Integrative Density Forecast and Uncertainty Quantification of Wind Power Generation

Jingxing Wang, Abdullah AlShelahi, Mingdi You, Eunshin Byon, *Member, IEEE*, and Romesh Saigal

***Abstract*—The volatile nature of wind power generation creates challenges in achieving secure power grid operations. It is, therefore, necessary to accurately predict wind power and its uncertainty quantification. Wind power forecasting usually depends on wind speed prediction and the wind-to-power conversion process. However, most current wind power prediction models only consider portions of the uncertainty. This paper develops an integrative framework for predicting wind power density, considering uncertainties arising from both wind speed prediction and the wind-to-power conversion process. Specifically, we model wind speed using the inhomogeneous Geometric Brownian Motion and convert the wind speed prediction density into the wind power density in a closed-form. The resulting wind power density allows quantifying prediction uncertainties through prediction intervals. To forecast the power output, we minimize the expected prediction cost with (unequal) penalties on the overestimation and underestimation. We show the predictive power of the proposed approach using data from multiple operating wind farms located at different sites.**

***Index Terms*—Inhomogeneous Geometric Brownian Motion, nonstationary process, power curve, wind farm**

I. Introduction

Unlike traditional fossil-based energy sources, wind power generation is severely affected by stochastic weather conditions [1], posing significant challenges in achieving secure power grid operations [2]. Thus, accurate forecasting of wind power generation and its uncertainty quantification becomes a critical component in several decision-making processes including unit commitment, economic dispatch, and reserve determination [3]. Wind power generation forecasts have been widely investigated in the literature (e.g., [4], [5]). Interestingly, many studies focus on generating point forecasts of wind power. However, due to the highly volatile and intermittent nature of wind power, probabilistic density forecasts become more crucial for decision-making, e.g., energy storage system sizing, in power system operations under large uncertainties [4], [6].

In providing probabilistic density forecasts, prediction uncertainties should be fully recognized [7]. Wind power depends heavily on wind speed and wind-to-power relationship called *the power curve* shown in Figure 1. Therefore, two major uncertainty sources need to be considered. The first is the uncertainty in predicting future wind speed, whereas the second uncertainty arises when the wind speed is converted to the wind power. Figure 1 illustrates the impact of uncertainties in both wind speed forecast and conversion process on the probabilistic density prediction. Due to the nonlinearity of power curves, the predictive wind speed distribution is not linearly translated into the probabilistic characteristics of wind power prediction. Such nonlinearity causes challenges in quantifying uncertainties in wind power predictions.

In this paper we devise a new integrative methodology for the wind power density forecast by translating the whole predictive wind speed density into the predictive power density forecast. In particular, we formulate the wind speed as a continuous stochastic process based on the inhomogeneous Geometric Brownian Motion (GBM). The inhomogeneous GBM is flexible in capturing nonstationary and highly volatile wind characteristics. We dynamically update the time-varying parameters in the inhomogeneous GBM model with the dual Kalman Filtering in order to characterize the nonstationary nature of wind speed. By applying the Ito's lemma [8] to the stochastic power curve, we then convert the predictive wind speed density to the predictive distribution of wind power.

The resulting closed-form density provides a comprehensive characterization of prediction uncertainties, including predictive intervals and quantiles. Besides, the predictive density allows us to assign different weights on overestimating and underestimating future generation. For example, wind farm operators may prefer to avoid penalties due to unsatisfied demand (or unsatisfied commitment) and thus, prefer underestimation to overestimation of future wind power outputs, while others may prefer overestimation to prevent salvage of excessively generated power [9], [10]. To accommodate such unequal

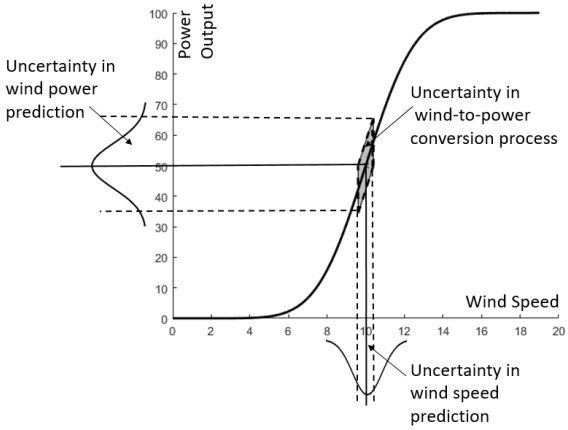

Figure 1: Uncertainties in Wind Power Output Prediction

Jingxing Wang, Abdullah AlShelahi, Eunshin Byon, and Romesh Saigal are with the Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109. Mingdi You is with Ford Motor Company, 22001 Michigan Ave, Dearborn, MI 48124. (corresponding author: Abdullah AlShelahi, email:shelahi@umich.edu). This work was supported by the National Science Foundation under Grants IIS-1741166, EECS-1709094 and CMMI-1536924.

penalties, we formulate an optimization problem to obtain the optimal point prediction that can minimize the expected prediction cost caused by over/underestimation, according to the operator's preference.

We summarize the main contributions as follows.

(1) We provide an integrative framework for wind power density forecast, incorporating uncertainties in both wind speed and the conversion process.
(2) The model parameters are adaptive to changes in the environment and in the power conversion process.
(3) The resulting predictive density takes a closed-form, providing rich information for probabilistic forecasting. It allows us to obtain the optimal prediction that can minimize the average prediction cost when the overestimation and underestimation penalties are different.

We apply the proposed approach to three datasets collected from actual operating wind farms. Our implementation results indicate that the proposed approach can successfully characterize the stochastic wind power process and provide prediction results in accordance with the wind farm operator's preference.

The remainder of this paper is organized as follows. Section II reviews relevant studies. Section III presents the proposed approach. Section IV shows the computational results on real datasets. Finally, we summarize the paper in Section V.

II. Literature Review

The fast increase in computational capabilities and data storage capacity has attracted much attention to data-driven prediction models. Several studies focus on wind speed forecasting using various statistical and machine learning methods, including time-series models such as the Auto-Regressive Moving Average (ARMA), Auto-Regressive Generalized Autoregressive with Conditional Heteroscedasticity (AR-GARCH) model, and Neural Networks (NN) [11], [12]. In [13], a continuous differential equation is developed to estimate spatially correlated wind speeds at multiple turbines in a wind farm. Then the obtained wind speed at each turbine is converted to wind power though a pre-specified power curve. Another commonly used model is the persistent model. It is a simple yet effective model that assumes that the next wind speed is similar to the current speed. Despite its simplicity, the persistent model provides promising prediction accuracy at some wind sites [9].

Various methods study the power curve estimation problem, using polynomial regression, splines, nonparametric regression, NNs, and support vector machines [14], [15], [16]. However, these studies do not consider wind speed forecasting uncertainties. Further, some recent studies provide point wind power forecasting. Abedinia et al. [17] propose a two-dimensional convolution NN to forecast wind power. Their model is trained based on particle swarm optimization. In [18], wind measurements are decomposed using an improved version of empirical mode decomposition. These decomposed measurements are fed into a forecasting model composed of a bagging NN combined with K-means clustering.

To provide probabilistic wind power forecasting, a number of studies simulate wind speed from the predictive density and convert the sampled wind speed to the power output using a static power curve. Taylor et al. [5] integrate wind speed predictions generated from multiple physics-based forecast models with different scenarios via ensemble forecasts. These forecasts are then used for providing wind speed density forecast. Recently, stochastic differential equations (SDEs) have been used to forecast wind power [19]. In [20], an SDE is developed to model the random variations of power processes, where an Ornstein-Uhlenbeck process is employed for modeling wind spend and memory less transformation. Loukatou et al. [21] describe the continuous-time wind speed with the Ornstein-Uhlenbeck GBM model to simulate the wind power trajectory using a deterministic power curve. Although this approach considers the uncertainties in predicting the wind speed, probabilistic characteristics and uncertainties in converting the wind speed to wind power are not addressed. Furthermore, as discussed in Section I, due to the nonlinearity in the wind-to-power conversion process, this approach does not provide the predictive wind power distribution in a closed-form.

Alternatively, some studies take wind historical information as covariates (or inputs) to estimate probabilistic characteristics of wind power [22], [23]. Based on NNs, Sideratos and Hatziargyriou [4] estimate quantiles of future wind power, whereas prediction intervals of wind power generation are constructed in [24]. In [25], a linear quantile regression with spline bases is employed to estimate quantiles of the forecast errors. In many of these studies, the predictive wind speed density is not used as input, but the point wind speed forecast and/or past observations are treated as covariates. Therefore, the prediction uncertainties of wind speed are not fully captured in these studies.

The last approach is to design a probabilistic wind power forecasting using wind power outputs only. Studies in this category employ various methods, such as quantile regression, lower upper bound estimation, Bayesian models, and kernel density estimation [26]. In [27], multivariate distribution modeling and probabilistic forecasts are integrated, while introducing an advanced method based on R-vine copula.

While these studies consider forecasting uncertainties to some extent, they have limited capabilities for quantifying complete probabilistic characteristics of future wind power. They account for portions of uncertainties or provide partial probabilistic information, such as quantiles or prediction intervals, instead of estimating the complete probabilistic density of future wind power. In this research, we fill the gap in the literature by collectively accounting for the uncertainties arising in both wind speed prediction and stochastic power conversion process and providing a predictive density of wind power in a closed-form.

III. Methodology

In this section, we first formulate the dynamics of the wind speed process and the wind-to-power conversion process. Next, we provide an optimization framework to forecast the future wind power output based on wind farm operator's preference on over- and underestimation, and finally present the implementation procedure.

A. Modeling Wind Speed Process

Considering the highly volatile and time-varying wind behavior, we characterize the dynamics of wind speed using the inhomogeneous GBM model [28]. Let $S(t)$ denote the true wind speed at time t. We model the stochastic process of $S(t)$ as

$$dS(t) = \mu_S(t)S(t)dt + \sigma_S(t)S(t)dW_S(t), \tag{1}$$

where $\mu_S(t)$ and $\sigma_S(t)$ capture the drift and volatility of the stochastic process, respectively, and both are time-dependent. $W_S(t)$ denotes a standard Brownian process with independent increments $\Delta W_S(t) = W_S(t+\Delta t) - W_S(t)$. These increments are normally distributed with mean 0 and variance Δt.

Let $X(t)$ denote $\ln S(t)$, i.e., $X(t) = \ln S(t)$. Given the underlying dynamics of $S(t)$ in (1), the dynamics of $X(t)$ can be represented as

$$d[X(t)] = \left[\mu_S(t) - \frac{1}{2}\sigma_S^2(t)\right] dt + \sigma_S(t)dW(t). \tag{2}$$

The detailed derivation of $X(t)$ is available in [8, Chap. 5].

Solving the SDE in (2) analytically is complicated but can be approximated by numerical discretization. By applying the Wagner-Platen expansion and the Euler discretization scheme [29] to (2), we obtain

$$X(t+\Delta t) = X(t) + \left[\mu_S(t) - \frac{1}{2}\sigma_S^2(t)\right] \Delta t + \sigma_S(t)\Delta W(t). \tag{3}$$

Then, it immediately follows that $X(t+\Delta t)$ in (3) follows a normal distribution as

$$X(t+\Delta t) \sim N\left(X(t) + \left[\mu_S(t) - \frac{1}{2}\sigma_S^2(t)\right] \Delta t, \sigma_S^2(t)\Delta t\right), \tag{4}$$

which implies that wind speed is log-normally distributed as

$$\begin{aligned} &\ln(S(t+\Delta t)) \\ &\sim N\left(\ln(S(t)) + \left[\mu_S(t) - \frac{1}{2}\sigma_S^2(t)\right] \Delta t, \sigma_S^2(t)\Delta t\right). \end{aligned} \tag{5}$$

Note that the wind speed distribution in (5) characterizes the stochastic dynamics of wind speed through the time-varying parameters, $\mu_S(t)$ and $\sigma_S(t)$. To estimate $\mu_S(t)$ and $\sigma_S(t)$, one should use wind measurements collected from a meteorological tower or turbine anemometers. However, the collected wind speed may have measurement errors and/or can be perturbed by disturbances such as wake effects [16]. Therefore, the true wind speed $S(t)$ is unobserved in practice. To incorporate such errors and disturbances, assume that the measured wind speed is a linear function of the unobserved true speed. Let $WS(t)$ denote the measured wind speed at time t and $Y(t) = ln(WS(t))$. Letting $X(t)(= ln(S(t))$ be a state variable perturbed by a normally distributed error term $z \sim N(0, \sigma_z^2)$ as follows.

$$Y(t) = X(t) + z. \tag{6}$$

Note that the dynamics of $X(t)$, governed by the linear SDE representation in (3), can be rewritten as

$$X(t+\Delta t) = X(t) + A\,\theta(t) + w(t), \tag{7}$$

where $A = \left(\Delta t, -\frac{1}{2}\Delta t\right)$, $\theta(t) = \left(\mu_S(t), \sigma_S^2(t)\right)^T$, and $w(t) \sim N(0, \sigma_S^2(t)\Delta t)$ is the process noise.

The equations in (6) and (7) together represent the linear state space model. Among several ways to estimate the model parameters in the linear state space model, we employ the Kalman filter due to its flexibility and strong performance in many applications [30], [31].

In particular, we employee the dual Kalman filtering procedure to estimate parameter vector $\theta(t)$ and state $X(t)$ [32]. To model the time-varying parameter $\theta(t)$, we assume that it drifts according to a two-dimensional Gaussian random walk process with covariance Q, i.e.,

$$\theta(t+\Delta t) = \theta(t) + \epsilon, \tag{8}$$

where $\epsilon \sim N(0, Q)$. We include the detailed procedure to update the parameters $\theta(t)$ and state $X(t)$ in Appendix.

B. Modeling Wind-to-Power Conversion Process

This section discusses how to convert the wind speed dynamics into the dynamics of wind power process. The relationship between the wind speed and the power generation can be quantified by the power curve function. A common approach to build the power curve is to include wind speed as a covariate. However, a recent study in [33] discusses the temporal correlation issue, implying that, in addition to wind speed, previous power output also need to be incorporated to better define the power curve. Let $F(t, S(t), P(t-\Delta t)))$ denote the power curve at time t, given the speed $S(t)$ and power $P(t-\Delta t)$. Here, $F(t, S(t), P(t-\Delta t)))$ can represent the power curve from a whole wind farm or a stand-alone turbine.

We model the power curve function $F(t, S(t), P(t-\Delta t))$ as a function of t (as well as $S(t)$ and $P(t-\Delta t)$) to incorporate the time-varying nature of power generation efficiency. This is because, in addition to the wind speed, the wind power output depends on many other environmental factors such as wind direction, humidity, and ambient temperature [15]. Moreover, turbines' age and degradation states of their components (e.g., blade, gearbox) also affect the generation efficiency. Including all of these additional factors, if not impossible, would make the power curve model overly complicated, and more importantly, it also needs to characterize the dynamics of each factor, as we did for wind speed in Section III-A. Instead, we consider the power curve as a function of wind speed and previous power output and let the power curve function itself time-varying. Our approach in modeling the power curve is flexible enough to employ a time-invariant power curve that only depends on inputs; in this case, the power curve function can be simply reduced to $F(t, S(t), P(t-\Delta t)) = F(S(t), P(t-\Delta t))$.

In modeling $F(t, S(t), P(t-\Delta t))$, any type of functions, e.g., parametric, semi-parametric such as splines [34], or nonparametric function [35], [36], can be employed as long as $F(t, S(t), P(t-\Delta t))$ satisfies some weak conditions. Suppose that $F(t, S(t), P(t-\Delta t))$ is twice differentiable over t, $S(t)$, and $P(t-\Delta t)$.

The power output $P(t)$ at time t is given by

$$P(t) = F(t, S(t), P(t-\Delta t)) + e(t), \tag{9}$$

where $e(t)$ is a random noise in the wind-to-power conversion process. We assume that $\Delta e(t) = e(t+\Delta t) - e(t)$

follows the normal distribution with mean 0 and variance $\sigma_F^2 F_S(t, S(t), P(t-\Delta t))\Delta t$, where $F_S(t, S(t), P(t-\Delta t))$ represents the first derivative of $F(t, S(t), P(t-\Delta t))$ over $S(t)$. In formulating the noise variance, we include $F_S(t, S(t), P(t-\Delta t))$, because the power conversion variability tends to be high when the power curve changes rapidly, which is mostly in the mid-speed range. For notational brevity, we will use F_S as an abbreviation of $F_S(t, S(t), P(t-\Delta t))$ in the subsequent discussion.

We first model the dynamics of the wind power process with any power curve function $F(t, S(t), P(t-\Delta t))$. Then, we derive the dynamics with a specific form for $F(t, S(t), P(t-\Delta t))$ to illustrate our approach.

1) Dynamics of Wind Power Process with General Power Curve Function: Given the wind speed process $S(t)$ in (1), the wind power process also follows the inhomogeneous GBM and its dynamics is modeled by

$$dP(t) = \mu_P(t)P(t)dt + \sigma_P(t)P(t)dW_P(t) \tag{10}$$

with

$$\begin{aligned}\mu_P(t) = &\frac{F_t + \mu_S(t)S(t)F_S + \frac{1}{2}\sigma_S^2(t)S(t)^2F_{SS}}{P(t)} \\ &+ \frac{\frac{1}{2}\sigma_P^2(t-\Delta t)P(t-\Delta t)^2F_{PP}}{P(t)} \\ &+ \frac{\mu_P(t-\Delta t)P(t-\Delta t)F_P}{P(t)},\end{aligned} \tag{11}$$

$$\sigma_P(t) = \frac{\sqrt{\sigma_S(t)^2S(t)^2F_S^2 + \sigma_F^2(t)F_S + \sigma_P^2(t-\Delta t)P^2(t-1)F_P^2}}{P(t)}, \tag{12}$$

where $W_P(t)$ denotes a standard Brownian process, F_t represents the first derivative of F over t, F_{SS} is the second derivative of F over S, and F_{PP} is the second derivative of F over P. Also, S, μ_S, and σ_S denote $S(t)$, $\mu_S(t)$, and $\sigma_S(t)$ in (1), respectively. We derive (10)-(12) using Ito's Lemma in [8, Chap. 4]. The detailed derivation is included in Appendix.

Note that $\mu_P(t)$ and $\sigma_P(t)$ in (11) and (12), respectively, depend on the parameters in $S(t)$ (i.e., μ_S, σ_S) and the power curve related functions (i.e., $F_t, F_S, F_{SS}, F_P, F_{PP}$), quantifying the integrated temporal correlation structure from the uncertainties in the wind speed process and wind-to-power conversion process. This result indicates that the stochastic dynamics of wind speed $S(t)$, together with the power curve function, is translated into the dynamics of power generation $P(t)$.

Following the similar procedure in (1)-(5), one can derive the distribution of wind power in a closed-form. Specifically, the power output $P(t+\Delta t)$ at time $t+\Delta t$ is log-normally distributed as

$$\begin{aligned}&\ln(P(t+\Delta t)) \\ &\sim N\left(\ln(P(t)) + \left[\mu_P(t) - \frac{1}{2}\sigma_P^2(t)\right]\Delta t, \sigma_P^2(t)\Delta t\right).\end{aligned} \tag{13}$$

2) Dynamics of Wind Power Process with Nonparametric Power Curve Function: As discussed earlier, the power curve $F(t, S(t), P(t-\Delta t))$ can be flexibly modeled using various functional forms. To illustrate, we employ the nonparametric adaptive power curve model [36] in our analysis. We explain only an outline of the nonparametric adaptive model in this study. For more detailed procedure, the reader is referred to [36].

In the nonparametric approach, the covariates are mapped into a feature space through a nonlinear mapping $(S(t), P(t-\Delta t)) \rightarrow \phi(S(t), P(t-\Delta t))$. Then $P(t)$ can be modeled by

$$\begin{aligned}P(t) &= F(t, S(t), P(t-\Delta t)) + e(t) \\ &= \omega_t^T\phi(S(t), P(t-\Delta t)) + e(t),\end{aligned} \tag{14}$$

where ω_t is a nonparametric regression coefficient vector at period t.

The coefficient vector ω_t is time-varying, so that the power curve $F(t, S(t), P(t-\Delta t))$ can be updated whenever a new sample is observed. Suppose that $\omega_{t-\Delta t}$ was estimated by $\hat{\omega}_{t-\Delta t}$ at time $t-\Delta t$ and we obtain newly observed data at time t. Then we estimate ω_t by solving the following optimization problem.

$$\min L = \frac{1}{2}\|\omega_t - \hat{\omega}_{t-\Delta t}\|^2 + \frac{1}{2}\gamma e(t)^2 \tag{15}$$

$$s.t. \quad P(t) = \omega_t^T\phi(S(t), P(t-\Delta t)) + e(t). \tag{16}$$

Here the first term in the objective function represents the change of the coefficient from $t-\Delta t$ to t. The second term regularizes the amount of update with the regularization parameter γ, balancing the coefficient change and quality of model fitting.

The inner product of $\phi()$'s is called a kernel function. Among many choices of the kernel function, we employ the Gaussian kernel due to its flexibility. Let $k(S(t_i), S(t_j)) = exp\{-[(S(t_i) - S(t_j)]^2/(2\delta_S)\}$ and $k(P(t_i), P(t_j)) = exp\{-[(P(t_i) - P(t_j)]^2/(2\delta_P)\}$ denote the inner products of $\phi(S(t_i))$ and $\phi(S(t_j))$ and $\phi(P(t_i))$ and $\phi(P(t_j))$, respectively, where δ_S and δ_P denote the parameters in the Gaussian kernel. Suppose there are n observations up to time t. Then $F(t, S(t), P(t-\Delta t))$ is updated by

$$\begin{aligned}\hat{F}(t, S(t), P(t-\Delta t)) = &\sum_{i=2}^{n}\lambda_i k(S(t), S(t-(n-i)\Delta t)) \\ &\times k(P(t-\Delta t), P(t-(n-i-1)\Delta t)),\end{aligned} \tag{17}$$

where λ_i is Lagrange multiplier corresponding to the equality constraint in (16).

Then the estimated power curve, $\hat{F}(t, S(t), P(t-\Delta t))$ in (17), can be plugged into the predictive distribution for $P(t+\Delta t)$ in (13). Specifically, to estimate $\mu_P(t)$ and $\sigma_P(t)$ in (11) and (12), respectively, we need to estimate F_t, F_S, F_{SS}, F_P, F_{PP} and σ_F, First, F_t can be estimated by taking the finite difference as

$$\begin{aligned}\hat{F}_t &= \frac{\partial F}{\partial t} \\ &= \frac{\hat{F}(t, S(t), P(t-\Delta t)) - \hat{F}(t-\Delta t, S(t), P(t-\Delta t))}{\Delta t}\end{aligned} \tag{18}$$

$$= \frac{\lambda_i k(S(t), S(t))k(P(t-\Delta t), P(t-\Delta t)}{\Delta t}. \tag{19}$$

Next, F_S and F_{SS}, which are the first and second derivatives of F over S, respectively, can be estimated by

$$\hat{F}_S = \frac{\partial F}{\partial S} \tag{20}$$
$$= \sum_{i=2}^{n} \lambda_i k(P(t-\Delta t), P(t-(n-i-1)\Delta t)) \frac{\partial k(S(t), S(t-(n-i)\Delta t))}{\partial S}$$
$$= \sum_{i=2}^{n} \lambda_i k(P(t-\Delta t), P(t-(n-i-1)\Delta t)) k(S(t), S(t-(n-i)\Delta t))$$
$$\cdot \left(-\frac{S(t) - S(t-(n-i)\Delta t)}{\delta_S} \right), \tag{21}$$

and

$$\hat{F}_{SS} = \frac{\partial^2 F}{\partial S^2} = \sum_{i=2}^{n} \lambda_i k(P(t-\Delta t), P(t-(n-i-1)\Delta t))$$
$$\cdot \frac{\partial^2 k(S(t), S(t-(n-i)\Delta t))}{\partial S^2}$$
$$= \sum_{i=2}^{n} \lambda_i k(P(t-\Delta t), P(t-(n-i-1)\Delta t)) k(S(t), S(t-(n-i)\Delta t))$$
$$\cdot \left(\frac{(S(t) - S(t-(n-i)\Delta t))^2}{\delta_S^2} - \frac{1}{\delta_S} \right), \tag{22}$$

Similarly, F_P and F_{PP}, which are the first and second partial derivatives of F over P, respectively, can be estimated by

$$\hat{F}_P = \frac{\partial F}{\partial P} \tag{23}$$
$$= \sum_{i=2}^{n} \lambda_i k(S(t), S(t-(n-i)\Delta t)) \frac{\partial k(P(t-\Delta t), P(t-(n-i-1)\Delta t))}{\partial P}$$
$$= \sum_{i=2}^{n} \lambda_i k(S(t), S(t-(n-i)\Delta t) k(P(t-\Delta t), P(t-(n-i-1)\Delta t))$$
$$\cdot \left(-\frac{P(t-\Delta t) - P(t-(n-i-1)\Delta t)}{\delta_P} \right), \tag{24}$$

and

$$\hat{F}_{PP} = \frac{\partial^2 F}{\partial P^2} = \sum_{i=2}^{n} \lambda_i k(S(t), S(t-(n-i)\Delta t))$$
$$\cdot \frac{\partial^2 k(P(t-\Delta t), P(t-(n-i)\Delta t - 1))}{\partial P^2}$$
$$= \sum_{i=2}^{n} \lambda_i k(S(t), S(t-(n-i)\Delta t)) k(P(t-\Delta t), P(t-(n-i-1)\Delta t))$$
$$\cdot \left(\frac{(P(t-\Delta t) - P(t-(n-i-1)\Delta t))^2}{\delta_P^2} - \frac{1}{\delta_P} \right). \tag{25}$$

Finally, to estimate σ_F in $\Delta e_t \sim N(0, \sigma_F^2 F_S(t, S(t), P(t-\Delta t))\Delta t)$, we use the sample standard deviation with the first n data points as follows.

$$\hat{\sigma}_F = \sqrt{\frac{1}{n-2} \sum_{i=2}^{n} \left(\frac{\Delta P(i\Delta t) - \Delta \hat{F}(i\Delta t, S(i\Delta t), P((i-1)\Delta t))}{\sqrt{\hat{F}_S(i\Delta t, S(i\Delta t) P((i-1)\Delta t)\Delta t}} \right)^2}, \tag{26}$$

where $\Delta P(i\Delta t) = P(i\Delta t) - P((i-1)\Delta t)$ and

$$\Delta \hat{F}(i\Delta t, S(i\Delta t), P((i-1)\Delta t))$$
$$= \hat{F}(i\Delta t, S(i\Delta t), P((i-1)\Delta t))$$
$$- \hat{F}((i-1)\Delta t, S((i-1)\Delta t, P((i-2)\Delta t)))) \tag{27}$$

By plugging the estimated parameters, $\hat{F}$, $\hat{F}_t$, $\hat{F}_S$, $\hat{F}_{SS}$, $\hat{F}_P$, $\hat{F}_{PP}$, and $\hat{\sigma}_F$ in (17)-(26) to $\mu_P(t)$ and $\sigma_P(t)$ in (11) and (12), we obtain the predictive distribution of power at $t + \Delta t$ in (13). Recall that other parameters associated with wind speed dynamics, i.e., μ_S and σ_S, are estimated from the dual Kalman filtering process discussed in Section III-A.

C. Uncertainty Quantification and Wind Power Prediction

The closed-form predictive distribution of wind power output in (13) provides comprehensive information to characterize prediction uncertainties such as the prediction interval and quantiles. Following the procedure discussed in [37], the $(1-\beta)100\%$ prediction interval for the power generation at time $t + \Delta t$ is given by

$$[exp(\mu' + \sigma' A), exp(\mu' + \sigma' B)] \tag{28}$$

where $\mu' = \ln(P(t) + \left[\mu_P(t) - \frac{1}{2}\sigma_P^2(t)\right]\Delta t$ and $\sigma' = \sigma_P(t)\sqrt{\Delta t}$, and A and B are the solution of

$$\begin{cases} \Phi(B) - \Phi(A) = 1 - \beta, \\ A + B = -2\sigma'. \end{cases} \tag{29}$$

Here $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal distribution.

The α-quantile Q_α such that $Pr(P(t+\Delta t) \leq Q_\alpha) = \alpha$ is obtained by

$$Q_\alpha = exp(\mu' + \sigma' \Phi^{-1}(\alpha)). \tag{30}$$

In particular, the median of $P(t+\Delta t)$ is given by $exp(\mu')$ for $\alpha = 0.5$.

The quantile information is critical in determining the prediction value. In time series analysis, quantities that represent a central tendency, e.g., mean or median, are typically used as a point forecast. Such forecast might not be accurate when the cost of underestimation and overestimation are different, as in wind power operations [9], [38]. Given the quantile, the power is flexibly estimated by penalizing under/overestimation differently.

Let p denote the predicted power output at time $t+\Delta t$. Let $f(x)$ is the probability density function (pdf) of the log-normal distribution described in (13) of the power output at $t + \Delta t$. The expected amount of underestimation and overestimation, denoted by $u(p; t + \Delta t)$ and $o(p; t + \Delta t)$, respectively, are given by

$$u(p; t+\Delta t) = E_{P(t+\Delta t)}[\max\{0, P(t+\Delta t) - p\}]$$
$$= \int_p^{+\infty} x f(x) dx, \tag{31}$$
$$o(p; t+\Delta t) = E_{P(t+\Delta t)}[\max\{0, p - P(t+\Delta t)\}]$$
$$= \int_{-\infty}^{p} x f(x) dx. \tag{32}$$

To predict the power output, one can minimize the expected cost due to possible under/overestimation. Therefore, the optimal p, denoted by p^*, is obtained by solving the following unconstrained optimization problem.

$$p^* = \arg\min_p \left(\alpha \cdot u(p; t+\Delta t) + (1-\alpha) \cdot o(p; t+\Delta t)\right) \tag{33}$$

where $\alpha \in [0, 1]$ represents the penalty to the underestimation. When the underestimation (overestimation) is more costly, α greater (less than) than 0.5 can be used. By taking the derivative of (33) with $u(p; t+\Delta t)$ and $o(p; t+\Delta t)$ and setting it equal to zero, it can be easily shown that the optimal value p^* is the αth quantile, i.e., $p^* = Q_\alpha$. The second derivative is positive, ensuring its optimality.

We provide the overview of the proposed approach in Fig. 2.

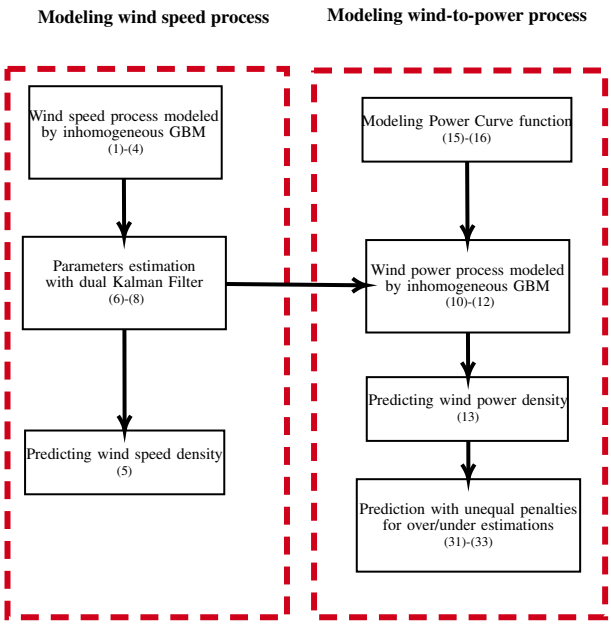

Figure 2: Overview of the proposed approach

D. Implementation Details

Algorithm 1 summarizes the procedure. In our implementation, we divide each wind farm dataset into training and testing sets. Let N_0 denote the number of observations in the training set. The parameters $\sigma_S(t)$ and $\mu_S(t)$ in the wind speed process, the error parameters (σ_z^2 in (6) and Q in (8)) in the dual Kalman filtering, and the power curve are initialized using the N_0 observations in the training set. In particular, to set the error parameters in the kalman filtering, we apply the validation technique to the N_0 data points and choose the values that minimize the prediction error [39]. Moreover, considering that $ln(S(t + \Delta t))$ is normally distributed as shown in (9), we use the sample mean and sample standard deviation of the measured wind speeds to initialize $\mu_S(N_0)$ and $\sigma_S(N_0)$ (see the lines #6-#8 in the algorithm).

The testing set contains 100 observations for each wind farm and is used for evaluating the prediction performance in each wind farm. In this prediction step we update (or filter) the model parameters whenever a new observation is obtained. In Algorithm 1, $\mu_S(t+1 \mid t)$, $\sigma_S(t+1 \mid t)$, and $S(t+1 \mid t)$ in line #12 denote the prior estimates of $\mu_S(t+1)$, $\sigma_S(t+1)$, and $S(t+1)$, respectively, from the Kalman filtering, whereas $\mu_S(t+1 \mid t+1)$, $\sigma_S(t+1 \mid t+1)$ and $S(t+1 \mid t+1)$ in the filtering step (lines #15-#18), correspond to their posterior estimates after observing wind speed $WS(t+1)$ and power $P(t+1)$ at time $t+1$; more detailed dual Kalman filtering procedures are included in Appendix.

Algorithm 1 Proposed Algorithm

1: **Initialization**
2: Initialize σ_z^2 and Q in (6) and in (8), respectively.
3: **for** $k = 2$ to N_0 **do**
4: $\quad r(k) \leftarrow \ln\big(WS(k)\big)/\ln\big(WS(k-1)\big)$.
5: **end for**
6: Obtain initial estimates of the σ_S and μ_S from (5) as follows:
7: $\sigma_S(N_0) \leftarrow std(r)$.
8: $\mu_S(N_0) \leftarrow mean(r) + \sigma_S^2(N_0)/2$.
9: Initialize the power curve function, $F(N_0, WS(N_0)), P(N_0 - 1))$, as discussed in [36].
10: **for** $t = N_0$ to ∞ **do**
11: $\quad$ **Prediction step**
12: $\quad$ Calculate $\mu_S(t+1 \mid t)$, $\sigma_S(t+1 \mid t)$, and $S(t+1 \mid t)$ from (36)-(39)
13: $\quad$ Use (11)-(12) to get $\mu_P(t, P)$ and $\sigma_P(t, P)$.
14: $\quad$ Solve (33) to predict the one-step ahead power output
15: $\quad$ **Filtering step**
16: $\quad$ Observe $WS(t+1)$ and $P(t+1)$.
17: $\quad$ Compute $\mu_S(t+1 \mid t+1)$, $\sigma_S(t+1 \mid t+1)$, and $S(t+1 \mid t+1)$ from (40)-(45).
18: $\quad$ Update the power curve function $F(t+1, S(t+1 \mid t+1), P(t \mid t))$.
19: **end for**

IV. Case Studies

We apply the proposed approach to real datasets collected from three operating wind farms, WF1, WF2, and WF3, summarized in Table I. Due to the data confidentiality required by the data providers, detailed information regarding each wind farm is omitted. Each dataset includes wind measurements and power outputs from the whole wind farm. In all wind farms, the power outputs are scaled to $[0, 100]$. In our implementation, we use 10-minute average wind speed and 10-minute average wind power as the only input and output, respectively.

Table I: Wind Farms Information

Dataset	WF1	WF2	WF3
Terrain	offshore	land-based	onshore
Number of turbines	about 35	240+	about 10
Total data size	1000	1000	650
Temporal resolution	10 minute	10 minute	10minute

A. Implementation Results

Figure 3 depicts the 50% and 90% prediction intervals in WF 1 testing set. Note that the upper bound is capped at 100 (the maximum normalized power output). The majority of the observations fall inside the prediction intervals, indicating that our approach can successfully capture the uncertainties. We can also observe that in general the more volatile the power output (i.e., when the power output changes rapidly), the wider the prediction intervals. For example, when t is about 95, the power output changes rapidly and the prediction intervals are wider, which represents larger prediction uncertainties. On the other hand, when the output is less volatile, e.g., when t is between 40 and 80, we obtain narrower intervals. We observe similar patterns in other wind farms but omit the results to save space.

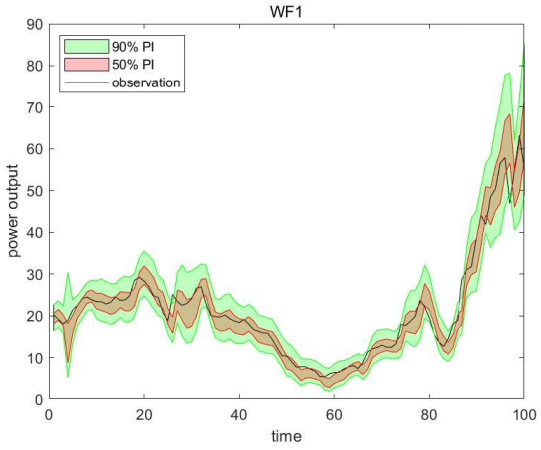

Figure 3: Power Output Prediction Intervals on WF1 Dataset

B. Comparison with Alternative Methods

We compare our approach with other alternative methods, including the persistent model, ARIMA, AR-GARCH model, NN with long short-term memory (LSTM) layers. In both ARIMA and AR-GARCH methods, the wind speed is assumed to follow a normal distribution. In implementing ARIMA, AR-GARCH, and LSTM NN, we use built-in functions in Matlab. We further consider the end-to-end process that formulates the wind power process directly without considering the wind speed forecasting and the conversion process. In the end-to-end process, the wind power is assumed to follow inhomogeneous GBM.

Each model order (or number of parameters) is chosen such that the Bayesian information criterion (BIC) is minimized. To determine the structure of LSTM, including the number of layers and the number of neurons, we apply the following validation technique. The networks are trained using 50% of the whole data set in each wind farm and the prediction performance is evaluated using the validation set consisting of 20% of the data set. We choose the best network structure with the lowest prediction error in the validation set. The models in these alternative methods are re-trained when a new observation is obtained. Once the wind speed at time $t+\Delta t$ is predicted using these approaches, the predicted wind speed is plugged into the power curve to get $\hat{P}(t+\Delta t)$. In all alternative methods except the end-to-end process, we apply the same adaptive non-parametric power curve discussed in Section III-B.

We first compare the point prediction results with the root mean squared error (RMSE) and mean absolute error (MAE). Table II summarizes the results in the testing sets, in comparison with other alternative methods. Overall, our proposed model yields superior performance compared to alternative approaches. For WF3 dataset, the end-to-end process provides slightly lower RMSE than our approach, however, the difference is insignificant.

Next, we evaluate the probabilistic prediction performance with multiple metrics. In order to see how different penalties on the overestimation and underestimation affect the prediction quality, we compute the following power curve error

Table II: Summary of RMSEs and MAEs.*

	WF1		WF2		WF3	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Proposed Approach	**3.02**	**1.86**	**1.72**	**1.30**	4.58	**2.90**
Persistent	3.40	2.34	2.34	1.83	6.26	4.33
ARIMA	4.76	3.65	4.00	3.02	6.77	4.67
AR-GARCH	4.56	3.24	3.13	2.39	6.07	4.14
LSTM NN	4.68	3.42	3.27	2.51	6.23	4.23
End-to-End Process	3.24	1.91	1.86	1.38	**4.49**	2.92

* Boldfaced values indicate the best performance.

(PCE) [10], also referred to as the pinball loss.

$$PCE(P(t),\hat{P}(t)) = \begin{cases} \alpha(P(t)-\hat{P}(t)), & \text{if } \hat{P}(t) < P(t) \\ (1-\alpha)(\hat{P}(t)-P(t)), & \text{otherwise.} \end{cases} \tag{34}$$

where $P(t)$ is the observed power at time t and $\hat{P}(t)$ is its predicted power from each method. Recall that in the proposed approach, we use the α-quantile of the predictive power output density as discussed in Section III. For fair comparison for probabilistic forecasting, in ARIMA and AR-GARCH, we also use the α-quantile of their predictive wind speed densities and plug the resulting α-quantile estimates to the power curve [9]. Note that the forecast values do not change with different α values in the persistent and LSTM NN method, because they do not provide predictive densities but only provide point predictions. Thus, we plug the point wind speed forecast into the power curve in these methods.

Table III summarizes the average PCE computed from multiple α values, α = 0.05, 0.10, $\cdots$, 0.95, for each method. The AR-GARCH generates lower PCEs than ARIMA, because it takes time-varying variance of wind speed into consideration. But PCEs from AR-GARCH are still higher than the proposed approach in all datasets. The LSTM NN also generates about two times higher PCEs than the proposed approach. Our approach consistently produces the lowest PCEs in all cases, indicating that our approach is superior in reflecting wind farm operators' prediction preference on overestimation and underestimation.

Table III: Average PCEs across Multiple Quantiles with α = 0.05, 0.10, $\cdots$, 0.95*

	Proposed Approach	Persis-tent	ARI-MA	AR-GARCH	LSTM NN	End-to-End Process
WF1	**0.75**	1.17	1.40	1.25	1.71	0.84
WF2	**0.51**	0.91	1.20	0.92	1.27	0.63
WF3	**1.17**	2.16	1.97	1.72	2.11	1.22

* Boldfaced values indicate the best performance.

To further assess probabilistic estimation performance, we employ the reliability measure [40]. To compute reliability, an indicator variable that compares an actual wind power $P(t)$ with its α-quantile forecast $\hat{P}(t)$ is obtained as

$$I(t) = \begin{cases} 1, & \text{if } P(t) \leq \hat{P}(t) \\ 0, & \text{if } P(t) > \hat{P}(t), \end{cases} \tag{35}$$

Then, we take the average of $I(t)$ in the dataset. Figure 4 depicts the reliability diagrams obtained from the predictive distributions of the proposed and alternative methods. The diagrams assess a number of quantiles in the testing data, where the x-axis represents the nominal probabilities of quantile forecasts. Each curve shows the deviation from "perfect reliability" as represented by the dash-dot line. Note that we do not include reliability diagrams from persistent and LSTM NN, because they do not provide the probabilistic forecasting. The curve corresponding to the proposed method is the close to the perfect reliability line. This result indicates that the proposed method generates the well-calibrated probabilistic performance. The end-to-end process also generate reasonably good reliability diagrams. However, as we will show in the subsequent discussion and Figure 5, its sharpness is much larger than ours.

Figure 5 shows the box plots of widths of 90% prediction intervals to evaluate sharpness of each method for the three wind farms. Here, sharpness assesses the quality of the prediction interval [41]. Small sharpness value is preferred, which implies narrow prediction intervals [42]. While ARIMA and AR-GARCH overall generate narrower prediction intervals than the proposed approach, their poor reliability diagrams in Figure 4 indicate that their probabilistic forecasting is not accurate. Notably, the proposed method provides more concentrated predictive distributions resulting in sharper forecasts, compared to the end-to-end process.

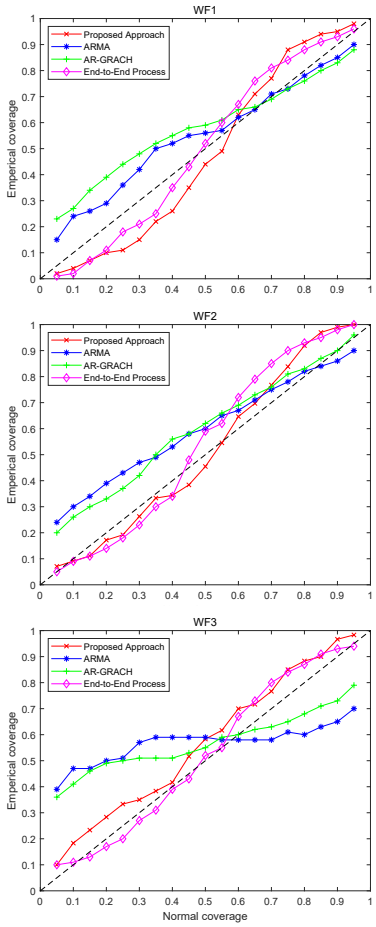

Figure 4: Reliability diagram.

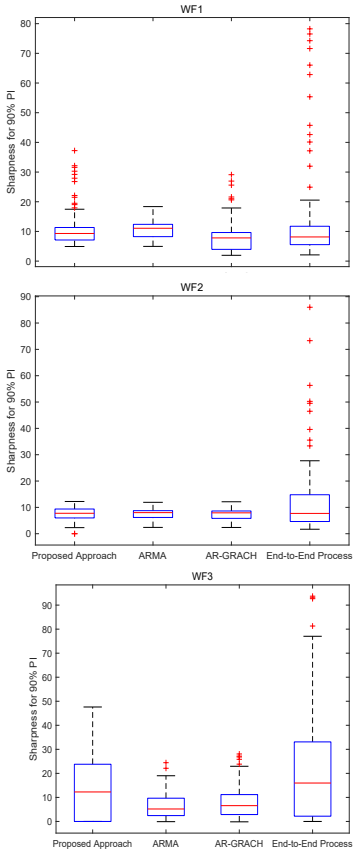

Figure 5: Box plots of the widths of 90% prediction intervals to assess the sharpness.

Overall, the proposed approach and end-to-end process show better point and probabilistic performance than other approaches. These results support the use of inhomeogenous GBM in capturing the volatile wind speed and power processes. Among these two, our approach leads to smaller point forecasting errors in terms of RMSE and MAE, smaller PCEs, and sharper forecasts. It reveals that both uncertainties in wind speed process and wind-to-power process contribute to the overall model performance.

C. Multi-step ahead forecasting

The focus of this study is to capture uncertainties in highly volatile wind process. In doing so, we update our parameters whenever new observation arrives. Thus, our approach is designed mainly for 1-step ahead forecasting. However, it would have been interesting to explore multi-ahead prediction and here, we describe the possibility for multi-step ahead prediction by simply modifying the proposed approach.

For the 2-step ahead forecasting, we train the model using every other data points, $S(t-4), S(t-2), S(t), \cdots$. With this set of observations, 2-step ahead prediction is estimated by applying our proposed method as if we are making one-step ahead prediction. The intuition behind this procedure is that observations can can be collected at longer time intervals as our model is continuous in nature. That is, our model captures the dynamics of longer jumps by only considering such observations. Longer term forecasting can be conducted in a similar manner.

We compare our proposed procedure with alternative methods for the 2-step ahead prediction in Table IV. Our approach provides the lowest 2-step ahead prediction errors in most cases.

Table IV: Two-Step ahead prediction results.*

		Proposed Approach	Persis-tent	ARI-MA	AR-GARCH	LSTM NN	End-to-End Process
WF1	RMSE	**4.17**	5.42	5.33	5.32	4.80	4.21
	MAE	3.00	3.82	4.10	4.04	3.50	**2.83**
WF2	RMSE	**3.30**	3.84	4.43	4.13	3.59	3.54
	MAE	**2.46**	2.92	3.36	3.11	2.71	2.75
WF3	RMSE	**6.83**	8.48	8.06	7.78	7.52	7.25
	MAE	**4.70**	5.90	5.76	5.56	5.23	4.71

* Boldfaced values indicate the best performance.

However, for three-step ahead forecasting, our approach generates worse results than LSTM NN. It is possibly due to the fact that our approach is highly adaptive to a volatile process and thus, for longer horizon forecasting, its strong adaptivity does not show advantage. Improving forecasting capability for longer horizon is left for future research.

V. Summary

We present a new integrative methodology for predicting the wind power density. The proposed approach collectively accounts for the uncertainties in wind speed process and wind-to-power conversion process and provides rich information for the probabilistic forecast through its closed-form prediction density. The closed-form density allows us to extract diverse information and to determine forecast, depending on the wind farm operator's preference on the overestimation and underestimation of future wind power outputs. This framework can minimize the overall costs associated with prediction errors.

Our approach is flexible enough to include other relevant information for prediction. For instance, it is known that purely data-driven approaches, which rely on temporal correlation with past data only, do not provide accurate predictions for medium-term forecasting, ranging from hours to days [9]. In such cases, numeric weather prediction (NWP) models can be employed. However, NWP models have inherent biases and require appropriate initial and boundary conditions [40]. We plan to extend our approach to adjust the bias of NWP projections and provide improved wind speed forecasting.

Further, this study does not consider spatial correlation (either turbine-to-turbine, or wind farm-to-wind farm correlation) [40], [43], [44]. We plan to extend our analysis to incorporate spatial correlations to make forecasts at multiple sites.

We believe that our approach could potentially benefit power grid operations. In the future, we will incorporate our prediction results into the optimization framework for solving decision-making problems such as economic dispatch. We also plan to apply the approach to predict the mechanical and structural load responses in the wind turbine system for the reliability analysis and maintenance optimization [45], [46]. The proposed methodology is also applicable to other engineering systems subject to nonstationary operating conditions, such as solar power systems [47].

References

[1] Hirth L. The market value of variable renewables: The effect of solar wind power variability on their relative price. *Energy Econ* 2013; 38: 218–236.

[2] Bouffard F, Galiana FD. Stochastic Security for Operations Planning with Significant Wind Power Generation. *IEEE Trans Power Syst* 2008; 23(2): 306 – 316.

[3] Zhang Y, Gatsis N, Giannakis GB. Robust Energy Management for Microgrids With High-Penetration Renewables. *IEEE Trans Sust Energy* 2013; 4(4): 944-953.

[4] Sideratos G, Hatziargyriou ND. Probabilistic Wind Power Forecasting Using Radial Basis Function Neural Networks. *IEEE Trans Power Syst* 2012; 27(4): 1788–1796.

[5] Taylor JW, McSharry PE, Buizza R. Wind Power Density Forecasting Using Ensemble Predictions and Time Series Models. *IEEE Trans Energy Convers* 2009; 24(3): 775–782.

[6] Oh E, Son SY. Theoretical energy storage system sizing method and performance analysis for wind power forecast uncertainty management. *Renew. Energy* 2020; 155: 1060-1069.

[7] Choe Y, Lam H, Byon E. Uncertainty Quantification of Stochastic Simulation for Black-box Computer Experiments. *Methodol Comput Appl* 2018; 20(4): 1155–1172.

[8] Bjork T. *Arbitrage Theory in Continuous Time*. Oxford University Press . 2009.

[9] Pourhabib A, Huang JZ, Ding Y. Short-term wind speed forecast using measurements from multiple turbines in a wind farm. *Technometrics* 2016; 58(1): 138–147.

[10] Hering AS, Genton MG. Powering Up With Space-Time Wind Forecasting. *J Am Stat Assoc* 2010; 105(489): 92-104.

[11] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Appl Energy* 2011; 88(4): 1405–1414.

[12] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. *Renew Sust Energ Rev* 2014; 32: 255 - 270.

[13] Ye L, Zhao Y, Zeng C, Zhang C. Short-term wind power prediction based on spatial model. *Renew. Energy* 2017; 101: 1067–1074.

[14] Yampikulsakul N, Byon E, Huang S, Shawn S, You M. Condition monitoring of wind turbine system with nonparametric regression-based analysis. *IEEE Trans Energy Convers* 2014; 29(2): 288-299.

[15] Lee G, Ding Y, Genton MG, Xie L. Power curve estimation with multivariate environmental factors for inland and offshore wind farms. *J Am Stat Assoc* 2015; 110(509): 56–67.

[16] You M, Byon E, Jin JJ, Lee G. When Wind Travels Through Turbines: A New Statistical Approach for Characterizing Heterogeneous Wake Effects in Multi-turbine Wind Farms. *IISE Trans* 2017; 49(1): 84-95.

[17] Abedinia O, Bagheri M, Naderi MS, Ghadimi N. A new combinatory approach for wind power forecasting. *IEEE Systems Journal* 2020; 14(3): 4614–4625.

[18] Abedinia O, Lotfi M, Bagheri M, Sobhani B, Shafie-Khah M, Catalão JP. Improved EMD-based complex prediction model for wind power forecasting. *IEEE Trans. on Sustain. Energy* 2020; 11(4): 2790–2802.

[19] Møller JK, Zugno M, Madsen H. Probabilistic forecasts of wind power generation by stochastic differential equation models. *Journal of Forecasting* 2016; 35(3): 189–205.

[20] Verdejo H, Awerkin A, Saavedra E, Kliemann W, Vargas L. Stochastic modeling to represent wind power generation and demand in electric power system based on real data. *appl Energy* 2016; 173: 283–295.

[21] Loukatou A, Howell S, Johnson P, Duck P. Stochastic wind speed modelling for estimation of expected wind power output. *appl energy* 2018; 228: 1328–1340.

[22] Hu J, Tang J, Lin Y. A novel wind power probabilistic forecasting approach based on joint quantile regression and multi-objective optimization. *Renew. Energy* 2020; 149: 141–164.

[23] Gilbert C, Browell J, McMillan D. Leveraging turbine-level data for improved probabilistic wind power forecasting. *IEEE Trans. on Sustain. Energy* 2019; 11(3): 1152–1160.

[24] Wan C, Xu Z, Pinson P, Dong ZY, Wong KP. Optimal Prediction Intervals of Wind Power Generation. *IEEE Trans Power Syst* 2014; 29(3): 1166-1174.

[25] Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy* 2006; 9: 95-108.

[26] Xue H, Jia Y, Wen P, Farkoush SG. Using of improved models of Gaussian Processes in order to Regional wind power forecasting. *Journal of Cleaner Production* 2020; 262: 121391.

[27] Wang Z, Wang W, Liu C, Wang Z, Hou Y. Probabilistic forecast for multiple wind farms based on regular vine copulas. *IEEE Trans Power Syst* 2017; 33(1): 578–589.

[28] Øksendal B. *Stochastic differential equations*. Springer . 2003.

[29] Platen E. An introduction to numerical methods for stochastic differential equations. *Acta Numer.* 1999; 8: 197–246.

[30] Haykin S. *Kalman filtering and neural networks*. 47. John Wiley & Sons . 2004.

[31] Sun B, Luh PB, Jia QS, O'Neill Z, Song F. Building energy doctors: An SPC and Kalman filter-based method for system-level fault detection in HVAC systems. *IEEE Trans Autom Sci and Eng* 2014; 11(1): 215–229.

[32] Wan EA, Nelson AT. Dual Kalman filtering methods for nonlinear prediction, smoothing, and estimation. *Advances in neural information processing systems* 1997; 9.

[33] Prakash A, Tuo R, Ding Y. The temporal overfitting problem with applications in wind power curve modeling. 2020.

[34] Lee G, Byon E, Ntaimo L, Ding Y. Bayesian spline method for assessing extreme loads on wind turbines. *Ann Appl Stat* 2013; 7(4): 2034-2061.

[35] Lee G, Ding Y, Xie L, Genton MG. A kernel plus method for quantifying wind turbine performance upgrades. *Wind Energy* 2015; 18(7): 1207–1219.

[36] Byon E, Choe Y, Yampikulsakul N. Adaptive learning in time-variant processes with application to wind power systems. *IEEE Trans Autom Sci Eng* 2016; 13(2): 997–1007.

[37] Dahiya RC, Guttman I. Shortest Confidence and Prediction Intervals for the Log-Normal. *Can J Stat* 1982; 10(4): 277–291.

[38] AlShelahi A, Wang J, You M, Byon E, Saigal R. Data-driven prediction for volatile processes based on real option theories. *International Journal of Production Economics* 2020; 226: 107605.

[39] Friedman J, Hastie T, Tibshirani R. *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in StatisticsSpringer, 2nd ed. . 2009.

[40] Jang Y, Byon E. Probabilistic Characterization of Wind Diurnal Variability for Wind Resource Assessment. *IEEE Trans. on Sustain. Energy* 2020; 11(4): 2535-2544. doi: 10.1109/TSTE.2020.2965444

[41] Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2007; 69(2): 243–268.

[42] Gneiting T. Probabilistic forecasting. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 2008: 319–321.

[43] You M, Byon E, Jin J, Lee G. When wind travels through turbines: A new statistical approach for characterizing heterogeneous wake effects in multi-turbine wind farms. *IISE Trans.* 2017; 49(1): 84–95.

[44] You M, Liu B, Byon E, Huang S, Jin JJ. Direction-Dependent Power Curve Modeling for Multiple Interacting Wind Turbines. *IEEE Trans. Power Syst.* 2018; 33(2): 1725–1733.

[45] Ko YM, Byon E. Condition-based joint maintenance optimization for a large-scale system with homogeneous units. *IISE Trans* 2017; 49(5): 493-504.

[46] Choe Y, Byon E, Chen N. Importance Sampling for Reliability Evaluation with Stochastic Simulation Models. *Technometrics* 2015; 57(3): 351–361.

[47] Choe Y, Guo W, Byon E, Jin J, Li J. Change-Point Detection on Solar Panel Performance Using Thresholded LASSO. *Qual Reliab Eng Int* 2016; 32(8): 2653-2665.

VI. APPENDIX

Dual Kalman Filtering Procedure

Recall that the parameter vector is $\theta(t) = [\mu_S(t), \sigma_S^2(t)]^T$ and state is $X(t)$. We use $\theta_2(t)$ for $\sigma_S^2(t)$. Let $\hat{X}(t \mid t)$ and $\hat{X}(t+\Delta t \mid t)$ denote the posterior and prior estimates of state variable $X(t)$ with their associated estimation error variances $P_X(t \mid t)$ and $P_X(t+\Delta t \mid t)$, respectively. Similarly, $\hat{\theta}(t \mid t)$ and $\hat{\theta}(t+\Delta t \mid t)$, respectively, denote the posterior and prior estimates of the parameter vector $\theta(t)$ and $P_\theta(t \mid t)$ and $P_\theta(t+\Delta t \mid t)$ represent the corresponding estimation error covariance matrices. We let $K_X(t)$ and $K_\theta(t)$ denote the Kalman gain associated with state and parameters filters at time t, respectively. Then the dual Kalman filtering proceeds as follows:

- Parameters prediction:

$$\hat{\theta}(t+\Delta t \mid t) = \hat{\theta}(t \mid t), \tag{36}$$
$$P_\theta(t+\Delta t \mid t) = P_\theta(t \mid t) + Q. \tag{37}$$

- State prediction:

$$\hat{X}(t+\Delta t \mid t) = \hat{X}(t \mid t) + A\,\hat{\theta}(t+\Delta t \mid t), \tag{38}$$
$$P_X(t+\Delta t \mid t) = P_X(t \mid t) + \Delta t\,\hat{\theta}_2(t+\Delta t \mid t). \tag{39}$$

- State filtering:

$$K_X(t+\Delta t) = P_X(t+\Delta t \mid t)\ \left[P_X(t+\Delta t \mid t) + \sigma_z^2\right]^{-1}, \tag{40}$$
$$\hat{X}(t+\Delta t \mid t+\Delta t) = \hat{X}(t+\Delta t \mid t) + K_X(t+\Delta t)\left[Y(t+\Delta t) - \hat{X}(t+\Delta t \mid t)\right], \tag{41}$$
$$P_X(t+\Delta t \mid t+\Delta t) = \left[I - K_X(t+\Delta t)\right] P_X(t+\Delta t \mid t). \tag{42}$$

- Parameters filtering:

$$K_\theta(t+\Delta t) = P_\theta(t+\Delta t \mid t)\,A^T\,\left[A\,P_\theta(t+\Delta t \mid t)\,A^T + \sigma_z^2\right]^{-1}, \tag{43}$$
$$\hat{\theta}(t+\Delta t \mid t+\Delta t) = \hat{\theta}(t+\Delta t \mid t) + K_\theta(t+\Delta t)\left[Y(t+\Delta t) - \hat{X}(t+\Delta t \mid t)\right], \tag{44}$$
$$P_\theta(t+\Delta t \mid t+\Delta t) = \left[I - K_\theta(t+\Delta t)\,A\right] P_\theta(t+\Delta t \mid t). \tag{45}$$

Then $\hat{X}(t+\Delta t \mid t)$, which is the posterior estimate of $X(t)$, is used to estimate $X(t)$ and similarly, $\hat{\theta}(t+\Delta t \mid t)$ for estimating $\mu_S(t)$ and $\sigma_S^2(t)$ in (5).

Derivation of $dP(t)$ in (10)

Based on Ito's Lemma [8, chap. 4] and assuming that $P(t-1)$ follows GBM, we obtain

$$dF(t,S(t),P(t-1)) = F_t dt + F_S dS(t) + F_P dP(t-1) + F_{SS}(dS(t))^2$$
$$+ F_{SP} dS(t) dP(t-1) + F_{PP}(dP(t-1))^2 = F_t + \mu_S(t)S(t)F_S$$
$$+ \mu_P(t-\Delta t)P(t-\Delta t)F_P + \frac{1}{2}\sigma_S^2(t)S(t)^2 F_{SS} + \frac{1}{2}\sigma_P^2(t-\Delta t)P^2(t-\Delta t)F_{PP}$$
$$+ \sigma_F(t)F_S S(t) dW_S(t) + \sigma_P(t-\Delta t)P(t-1)F_P dW_P(t-\Delta t)$$

Note that during time t to $t+\Delta t$, the jump value is given by

$$\begin{aligned}\Delta P(t) &= P(t+\Delta t) - P(t)\\ &= F(t+\Delta t, S(t+\Delta t), P(t)) - F(t,S(t),P(t-\Delta t)) + e(t+\Delta t) - e(t)\\ &= \Delta F(t,S(t),P(t-\Delta t)) + \Delta e_t,\end{aligned}$$

where Δe_t is assumed to follow the normal distribution with mean 0 and variance $\sigma_F^2 F_S(t,S(t),P(t-\Delta t))\Delta t$, i.e., $\Delta e_t \sim N(0, \sigma_F^2 F_S(t,S(t),P(t-\Delta t))\Delta t)$. Or equivalently,

$$de(t) = \sigma_F \sqrt{F_S(t,S(t),P(t-\Delta t))} dW_e(t),$$

where $dW_e(t)$ denotes a standard Brownian process. Taking the errors in the power curve into account, we have $\Delta P(t) = \Delta F(t,S(t),P(t-\Delta t) + \Delta e_t$ with $\Delta e_t \sim N(0, \sigma_F^2 F_S^2(t)\Delta t)$. Therefore, the dynamic of $P(t)$ becomes

$$dP(t) = dF(t,S(t),P(t-\Delta t)) + de(t)$$
$$= \frac{F_t + \mu_S(t)S(t)F_S + \frac{1}{2}\sigma_S^2(t)S(t)^2 F_{SS} + \mu_P(t-\Delta t)P(t-\Delta t)F_P}{P(t)} P(t)dt$$
$$+ \frac{\frac{1}{2}\sigma_P^2(t-\Delta t)P(t-\Delta t)^2 F_{PP}}{P(t)} P(t)dt + \sigma_S(t)S(t)F_S dW_S(t)$$
$$+ \sigma_P(t-\Delta t)P(t-\Delta t)F_P dW_P(t-\Delta t) + \sigma_F\sqrt{F_S(t,S(t),P(t-\Delta t))} dW_e(t),$$

where $W_P(t-1)$, $W_S(t)$ and $W_e(t)$ are three independent Brownian motions, which leads to

$$dP(t) = \mu_P(t,P)P(t)dt + \sigma_P(t,P)P(t)dW_P(t),$$

where $\mu_P(t)$ and $\sigma_P(t)$ are, respectively, given by

$$\mu_P(t) = \frac{F_t + \mu_S(t)S(t)F_S + \frac{1}{2}\sigma_S^2(t)S(t)^2 F_{SS}}{P(t)}$$
$$+ \frac{\frac{1}{2}\sigma_P^2(t-\Delta t)P(t-\Delta t)^2 F_{PP}}{P(t)} + \frac{\mu_P(t-\Delta t)P(t-\Delta t)F_P}{P(t)},$$
$$\sigma_P(t) = \frac{\sqrt{\sigma_S^2(t)S(t)^2F_S^2 + \sigma_F^2(t)F_S + \sigma_P^2(t-\Delta t)P^2(t-1)F_P^2}}{P(t)}.$$

PLACE PHOTO HERE

Abdullah AlShelahi received his B.S. degree in Industrial and Management Systems Engineering from Kuwait University, and M.S. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan-Ann Arbor in 2014 and 2019, respectively. He was an Advanced Analytics Researcher at General Motors in the Advanced Analytics Center of Expertise. Currently, he is a faculty member at Kuwait University and a Quantitative Modeler and Machine Learning Engineer at JPMorgan Chase. His current research interests include data analytics, machine learning, deep learning, and financial engineering.

PLACE PHOTO HERE

Mingdi You received the B.S. degree in mathematics from Southeast University, Nanjing, China, and the M.S. and Ph.D degrees in Industrial & Operations Engineering from the University of Michigan, Ann Arbor, MI, USA. He is now a senior data scientist at Aetna. His research interest includes applied statistics and spatial data analytic and reliability engineering with applications on renewable energy.

PLACE PHOTO HERE

Eunshin Byon (S'09-M'21) is an associate professor in the Department of Industrial and Operations Engineering at the University of Michigan. She received her B.S. and M.S. in Industrial and Systems Engineering from the Korea Advanced Institute of Science and Technology (KAIST) and Ph.D. in Industrial and Systems Engineering from Texas A&M University. Her research interests include optimizing operations of renewable systems, data analytics, quality and reliability engineering, and uncertainty quantification. She is a member of IIE, INFORMS, and IEEE.

PLACE PHOTO HERE

Jingxing Wang is currently a data scientist in Ford Motor Company. He received his B.S. degree in Computational Mathematics from Peking University, and M.S. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan-Ann Arbor. His research focuses on data analytics, stochastic process, optimization, and other models in renewable energy, healthcare, marketing and sales.

PLACE PHOTO HERE

Romesh Saigal is a Professor of Industrial and Operations Engineering at the University of Michigan, Ann Arbor, Michigan. He received the BTech and MTech degrees from the Indian Institute of Technology, Kharagpur and his PhD degree from the University of California at Berkeley. He has been on the faculty of the Haas School of Business, University of California Berkeley and The Department of Industrial Engineering and Management Sciences at Northwestern University, Evanston. He teaches courses in Optimization, Stochastic Processes and Financial Engineering, and his recent research involves applications of financial engineering tools in managing risks in transportation and energy sectors.