60	1.91	9%
	PER	1.49	1.71	2.06	4.78	2.51	31%
	ARMA	1.67	1.44	1.76	3.51	2.09	17%
	SVM	1.43	1.44	1.71	3.60	2.05	15%
	HYB	**0.99**	**1.18**	1.54	**3.26**	**1.74**	

Table 2.5: RMSE of wind power predictions. Bold-faced values indicate the best performance. The percentage improvements are the error inflation rate relative to HYB.

Period	h	ASYM	SEP1	SEP2	PER	ARMA	SVM	HYB
1	1	**0.10**	0.13	0.11	0.13	0.17	0.16	**0.10**
	2	**0.17**	0.21	0.19	0.20	0.24	0.23	**0.17**
	3	0.37	0.39	0.38	0.38	0.44	0.43	**0.36**
	4	**0.40**	0.42	0.42	0.42	0.48	0.45	**0.40**
	Av.	0.29	0.31	0.30	0.31	0.36	0.34	**0.28**
	%	1%	7%	5%	6%	19%	16%	
2	1	0.13	0.31	0.17	0.19	0.22	0.29	**0.12**
	2	0.25	0.36	0.27	0.30	0.33	0.43	**0.24**
	3	**0.24**	0.30	0.25	0.28	0.32	0.41	**0.24**
	4	–	–	–	–	–	–	–
	Av.	0.22	0.33	0.23	0.26	0.29	0.38	**0.20**
	%	4%	36%	11%	21%	29%	45%	
3	1	**0.10**	0.14	0.11	0.12	0.13	0.13	**0.10**
	2	0.12	0.15	0.14	0.12	0.16	0.14	**0.11**
	3	0.12	0.14	0.14	0.14	0.16	0.15	**0.11**
	4	0.17	0.19	0.19	0.18	0.21	0.20	**0.16**
	Av.	0.13	0.16	0.15	0.14	0.17	0.16	**0.12**
	%	3%	17%	13%	9%	22%	18%	
4	1	**0.17**	0.34	0.18	0.25	0.27	0.22	**0.17**
	2	0.19	0.31	0.20	0.28	0.23	0.23	**0.18**
	3	**0.24**	0.33	0.26	0.33	0.28	0.28	**0.24**
	4	0.42	0.46	0.39	0.62	0.41	0.42	**0.35**
	Av.	0.25	0.36	0.26	0.38	0.30	0.29	**0.23**
	%	8%	34%	9%	37%	20%	17%	

these findings, an enhanced procedure for short-term wind forecasting was devised and shown to outperform the commonly used forecast methods based on persistence, autoregressive, support vector machines, and separable spatio-temporal models.

3. THE CALIBRATED REGIME-SWITCHING METHOD FOR SHORT-TERM WIND SPEED AND POWER FORECASTING *

On a wind farm, local wind conditions exhibit sizeable variations at a fine temporal resolution. Existing regime-switching methods attempt to capture these variations by fitting statistical models that are conditioned on a finite set of regimes or states, thus producing regime-dependent forecasts. Most of these methods 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. We propose a statistical model called "*the calibrated regime-switching model*". The essence of the model is to introduce 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. In addition to regime-switching dynamics, the proposed model also accounts for salient physical features of wind fields: spatio-temporal dependencies, transport effect of wind, and nonstationarity. Using one year of turbine-specific wind data, the calibrated regime-switching method is shown to offer a wide margin of improvement over existing forecasting methods in terms of both wind speed and power.

3.1 Regime-switching modeling for wind forecasting

Regime-switching dynamics refer to the variation of the wind characteristics over a spectrum of physical states or regimes. In the past decade, there has been a growing recognition for *regime-switching* methods for short-term wind forecasting [38]. The essence of these regime-switching approaches is to fit statistical models that are conditioned on a finite set of wind regimes or states, and hence, produce regime-dependent forecasts.

A wide array of existing 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.

*Reprinted, with permission, from [26].

One such approach is the regime-switching autoregressive method in [88], which is presented in (1.8). Several regime-switching models have been proposed in the past decade, mostly sharing the same concept as the model in (1.8), but with key differences in how regimes are defined, whether using lagged values of wind speed, power, direction, precipitation, temperature, or other exogenous variables [38, 71, 40, 87, 11].

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, we propose a statistical forecasting model titled: "the calibrated regime-switching (CRS) model." 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 [5, 68, 3, 41, 10], 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. In the case study, the CRS approach is shown to have 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 3.4.

3.2 Preliminary data analysis

Preliminary analysis on the wind farm data described in Section 1.5 suggests that the data exhibit dependencies across space and time. The first five panels of Figure 3.1 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 [24, 51, 69]. The quick decay of temporal correlation hints to the irrelevance of longer-memory effects to short-term prediction.

The last panel of Figure 3.1 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

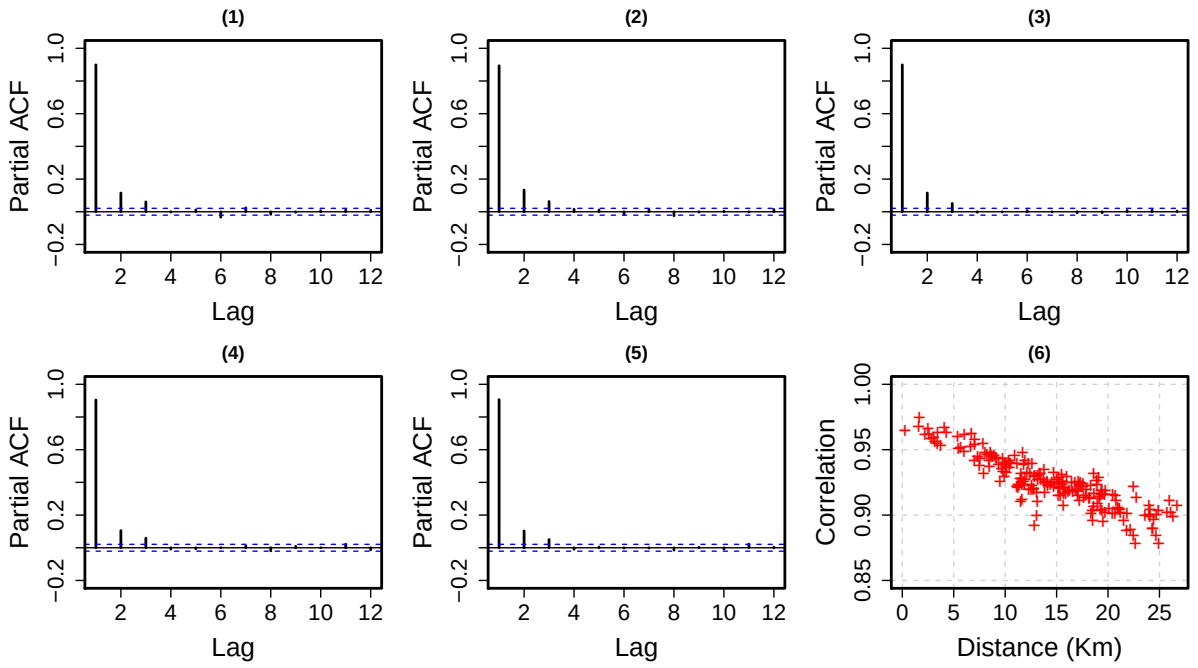


Figure 3.1: Panels (1)-(5): PACFs of the wind speeds at five wind turbines. The x-axis is the time lag in hours. Panel (6): Correlations between wind speeds at an arbitrarily chosen turbine and those at the remaining 199 turbines at zero-hour lag against the separating distances.

wind farm, ranging from 0.88 to 0.97. 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 [7, 71, 25], 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, two univariate change point tests are conducted on the first 30 days of spatially averaged data of wind speed and wind direction. For the wind speed variable, a change point detection test with binary segmentation is implemented using the package `changepoint` in `R` [52]. For the wind direction, which is a circular variable, a binary segmented version of the circular change point detection in [47] is

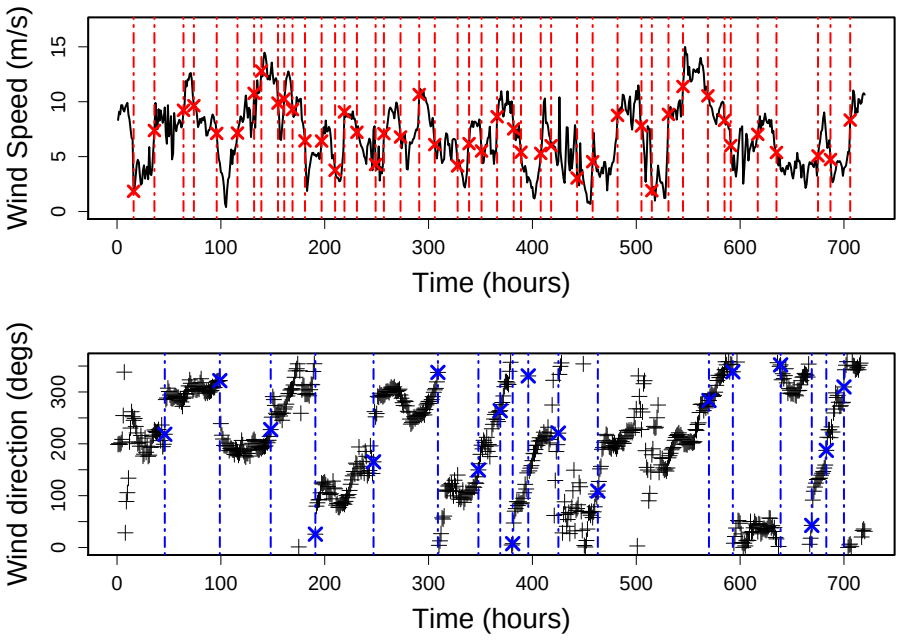


Figure 3.2: Top panel: change points in one month of spatially averaged wind speed data. Bottom panel: change points in one month of spatially averaged wind direction data. The span of the x-axis is a month or 720 hours.

implemented using the package `circular` [2]. Figure 3.2 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, 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 [66]. The system's short memory noted in the PACFs of Figure 3.1 is perhaps an implication of this dynamic behavior.

3.3 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. Similar to the notation established in Section 1 of this dissertation, 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.

3.3.1 Regime identification

There are several ways to define regimes in the wind simulation and forecasting literature [51, 50]. Clustering-based methods can be run on pre-specified meteorologic variables to identify regimes corresponding to physically interpretable weather patterns [11]. In other instances, regimes can be defined implicitly through latent variables such as in hidden Markov models [5]. 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, two wind regimes were proposed in [38] based on wind direction to reflect the alternation of westerly and easterly winds in the U.S. Pacific Northwest [38].

Here, we follow an approach similar to [38], 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 3.4.2, 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 3.3 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 3.3 is estimated by using the binning method [46], a common industrial practice. The binning estimates are shown in Figure 3.3 as the red triangles.

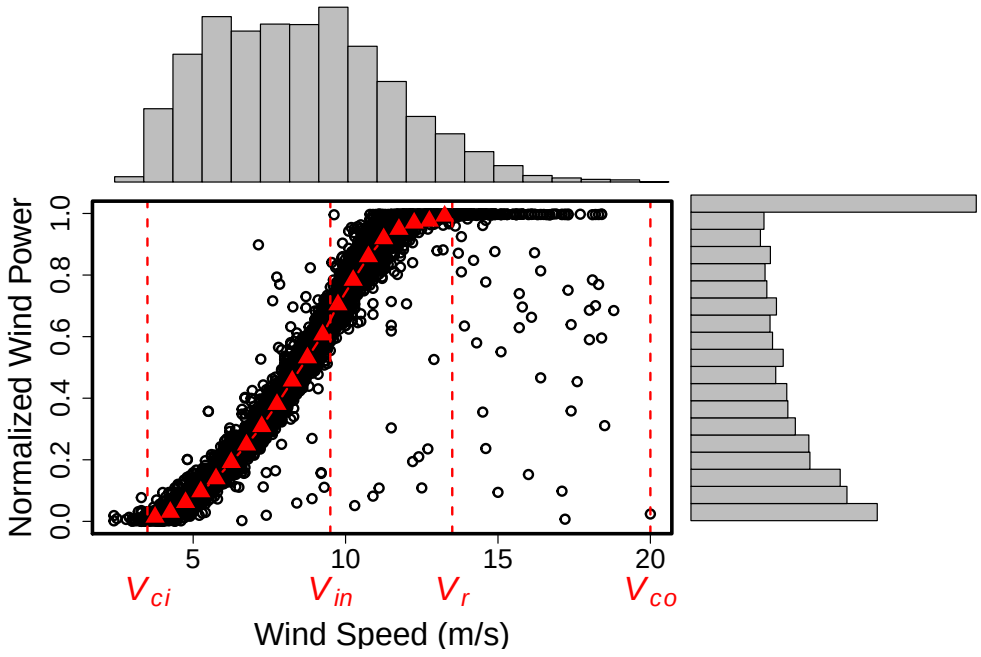


Figure 3.3: Normalized power versus speed values, histograms of wind speed and power, for one of the turbines on the wind farm. V_{ci}: cut-in speed, V_{in}: inflection point, V_r: rated speed and V_{co}: cut-out speed.

Four physically meaningful wind speed values define 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 $20\,\mathrm{m/s}$, respectively, whereas V_{in} is estimated by [44] for modern turbines to be around $9.5\,\mathrm{m/s}$.

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 $20\,\mathrm{m/s}$ for our data, V_{co} most commonly takes the value of $25\,\mathrm{m/s}$ in commercial turbines. Extending V_{co} from $20\,\mathrm{m/s}$ to $25\,\mathrm{m/s}$, however, does not affect our wind regime definition.

With respect to wind direction, the left panel of Figure 3.4 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). The right panel of Figure 3.4 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 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 3.5 shows the time series of the first 30 days of $\bar{Y}(j)$, $d(j)$, and $r(j)$. In Section

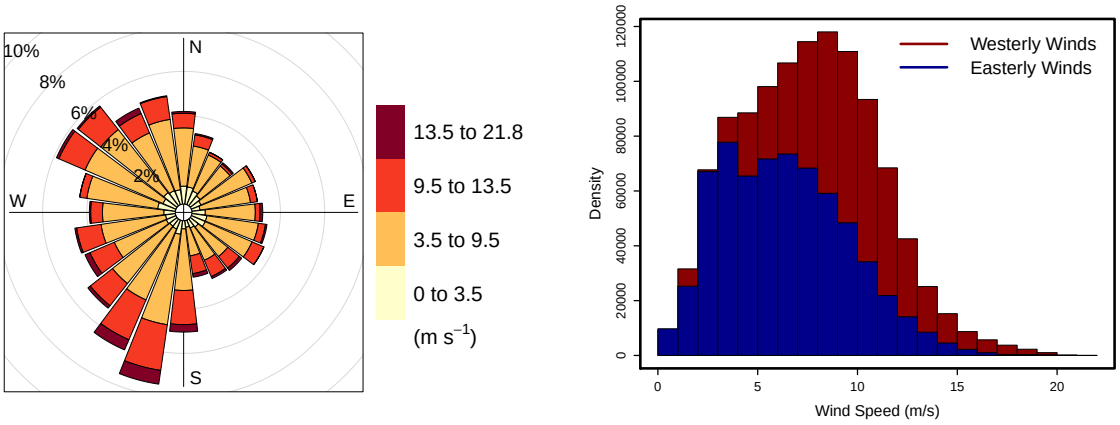


Figure 3.4: Left panel: Rose plot of spatially averaged wind speeds and directions. Right panel: Histogram of wind speeds for westerly versus easterly winds.

3.4, 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.

$$r(j) = \begin{cases} 1, \ \bar{Y}(j) \in [0, 3.5) & \& \ d(j) \in [0°, 180°) \Rightarrow \text{Low speed, easterly wind} \\ 2, \ \bar{Y}(j) \in [0, 3.5) & \& \ d(j) \in [180°, 360°) \Rightarrow \text{Low speed, westerly wind} \\ 3, \ \bar{Y}(j) \in [3.5, 9.5) & \& \ d(j) \in [0°, 180°) \Rightarrow \text{Moderate speed, easterly wind} \\ 4, \ \bar{Y}(j) \in [3.5, 9.5) & \& \ d(j) \in [180°, 360°) \Rightarrow \text{Moderate speed, westerly wind} \\ 5, \ \bar{Y}(j) \in [9.5, 20 \text{ or } 25) & \& \ d(j) \in [0°, 180°) \Rightarrow \text{High speed, easterly wind} \\ 6, \ \bar{Y}(j) \in [9.5, 20 \text{ or } 25) & \& \ d(j) \in [180°, 360°) \Rightarrow \text{High speed, westerly wind} \end{cases} \tag{3.1}$$

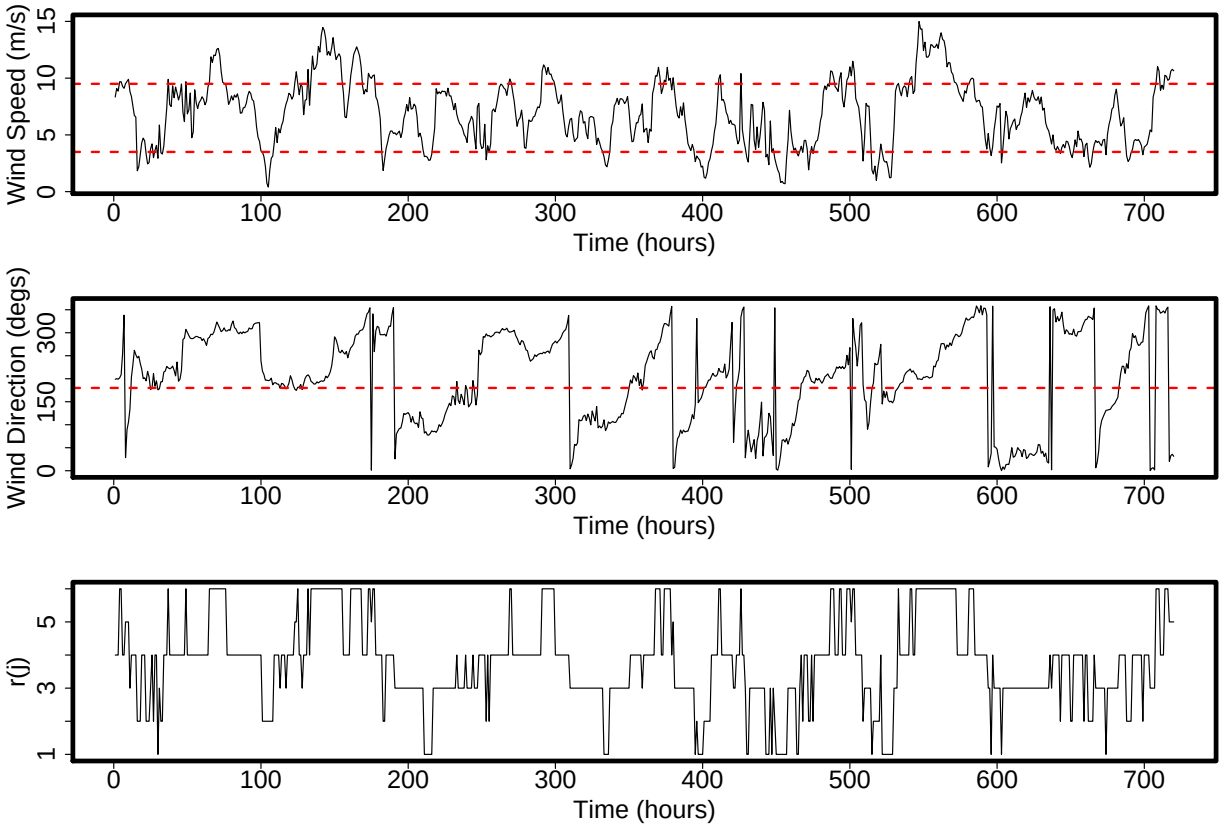


Figure 3.5: Top panel: Spatially averaged wind speed data (solid line), along with the wind speed regime thresholds at 3.5 and 9.5 m/s (dashed line). Middle panel: spatially averaged wind direction data (solid line), along with the wind direction regime threshold value at 180° (dashed line). Bottom panel: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

3.3.2 The CRS method

We assume that we have at hand a statistical model, denoted by $\mathcal{M}$. In the CRS formulation, $\mathcal{M}$ is a reactive regime-switching model, which, conditioned on $r(t)$, can produce regime-dependent spatio-temporal forecasts, where $\hat{Y}_i^{r(t)}(t+h)$ 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 3.4, 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 $\hat{Y}_i^{r(t)}(t+h)$. Hereinafter, we refer to $c_i(t+h|r(t))$ as the regime-dependent *forecast calibration*, and the quantity $\tilde{Y}_i(t+h) = \hat{Y}_i^{r(t)}(t+h) + c_i(t+h|r(t))$ as the *calibrated forecast*. The essence of the CRS approach can be better illustrated in Figure 3.6, where the goal of the calibration is to learn how and when $Y_i(t)$ is going to change, and adjust the forecast at $t+h$ accordingly.

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

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

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

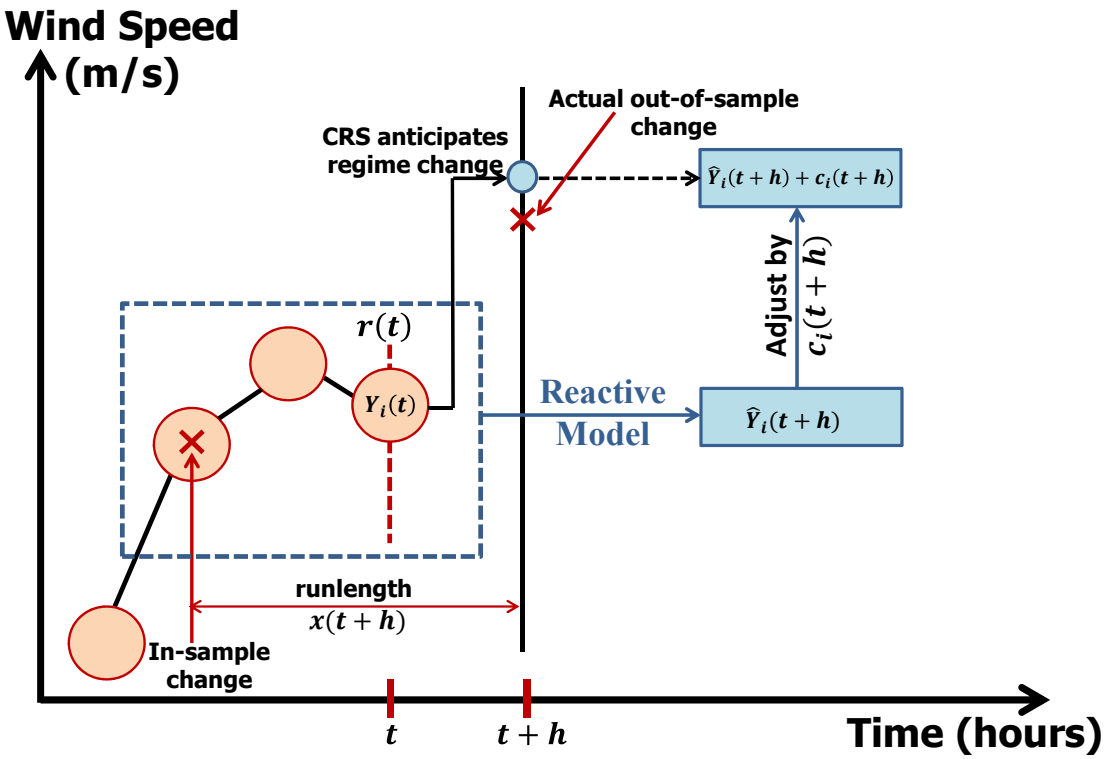


Figure 3.6: Illustration of the forecasting calibration for out-of-sample changes.

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 [1, 72], 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 distribution 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 Ψ^k. The parameters in Ψ^k are estimated from the data and continuously updated in a rolling forward forecasting scheme. Other selections of calibration functions are discussed in Section 3.4.

The estimation procedure of parameters in Ψ^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 ℓ-th roll, the data observed up to time t^ℓ 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 ℓ-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 $t+h$, $\hat{Y}_i^{r^\ell(t)}(t+h)$, 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),

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

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 $\{\Psi^k\}_{k=1}^R$.

3.3.3 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 Θ^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 Θ^k using a subset of historical data and obtain raw forecasts from $t+1$ till $t+H$. As such, the parameters in Θ^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 Ψ^k.

Afterwards, we proceed to Phase III, where the same rolling mechanism is performed. Specifically, at the present time t, we estimate Θ^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 Ψ^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 Ψ^k by re-solving (3.4), given the newly revealed observations. As such, similar to Θ^k, the parameters Ψ^k are also regime-specific and time-varying. The cycle is repeated until the forecasts for the remaining 11 months are produced. Figure 3.7 provides additional details for understanding the implementation procedure.

3.4 Case Study: application to data from an onshore wind farm

We conduct a case study of the CRS method using the wind farm data explained in Section 3.2. 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 [66, 82].

3.4.1 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 a similar form to what was presented in (1.1), where the entries of the covariance matrix $\mathbf{\Sigma}$ are mod-

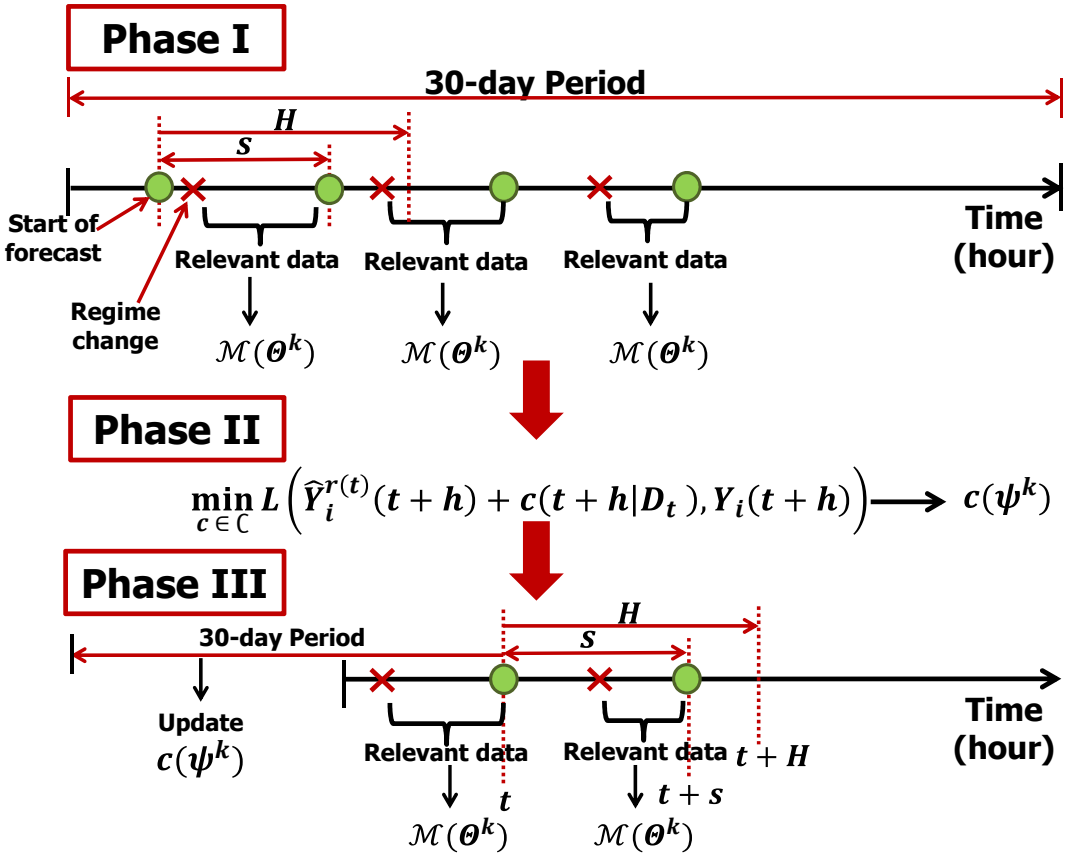


Figure 3.7: Steps and notations in the implementation of the CRS approach. For this research, $s = 6$ hours is the rolling horizon, while $H = 12$ is the forecasting horizon.

eled 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" [33, 78, 36, 49]. Asymmetry implies that a dominant wind direction causes a discrepancy between a stronger, along-wind spatio-temporal correlation and a weaker, span-wind correlation. Recently, it has been shown that spatio-temporal asymmetry in local wind fields exist and is flow dependent [25]. To account for possible flow dependent asymmetries and following [36] and [25], we model $K(\cdot, \cdot)$ as a convex combination of two components: a fully symmetric nonseparable model, denoted by C_1, and an asymmetric model, denoted by C_2, as in (3.5).

$$K(\mathbf{u}, w) = \sigma^2\big[(1-\lambda)C_1(\mathbf{u}, w) + \lambda C_2(\mathbf{u}, w)\big] + \eta \mathbb{1}_{\{||\mathbf{u}||=|w|=0\}}, \tag{3.5}$$

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 nonseparable model proposed by [33]:

$$C_1(\mathbf{u}, w) = \frac{1-\delta}{1+\alpha|w|^2}\left[\exp\left(-\frac{c||\mathbf{u}||}{(1+\alpha|w|^2)^{\frac{\beta}{2}}}\right) + \frac{\delta}{1-\delta}\mathbb{1}_{\{||\mathbf{u}||=0\}}\right], \tag{3.6}$$

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:

$$C_2(\mathbf{u}, w) = \mathbb{E}_{\mathbf{V}}\zeta(||\mathbf{u} - \mathbf{V}w||), \tag{3.7}$$

where $\mathbf{V}$ is a random vector in $\mathbb{R}^2$ and $\zeta(\cdot)$ is a valid spatial covariance function [33, 36].

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 [75], 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 [36]. As such, the model in (3.7) is re-written as in (3.8):

$$C_2(\mathbf{u}, w) = \frac{1}{\sqrt{|\mathbf{1}_{2\times 2} + w^2\mathbf{D}|}} \exp\big[-(\mathbf{u} - w\boldsymbol{\mu})^T(\mathbf{1}_{2\times 2} + w^2\mathbf{D})^{-1}(\mathbf{u} - w\boldsymbol{\mu})\big], \tag{3.8}$$

where $|\cdot|$ in (3.8) denotes the matrix determinant.

To produce regime-dependent forecasts, we estimate the parameters in (3.5) 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 3.3.3, we naturally overcome temporal nonstationarity [29, 20, 69], 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 (3.5) 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 3.1. We also impose a minimum of 2 time lags in history to ensure a reliable estimation of the parameters in (3.5). 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 [29]. We define three sub-regions of wind turbines based on their proximity to the three masts, as shown in Figure 1.2. Within each sub-region, we fit the stationary spatio-temporal model of (3.5) 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 (3.5), we can obtain spatio-temporal ordinary kriging-based predictions.

3.4.2 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 $180^\circ - 45^\circ$ to $180^\circ + 45^\circ$ with 15° increment, and set $r_1^d = 360^\circ - r_2^d$. 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 $\{45^\circ, 225^\circ\}$ for wind direction, resulting in $R = 6$ regimes.

Figure 3.8 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). 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 3.3.3, we allow these functions to change with time by continuously re-estimating Ψ^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 $h = 1, 2$.

3.4.3 Forecasting results

The rolling mechanism described in Section 3.3.3 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 (3.5). Similar to ASYM is SEP, except that for that model, we set $\beta = \lambda = 0$ in (3.5) 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

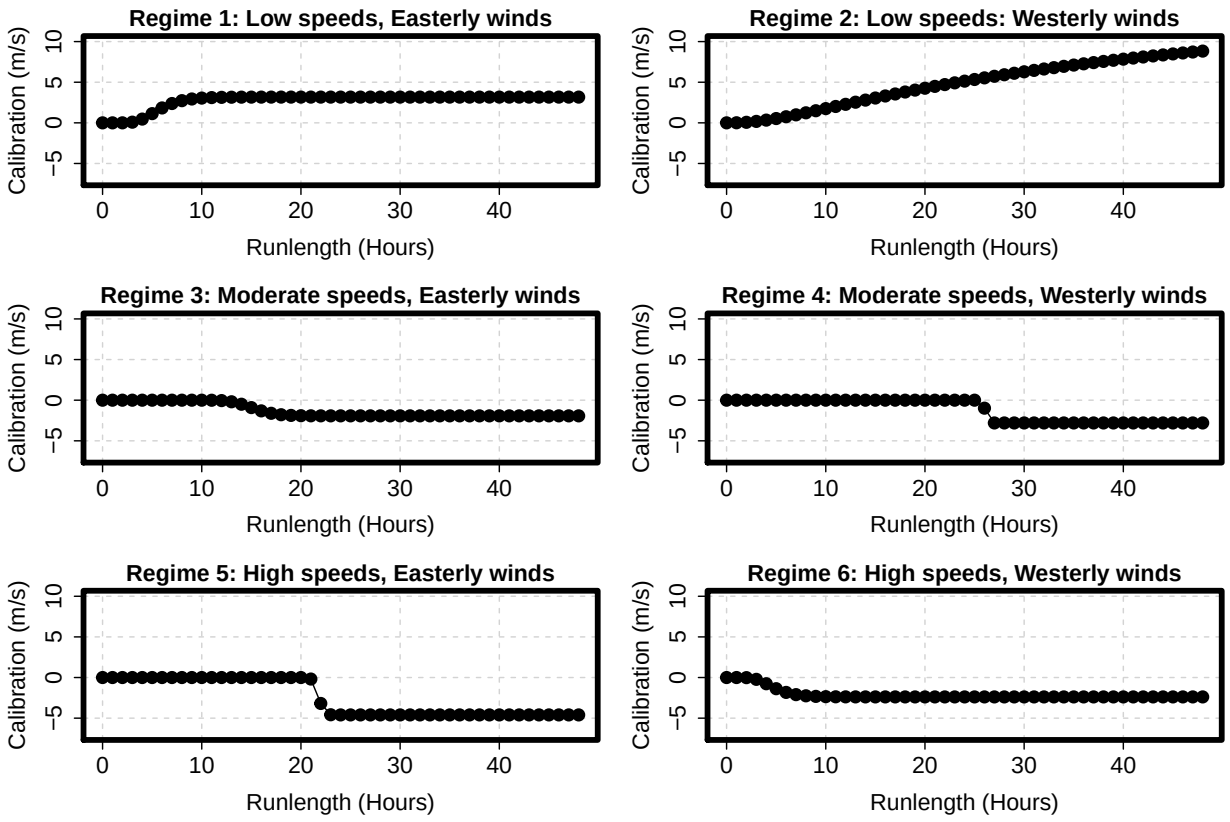


Figure 3.8: Estimated calibration functions using Phase I data for the six regimes.

in RSAR are low-order autoregressive models like AR(1). Low-order AR models are common choices in the wind forecasting literature [43, 69]. 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: $\hat{Y}_i(t+h) = \sum_{k=1}^{R} \hat{Y}_i^k(t+h) Pr(r(t+h) = k)$, where $\hat{Y}_i^k(t+h)$ 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 [69], 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 3.9 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.

We evaluate the overall forecast accuracy over all testing instances using the Mean Absolute Error (MAE), as expressed in (1.11). We note that MAE as a loss measure suggests the use of the

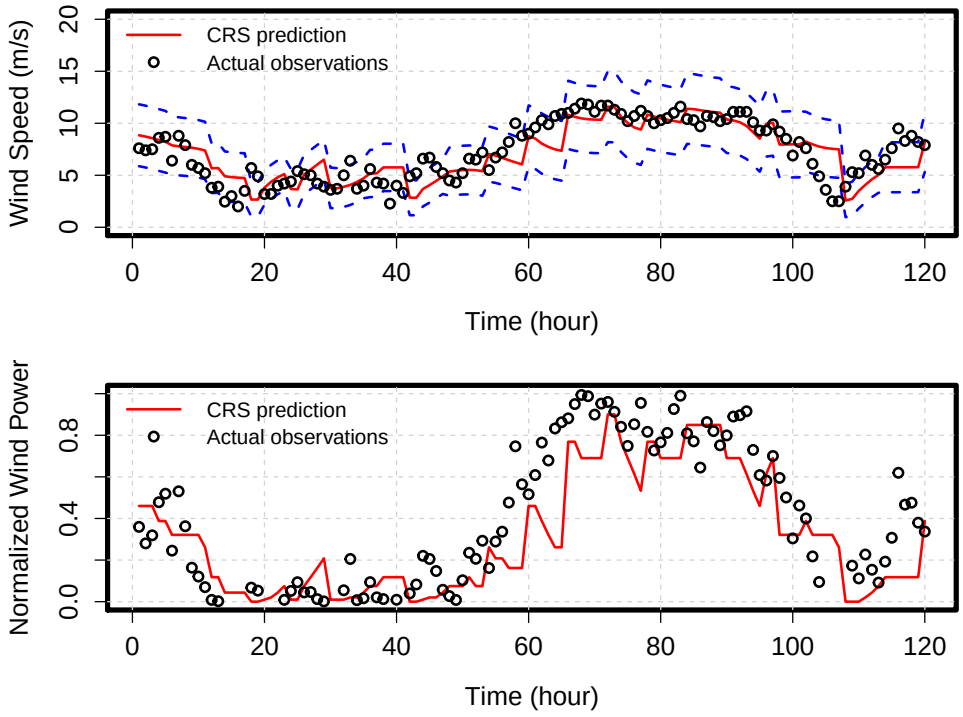


Figure 3.9: Top panel: 6-hour ahead forecasts of wind speeds at the chosen turbine for five days starting on November 23, 2010. Point forecasts shown by the solid red line, along with ± 1 standard deviation as the dashed blue lines. Actual observations shown as black circles. Similarly, the bottom panel shows the corresponding power forecasts and the actual observations.

medians of the predictive distributions as optimal point forecasts [34]. Since the mean and median coincide under a Gaussian predictive distribution, the point forecasts used in (1.11) 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 (1.11). The resulting MAE values computed using the 11-month of testing data are presented in Table 3.1.

In addition to MAE, we also evaluate the wind power forecasts using the PCE loss measure introduced in (1.12). Under the loss function of (1.12), it has been shown that the g-th quantile of the predictive distribution is an optimal point forecast [35]. For ASYM, SEP, CRS and RSAR, the g-th quantile is directly computed from the resulting Gaussian predictive distribution and is used as

Table 3.1: 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

the input point forecast to the PCE loss function in (1.12). 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 (1.12). In Table 3.2, 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 $g = 0.73$, which is the value suggested in [67]. We stress that when computing MAE in (1.11) and PCE in (1.12) for the CRS approach, $\hat{Y}_i(t+h)$ is the calibrated forecast, i.e. $\tilde{Y}_i(t+h)$.

Table 3.1 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 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

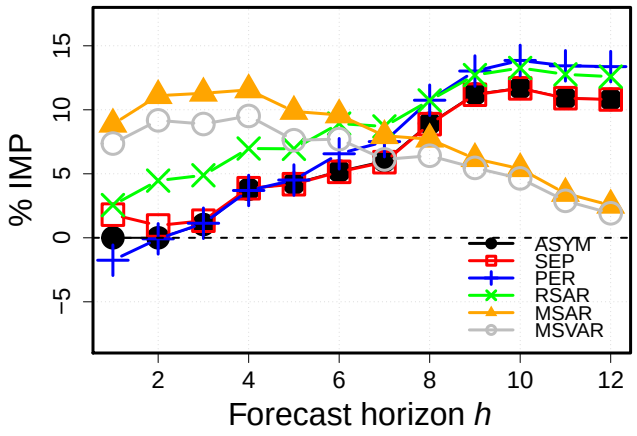


Figure 3.10: Percentage improvements in MAE of CRS over competing approaches in wind speed.

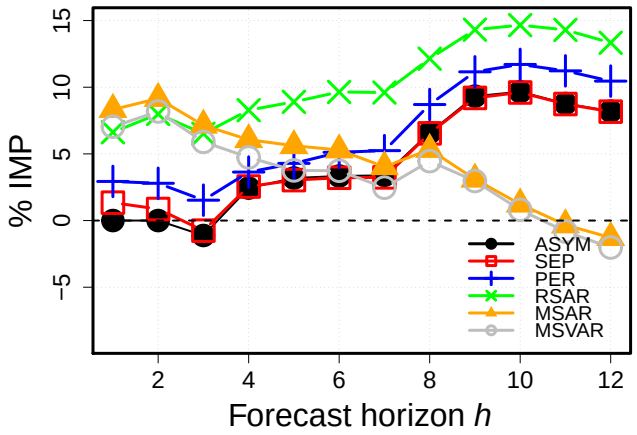


Figure 3.11: Percentage improvements in MAE of CRS over competing approaches in wind power.

competitive performance.

Figure 3.10 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 3.12 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 $\hat{Y}_i(t+h)$ cross the same speed threshold, we label that as a true positive; while on the other hand, when neither $Y_i(t+h)$ nor $\hat{Y}_i(t+h)$ crosses any speed threshold, we label it as a true negative. In the left panel of Figure 3.12, we plot the true positive rate (TPR) of CRS versus MSVAR for $h = 3, 4, \ldots, 12$. Similarly, the right panel of Figure 3.12 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 3.10. 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 3.1; the CRS approach outperforms the competing models for power prediction for most forecasting horizons. Its

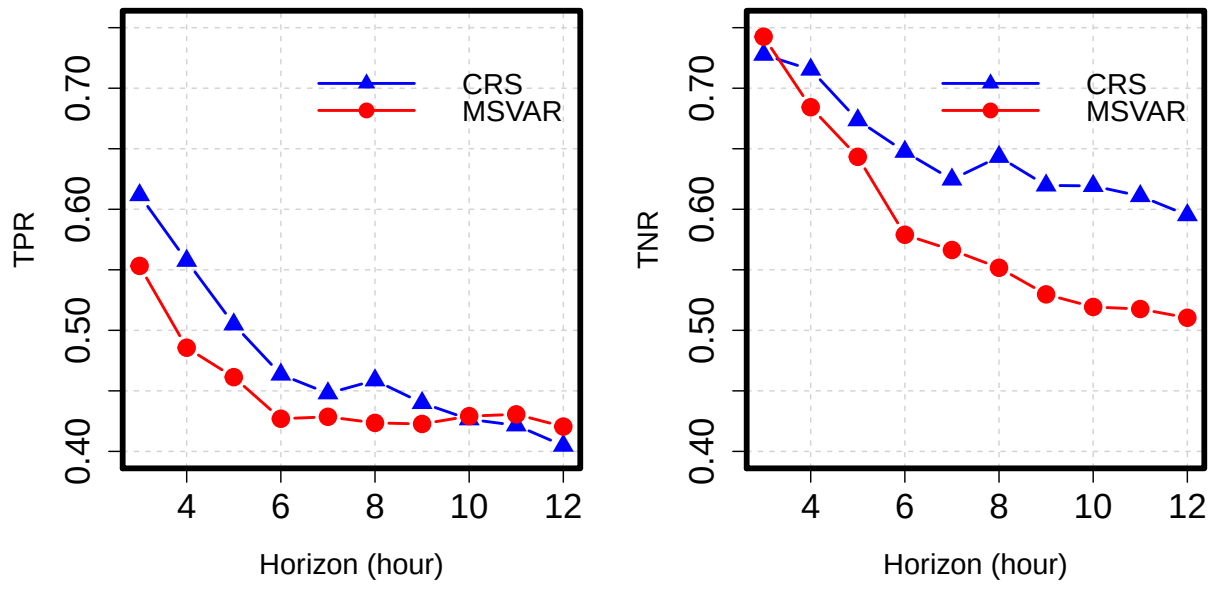


Figure 3.12: Left panel: the true positive rate (TPR). Right panel: the true negative rate (TNR). Comparisons are between CRS (blue triangles) and MSVAR (red circles) for $h = 3, ..., 12$.

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 wind power forecast, the performance of MS approaches in the end surpass that of CRS at $h = 11$. The percentage improvements shown in Figure 3.11 are somewhat different from their counterparts in Figure 3.10; the difference is mainly due to the nonlinear speed-to-power conversion.

In Table 3.2, 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

Table 3.2: Average PCE values across all horizons. Bold-faced values indicate best performance. * corresponds to the value suggested in [67].

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**

of view. Similar to [40], we implement the large sample test first introduced by [21] 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 3.13 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 3.13, are supposed to be one; these p-values are trivial and thus not shown.

3.4.4 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 [66, 37]. 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 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.

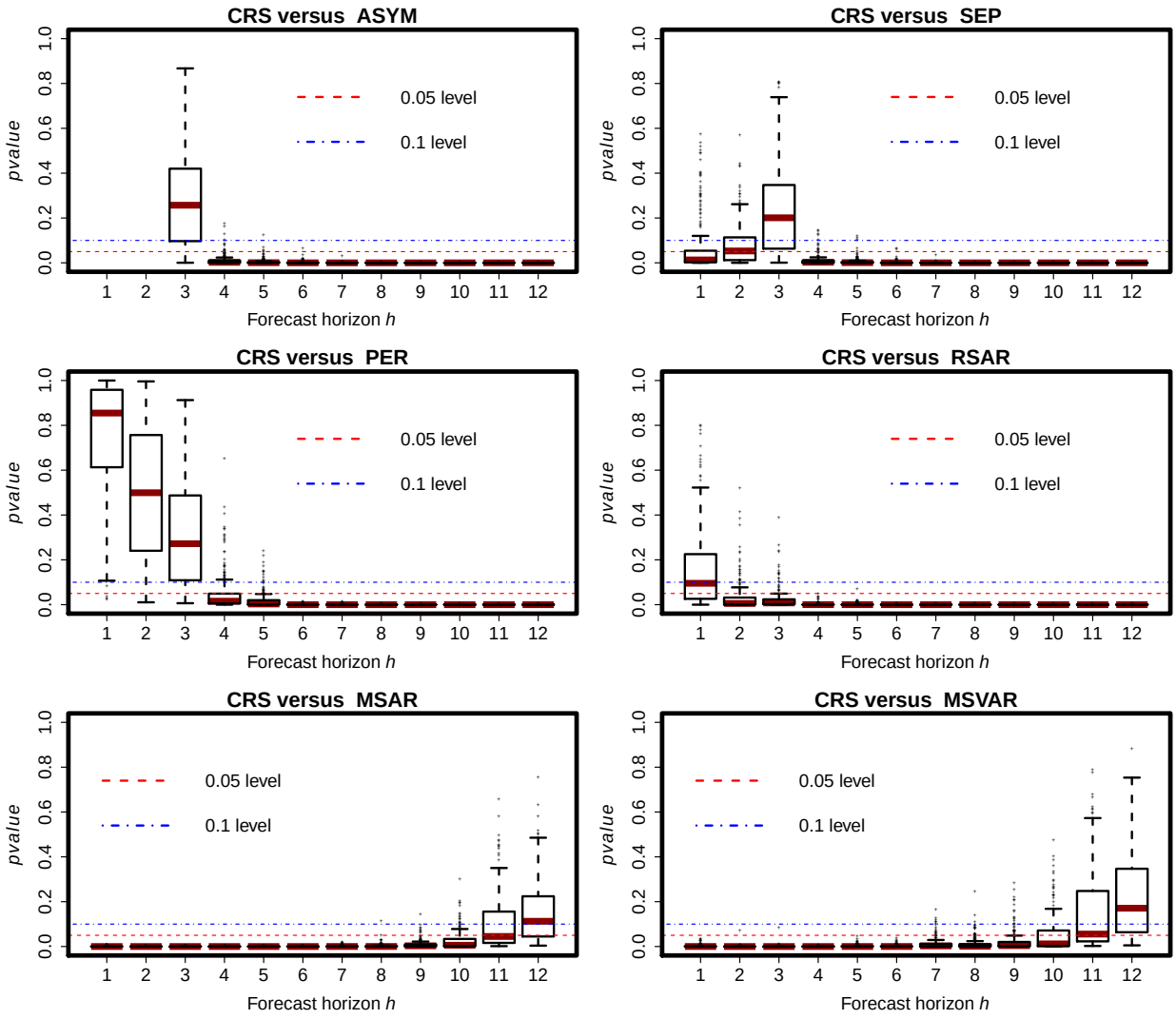


Figure 3.13: Boxplots of p-values generated from conducting the one-sided test of [21] to compare the turbine-specific CRS forecasts against those from the competing models at different horizons.

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 [38, 69]. 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 3.14 the forecast at one of the turbines from an arbitrarily selected forecasting roll. In the left panel, it shows that the calibrated forecasts are able to pick up an out-of-sample change in wind speed, in contrast to the reactive ASYM model, which fails to do so, as it solely relies on extrapolating 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.

A standard evaluation measure in probabilistic forecasting is the continuous ranked probability score (CRPS). In Table 3.3, 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.

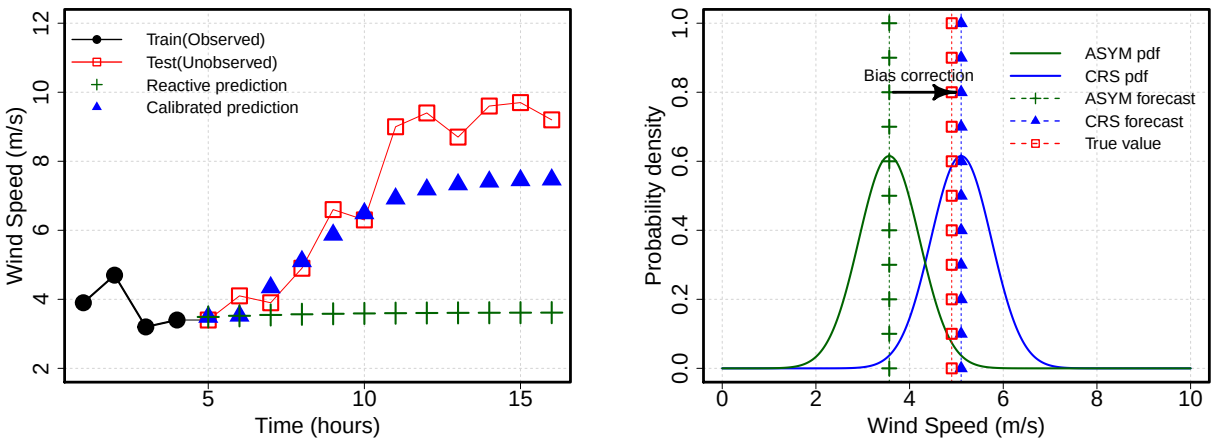


Figure 3.14: Left panel: CRS versus ASYM predictions for one turbine from an arbitrarily selected forecasting roll. Calibrated predictions shown without imposing the bounds discussed in Section 3.4.2. Right panel: Predictive distributions for the point forecast at $h = 4$. Absolute errors of point predictions at $h = 4$ are, respectively, 0.20 and 1.33 for CRS and ASYM.

Table 3.3: 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**

3.5 Conclusion

We have proposed 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 research 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.

4. SUMMARY AND CONCLUSIONS

In this Section, the overall contributions of this dissertation are summarized. Potential future extensions beyond the scope of this dissertation are also outlined.

4.1 Summary

The overarching theme of this dissertation is to develop a set of useful analytics, statistical models and tools tailored for wind energy applications. Guided by the physics of wind, two sets of physics-informed statistical models and analytics are proposed to capture salient physical properties of wind fields, namely: stochasticity, spatial and temporal dependencies, transport effect of wind, and regime-switching dynamics. These methods are used to make short-term, turbine-specific forecasts of the wind speed and wind power. The forecasts are evaluated using common performance metrics in the wind forecasting literature against a wide set of existing approaches. The value of these spatio-temporal forecasts stems from their importance for several operational decision-making problems at the turbine, farm, and grid levels, as outlined in Section 1.

At the heart of all of these methods, the wind field on a farm is modeled as a stochastic spatio-temporal process. Two physical features unique to local wind fields are central to the analytics and models proposed in this dissertation—the transport effect of wind and the regime-switching dynamics, which are addressed in Sections 2 and 3 of this dissertation, respectively.

The transport effect refers to a physical property unique to wind and other environmental processes in which the process under study has a prevailing flow over time in a certain dominant direction. In Section 2 of this dissertation, a statistical tool, called "*the spatio-temporal lens,*" is proposed to unearth the influence of this physical property on the generation of complex spatio-temporal correlations, interactions and patterns. The implementation of this statistical tool on a yearlong operational wind farm data suggests that the assumptions of separability and symmetry, commonly adopted in the literature, have to be avoided when modeling local wind fields. In other words, strong signs of nonseparability and asymmetry have been detected in the wind farm data,

specifically attributed to the transport effect of wind. Motivated by these findings, a physically justifiable nonseparable asymmetric model is proposed, which takes into account the transport effect of wind by characterizing the spatial and temporal dependence of local wind fields in tandem, rather than separately as traditionally performed in the literature. The model demonstrated promising improvements in predictive accuracy relative to a set of existing wind forecasting approaches.

Section 3 addresses another physical feature of wind fields: regime-switching dynamics, which refer to the fast-changing variations of the wind characteristics over a spectrum of physical states or regimes. The ability to account for these fast-changing variations are of critical importance to wind farm and power grid operators, as these variations in the wind characteristics often correspond to severe ramps in the associated power generation levels. These sudden ramps represent a serious difficulty to wind farm practitioners and power system operators in their operational activities such as determining hourly market participation, dispatch, balancing and reserve quantification.

In light of that, an advanced statistical model called "*the calibrated regime-switching (CRS) model*" is proposed. The model is intended to take into account in- and out-of-sample changes in the wind behavior. The essence of the proposed model is to introduce an action of calibration to a base spatio-temporal model to assist in anticipating potential out-of-sample changes. The calibration is modeled as a parametric function of two indicators that are shown to be able predictors of out-of-sample variations: the observed wind regime at the time of forecast, as well as the runlength, which is the time elapsed since the most recent regime change.

In Section 3, the CRS model is tested, in a rolling forward forecasting fashion, on a massive validation set of more than $3,000,000$ data instances, corresponding to 11 months of data. The proposed model is shown to offer double-digit, statistically significant improvements in wind speed and power forecasting relative to 6 existing approaches, ranging between time-series, to spatio-temporal, to regime-switching models. The adoption of the model in a probabilistic forecasting framework is also discussed.

4.2 Future work

Wind is a complex physical process, and up till now, the research is ongoing to understand the complex physical properties of wind at a multitude of spatial and temporal scales. In this dissertation, we have sought a balance between a purely data-driven, "Blackbox", approach on one hand, and a purely physics-based approach on the other. By developing statistical, yet physics-guided models, substantial improvements over purely data-driven approaches have been demonstrated.

Two closely related questions remain to be addressed: (i) *How can the models developed in this dissertation complement the existing physics-based approaches?*, (ii) *How to incorporate additional physical features in our models?* An answer to the first question would yield an important insight to the research community, as well as the wind industry, regarding the situations in which statistical models outperform physics-based models, and vice-versa, and when a combination of both approaches would be desirable. This issue, as outlined in Section 1 of this dissertation, is not yet settled in the literature, despite the general consensus that statistical approaches are often favorable at finer spatial and temporal resolutions then their physics-based counterparts.

For the latter question, a combination of expertise in atmospheric sciences, together with in-depth knowledge of statistical modeling, is required to unravel intrinsic relationships between different environmental variables. A better understanding of the effect of such environmental variables can enrich our ability to develop models that can leverage the rich measurements about related environmental processes to benefit the forecasts about the wind speed and wind power. The adoption of a multivariate statistical framework, which not only characterizes the within-process dependencies, but also the cross-process correlations and interactions, may play a critical role in predicting the wind resource at a multitude of spatial and temporal scales.

There are also some specific future directions related to the CRS approach which 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 3.3.2, is also a matter of promising future research.

Finally, it is worth noting that a lot of the ideas proposed in this dissertation can be exported—with necessary domain-specific adaptations—to other areas of application. For instance, applications in which an environmental dynamic process is involved, like water quality prediction, solar forecasting and others, can benefit from the ideas and methods proposed in this dissertation.

REFERENCES

[1] ADAMS, R. P., AND MACKAY, D. J. Bayesian online changepoint detection. *arXiv: 0710.3742* (2007).

[2] AGOSTINELLI, C., AND LUND, U. R package circular: circular statistics (version 0.4-7). https://r-forge.r-project.org/projects/circular.

[3] AILLIOT, P., BESSAC, J., MONBET, V., AND PENE, F. Non-homogeneous hidden markov-switching models for wind time series. *Journal of Statistical Planning and Inference 160* (2015), 75–88.

[4] AILLIOT, P., AND MONBET, V. Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software 30* (2012), 92–101.

[5] AILLIOT, P., MONBET, V., AND PREVOSTO, M. An autoregressive model with time-varying coefficients for wind fields. *Environmetrics 17*, 2 (2006), 107–117.

[6] ALEXIADIS, M., DOKOPOULOS, P., AND SAHSAMANOGLOU, H. Wind speed and power forecasting based on spatial correlation models. *IEEE Transactions on Energy Conversion 14*, 3 (1999), 836–842.

[7] ALEXIADIS, M., DOKOPOULOS, P., SAHSAMANOGLOU, H., AND MANOUSARIDIS, I. Short-term forecasting of wind speed and related electrical power. *Solar Energy 63*, 1 (1998), 61–68.

[8] AMJADY, N., KEYNIA, F., AND ZAREIPOUR, H. Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization. *IEEE Transactions on Sustainable Energy 2*, 3 (2011), 265–276.

[9] BESSA, R. J., MIRANDA, V., BOTTERUD, A., WANG, J., AND CONSTANTINESCU, E. M. Time adaptive conditional kernel density estimation for wind power forecasting. *IEEE Transactions on Sustainable Energy 3*, 4 (2012), 660–669.

[10] BESSAC, J., AILLIOT, P., CATTIAUX, J., AND MONBET, V. 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 (2016), 1–16.

[11] BROWELL, J., DREW, D., AND PHILIPPOPOULOS, K. Improved very short-term spatio-temporal wind forecasting using atmospheric regimes. *Wind Energy 21*, 11 (2018), 968–979.

[12] BROWN, B. G., KATZ, R., AND MURPHY, A. Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology 23*, 8 (1984), 1184–1195.

[13] BYON, E., AND DING, Y. Season-dependent condition-based maintenance for a wind turbine using a partially observed Markov decision process. *IEEE Transactions on Power Systems 25* (2010), 1823–1834.

[14] BYON, E., NTAIMO, L., AND DING, Y. Optimal maintenance strategies for wind turbine systems under stochastic weather conditions. *IEEE Transactions on Reliability 59*, 2 (2010), 393–404.

[15] CADENAS, E., AND RIVERA, W. Wind speed forecasting in three different regions of mexico, using a hybrid ARIMA-ANN model. *Renewable Energy 35*, 12 (2010), 2732–2738.

[16] CRESPO, A., HERNÁNDEZ, J., AND FRANDSEN, S. Survey of modelling methods for wind turbine wakes and wind farms. *Wind Energy 2*, 1 (1999), 1–24.

[17] CRESSIE, N., BURDEN, S., DAVIS, W., KRIVITSKY, P., MOKHTARIAN, P., SUESSE, T., AND ZAMMIT-MANGION, A. Capturing multivariate spatial dependence: model, estimate and then predict. *Statistical Science 30*, 2 (2015), 170–175.

[18] CRESSIE, N., AND HUANG, H. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association 94*, 448 (1999), 1330–1340.

[19] CRESSIE, N., AND WIKLE, C. *Statistics for Spatio-temporal Data*. John Wiley & Sons, 2011.

[20] DE LUNA, X., AND GENTON, M. G. Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica 15* (2005), 547–568.

[21] DIEBOLD, F., AND MARIANO, R. Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*, 3 (1995), 253–263.

[22] DOE. Wind vision: A new era for wind power in the United States. U.S. Department of Energy, Washington D.C.. Available at: https://www.energy.gov/eere/wind/maps/wind-vision.

[23] DOWELL, J., AND PINSON, P. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid 7*, 2 (2016), 763–770.

[24] ERDEM, E., AND SHI, J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy 88*, 4 (2011), 1405–1414.

[25] EZZAT, A. A., JUN, M., AND DING, Y. Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting. *IEEE Transactions on Sustainable Energy 9*, 3 (2018), 1437–1447.

[26] EZZAT, A. A., JUN, M., AND DING, Y. Spatio-temporal short-term wind forecast: a calibrated regime-switching method. *The Annals of Applied Statistics* (2019). Accepted. To appear.

[27] FRIEDMAN, J., HASTIE, T., AND TIBSHIRANI, R. *The Elements of Statistical Learning*, vol. 1. Springer, New York, 2001.

[28] FRYZLEWICZ, P. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*, 6 (2014), 2243–2281.

[29] FUENTES, M. A high frequency kriging approach for non-stationary environmental processes. *Environmetrics 12*, 5 (2001), 469–483.

[30] FUENTES, M. Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference 136*, 2 (2006), 447–466.

[31] GIEBEL, G., BROWNSWORD, R., KARINIOTAKIS, G., DENHARD, M., AND DRAXL, C. The state-of-the-art in short-term prediction of wind power: A literature overview, 2nd edition. Tech. rep., ANEMOS.plus, 2011.

[32] GIRARD, R., AND ALLARD, D. Spatio-temporal propagation of wind power prediction errors. *Wind Energy 16*, 7 (2013), 999–1012.

[33] GNEITING, T. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association 97*, 458 (2002), 590–600.

[34] GNEITING, T. Making and evaluating point forecasts. *Journal of the American Statistical Association 106*, 494 (2011), 746–762.

[35] GNEITING, T. Quantiles as optimal point forecasts. *International Journal of Forecasting 27*, 2 (2011), 197–207.

[36] GNEITING, T., GENTON, M., AND GUTTORP, P. Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistical Methods for Spatio-Temporal Systems*, B. Finkenstadt, L. Held, and V. Isham, Eds. Chapman & Hall/CRC, 2007, ch. 4.

[37] GNEITING, T., AND KATZFUSS, M. Probabilistic forecasting. *The Annual Review of Statistics and its Application 1* (2014), 125–151.

[38] GNEITING, T., LARSON, K., WESTRICK, K., GENTON, M., AND ALDRICH, E. Calibrated probabilistic forecasting at the stateline wind energy center. *Journal of the American Statistical Association 101*, 475 (2006), 968–979.

[39] HASLETT, J., AND RAFTERY, A. E. Space-time modelling with long-memory dependence: Assessing ireland's wind power resource. *Journal of the Royal Statistical Society, Series C (Applied Statistics) 38*, 1 (1989), 1–50.

[40] HERING, A., AND GENTON, M. Powering up with space-time wind forecasting. *Journal of the American Statistical Association 105* (2010), 92–104.

[41] HERING, A. S., KAZOR, K., AND KLEIBER, W. A markov-switching vector autoregressive stochastic wind generator for multiple spatial and temporal scales. *Resources 4*, 1 (2015), 70–92.

[42] HU, J., WANG, J., AND ZENG, G. A hybrid forecasting approach applied to wind speed time series. *Renewable Energy 60* (2013), 185–194.

[43] HUANG, Z., AND CHALABI, Z. Use of time-series analysis to model and forecast wind speed. *Journal of Wind Engineering and Industrial Aerodynamics 56*, 2 (1995), 311–322.

[44] HWANGBO, H., JOHNSON, A. L., AND DING, Y. A production economics analysis for quantifying the efficiency of wind turbines. *Wind Energy 20* (2017), 1501–1513.

[45] HWANGBO, H., JOHNSON, A. L., AND DING, Y. Spline model for wake effect analysis: Characteristics of a single wake and its impacts on wind turbine power generation. *IISE Transactions 50* (2018), 112–125.

[46] IEC-INTERNATIONAL ELECTROTECHNICAL COMMISSION, IEC-61400-12, GENEVA, SWITZERLAND. Wind turbines-part 12-1: Power performance measurements of electricity producing wind turbines.

[47] JAMMALAMADAKA, S., AND SENGUPTA, A. *Topics in Circular Statistics*. World scientific, 2001.

[48] JEON, J., AND TAYLOR, J. W. Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association 107*, 497 (2012), 66–79.

[49] JUN, M., AND STEIN, M. An approach to producing space-time covariance functions on spheres. *Technometrics 49*, 4 (2007), 468–479.

[50] KAZOR, K., AND HERING, A. S. 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*, 2 (2015), 192–217.

[51] KAZOR, K., AND HERING, A. S. The role of regimes in short-term wind speed forecasting at multiple wind farms. *Stat 4*, 1 (2015), 271–290.

[52] KILLICK, R., AND ECKLEY, I. changepoint: An r package for changepoint analysis. *Journal of Statistical Software 58*, 3 (2014), 1–19.

[53] KUSIAK, A., AND LI, W. Estimation of wind speed: A data-driven approach. *Journal of Wind Engineering and Industrial Aerodynamics 98* (2010), 559–567.

[54] LANGE, M., AND FOCKEN, U. *Physical Approach to Short-term Wind Power Prediction*. Springer, Berlin, 2006.

[55] LEE, G., DING, Y., GENTON, M., AND XIE, L. Power curve estimation with multivariate environmental factors for inland and offshore wind farms. *Journal of the American Statistical Association 110*, 509 (2015), 56–67.

[56] LEW, D., MILLIGAN, M., JORDAN, G., AND PIWKO, R. The value of wind power forecasting. In *91st American Meteorological Society Annual Meeting, NREL/CP5500-50814* (2011). Available at: www.nrel.gov/docs/fy11osti/50814.pdf.

[57] LI, B., GENTON, M., AND SHERMAN, M. A nonparametric assessment of properties of space-time covariance functions. *Journal of the American Statistical Association 102*, 478 (2007), 736–744.

[58] LIU, X., YEO, K., AND KALAGNANAM, J. A statistical modeling approach for spatio-temporal degradation data. *Journal of Quality Technology 50*, 2 (2018), 166–182.

[59] MAHMOUDI, M., EZZAT, A. A., AND ELWANY, A. Layerwise anomaly detection in laser powder-bed fusion metal additive manufacturing. *Journal of Manufacturing Science and Engineering 141*, 3 (2019), 031002–1 – 031002–13. doi 10.1115/1.4042108.

[60] MARDIA, K. V. *Statistics for Directional Data*. Academic Press, 1972.

[61] MIAO, H., YANG, L., ZHANG, J., AND VITTAL, V. A spatio-temporal analysis approach for short-term forecast of wind farm generation. *IEEE Transactions on Power Systems 29*, 4 (2014), 1611–1622.

[62] MILLIGAN, M., SCHWARTZ, M., AND WAN, Y. Statistical wind power forecasting models: Results for US wind farms. Tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO., 2003.

[63] MITCHELL, M., GENTON, M., AND GUMPERTZ, M. Testing for separability of space-time covariances. *Environmetrics 16*, 8 (2005), 819–831.

[64] MOHANDES, M., HALAWANI, T., REHMAN, S., AND HUSSAIN, A. A. Support vector machines for wind speed prediction. *Renewable Energy 29*, 6 (2004), 939–947.

[65] PARK, M., AND FUENTES, M. New classes of asymmetric spatial-temporal covariance models. Tech. Rep. 2584, North Carolina State University, Department of Statistics, 2006.

[66] PINSON, P. Wind energy: forecasting challenges for its operational management. *Statistical Science 28*, 4 (2013), 564–585.

[67] PINSON, P., CHEVALLIER, C., AND KARINIOTAKIS, G. N. Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems 22*, 3 (2007), 1148–1156.

[68] PINSON, P., CHRISTENSEN, L., MADSEN, H., SØRENSEN, P. E., DONOVAN, M. H., AND JENSEN, L. E. Regime-switching modelling of the fluctuations of offshore wind generation. *Journal of Wind Engineering and Industrial Aerodynamics 96*, 12 (2008), 2327–2347.

[69] POURHABIB, A., HUANG, J., AND DING, Y. Short-term wind speed forecast using measurements from multiple turbines in a wind farm. *Technometrics 58*, 1 (2016), 138–147.

[70] RASMUSSEN, C. E., AND WILLIAMS, C. K. *Gaussian Processes for Machine Learning*. MIT press Cambridge, 2006.

[71] REIKARD, G. Using temperature and state transitions to forecast wind speed. *Wind Energy 11*, 5 (2008), 431–443.

[72] SAATÇI, Y., TURNER, R., AND RASMUSSEN, C. E. Gaussian process change point models. In *Proceedings of the 27th International Conference on Machine Learning, (ICML-10)* (2010), pp. 927–934.

[73] SANTAMARÍA-BONFIL, G., REYES-BALLESTEROS, A., AND GERSHENSON, C. Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy 85* (2016), 790–809.

[74] SANTOS, R. A. *Damage Mitigating Control for Wind Turbines*. PhD thesis, University of Colorado at Boulder, 2007.

[75] SCHLATER, M. Some covariance models based on normal scale mixtures. *Bernoulli 16*, 3 (2010), 780–797.

[76] SIDERATOS, G., AND HATZIARGYRIOU, N. D. Probabilistic wind power forecasting using radial basis function neural networks. *IEEE Transactions on Power Systems 27*, 4 (2012), 1788–1796.

[77] SONG, Z., JIANG, Y., AND ZHANG, Z. Short-term wind speed forecasting with markov-switching model. *Applied Energy 130* (2014), 103–112.

[78] STEIN, M. Space-time covariance functions. *Journal of the American Statistical Association 100*, 469 (2005), 310–321.

[79] TATSU, J., PINSON, P., TROMBE, P., AND MADSEN, H. Probabilistic forecasts of wind power generation accounting for geographically dispersed information. *IEEE Transactions on Smart Grid 5*, 1 (2014), 480–489.

[80] TRAITEUR, J. J., CALLICUTT, D. J., SMITH, M., AND ROY, S. B. A short-term ensemble wind speed forecasting system for wind power applications. *Journal of Applied Meteorology and Climatology 51*, 10 (2012), 1763–1774.

[81] WAN, Y., MILLIGAN, M., AND PARSONS, B. Output power correlation between adjacent wind power plants. *Journal of Solar Energy Engineering 125*, 4 (2003), 551–555.

[82] XIE, L., GU, Y., ZHU, X., AND GENTON, M. Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Transactions on Smart Grid 5*, 1 (2014), 511–520.

[83] YAN, J., LI, K., BAI, E., DENG, J., AND FOLEY, A. Hybrid probabilistic wind power forecasting using temporally local gaussian process. *IEEE Transactions on Sustainable Energy 7*, 1 (2016), 87–95.

[84] YANG, L., HE, M., ZHANG, J., AND VITTAL, V. Support-vector-machine-enhanced markov model for short-term wind power forecast. *IEEE Transactions on Sustainable Energy 6*, 3 (2015), 791–799.

[85] YOU, M., BYON, E., JIN, J., AND LEE, G. When wind travels through turbines: A new statistical approach for characterizing heterogeneous wake effects in multi-turbine wind farms. *IISE Transactions 49*, 1 (2017), 84–95.

[86] ZHANG, Y., WANG, J., AND WANG, X. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews 32* (2014), 255–270.

[87] ZHU, X., GENTON, M., GU, Y., AND XIE, L. Space-time wind speed forecasting for improved power system dispatch. *Test 23*, 1 (2014), 1–25.

[88] ZWIERS, F., AND VON STORCH, H. Regime-dependent autoregressive time series modeling of the southern oscillation. *Journal of Climate 3*, 12 (1990), 1347–1363.