A Spatio-temporal Analysis Approach for Short-term Forecast of Wind Farm Generation

Miao He, *Member, IEEE,* Lei Yang, *Member, IEEE,* Junshan Zhang, *Fellow, IEEE,* and Vijay Vittal, *Fellow, IEEE*

***Abstract*—In this paper, short-term forecast of *wind farm generation* is investigated by applying spatio-temporal analysis to extensive measurement data collected from a large wind farm where multiple classes of wind turbines are installed. Specifically, using the data of the wind turbines' power outputs recorded across two consecutive years, graph-learning based spatio-temporal analysis is carried out to characterize the statistical distribution and quantify the level crossing rate of the wind farm's aggregate power output. Built on these characterizations, finite-state Markov chains are constructed for each epoch of three hours and for each individual month, which accounts for the diurnal non-stationarity and the seasonality of wind farm generation. Short-term distributional forecasts and a point forecast are then derived by using the Markov chains and ramp trend information. The distributional forecast can be utilized to study stochastic unit commitment and economic dispatch problems via a Markovian approach. The developed Markov-chain-based distributional forecasts are compared with existing approaches based on high-order autoregressive models and Markov chains by uniform quantization, and the devised point forecasts are compared with persistence forecasts and high-order autoregressive model-based point forecasts. Numerical test results demonstrate the improved performance of the Markov chains developed by spatio-temporal analysis over existing approaches.**

***Index Terms*—Short-term wind power forecast, distributional forecast, point forecast, wind farm, graphical learning, spatio-temporal analysis, Markov chains.**

NOMENCLATURE

t	time index of measurement data
m	index of wind turbine class and the corresponding meteorological tower (MET)
M	number of wind turbine classes within the wind farm
C_m	wind turbine class m
N_t	number of measurement data
N_m	number of wind turbines in C_m
H_m	MET for C_m
r_m	wind turbine co-located with H_m in C_m
$W_m(t)$	wind speed measured at H_m
$P_i(t)$	power output of wind turbine i
$U_m(\cdot)$	power curve of C_m, which maps $W_m(t)$ to $P_i(t)$, $\forall i \in C_m$
$P_{ag,m}(t)$	aggregate power output of C_m
$P_{ag}(t)$	aggregate power output of the wind farm
$P_{ag}^{\max}$	rated capacity of the wind farm
$\overline{m}$	index of the reference MET
$d_m(i)$	'distance' from node i to the root of the minimal spanning tree of C_m
α_m	linear regression coefficient for the parent-child turbine pairs of C_m
β_m	linear regression coefficient for $W_m(t)$ as an affine function of $W_{\overline{m}}(t)$
$G_{pw}(\cdot)$	'power curve' of the wind farm, which maps $W_{\overline{m}}(t)$ to $P_{ag}(t)$
Γ	wind farm generation level
γ	wind speed level
$f_X(\cdot)$	probability density function (PDF) of X
$F_X(\cdot)$	cumulative density function (CDF) of X
$L_X(\cdot)$	level crossing rate (LCR) function of X
$\mathcal{N}$	standard normal random variable
$W_{\overline{m}}^{\mathcal{N}}(t)$	Gaussian transformation of $W_{\overline{m}}(t)$
ϕ	regression coefficient of the first-order autoregressive (AR(1)) model
$\epsilon(t)$	white noise of the AR(1) model
σ_ϵ	variance of $\epsilon(t)$
$\mathcal{S}$	state space of Markov chain (MC)
N_s	number of states in $\mathcal{S}$
S_k	state k in $\mathcal{S}$, $k \in \{1, \cdots, N_s\}$
τ_k	average duration of state S_k
$P_{ag,k}$	representative generation level of state S_k
Q	transition matrix of Markov chain
n_{ij}	number of transitions from S_i to S_j encountered in the measurement data
$Pr(A)$	probability of an event A
$\mathbf{E}[X\|Y]$	conditional expectation of X given Y
argmin	argument of the minimum

Manuscript received March 15, 2013; revised June 20, 2013, September 19, 2013 and November 21, 2013; accepted December 20, 2013. This work was supported in part by the US National Science Foundation under grant CPS-1035906 and CNS-1218484, in part by the DTRA grant HDTRA1-09-1-0032, and in part by the Power System Engineering Research Center. Paper no. TPWRS-00319-2013.

M. He is with Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, TX 79401, USA (e-mail: Miao.He@ttu.edu); L. Yang, J. Zhang and V. Vittal are with School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA (e-mail: Lei.Yang.7@asu.edu; Junshan.Zhang@asu.edu: Vijay.Vittal@asu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2014.xxxxxxx

I. Introduction

A critical aspect in meeting the renewable portfolio standard (RPS) adopted by many states in the U.S. includes the integration of renewable energy sources, such as wind and solar [1]. Given the fact that the power outputs of wind turbines are highly dependent on wind speed, the power generation of a wind farm varies across multiple timescales of power system planning and operations. With increasing penetration into bulk power systems, wind generation has posed significant challenges for reliable system operations, because of its high variability and non-dispatchability [2]. Specifically, one key complication arises in terms of committing and dispatching conventional generation resources, when the short-term forecast of wind farm generation is not accurate. Currently, wind generation forecast for an individual wind farm typically has an error of 15% to 20% [3], in sharp contrast to the case of load forecast. When the actual wind generation is above the forecasted value, i.e., more conventional generation capacity has been committed than needed, it could result in less efficient set points for thermal units. In some cases, wind generation may need to be curtailed [4]. On the flip side of the coin, when the actual wind generation is less than the forecasted value, costly ancillary services and fast acting reserves have to be called upon. Therefore, it is imperative to develop accurate forecast approaches for wind farm generation.

State-of-the-art short-term wind power forecast approaches include time-series models (e.g., autoregressive models [5], Kalman filtering [6]), Markov chains [7], [8], and data mining [9], [10]. A comprehensive literature review on wind power forecast can be found in [11] and [12]. Time-series models and data mining-based regression models, while being able to provide continuous-value wind power forecast, could suffer from high computational complexity. Compared to other forecast models, finite-state Markov chains strike a good balance between complexity and modeling accuracy. In particular, the transition probability matrix of Markov chains, which is used to provide distributional forecasts and point forecasts, can be learned from historical data (e.g., by using the maximum likelihood estimation technique [7]); when new data points are available online, it is also easy to update the transition probability matrix. It is worth noting one of state-of-the-art forecasting approaches is to utilize empirical distributions and the rich statistical information extracted from historical data (see [13], [14] and the references therein). Generally, empirical distribution of wind power data is non-Gaussian [15]. In [16], the logit transform is carried out as preprocessing, so that such a bounded time series can be studied by using autoregressive models in a Gaussian framework. In this paper, finite-state Markov chains are utilized to model the bounded wind power time series with a general probability distribution. It is worth noting that finite-state Markov chains inherently have bounded support, and the stationary distribution of a Markov chain can be general. Despite the appealing features of Markov chains, there is no existing studies to systematically design the state space of Markov chains for wind power. The proposed approach in this paper addresses this issue by developing a general spatio-temporal analysis framework.

In this paper, Markov-chain-based stochastic models for wind farm generation are developed for different seasons and for different epochs of the day across the whole year. From these Markov-chain-based stochastic models, short-term distributional forecasts and point forecasts of wind farm generation are obtained. The information used for forecasts includes both historical data and real-time data (the present wind farm generation). With a forecasting lead time of 10 min (or larger), these Markov-chain-based forecasts could be utilized for a variety of power system operation functions. An overview of the main contributions of this work is presented below.

A. Summary of Main Results and Contributions

One key observation of this study is the wind farm spatial dynamics, i.e., *the power outputs of wind turbines within the same wind farm can be quite different, even if the wind turbines are of the same class and physically located close to each other*. The disparity in the power outputs of wind turbines may be due to the wake effect of wind speed, diverse terrain conditions, or other environmental effects. Motivated by this observation, graph-learning based spatial analysis is carried out to quantify the statistical distribution of *wind farm generation*, with rigorous characterization of wind farm spatial dynamics. Then, time series analysis is applied to quantify the level crossing rate (LCR) of the wind farm's aggregate power output. Finite-state Markov chains are then constructed, with the state space and transition matrix designed to capture both the spatial and temporal dynamics of the wind farm's aggregate power output. Based on [17], the distributional forecasts and the point forecasts of wind farm generation are provided by using the Markov chains and ramp trend information. In this work, another finding of independent interest is that the tail probability of wind farm's aggregate power output exhibits a 'power-law' decay with an exponential cut-off, where the power-law part has a much heavier tail than the Gaussian distribution. This indicates that one cannot simply apply the central limit theorem (CLT) to characterize the aggregate power output, because of the strong correlation across the power outputs of wind turbines within a wind farm.

The main contributions of this study are summarized below:

- A general spatio-temporal analysis framework is developed, in which the spatial and temporal dynamics of wind farm generation are characterized by analytically quantifying the statistical distribution and the LCR.
- Built on the results of spatio-temporal analysis, a systematic approach for designing the state space of the Markov chain is introduced.
- By modeling variable wind power as a Markov chain, stochastic unit commitment and economic dispatch problems can be studied by using Markovian state-space approaches instead of scenario-based approaches [18], [19]. Thus, the complexity induced by exponentially-growing scenarios of scenario-based approaches can be mitigated. *Therefore, this study is a timely contribution to the recent efforts on wind generation integration that involve Markov-chain-based stochastic optimizations*.

The rest of the paper is organized as follows. A few critical observations from the measurement data are first discussed in

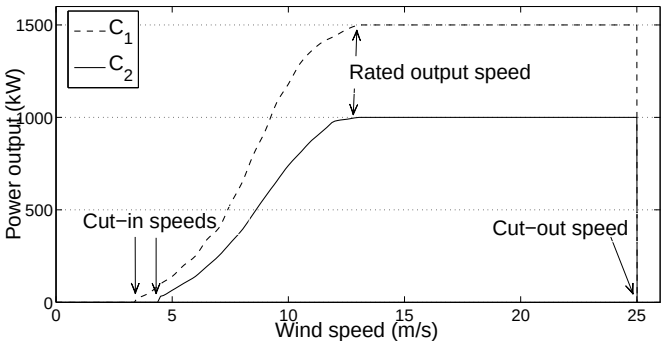


Fig. 1. Power curves for wind turbines from classes C_1 and C_2.

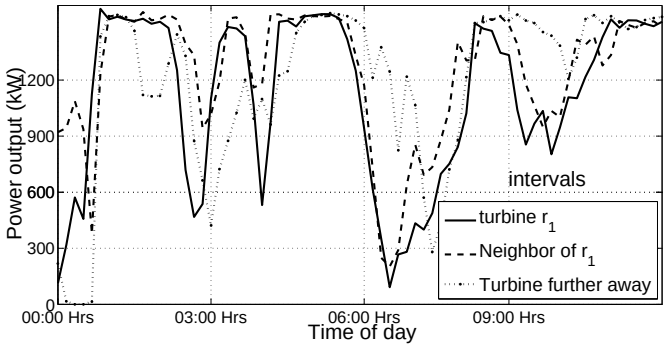


Fig. 2. Power outputs of three wind turbines in C_1.

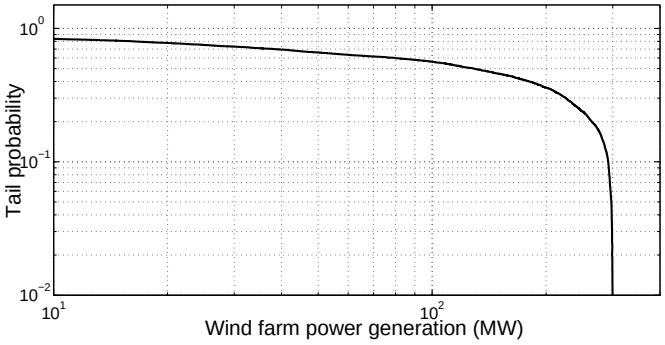


Fig. 3. Tail probability of the wind farm's aggregate power output.

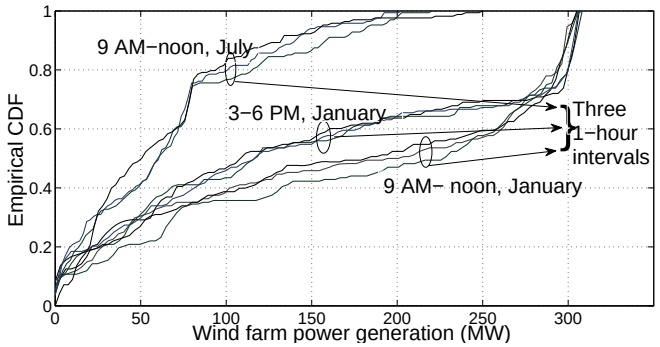


Fig. 4. Empirical distributions of wind farm generation over various 1-hour intervals of different epochs of the day and different months.

Section II. Spatio-temporal analysis and the design of Markov chains are presented in Section III. Section IV discusses the proposed Markov-chain-based forecast approach and numerical examples. Conclusions are provided in Section V.

II. Available Data and Key Observations

In this paper, spatio-temporal analysis is carried out for a large wind farm with a rated capacity of $P_{ag}^{\max}$=300.5MW. There are $M = 2$ classes of wind turbines in this wind farm, with $N_1 = 53$ and $N_2 = 221$, respectively. The power curves of the two turbine classes are provided in Fig. 1. For each class C_m, a meteorological tower (MET) H_m is deployed and co-located with a wind turbine, denoted by r_m. The power outputs of all wind turbines and the wind speeds measured at all METs are recorded every 10 minutes for the years 2009 and 2010. From the measurement data, several key observations can be made as follows.

A. Spatial Dynamics of Wind Farm

A critical observation from the measurement data is that the power outputs of wind turbines within the wind farm can be quite different. Fig. 2 illustrates the power outputs of three wind turbines in C_1. It is clear that the power outputs are not equal, despite the geographic proximity of r_1 and its nearest neighbor (the disparity in the power outputs of the wind turbines belonging to C_2 has also been observed; the plots are not included for the sake of brevity). This disparity has been largely neglected in the existing literature.

Although the variable power outputs of wind turbines are not identical, it is reasonable to assume that they follow the same probability distribution if the wind turbines are of the same class. A natural question here is whether the CLT, either the classic CLT or the generalized CLT, can be applied to characterize the probability distribution of the aggregate power output of a large number of wind turbines. To this end, the tail probability distribution of the wind farm's aggregate power output is examined and plotted in Fig. 3. As illustrated in Fig. 3, the tail probability demonstrates a 'power-law' decay with an exponential cut-off and the power-law part has a much heavier tail than the Gaussian distribution. It is useful to note that this kind of tail behavior has been observed in many natural phenomena (e.g., size of forest fires) that have strong component-wise correlations [20]. Because of the strong correlation between the power outputs of wind turbines, particularly from adjacent wind turbines, the classic CLT cannot be applied to characterize the probability distribution of the wind farm's aggregate power output. In fact, even the 'CLT under weak dependence' cannot be directly applied, despite the fact that the correlation between the power outputs of wind turbines weakens with the distance between them (the 'mixing distance'). Hence, the probability distribution of the wind farm's aggregate power output cannot be characterized using the classic CLT; and it may not even be governed by stable laws [21]. With this insight, the proposed approach resorts to graphical learning methods to model the dependence structure in the power outputs of individual wind turbines and carries out spatio-temporal analysis accordingly.

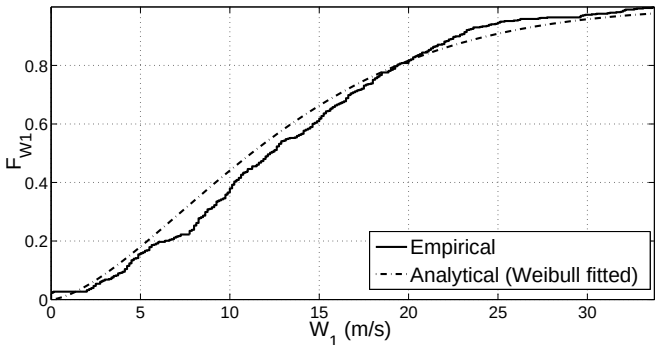


Fig. 5. Weibull-fitted CDF (λ=11.37, k=1.54) and empirical CDF of W_1 for the 9 AM-noon epoch of January 2009.

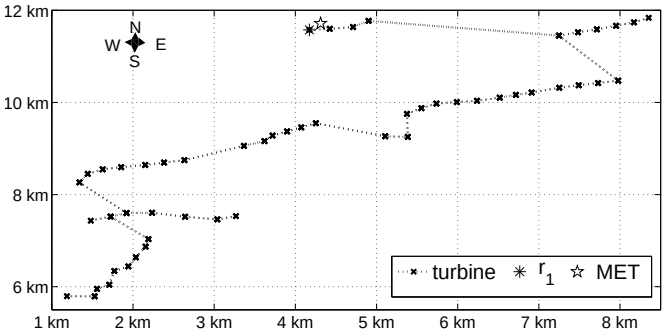


Fig. 6. MST of C_1 (with distance to the southwest corner of the wind farm).

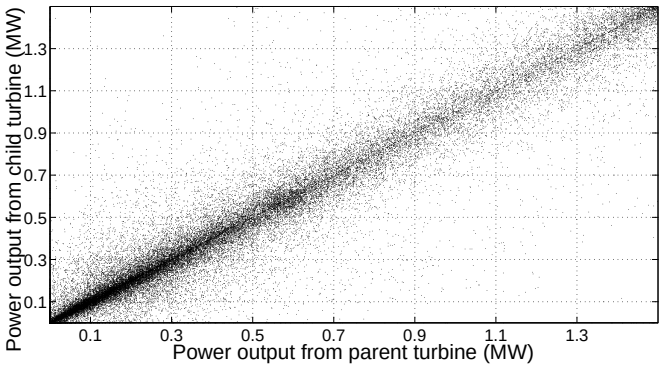


Fig. 7. Power outputs of parent-child turbine pairs of C_1 for the 9 AM-noon epoch of January 2009.

B. Diurnal Non-Stationarity and Seasonality

Another key observation, as illustrated in Fig. 4, is the diurnal non-stationarity and the seasonality of wind farm generation. Specifically, it is observed that within each three-hour epoch, the probability distributions of wind farm generation over three consecutive 1-hour intervals are consistent. However, these CDFs from different epochs of three hours and different seasons can be quite different, indicating the non-staionarity of wind farm generation. Due to the non-stationary (empirical) distributions of wind farm generation, the distributional forecasts and point forecasts of wind farm generation, together with the developed models (Weibull distributions and Markov chains) used to derive distributional forecasts, can have quite different parameters for different months and different epochs. Therefore, it is necessary to develop forecast models *separately* for each month and each epoch (three hours for the wind farm considered here). Further, when estimating the parameters of Weibull distributions and Markov chains, relevant historical data, i.e., the historical data from the same month and the same epoch, can be used.

In what follows, data-driven analysis is carried out to characterize the spatial and temporal dynamics of the wind farm's aggregate power output. The data of the year 2009 is used in spatio-temporal analysis to guide the design of Markov chains, and the data of the year 2010 is used to assess the accuracy of the forecast provided by the proposed Markov-chain-based approach. Specifically, the 9 AM-noon epoch of January 2009 is used as an illustrative example in the following spatio-temporal analysis, since *this epoch exhibits the richest spatio-temporal dynamics, in the sense that the wind farm's aggregate power output during this epoch takes values ranging from 0 to the wind farm's rated capacity and exhibits the highest variability over time (quantified by LCR).*

C. Weibull Distribution of Wind Speed

In the existing literature, wind speed is usually characterized using Weibull distributions [22]. In this work, it is observed from the measurement data that the wind speed W_m at each MET within the wind farm closely follows a Weibull distribution during each epoch, the probability density function (PDF) of which is given by:

$$f_{W_m}(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp^{-(x/\lambda)^k}, \quad \forall x \geq 0, \qquad (1)$$

where k is the shape parameter and λ is the scale parameter. The fitted cumulative density function (CDF) and the empirical CDF of W_1 for the 9 AM-noon epoch of January 2009 are plotted in Fig. 5. The match between the empirical CDF and the fitted CDF suggests that the fitted Weibull distribution with the two parameters k and λ estimated from wind speed measurements can be utilized to analytical quantify wind speed dynamics. Under the developed spatio-temporal analysis framework, the fitted Weibull distributions of wind speed are also critical to the analytical characterizations of both the statistical distribution and the LCR of wind farm generation. The application of the fitted Weibull distributions of wind speed in the spatial analysis and the temporal analysis will be discussed in Section III.A and Section III.B, respectively.

III. Spatio-temporal Analysis of Wind Farm Generation

A. Spatial Analysis and Statistical Characterization

A key objective of spatial analysis is to characterize the statistical distribution of $P_{ag}(t)$. To this end, regression analysis is applied to the measurement data of each turbine's power output, so that $P_{ag}(t)$ could be expressed in terms of wind speed. Then, the analytical CDF of $P_{ag}(t)$ can be obtained from the fitted Weibull CDF of wind speed. In what follows, the key steps of spatial analysis are provided in detail.

Using the geographical information of wind turbine locations, a minimal spanning tree (MST) with r_m as the root node is constructed for each class C_m by using Prim's algorithm [23], as illustrated in Fig. 6. For each wind turbine i in C_m,

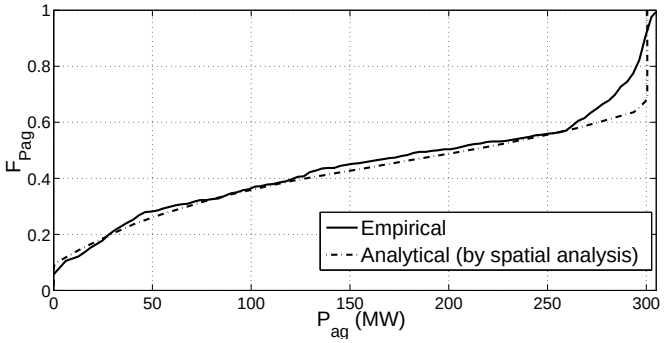


Fig. 8. CDF of $P_{ag}(t)$ for the 9 AM-noon epoch of January 2009.

there exists only one path from r_m to i in the MST of C_m. Define the node which is closest to i along this path as the 'parent' node of i. Another key observation from the measurement data is that an affine relationship exists between the power outputs of the parent-child turbine pairs for each class, with the case of C_1 illustrated in Fig. 7. Therefore, a coefficient α_m is introduced for C_m, and the linear regression model $P_k(t)=\alpha_m P_j(t)$ is used for each parent-child turbine pair (j,k) in C_m accordingly. Further, define $d_m(i)$ as the number of the nodes (excluding node i) along the path from r_m to node i, then the linear regression model $P_i(t)=\alpha_m^{d_m(i)} P_{r_m}(t)$ can be used for any wind turbine i in C_m. The value of α_m is determined by applying the minimum mean square error (MMSE) principle to the aggregate power output of C_m, as follows:

$$\alpha_m = \underset{\alpha}{\operatorname{argmin}} \frac{1}{N_t} \sum_t (P_{ag,m}(t) - \sum_{i \in C_m} \alpha^{d_m(i)} P_{r_m}(t))^2. \quad (2)$$

Similarly, an affine relationship between the wind speeds is also observed from the measurement data. For convenience, H_1 is chosen as the reference MET, i.e., $\overline{m}$=1. Then, the linear regression models for wind speeds are given by $W_m(t)=\beta_m W_{\overline{m}}(t)$, where β_m is solved using the MMSE principle as follows:

$$\beta_m = \underset{\beta}{\operatorname{argmin}} \frac{1}{N_t} \sum_t (W_m(t) - \beta W_{\overline{m}}(t))^2. \quad (3)$$

Using $P_{r_m}(t)=U_m(W_m(t))$, the aggregate power output of the wind farm could be characterized as follows:

$$P_{ag}(t) = \sum_m P_{ag,m}(t) = \sum_m \sum_{i \in C_m} \alpha_m^{d_m(i)} U_m(\beta_m W_{\overline{m}}(t))$$

$$\triangleq G_{pw}(W_{\overline{m}}(t)). \quad (4)$$

Due to the monotone characteristics of $U_m(\cdot)$, $G_{pw}(\cdot)$ is monotonically increasing. Therefore, the analytical CDF of $P_{ag}(t)$ can be obtained from the fitted Weibull distribution of $W_{\overline{m}}(t)$, given by $F_{P_{ag}}(\cdot)= F_{W_{\overline{m}}}(G_{pw}^{-1}(\cdot))$. The analytical CDF and the empirical CDF of $P_{ag}(t)$ for the considered epoch are illustrated in Fig. 8.

It is worth noting that the linear regression models with homogeneous regression coefficients used here are motivated by the observation from the measurement data. The above regression analysis could be generalized by applying more general regression analysis methods. For example, each parent-child turbine pair can have a different linear regression coefficient or the parent-child turbine pairs can be analyzed by using different regression models.

B. Temporal Analysis and LCR Quantification

During each epoch, both the wind speed $W_{\overline{m}}(t)$ and the wind farm generation $P_{ag}(t)$ could be regarded as stationary stochastic processes. The LCR of a stochastic process is formally defined as the number of instances per unit time that the stochastic process crosses a level in only the positive/negative direction [24]. Intuitively, $L_{P_{ag}}(\cdot)$ quantifies how frequently $P_{ag}(t)$ transits between different generation levels. It will be apparent soon that $L_{P_{ag}}(\cdot)$, together with the statistical characterization $F_{P_{ag}}(\cdot)$, is critical in designing the state space representation of the Markov chains used for wind farm generation forecast.

It is useful to note that due to the discontinuity in $F_{P_{ag}}(\cdot)$, as illustrated in Fig. 8, a smooth Gaussian transformation for $P_{ag}(t)$ is unattainable. Hence, the LCR of wind speed is first characterized. In order to quantify $L_{P_{ag}}(\cdot)$ analytically, $L_{W_{\overline{m}}}(\cdot)$ is first derived and converted to $L_{P_{ag}}(\cdot)$ by using the mapping defined in (4). To this end, autoregressive analysis is applied to $W_{\overline{m}}(t)$. As argued in [25], autoregressive analysis preceded by transforming the stationary non-Gaussian process $W_{\overline{m}}(t)$ to a Gaussian process can result in a better fit, compared with fitting to an autoregressive model directly. Therefore, $W_{\overline{m}}(t)$ is transformed to a standard normal random variable, given by

$$W_{\overline{m}}^{\mathcal{N}}(t) = F_{\mathcal{N}}^{-1}(F_{W_{\overline{m}}}(W_{\overline{m}}(t))), \quad (5)$$

A first-order autoregressive (AR(1)) model [26] is then fitted to $W_{\overline{m}}^{\mathcal{N}}(t)$:

$$W_{\overline{m}}^{\mathcal{N}}(t) = \phi W_{\overline{m}}^{\mathcal{N}}(t-1) + \epsilon(t), \quad (6)$$

where the white noise term is modeled as a zero-mean Gaussian random variable $\epsilon(t) \sim \mathcal{N}(0, \sigma_\epsilon^2)$. It is worth noting that the above AR(1) model is not used for short-term wind speed prediction. Instead, it is used to quantify the LCR of wind speed. The parameters ϕ and σ_ϵ of the above AR(1) model can be estimated by solving the Yule-Walker equations [26]. Then, the LCR of $W_{\overline{m}}^{\mathcal{N}}(t)$ for a specific wind speed level γ (γ>0) can be calculated using the following steps:

$$\begin{aligned} & L_{W_{\overline{m}}^{\mathcal{N}}}(\gamma) \\ &= \int_{-\infty}^{\gamma} \Pr(W_{\overline{m}}^{\mathcal{N}}(t) > \gamma | W_{\overline{m}}^{\mathcal{N}}(t-1) = w) f_{\mathcal{N}}(w) \mathrm{d}w \\ &= \int_{-\infty}^{\gamma} \Pr(\epsilon(t) > \gamma - \phi w) f_{\mathcal{N}}(w) \mathrm{d}w \\ &= \int_{-\infty}^{\gamma} \left(1 - F_{\mathcal{N}}\left(\frac{\gamma - \phi w}{\sigma_\epsilon}\right)\right) f_{\mathcal{N}}(w) \mathrm{d}w. \end{aligned} \quad (7)$$

Then, $L_{W_{\overline{m}}}(\cdot)$ can be obtained from $L_{W_{\overline{m}}^{\mathcal{N}}}(\cdot)$ using the inverse mapping of the strictly increasing function defined in (5). Further, using the monotonically increasing function defined in (4), the LCR of $P_{ag}(t)$ for a specific wind farm generation level Γ ($\Gamma \in (0, P_{ag}^{\max}]$) is given by:

$$L_{P_{ag}(t)}(\Gamma) = L_{W_{\overline{m}}^{\mathcal{N}}}(F_{\mathcal{N}}^{-1}(F_{W_{\overline{m}}}(G_{pw}^{-1}(\Gamma)))). \quad (8)$$

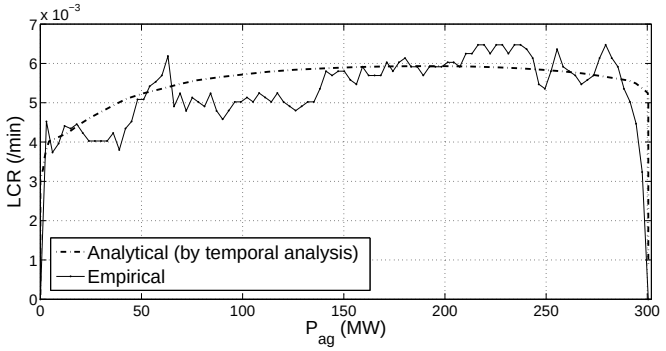


Fig. 9. LCR of $P_{ag}(t)$ for the 9 AM-noon epoch of January 2009.

The procedure presented above completes the characterization of the analytical LCR of $P_{ag}(t)$ for an arbitrary epoch. The analytical LCR and the empirical LCR of $P_{ag}(t)$ for the 9 AM-noon epoch of January 2009 are illustrated in Fig. 9.

C. Markov Chain Model for Spatio-temporal Wind Power

A critical step in developing the Markov-chain-based forecast approach is to capture the statistical distribution and the temporal dynamics of $P_{ag}(t)$ during each epoch using a Markov chain with the following characteristics:

- The Markov chain has finite states. Specifically, state S_k (k=1,$\cdots$,N_s) corresponds to a specific range of generation levels $[\Gamma_k,\Gamma_{k+1})$, with Γ_1=0 and Γ_{N_s+1}=$P_{ag}^{\max}$.
- The Markov chain is discrete-time and of order 1.

The above characteristics are adopted to make the Markov chains practical for forecasting applications, so that forecast is made based on the most recent 10-min data only.

The objective of the Markov chain design is to determine the generation levels Γ_k (k=1,$\cdots$,N_s+1) that defines the states, the transition matrix Q, and the representative generation level $P_{ag,k}$ for each state k. The procedure developed in [24] is utilized to design the state space. First, define τ_k as the average duration for which $P_{ag}(t)$ stays in S_k, given by:

$$\tau_k = \frac{F_{P_{ag}}(\Gamma_{k+1}) - F_{P_{ag}}(\Gamma_k)}{L_{P_{ag}}(\Gamma_{k+1}) + L_{P_{ag}}(\Gamma_k)}, \tag{9}$$

where $F_{P_{ag}}(\cdot)$ is the analytical CDF of $P_{ag}(t)$ that was characterized in spatial analysis, and $L_{P_{ag}}(\cdot)$ is the analytical LCR of $P_{ag}(t)$ derived in temporal analysis. Note that τ_k plays a critical role in the Markov chain model and determines how well the stochastic process $P_{ag}(t)$ is captured:

- A smaller value of τ_k suggests that $P_{ag}(t)$ is more likely to switch out of the state S_k within a 10-min slot, i.e., non-adjacent transitions are more likely to occur, *and hence the transitional behaviors of $P_{ag}(t)$ are not sufficiently captured by the discrete-time Markov chain.*
- If the values of τ_k (k=1,$\cdots$,N_s) are too large, there would be fewer states, indicating that the quantization by the Markov chain is too crude, and the corresponding forecast would be less accurate.

Therefore, a key objective of state space design is to make each of τ_k (k=1,$\cdots$,N_s) fall into a reasonable range [24]. However, it is challenging to achieve this design goal, especially when the closed-form expressions of $F_{P_{ag}}(\cdot)$ and $L_{P_{ag}}(\cdot)$ are unattainable. A practical solution adopted here is to introduce a constant τ and find the N_s-1 variables $\{\Gamma_2,\Gamma_3,\cdots,\Gamma_{N_s}\}$ by solving (9) numerically with τ_k=τ, $\forall k\in\{1,\cdots,N_s-1\}$. Once the state space $\mathcal{S}$ is designed, the transition probabilities can be estimated following the approach proposed in [7]. Specifically, the probability of a transition from S_i to S_j is given by

$$Q_{i,j} = \frac{n_{ij}}{\sum_{k=1}^{N_s} n_{ik}},\ i,j \in \{1,\cdots,N_s\}, \tag{10}$$

The representative generation level for each state S_k, $k\in\{1,\cdots,N_s\}$, is determined using the MMSE principle, given by (the time index of $P_{ag}(t)$ is dropped for simplicity):

$$P_{ag,k} = \underset{P_k}{\operatorname{argmin}}\, \mathbf{E}\left[(P_k - P_{ag})^2 | P_{ag} \in [\Gamma_k, \Gamma_{k+1})\right], \tag{11}$$

Then, the representative generation level is given by:

$$P_{ag,k} = \frac{\int_{\Gamma_k}^{\Gamma_{k+1}} x f_{P_{ag}}(x)\mathrm{d}x}{F_{P_{ag}}(\Gamma_{k+1}) - F_{P_{ag}}(\Gamma_k)}. \tag{12}$$

The above procedure is applied to the 9 AM-noon epoch of January 2009, by choosing τ=2 min. The boundaries for

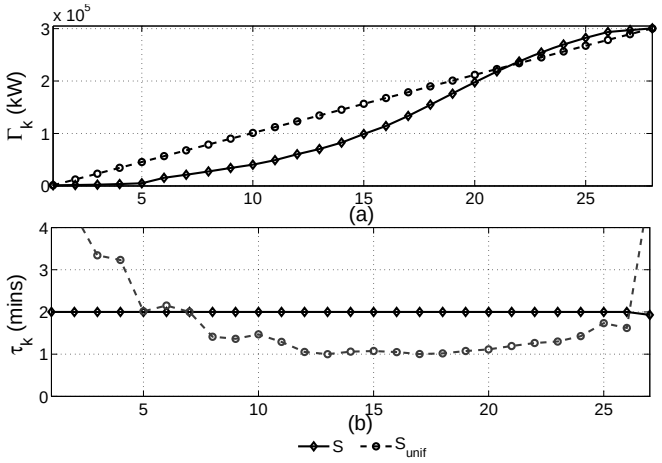


Fig. 10. Boundaries and average duration for each state of the Markov chain for the 9 AM-noon epoch of January 2009.

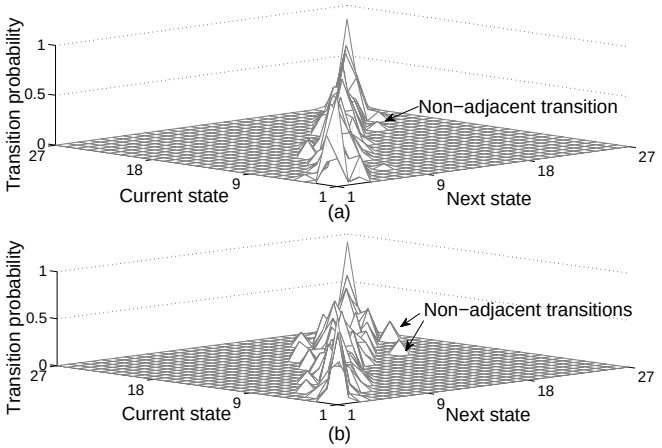


Fig. 11. Transition matrix (a) by spatio-temporal analysis (b) by uniform quantization, for the 9 AM-noon epoch of January 2009.

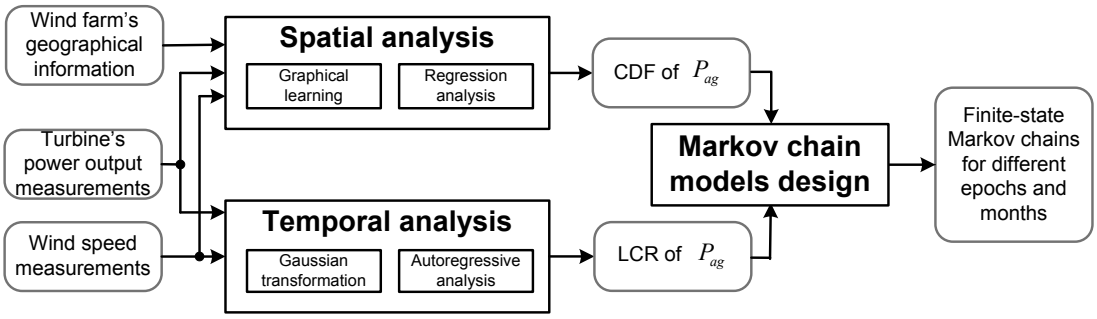


Fig. 12. Offline spatio-temporal analysis (carried out for each epoch and each month by using historical measurement data).

each state are illustrated in Fig. 10(a), and the corresponding transition probabilities are plotted in Fig. 11(a). In [7], [8], the Markov chain for wind power (not in the context of *wind farm* generation) is obtained by uniform quantization. By choosing $\Gamma_{k+1}=P_{ag}^{\max}k/N_s$, $\forall k\in\{1,\cdots,N_s-1\}$, the resultant state space, denoted by $\mathcal{S}_{\text{unif}}$, is compared with $\mathcal{S}$. From Fig. 10(b), it is clear that higher values of τ_k are achieved for most of the states in $\mathcal{S}$. Hence, fewer non-adjacent transitions are incurred by $\mathcal{S}$, as can be seen from Fig. 11.

IV. Markov-Chain-based Short-term Forecast of Wind Farm Generation

As illustrated in Fig. 12 and Fig. 13, the proposed approach for short-term wind farm generation forecasting consists of two major steps: offline spatio-temporal analysis and online forecasting. These two steps utilize two types of information to provide both distributional forecasts and points forecasts. Specifically, in offline spatio-temporal analysis, the procedures presented in Section III are carried out on historical data of turbines' power output and wind speed, for each epoch and each month, to build multiple Markov chains by capturing the statistical characteristics from the historical data. It is worth noting that Weibull parameter estimation is part of spatio-temporal analysis. The inputs to the spatial analysis sub-step are the wind farm's geographical information and historical data of each wind turbine's power output. Historical data of wind turbines' power output and wind speed is used by the temporal analysis sub-step. In online forecasting, the Markov chains obtained are applied to the real-time measurement of wind farm generation to provide both distributional forecasts and point forecasts. Specifically, the transition probabilities of Markov chains determine the *conditional* probability distribution of future wind power $\hat{P}_{ag}(t+1)$, i.e., the probability distribution of $\hat{P}_{ag}(t+1)$ conditioned on the real-time wind power measurement $P_{ag}(t)$.

In what follows, short-term distribution forecasts and point forecasts are first derived by using the three inputs to the online forecasting step: 1) the Markov chain developed for the present epoch and month, 2) the wind farm's present aggregate power output $P_{ag}(t)$ at time t, and 3) short-term complementary information that can be utilized to enhance forecasting (e.g., ramp trend information). Then, the developed forecasting methods, with the parameters of the Markov chain models computed by using 2009 measurement data, are tested on the corresponding 2010 measurement data. For example, the forecasting method with the Markov chain developed based on the measurement data in the 9 AM-noon epochs of January 2009 will be applied to the measurement data in the 9 AM-noon epochs of January 2010 only.

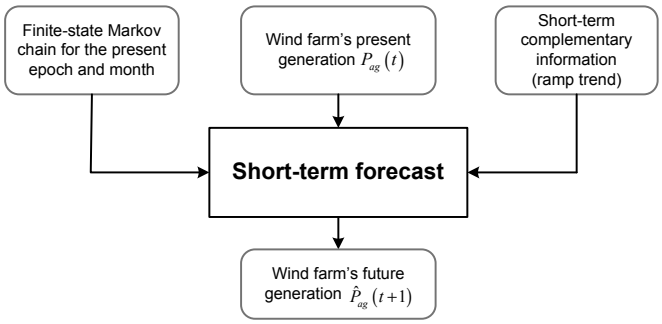


Fig. 13. Online short-term forecasting.

A. Short-term Distributional Forecasts and Point Forecasts

To derive a short-term forecast by using the Markov chain, it is worth noting that some complementary information can be utilized. One such complementary information is the ramp trend of wind farm generation. It is observed from available data that wind farm generation usually increases or decreases for several consecutive time-slots. Therefore, the ramp trend can be used to "steer" the transition of the Markov chain.

1) Distributional Forecasts: Given the current 10-min wind farm generation data $P_{ag}(t)$, the state of the Markov chain at time t, denoted by $S(t)$, is determined by searching for a state k_0 so that $P_{ag}(t)\in[\Gamma_{k_0},\Gamma_{k_0+1})$. Thus, $S(t+1)$ and hence $P_{ag}(t+1)=P_{ag,S(t+1)}$ are random variables that depend on the transition matrix Q, $S(t)$ and $R(t)$. Further, let $R(t)=-1$ denote a decreasing trend, and $R(t)=1$ for the non-decreasing case. Then, the distributional forecast is given by

$$\Pr(P_{ag}(t+1)=P_{ag,j}|S(t),R(t))=\begin{cases}\dfrac{Q_{k_0,j}}{\sum_{k\geq k_0}^{N_s}Q_{k_0,k}}, & \text{if } R(t)=1 \text{ and } j\geq k_0\\ \dfrac{Q_{k_0,j}}{\sum_{k=1}^{k_0-1}Q_{k_0,k}}, & \text{if } R(t)=-1 \text{ and } j<k_0\\ 0, & \text{otherwise.}\end{cases} \tag{13}$$

2) Point Forecasts: From the above distributional forecast, a point forecast can be derived by using the MMSE principle:

$$\hat{P}_{ag}(t+1) = \underset{P_{ag}}{\operatorname{argmin}} \mathbf{E}\left[(P_{ag} - P_{ag,S(t+1)})^2 | S(t), R(t)\right] \tag{14}$$

Then, the solution to the above problem is given by:

$$\hat{P}_{ag}(t+1) = \begin{cases} \dfrac{\sum_{k\geq k_0}^{N_s} P_{ag,k} Q_{k_0,k}}{\sum_{k\geq k_0}^{N_s} Q_{k_0,k}}, & \text{if } R(t) = 1 \\ \dfrac{\sum_{k=1}^{k_0-1} P_{ag,k} Q_{k_0,k}}{\sum_{k=1}^{k_0-1} Q_{k_0,k}}, & \text{if } R(t) = -1 \end{cases} \tag{15}$$

which is exactly the mean value of the Markov chain conditioned on the currently observed state and the ramp trend.

B. Numerical Examples

1) Distributional Forecasts: The continuous rank probability score (CRPS) is utilized to quantitatively assess the performance of Markov-chain-based distributional forecasts, given by:

$$CRPS = \frac{1}{N_t} \sum_t \int_0^{P_{ag}^{\max}} \left(\hat{F}(x) - H(x - P_{ag}(t))\right)^2 dx, \tag{16}$$

where N_t is the total number of data points, $\hat{F}(x)$ is the CDF obtained by using the Markov-chain-based distributional forecast, and $H(x - P_{ag}(t)$ is the unit step function, which takes value 0 when $x<P_{ag}(t)$ and takes value 1 when $x\geq P_{ag}(t)$. Basically, *a higher CRPS value suggests that the distributional forecast is less accurate*. By using the above definition, the CRPS value of the Markov-chain-based distributional forecast over all the 52560 (365*24*6) data points of the year 2010 is calculated. The CRPS of the Markov-chain-based distributional forecast over the data points of the year 2010 is provided in Table. I. Since one main objective of this work is to develop Markov-chain-based distributional forecasting models, the Markov chain developed by the existing approach [7], [8] (uniform quantization) is used as a benchmark. The Markov chain developed by the proposed spatio-temporal analysis with the design parameter τ=2 (column 3 of Table. I) has a CRPS that is 13% less than that of the benchmark Markov chain that has the same number of states designed by uniform quantization (column 2 of Table. I). By reducing the design parameter τ to 1, the forecasting performance of the Markov chain developed by the proposed spatio-temporal analysis (column 4 of Table. I) is further improved.

The proposed Markov-chain-based distributional forecasts are also compared with the distributional forecasts based on high-order AR models. Here, two high-order AR models with a truncated Gaussian distribution and a truncated log-normal distribution are considered. The high-order AR model with a Gaussian distribution is adopted from [5] by considering one regime, and then the support of the Gaussian distribution is truncated into $[0, P_{ag}^{\max}]$. The procedure for building AR models with truncated log-normal distributions

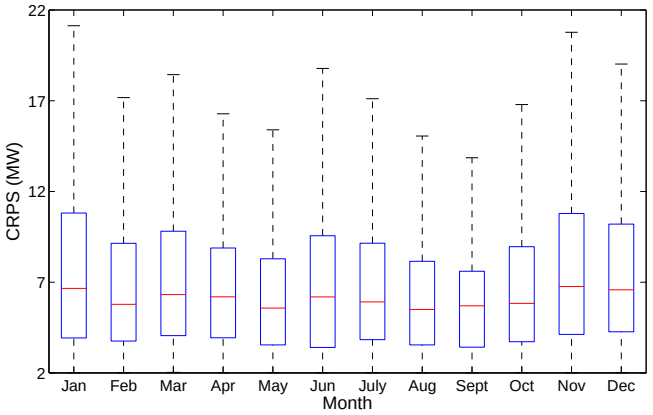


Fig. 14. Statistics of CRPS over all months of the year 2010.

TABLE I
CRPS OF DISTRIBUTIONAL FORECASTS OVER THE TESTING DATA POINTS OF THE YEAR 2010.

	MC (unif.)	MC (τ=2)	MC (τ=1)	AR (Gaussian)	AR (Log-normal)
CRPS	7.14 MW	6.27 MW	6.09 MW	6.89 MW	6.54 MW

can be found in [16]. Specifically, the order of the AR models are determined by using the partial autocorrelation functions of the wind power time series [27]. Then, the recursive least square algorithm [27] is applied to calculate the regressive coefficients, the predicted wind power $\hat{P}_{ag}(t)$ (the point forecast of the AR model), and the variance of innovation $\mathbf{C}$. Finally, by using $\hat{P}_{ag}(t)$ as the mean and $\mathbf{C}$ as the variance of a Gaussian distribution or a log-normal distribution which is truncated into $[0, P_{ag}^{\max}]$, the wind power distributional forecasts can be obtained. The CRPS values of the distributional forecasts based on high-order AR models are calculated by using (16), and are shown in Table I. It can be seen from Table I that the Markov-chain-based distributional forecasts with the design parameter τ=1 (column 4 of Table I) achieves a CRPS value that is 11.6% and 6.9% lower than those of the AR-based distributional forecasts (column 5 and column 6 of Table I), respectively. The reason for this improvement of Markov-chain-based distributional forecast is that the conditional probability distributions provided by Markov chains do not assume the shape of the distribution (and thus can be regarded as "non-parametric" distributional forecasts in literature [13]). Therefore, by using the transition probability estimated from historical data, Markov chains can provide more accurate distributional forecasts than those based on assumed parametric distributions (e.g., Gaussian, β and log-normal distributions). The superiority of non-parametric distributional forecasts over parametric ones is also discussed in [13] and references therein. In summary, the improvement of the developed Markov-chain-based approach over other approaches can be attributed to the rigorous design of Markov chains and transition probabilities, which in turn utilizes the analytical results from spatio-temporal analysis.

To further examine the performance of the developed Markov-chain-based distributional forecasting method over different epochs and different month, the median and per-

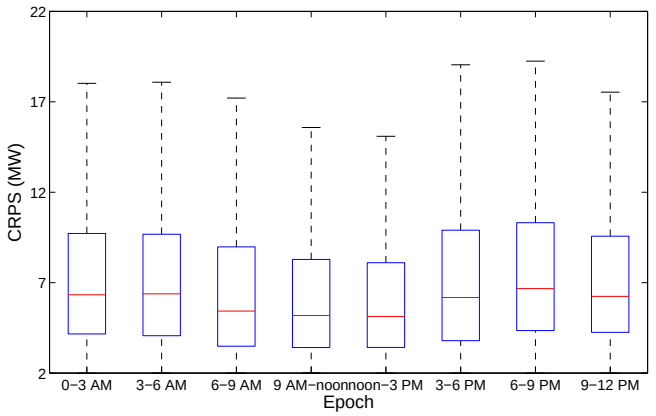


Fig. 15. Statistics of CRPS over all 8 epochs of the year 2010.

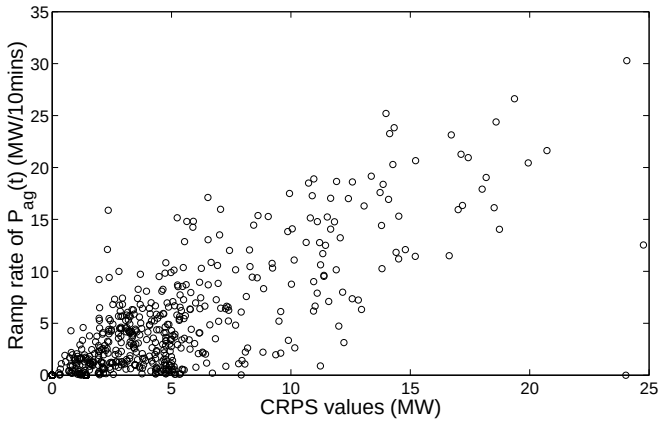


Fig. 16. Correlation between the ramp rates of $P_{ag}(t)$ and the CRPS values of distributional forecast.

centiles of the CRPS values over the data points for each month or each epoch is computed. In the box plots of Fig. 14 and Fig. 15, the central bar in a box represents the median value of the CRPS values over all data points that fall into a specific epoch or a specific month. The top edge and bottom edge of a box represent the 25th and 75th percentiles, respectively. The top bar and bottom bar correspond to the extremes calculated from 1.5 interquartile ranges. It is observed from Fig. 15 that the medians and standard deviations of the CRPS values are a little higher during afternoon-night epochs. Fig. 14 shows that the medians of the CRPS values have little variability across different months, and the standard deviations of the CRPS values are slightly higher across the winter season. Another key observation from the results of numerical experiments is that *the CRPS of the Markov-chain-based distributional forecast over a realized data points $P_{ag}(t)$ is highly dependent on the ramp rate of $P_{ag}(t)$ at time t*. Here, the ramp rate of $P_{ag}(t)$ is defined as the absolute value of the change in the wind farm generation in a 10-min slot. For example, the ramp rate of $P_{ag}(t)$ at time t is given by $|P_{ag}(t)-P_{ag}(t-1)|$. By using the data points of the year 2010, the corresponding pairs of ramp rates and CRPS values are plotted in Fig. 16. It is observed that the ramp rates of $P_{ag}(t)$ and the CRPS values of the Markov-chain-based distributional forecast over a realized data points $P_{ag}(t)$ follows a *positive correlation*. The above observation also explains the 'phase transition' from the noon-3 PM epoch to the 3-6 PM in Fig. 15, i.e., the increased wind ramp caused by the sudden change in diurnal heating/vertical mixing conditions [28]. In summary, the statistics (especially the median value) of the CRPS values vary slightly differently over different months and epochs, which suggests that the developed Markov-chain-based distributional forecasting methods deliver consistent forecasting performance across the entire year.

Further, three episodes of prediction intervals are plotted to better illustrate the developed Markov-chain-based distributional forecasts. According to the above observation, three representative time periods are chosen: 1) the 0-3 AM epoch of January 23rd, 2) the 3-6 PM epoch of January 16th, and 3) the 3-6 PM epoch of April 16th. The first period in January 23rd is chosen since it has much higher average ramp rate than

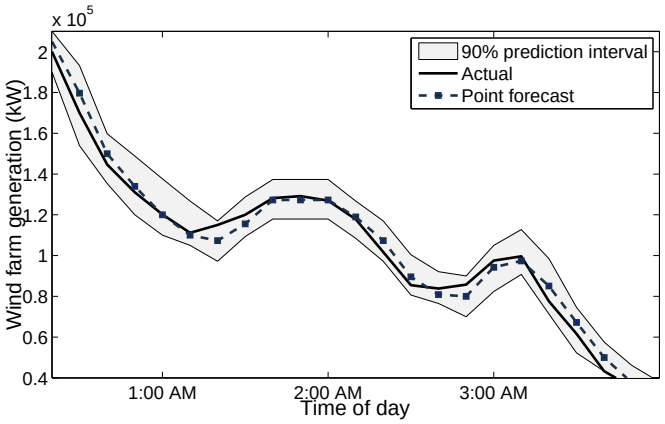


Fig. 17. 10-min distributional forecasts on January 23rd, 2010.

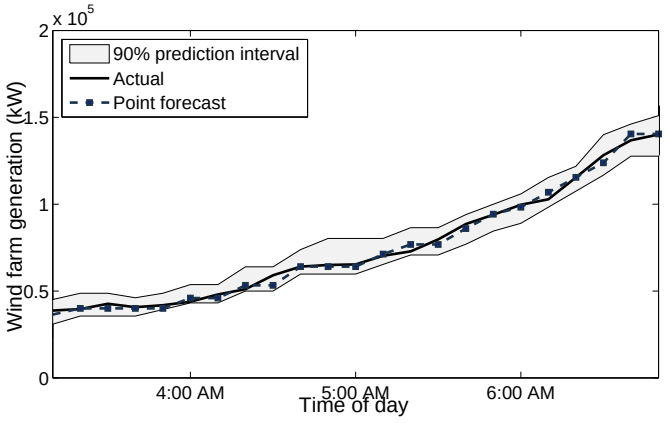


Fig. 18. 10-min distributional forecasts on January 16th, 2010.

other January days, and the 0-3 AM epoch experienced a large down-ramp from 75% to 25% of the rated capacity. The second period is chosen because January and the 3-6 PM epoch have the highest median CRPS value (i.e., least accurate forecasts), and the CRPS value of January 16th is mostly close to the corresponding median value. The third period is chosen due to similar reasons as the second period, except that April is the month that has the least CRPS values. Fig. 17-19 illustrate the 90% prediction intervals obtained by the developed Markov-

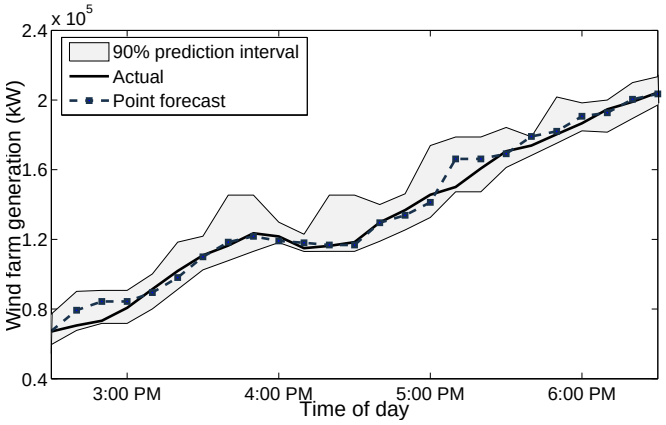


Fig. 19. 10-min distributional forecasts on April 16th, 2010.

TABLE II
10-MIN POINT FORECAST ERROR OF WIND FARM GENERATION (ALL TEST DATA OF THE YEAR 2010 IS USED).

Error	Persistence	MC (unif)	MC (τ=2)	MC (τ=1)	AR
MAE	6.98 MW	7.14 MW	6.83 MW	6.62 MW	6.38 MW
MAPE	7.31 %	7.48 %	7.15 %	6.93 %	6.68 %
RMSE	11.18 MW	11.58 MW	10.89 MW	10.56 MW	10.25 MW

TABLE III
10-MIN POINT FORECAST ERROR OF WIND FARM GENERATION OVER THE PERIOD SHOWN IN FIG. 17.

Error	Persistence	MC (unif)	MC (τ=2)	MC (τ=1)	AR
MAE	13.26 MW	13.83 MW	9.97 MW	9.59 MW	9.39 MW
MAPE	11.61 %	12.1 %	8.73 %	8.4 %	8.22 %
RMSE	15.81 MW	16.26 MW	12.81 MW	12.23 MW	11.94 MW

TABLE IV
10-MIN POINT FORECAST ERROR OF WIND FARM GENERATION OVER THE PERIOD SHOWN IN FIG. 18.

Error	Persistence	MC (unif)	MC (τ=2)	MC (τ=1)	AR
MAE	4.6 MW	4.71 MW	4.54 MW	4.32 MW	4.28 MW
MAPE	6.28 %	6.43 %	6.2 %	5.9 %	5.84 %
RMSE	6.16 MW	6.32 MW	6.09 MW	5.91 MW	5.86 MW

TABLE V
10-MIN POINT FORECAST ERROR OF WIND FARM GENERATION OVER THE PERIOD SHOWN IN FIG. 19.

Error	Persistence	MC (unif)	MC (τ=2)	MC (τ=1)	AR
MAE	6.02 MW	6.31 MW	4.95 MW	4.81 MW	4.73 MW
MAPE	4.64 %	4.86 %	3.82 %	3.71 %	3.65 %
RMSE	6.86 MW	7.17 MW	5.73 MW	5.41 MW	5.23 MW

chain-based distributional forecasts. It is observed at for all three representative periods, the realized wind farm generation reasonably lies in the 90% prediction intervals.

2) Point Forecasts: By comparing the point forecast $\hat{P}_{ag}(t)$ with the actual wind farm generation $P_{ag}(t)$, forecast errors are quantified by mean absolute error (MAE), defined as

$$\text{MAE} = \frac{1}{N_t}\sum\nolimits_t |P_{ag}(t) - \hat{P}_{ag}(t)|, \tag{17}$$

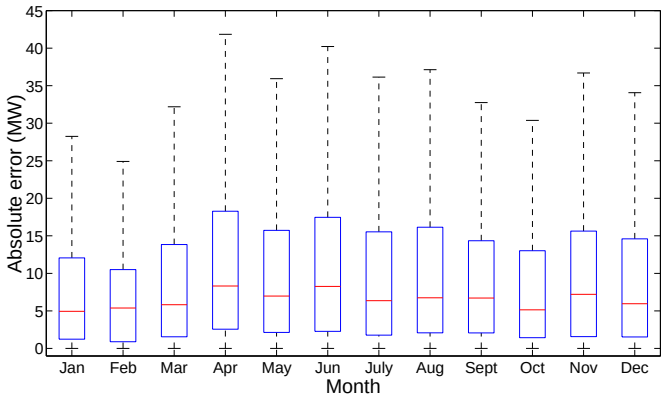


Fig. 20. Statistics of absolute error over all months of the year 2010.

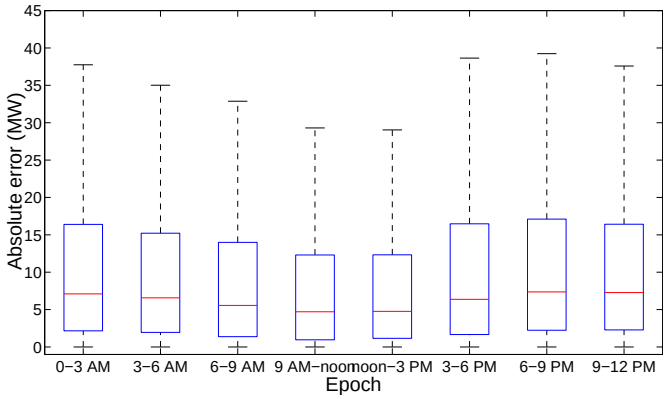


Fig. 21. Statistics of absolute error over all 8 epochs of the year 2010.

mean absolute percentage error (MAPE), defined as

$$\text{MAPE} = \frac{\sum_t |P_{ag}(t) - \hat{P}_{ag}(t)|}{\sum_t P_{ag}(t)}, \tag{18}$$

and root mean square error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{\sum_t |P_{ag}(t) - \hat{P}_{ag}(t)|^2}{N_t}}. \tag{19}$$

Besides AR models, two point forecast approaches are used as benchmark:

- persistence forecast [29]: $\hat{P}_{ag}(t+1)$=$P_{ag}(t)$;
- forecast by Markov Chain with uniform quantization.

The proposed Markov-chain based forecast method is compared with several state-of-the-art approaches. Specifically, the wind power data used for forecast is first mapped to the state space designed by following the procedure in Section III-C. Then, point forecasts are obtained by using the representative generation levels of corresponding states. The test results by using the data for the year 2010 and the three selected epochs are provided in Table II-V, respectively. It is observed that the Markov chains based on uniform quantization give less accurate forecast than persistence forecast. This can be attributed to the uniform quantization not considering the spatio-temporal dynamics of wind farm generation. Also note that the proposed Markov-chain-based forecast approach has improved accuracy compared to the persistence forecast approach, and

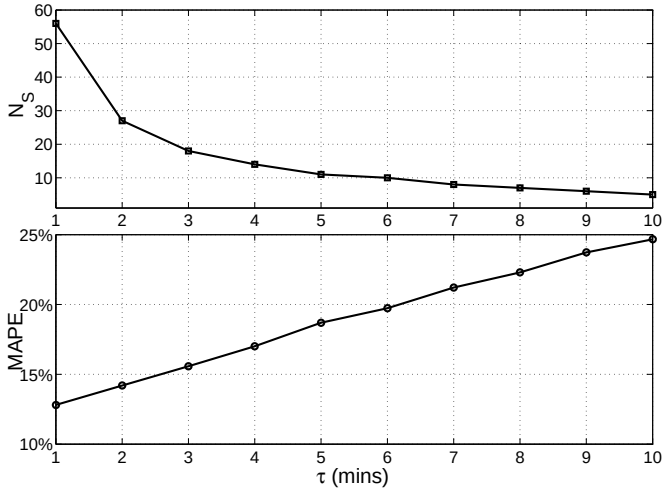


Fig. 22. Number of states and the forecast error of Markov chains at various τ for the January 9 AM-noon epoch).

comparable accuracy to the AR-based approach. Further, the statistics of the absolute error of the Markov-chain-based point forecasts over different months and different epochs are illustrated in Fig. 20 and Fig. 21, respectively. It can be seen from Fig. 20 and Fig. 21 that the developed Markov-chain-based point forecasting methods perform consistently across the entire year.

Another key observation from Table IV and Table II is that smaller values of τ leads to higher forecast accuracy of the Markov chains, at the cost of higher complexity of the Markov chains (in terms of the number of states). The trade-off between the forecast accuracy and the complexity of the Markov chain for the 9 AM-noon epoch of January 2010 is illustrated in Fig. 22.

From the results presented above, it can be seen that the proposed distributional forecast approach outperforms the high-order AR-based distributional forecasts with Gaussian and log-normal distributions. This is because the proposed spatio-temporal analysis extracts from historical data the rich statistical information of wind farm generation, and accordingly the corresponding Markov chain models can provide more accurate distributional forecasts than AR-based models with assumed Gaussian and log-normal distributions. Further, the proposed point forecasts have a lightly higher mean absolute error (MAE) than those of high-order AR-based forecasts. However, note that one main objective of this study is to develop Markov-chain-based distributional forecasts that can be used for economic dispatch in the presence of wind generation uncertainty [18], [19], in which a good balance is needed between computational complexity and modeling accuracy. Here, computational complexity involves both the computational effort for building and utilizing the forecasting models to provide distributional forecasts and the computational effort for solving stochastic economic dispatch problems by using these distributional forecasts. Therefore, compared to AR-based distributional forecast methods, the developed Markov chain models are more suitable for stochastic economic dispatch, because the computational burden of using continuous distributions of AR-based forecasts for stochastic economic dispatch would be significantly higher. Moreover, even though the computational effort of using AR-based distributional forecasts can be reduced by applying quantization (i.e., 0-300 MW quantized into 50-70 states for the cases in this study) and scenario reduction, the quantization error would cause the quantized AR-based forecasts to be even less accurate than the proposed Markov-chain-based forecasts. In summary, the proposed Markov-chain-based distributional forecast approach achieves higher accuracy than existing approaches, and the well-balanced complexity and accuracy of Markov chain models make them an ideal tool to study stochastic economic dispatch problems.

V. Conclusion

A general spatio-temporal analysis framework is developed for wind farm generation forecast, in which finite-state Markov chain models are derived. The state space, transition matrix and representative generation levels of the Markov chains are optimized by using a systematic approach. The short-term distributional forecast and point forecast are derived by using the Markov chains and the ramp trend information. One main contribution of this study is that the distributional forecast can be directly integrated into the problems of unit commitment and economic dispatch with uncertain wind generation, so that these problems can be studied in a general Markov-chain-based stochastic optimization framework.

In a related work [18], we are investigating power system economic dispatch with wind farm generation by utilizing a realistic test system and the Markov-chain-based distributional forecasts of wind farm generation. The distributional forecasts of wind farm generation are integrated into a stochastic programming framework of multi-period economic dispatch, so as to optimize the dispatch decisions over the operating horizon. The impact of the forecast errors of wind farm generation on economic dispatch is also studied.

Acknowledgement

The authors are grateful to National Renewable Energy Laboratory (NREL) and Xcel Energy for providing the data used in this study.

References

[1] K. S. Cory and B. G. Swezey, "Renewable portfolio standards in the states: balancing goals and implementation strategies," *NREL Technical Report TP-670-41409*, Dec. 2007.

[2] E. A. DeMeo, G. A. Jordan, C. Kalich, J. King, M. R. Milligan, C. Murley, B. Oakleaf, and M. J. Schuerger, "Accommodating wind's natural behavior," *IEEE Power Energy Mag.*, vol. 5, pp. 59–67, Nov.-Dec. 2007.

[3] D. Lew, M. Milligan, G. Jordan, and R. Piwko, "The value of wind power forecasting," *NREL Conference Paper CP-5500-50814*, Apr. 2011.

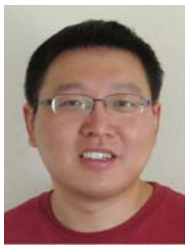
[4] S. Fink, C. Mudd, K. Porter, and B. Morgenstern, "Wind energy curtailment case studies." NREL Subcontract Report SR-550-46716, Oct. 2009.

[5] P. Pinson and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models," *Journal of Forecasting*, vol. 31, no. 4, pp. 281–313, 2012.

[6] F. Cassola and M. Burlando, "Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output," *Applied Energy*, vol. 99, pp. 154–166, 2012.

[7] G. Papaefthymiou and B. Klockl, "MCMC for wind power simulation," *IEEE Trans. on Energy Convers.*, vol. 23, pp. 234–240, Mar. 2008.
[8] A. Carpinone, R. Langella, A. Testa, and M. Giorgio, "Very short-term probabilistic wind power forecasting based on Markov chain models," in *Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on*, pp. 107–112, June 2010.
[9] S. Santoso, M. Negnevitsky, and N. Hatziargyriou, "Data mining and analysis techniques in wind power system applications: abridged," in *Power Engineering Society General Meeting, 2006. IEEE*, pp. 1–3, 2006.
[10] A. Kusiak, H. Zheng, and Z. Song, "Wind farm power prediction: a data-mining approach," *Wind Energy*, vol. 12, no. 3, pp. 275–293, 2009.
[11] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, *The State of the Art in Short-Term Prediction of Wind Power - A Literature Overview*. ANEMOS.plus, 2011. [Online] Available: http://www.anemos-plus.eu/images/pubs/deliverables/aplus.deliverable_d1%.2.stp_sota_v1.1.pdf.
[12] C. Monteiro, H. Keko, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, "A quick guide to wind power forecating: state-of-the-art 2009." [Online] Available: http://www.dis.anl.gov/pubs/65614.pdf, 2009.
[13] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845–1856, 2010.
[14] "NSF Initiative on Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)." Online [Available]:http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm#toc.
[15] A. Lau and P. McSharry, "Approaches for multi-step density forecasts with application to aggregated wind power," *Ann. Appl. Stat.*, vol. 4, no. 3, pp. 1311–1341, 2010.
[16] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 61, no. 4, pp. 555–576, 2012.
[17] S. Murugesan, J. Zhang, and V. Vittal, "Finite state Markov chain model for wind generation forecast: a data-driven spatio-temporal approach," *Innovative Smart Grid Technologies, IEEE PES*, pp. 1–8, Jan. 2012.
[18] M. He, L. Yang, J. Zhang, and V. Vittal, "Spatio-temporal analysis for smart grids with wind generation integration," in *Computing, Networking and Communications (ICNC), 2013 International Conference on*, pp. 1107–1111, 2013.
[19] P. Luh, Y. Yu, B. Zhang, E. Litvinov, T. Zheng, F. Zhao, J. Zhao, and C. Wang, "Grid Integration of Intermittent Wind Generation: a Markovian Approach." in press, *IEEE Trans. Smart Grids*.
[20] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.
[21] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance (Stochastic Modeling Series)*. Chapman and Hall/CRC, 1994.
[22] E. S. Tackle and J. M. Brown, "Note on the use of Weibull statistics to characterize wind speed data," *Journal Appl. Meteorol*, vol. 17, pp. 556–559, 1978.
[23] R. C. Prim, "Shortest connection networks and some generalizations," *Bell System Technical Journal*, vol. 36, pp. 1389–1401, 1957.
[24] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. on Commun.*, vol. 47, pp. 1688–1692, Nov. 1999.
[25] D. Kugiuntzis and E. Bora-Senta, "Gaussian analysis of non-Gaussian time series," *Brussels Economic Review*, vol. 53, no. 2, pp. 295–322, 2010.
[26] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control, 2nd ed*. San Francisco: Holden-Day, 1976.
[27] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Wiley, 1996.
[28] NERC IVGTF Task 2.1 report: Variable Generation Power Forecasting for Operations. www.nerc.com/docs/pc/ivgtf/Task2-1(5.20).pdf, May 2010.
[29] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen, "Standardizing the performance evaluation of short-term wind power prediction models," *Wind Engineering*, vol. 29, no. 6, pp. 475–489, 2005.

Miao He (S'08) received the B.S. degree from Nanjing University of Posts and Telecommunications, in 2005, and the M.S. degree from Tsinghua University, in 2008, and the Ph.D. degree from Arizona State University, in 2013. Currently, he is an Assistant Professor at Texas Tech University. His research is focused on stochastic modeling and data analytics for smart grids, wind power systems and cyber physical systems.



Lei Yang (M'13) received the B.S. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, in 2005 and 2008, respectively, and the Ph.D. degree from the School of ECEE at Arizona State University, Tempe, in 2012. He has been an Assistant Research Professor with the School of ECEE at Arizona State University since 2013. His research interests include stochastic optimization and big data analytics for renewable energy integration, grid integration of plug-in electric vehicle, networked control of cyber-physical systems, modeling and control of power systems, network security and privacy, network optimization and control, and cognitive radio.

Junshan Zhang (F'12) received his Ph.D. degree from the School of ECE at Purdue University in 2000. He joined the EE Department at Arizona State University in August 2000, where he has been Professor since 2010. His research interests include communications networks, cyber-physical systems with applications to smart grid, stochastic modeling and analysis, and wireless communications.

Dr. Zhang is a recipient of the ONR Young Investigator Award in 2005 and the NSF CAREER award in 2003. He received the Outstanding Research Award from the IEEE Phoenix Section in 2003.

Vijay Vittal (S'78-F'97) received the B.E. degree in electrical engineering from the B.M.S. College of Engineering, Bangalore, India, in 1977, the M.Tech. degree from the India Institute of Technology, Kanpur, India, in 1979, and the Ph.D. degree from Iowa State University, Ames, in 1982.

Currently, he is the Ira A. Fulton Chair Professor in the Electrical Engineering Department, Arizona State University, Tempe.

Dr. Vittal received the 1985 Presidential Young Investigator Award and the 2000 IEEE Power Engineering Society Outstanding Power Engineering Educator Award. He is a member of the National Academy of Engineering.