A Nonparametric Bayesian Framework for Short-Term Wind Power Probabilistic Forecast

Wei Xie, Pu Zhang, Rong Chen, Zhi Zhou *Member, IEEE*

***Abstract*—To improve the energy system resilience and economic efficiency, the wind power as a renewable energy starts to be deeply integrated into smart power grids. However, the wind power forecast uncertainty brings operational challenges. In order to provide a reliable guidance on operational decisions, in this paper, we propose a short-term wind power probabilistic forecast. Specifically, to model the rich dynamic behaviors of underlying physical wind power stochastic process occurring in various meteorological conditions, we first introduce an infinite Markov switching autoregressive model. This nonparametric time series model can capture the important properties in the real-world data to improve the prediction accuracy. Then, given finite historical data, the posterior distribution of flexible forecast model can correctly quantify the model estimation uncertainty. Built on it, we develop the posterior predictive distribution to rigorously quantify the overall forecasting uncertainty accounting for both inherent stochastic uncertainty and model estimation error. Therefore, the proposed approach can provide accurate and reliable short-term wind power probabilistic forecast, which can be used to support smart power grids real-time risk management.**

***Index Terms*—Bayesian nonparametric approach, probabilistic forecast, predictive distribution, prediction interval, wind power.**

NOMENCLATURE

A. Abbreviation

AR	Autoregressive
BELM	Bootstrap-based Extreme Learning Machine
DP	Dirichlet Process
HDP	Hierarchical Dirichlet Process
IHMM	Infinite Hidden Markov Model
IMSAR	Infinite Markov Switching Autoregressive
MSAR	Markov Switching Autoregressive
PI	Prediction Interval
TVQR	Time-varying Quantile Regression

B. Wind Power Data and Variables

T	Index of current time period or the size of historical data
$\mathbf{x}_{[1:T]}$	Historical wind power data stream
$\mathbf{x}_{[T+1:T+\tau]}$	Wind power forecast in next τ time periods
s_t	Latent meteorological state at time period t

C. IMSAR Forecast Model Parameters

p	Order of AR time series kernel
$\boldsymbol{\theta}_i$	AR time series kernel parameters in state i
$\boldsymbol{\psi}(i), \sigma^2(i)$	Components of AR parameters $\boldsymbol{\theta}_i$
$G_{\boldsymbol{\theta}}$	Base DP prior for $\boldsymbol{\theta}$
α	Concentration parameter of the global DP
G_0	Global DP prior for state transition
η	Concentration parameter of the state-conditional DPs
G_i	State-conditional DP prior for transition probability distribution from state i.
$\boldsymbol{\pi}$	Global probability measure on the states
$\boldsymbol{p}_i$	Transition probability from state i

D. Hyperparameters of the Priors for AR parameters

a, b	Hyper-parameters of Gamma prior for $\sigma^2(i)$
μ_ψ, σ_ψ	Hyper-parameters of Normal prior for $\boldsymbol{\psi}(i)$

E. Variables and Data for Bayesian Inference

K	Number of active states which are visited by historical wind power data
$\boldsymbol{m}$	Vector of visits to active states
m_i	Number of visits to state i
$\boldsymbol{N}$	Transition matrix with $N_{i,j}$ recording the number of transitions from active state i to j for $i, j = 1, \ldots, K$
T_i	Set of time indexes when $s_t = i$
$\boldsymbol{x}_{[T_i]}$	All wind power data occurring at state i
$\tilde{\boldsymbol{x}}_t$	Regressors for x_t
$\boldsymbol{X}(i)$	Regressor matrix for data at state i

F. Posterior Predictive Distribution and Samples

$F^{(b)}$	The b-th posterior sample of statistical model for wind power
$\Theta^{(b)}$	AR parameters for states in $F^{(b)}$
$\boldsymbol{P}^{(b)}$	State transition probability matrix in $F^{(b)}$
$q_\gamma(x_{T+h})$	The γ-th quantile of the predictive distribution for wind power at $T+h$
$\text{PI}(x_{T+h})$	Prediction Interval for wind power at $T+h$
$x_{T+h}^{(I)}$	The I-th sample of X_{T+h}
$\widehat{\text{PI}}(x_{T+h})$	Estimated PI for wind power at $T+h$

G. Evaluation of Forecasting Performance

f_t	Predictive distribution for wind power x_t

Corresponding author: W. Xie.

Wei Xie is with the Department of Mechanical and Industrial Engineering, Northeastern University, 360 Huntington Avenue, 334 SN, Boston, MA 02115 USA (e-mail: xieweirosie6@gmail.com).

P. Zhang is with the Department of Industrial and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (email: zhangp5@rpi.edu).

R. Chen is with the Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854-8019 USA (email:rongchen@stat.rutgers.edu).

Z. Zhou is with Argonne National Laboratory, Argonne, IL 60439 USA (email: zzhou@anl.gov).

$SC(f_t, x_t)$	Skill score of the predictive distribution
$\overline{\text{SC}}(h)$	Average skill score for h-step forecasting

I. Introduction

With the desire of energy system resilience and clean energy, smart power grids with distributed energy resources get increasing attention. According to the report from American Wind Energy Association [1], wind energy starts to be widely used in the power systems. During the year of 2013, U.S. wind energy provides enough electricity to power the equivalent of over 18 million homes. While the application for wind power is in a rapid growth, due to the uncertainty in wind power forecast, there are concerns on the reliability and cost. The operational decisions, i.e., generator scheduling, electricity pricing and trading, depend on the short-term wind power forecast ranging from 1 - 48 hours [2], [3]. Inaccurate quantification of forecast uncertainty can lead to low system reliability and high operational cost [4]. *Thus, it is important to improve the short-term wind power forecast accuracy and correctly quantify the forecast uncertainty*.

There are various approaches proposed to quantify the uncertainty for short-term wind power forecast; see [3], [2] for a detailed review. In general, the forecast uncertainty of wind power can be represented by a probability distribution or a set of quantiles. The probability distribution is preferred to support stochastic optimization for real-time risk and reliability management [3], [5]. Thus, in this paper, we focus on developing the probability distribution that can rigorously characterize the overall wind power forecast uncertainty.

The dynamic behaviors of wind power could have significant fluctuations at hourly time scales, which can be caused by atmospheric changes, such as weather front and rain showers [6], [7]. To faithfully capture the important properties in the real data and improve the wind power forecast, it is desirable to model the evolution of meteorological conditions and estimate the dynamics of wind power under different conditions.

Motivated by the regime-switching models [8], [9], [10], [11], [12], [13], [14], in this paper, we propose a nonparametric time series forecast model, called Infinite Markov Switching Autoregressive (IMSAR), for the stochastic process of wind power generation. The latent meteorological changes are modeled by a Markov state transition process. At each state, the local dynamic behaviors of wind power are modeled with Autoregressive (AR) time series. Notice that the introduction of latent regime/state variable is to absorb the unknown and unobservable meteorological factors into the model. Since the states separate observations occurring under different meteorological conditions into different groups, the estimation error of latent state number induced by using restrict parametric regime-switching models can impact on correctly modeling both global and local dynamic behaviors, which can reduce the wind power prediction accuracy and lead to an unreliable quantification on the forecast uncertainty. Differing with existing parametric models, i.e., MSAR, our approach can automatically adapt the model complexity to the real-world data. Thus, it can faithfully capture the rich dynamic behaviors, and improve the prediction accuracy.

The forecast model characterizing the physical *stochastic uncertainty* of wind power generation is estimated by finite historical data. It can cause the model estimation error, called *model uncertainty*. Thus, there are two types of uncertainties in wind power prediction, including stochastic and model uncertainties. To provide the reliable probabilistic forecast, both sources of uncertainties should be considered.

Thus, in this paper, we first propose the nonparametric IMSAR that can faithfully capture the important properties in the wind power data. Then, given finite historical data, the posterior distribution of *flexible* forecast model can quantify the model selection and parameter estimation uncertainty. After that, the posterior predictive distribution of wind power is developed to rigorously characterize the forecast uncertainty accounting for both inherent stochastic uncertainty of wind power generation and model uncertainty. We propose an efficient sampling procedure to generate scenarios from the predictive distribution, which can be used in the stochastic programming to find real-time operational decisions hedging against all sources of uncertainty.

The contributions of this paper are described as follows.

- We present a nonparametric time series statistical model for wind power generation. It can faithfully capture the rich properties in the historical data streams, including nonstationarity, serial dependence, skewness and multi-modality.
- Then, the posterior distribution of flexible model for wind power allows us to quantify the model uncertainty, including the model selection and parameter estimation error.
- The overall wind power forecast uncertainty is characterized by the posterior predictive distribution accounting for both model estimation error and underlying stochastic uncertainty of wind power generation. An efficient Bayesian inference and sampling procedure is provided to generate scenarios from the posterior predictive distribution. Our approach also delivers percentile prediction intervals (PIs) for wind power forecast.

The structure of this paper is as follows. In Section II, we briefly review the related existing approaches on short-term wind power prediction. We then formally introduce the Bayesian nonparametric forecasting framework and the sampling procedure for probabilistic forecast in Section III. In Section IV, the case studies over wind power data demonstrate that our approach has the promising performance.

II. Background

The methodologies developed for wind power forecasting is reviewed by [2], [3]. Here, we provide a brief review of the statistical methods for short-term wind power probabilistic forecast related to our study. The existing approaches can be categorized into machine learning based and time series based approaches. As the development in artificial intelligence, many machine learning algorithms were proposed for wind power forecast, including quantile regression [15], [16], [17], [18], time varying quantile regression [19], artificial neural network [20], [21], [22], [23], support vector machines [24], [25] and so on.

Another category of statistical methods are the time series based approaches. Conventional Autoregressive (AR) models and Autoregressive Moving Average (ARMA) models are widely used in wind power forecasting. For example, [26] used ARMA to predict hourly average wind speed up to 10 hours in advance. [27] proposed a time series model to capture non-linear and double-bounded nature of wind power data. A logit-normal transformation was developed for the wind power bounded between 0 to 1, and the location and scale parameters are modeled by AR or Conditional Parametric AR processes.

Since the dynamic behaviors of wind power change due to the fluctuations of meteorological state, the studies in [7], [12] proposed a Markov-switching autoregressive (MSAR) approach. It models the state transition with a Markov process having finite states, and models the local dynamic behaviors at each state with AR process. [14] further introduced a MSAR model with seasonal components to capture seasonal and inter-annual fluctuations in wind power process.

In addition, since the forecast model is estimated by finite historical data with error, [28] used the Bayesian Model Averaging (BMA) approach to quantify the model uncertainty, where a few Gamma distributions are selected as the candidate model for wind power. [22] proposed the extreme learning machine (ELM) and used the bootstrap to quantify the model parameter estimation error. Given the total variance, including the model uncertainty variance and data noise variance, the prediction interval is constructed based the normality assumption on the prediction distribution. [23] proposed to use intervals quantifying the parameter estimation uncertainty for the neural network based predictor, which does not provide a full predictive distribution for wind power.

III. A Bayesian Nonparametric Probabilistic Forecast Framework

Our approach is motivated by parametric MSAR approaches [8], [7], [12]. Considering that the number of underlying meteorological states is unknown and also there exists the model estimation uncertainty for wind power forecast, we propose a nonparametric Bayesian probabilistic forecasting framework. Specifically, motivated by the infinite Hidden Markov model [29], we first present the nonparametric forecast model, called IMSAR, in Section III-A, which can adapt the model complexity to capture the important properties in the wind power data. Then, in Section III-B, given finite historical data, we derive the posterior distribution quantifying the model estimation uncertainty and further develop the posterior predictive distribution to quantify the forecast uncertainty accounting for both the inherent stochastic uncertainty of wind power generation and the forecast model estimation error. In Section III-C, we provide a sampling procedure to efficiently generate scenarios for wind power probabilistic forecasting and further deliver a PI.

A. IMSAR Nonparametric Forecast Model for Wind Power

Let x_t and s_t be the wind power and latent meteorological state at time period t. In the IMSAR forecast model, denoted by F, we model the dependent stochastic processes $\{x_t\}$ and $\{s_t\}$ simultaneously. For the short-term wind power forecast, the state in next time period highly depends on the current state. Motivated by [13], we model $\{s_t\}$ as a Markov process. Without strong prior information on the underlying meteorological states, an infinite hidden Markov model (IHMM) models the state transition. At each state, the dynamic behaviors of wind power is modeled by AR time series. Thus, at any time period t, given the historical wind power data, denoted by $\mathbf{x}_{[1:t-1]} = (x_1, \ldots, x_{t-1})$, the probabilistic forecast for x_t is

$$\begin{aligned} & f(x_t|\mathbf{x}_{[1:t-1]}, F) \\ & = \sum_{i=1}^{+\infty} p(s_t = i|\mathbf{x}_{[1:t-1]}) h(x_t|\mathbf{x}_{[1:t-1]}, \boldsymbol{\theta}_{s_t}, s_t = i) \quad (1) \end{aligned}$$

with the conditional probability of $s_t = i$

$$p(s_t = i|\mathbf{x}_{[1:t-1]}) = \sum_{j=1}^{+\infty} p(s_t = i|s_{t-1} = j) p(s_{t-1} = j|\mathbf{x}_{[1:t-1]})$$

where $h(\cdot)$ denotes the AR time series kernel specified by parameters $\boldsymbol{\theta}_i$ characterizing the dynamic behaviors of wind power when $s_t = i$.

The nonparametric IMSAR can automatically adapt the model complexity to the wind power data. It can capture the rich properties, including non-stationarity, multi-modality, skewness, tail and serial dependence.

To support the inference and implementation, a hierarchical Dirichlet process (HDP) is used to represent IHMM; see the introduction on HDP in [30]. The prior for state transition probabilities is modeled by the global DP in the *hierarchical model*, denoted by $G_0 \sim \text{DP}(\alpha, G_{\boldsymbol{\theta}})$, where α is the concentration parameter and $G_{\boldsymbol{\theta}}$ represents the prior distribution for parameters $\boldsymbol{\theta}$. Based on the definition of DP in [31], the random distribution G_0 over any finite measurable partitions $B_1, \ldots, B_r$ of the space of $\boldsymbol{\theta}$ follows a Dirichlet distribution, $\big(G_0(B_1), \ldots, G_0(B_r)\big) \sim \text{Dirichlet}\big(\alpha G_{\boldsymbol{\theta}}(B_1), \ldots, \alpha G_{\boldsymbol{\theta}}(B_r)\big)$. We can write $G_0 = \sum_{\ell=1}^{+\infty} \pi_\ell \delta_{\boldsymbol{\theta}_\ell}$, where π_ℓ is the probability staying in state ℓ which has the wind power dynamic behaviors specified by parameters $\boldsymbol{\theta}_\ell$, and $\delta_{\boldsymbol{\theta}_\ell}$ denotes a Dirac function at $\boldsymbol{\theta}_\ell$. Then, since underlying meteorological conditions are shared by state variables s_t with $t = 1, 2, \ldots$, a set of *state-conditional DPs*, $G_i|G_0, \eta \sim \text{DP}(\eta, G_0)$, is used to model the prior transition probabilities from current state i to the next state for IHMM, which is statistically centering around G_0 with η measuring the concentration.

Since the real-world data are assumed to be one sample path realization from the physical stochastic process $\{x_t\}$, the current state s_t could take different values. Up to the time t, suppose that there are K meteorological states visited by $\mathbf{x}_{[1:t]}$, called *active states*. The next state s_{t+1} can move to any of these K active states or a new state. The set of potential next state is shared by transitions from different values of current state s_t, and the prior of global transition probabilities $\boldsymbol{\pi} = \{\pi_j\}_{j=1}^{K+1}$ of $s_{t+1} = j$ for $j = 1, \ldots, K+1$ does not depend on the meteorological conditions of the current state. As $K \to +\infty$, the prior of global probability measure becomes a stick-breaking process, $\boldsymbol{\pi} \sim \text{Stick}(\alpha)$; see [32].

Then, the state-conditional transition probability from state $s_t = i$ to next state, denoted by $\boldsymbol{p}_i = \{p_{ij}\}_{j=1}^{K+1}$ for $i = 1, \ldots, K$, has the prior belief modeled by,

$$\boldsymbol{p}_i|\boldsymbol{\pi} \sim \text{Dirichlet}(\eta\pi_1, \ldots, \eta\pi_K, \eta\pi_{K+1}) \tag{2}$$

with η controlling the impact of the global probability measure $\boldsymbol{\pi}$ on each transition distribution $\boldsymbol{p}_i$. As $K \to +\infty$, the prior can be written as $\boldsymbol{p}_i|\boldsymbol{\pi} \sim \text{DP}(\eta, \boldsymbol{\pi})$.

Therefore, in the IMSAR, IHMM models the underlying state transition and AR time series kernel models the dynamic behaviors of wind power at each state. With the HDP representation for IHMM, the IMSAR model F is written as

$$\begin{aligned}
&x_t = \phi_0(s_t) + \sum_{j=1}^{p} \phi_j(s_t) x_{t-j} + \sigma(s_t)\epsilon_t \\
&\boldsymbol{\theta}_{s_t} \equiv \{\phi_0(s_t), \ldots, \phi_p(s_t), \sigma^2(s_t)\} \sim G_{\boldsymbol{\theta}} \\
&s_t|s_{t-1}, \{\boldsymbol{p}_i\}_{i=1}^{+\infty} \sim \boldsymbol{p}_{s_{t-1}} \\
&\boldsymbol{p}_i|\boldsymbol{\pi} \sim \text{DP}(\eta, \boldsymbol{\pi}) \\
&\boldsymbol{\pi} \sim \text{Stick}(\alpha)
\end{aligned} \tag{3}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$ with $\mathcal{N}(a, b^2)$ denoting a normal distribution with mean a and variance b^2. At each state s_t, the AR process with parameters $\boldsymbol{\theta}_{s_t}$ models the local dynamic behaviors of wind power, and $\boldsymbol{\theta}_{s_t}$ has the prior $G_{\boldsymbol{\theta}}$, as shown in the first two equations in (3). The state transition model for $\{s_t\}$ has the HDP prior as shown in the last three equations.

B. Forecast Uncertainty Quantification

At the current time, denoted by T, the underlying stochastic model for $\{x_t\}$ is unknown and estimated by finite historical data, denoted by $\mathbf{x}_{[1:T]} = (x_1, \ldots, x_T)$. By applying the Bayes rule, the model uncertainty is quantified by the posterior distribution,

$$p(F|\mathbf{x}_{[1:T]}) \propto p(F)p(\mathbf{x}_{[1:T]}|F),$$

where $p(F)$ and $p(\mathbf{x}_{[1:T]}|F)$ represent the prior belief on the model and the likelihood of wind power data.

When the model estimate F is used for the τ-step ahead probabilistic forecast by $p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]}, F)$, the posterior predictive distribution,

$$\begin{aligned}
&p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]}) \\
&\quad = \int p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]}, F)p(F|\mathbf{x}_{[1:T]})dF,
\end{aligned} \tag{4}$$

can characterize the overall forecast uncertainty with $p(F|\mathbf{x}_{[1:T]})$ quantifying the model estimation error and $p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]}, F)$ quantifying the prediction uncertainty induced by the inherent stochastic uncertainty of wind power generation. The dynamic behaviors learning through $p(F|\mathbf{x}_{[1:T]})$ can be used to improve the prediction accuracy.

Given the historical data $\mathbf{x}_{[1:T]}$, suppose that there are K active states. Let $T_i = \{t : s_t = i\}$ be all the time periods with $s_t = i$, let $\boldsymbol{m} = (m_1, \ldots, m_K)$ with m_i recording the number of visits to state i, and let $\boldsymbol{N}$ denote the transition matrix with $N_{i,j}$ recording the number of transitions from state i to state j for $i, j = 1, \ldots, K$.

Then, we derive the conditional posterior distributions for the stochastic model of $\{x_t\}$ in (3), which will be used in the sampling procedure in Section III-C to generate the posterior samples from $p(F|\mathbf{x}_{[1:T]})$. Let the prior for $\boldsymbol{\phi}(i) = [\phi_0(i), \phi_1(i), \ldots, \phi_p(i)]^\top$ with $i = 1, \ldots, K$ to be a multivariate Normal distribution with mean $\boldsymbol{\mu}_\phi = \mu_\phi \mathbf{1}_{(p+1)\times 1}$ and covariance matrix $\Sigma_\phi = \sigma_\phi^2 \mathbf{I}_{(p+1)\times(p+1)}$, and let the prior for $\sigma^2(i)$ to be an Inverse Gamma distribution, denoted by $\sigma^2(i) \sim \text{IG}(\frac{a}{2}, \frac{b}{2})$, which specifies $G_{\boldsymbol{\theta}}$ in (3), where μ_ϕ, σ_ϕ, a and b are hyper-parameters, and $\mathbf{I}$ is an identity matrix. Let $\tilde{\boldsymbol{x}}_t = [1, x_{t-1}, \ldots, x_{t-p}]$. Let $\boldsymbol{x}_{[T_i]}$ denote a $(m_i \times 1)$ vector of x_t with $t \in T_i$, and $\boldsymbol{X}(i)$ denote the corresponding $m_i \times (p+1)$ regressors matrix having each row to be $\tilde{\boldsymbol{x}}_t$ with $t \in T_i$.

By following [33], when $s_t = i$, the conditional posteriors for the AR parameters $\boldsymbol{\phi}(i)$ and $\sigma^2(i)$ are,

$$\begin{aligned}
\boldsymbol{\phi}(i)|\sigma(i), \mathbf{x}_{[1:T]} &\sim \mathcal{N}(\mathbf{A}^{-1}\mathbf{B}, \mathbf{A}^{-1}) \\
\sigma^2(i)|\boldsymbol{\phi}(i), \mathbf{x}_{[1:T]} &\sim \text{IG}\left(\frac{a+m_i}{2}, \frac{b+c}{2}\right)
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
&\mathbf{A} = \Sigma_\phi^{-1} + \boldsymbol{X}(i)^\top \boldsymbol{X}(i)/\sigma^2(i) \\
&\mathbf{B} = \Sigma_\phi^{-1}\boldsymbol{\mu}_\phi + \boldsymbol{X}(i)^\top \boldsymbol{x}_{[T_i]}/\sigma^2(i) \\
&c = \boldsymbol{x}_{[T_i]}^\top \boldsymbol{x}_{[T_i]} - 2\boldsymbol{\phi}(i)^\top \boldsymbol{X}(i)^\top \boldsymbol{x}_{[T_i]} + \boldsymbol{\phi}(i)^\top \boldsymbol{X}(i)^\top \boldsymbol{X}(i)\boldsymbol{\phi}(i).
\end{aligned}$$

In the case studies, we use the non-informative priors with hyper-parameters $a = 1, b = 1$ and $\mu_\phi = 0, \sigma_\phi = 100$.

Given the HDP representation in (3), the conditional posterior for the historical states $\mathbf{s}_{[1:T]} = (s_1, \ldots, s_T)$ is derived by following [34],

$$\begin{aligned}
&p(s_t = i|\mathbf{x}_{[1:T]}, \boldsymbol{\pi}, s_{t-1}, s_{t+1}, \boldsymbol{\theta}_i) \\
&= C_0 p(s_t = i|\boldsymbol{\pi}, s_{t-1})p(s_{t+1}|\boldsymbol{\pi}, s_t = i)p(x_t|s_t = i, \boldsymbol{\theta}_i, \mathbf{x}_{[1:t-1]}) \\
&= \begin{cases} C_0 \frac{\eta\pi_i + N_{s_{t-1},i}}{\eta + m_{s_{t-1}}} \frac{\eta\pi_{s_{t+1}} + N_{i,s_{t+1}}}{\eta + m_i} f(x_t|\boldsymbol{\theta}_i), & 1 < t < T \\ C_0 \frac{\eta\pi_i + N_{s_{t-1},i}}{\eta + m_{s_{t-1}}} f(x_t|\boldsymbol{\theta}_i), & t = T \end{cases}
\end{aligned} \tag{6}$$

and $p(s_1 = 1) = 1$, where

$$f(x_t|\boldsymbol{\theta}_i) = \frac{1}{\sqrt{2\pi}\sigma(i)} \exp\left[-\frac{\left(x_t - \phi_0(i) - \sum_{j=1}^{p} \phi_j(i) x_{t-j}\right)^2}{2\sigma^2(i)}\right]$$

and C_0 is a normalizing constant shared by all $s_t = i$ to guarantee that $\sum_{i=1}^{K+1} p(s_t = i|\mathbf{x}_{[1:T]}, \boldsymbol{\pi}, s_{t-1}, s_{t+1}, \boldsymbol{\theta}_i) = 1$.

Given the prior $\boldsymbol{\pi} \sim \text{Stick}(\alpha)$, the conditional posterior for $\boldsymbol{\pi}$ is derived by following [30],

$$\boldsymbol{\pi}|\mathbf{s}_{[1:T]} \sim \text{Dirichlet}(m_1, \ldots, m_K, \alpha). \tag{7}$$

Then, given $\boldsymbol{\pi}$ and the prior in (2), the transition probability $\boldsymbol{p}_i$ for $i = 1, \ldots, K$ has the conditional posterior distribution

$$\boldsymbol{p}_i|\boldsymbol{\pi} \sim \text{Dirichlet}(\eta\pi_1 + N_{i,1}, \ldots, \eta\pi_K + N_{i,K}, \eta\pi_{K+1}). \tag{8}$$

According to [30], in the case studies, we set $\alpha = 1$ and $\eta = 1$ so that the priors do not have an obvious impact on the conditional posterior distributions in (7) and (8).

C. A Sampling Procedure for Probabilistic Forecast

In this section, we propose a sampling procedure to generate

scenarios for wind power probabilistic forecast. Given the historical data $\mathbf{x}_{[1:T]}$, we first provide a Gibbs sampling procedure in Algorithm 1 to generate posterior samples, $F^{(b)} \sim p(F|\mathbf{x}_{[1:T]})$ with $b = 1, \ldots, B$, quantifying the forecast model estimation uncertainty; see [35] for the introduction of Bayesian inference and Gibbs sampling. Then, by drawing samples of wind power from $p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]}, F^{(b)})$, we generate scenarios from the posterior predictive distribution $p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]})$ quantifying the overall forecast uncertainty and further construct a PI by following Algorithm 2.

Algorithm 1: Gibbs Sampling Procedure to Generate Posterior Samples for the IMSAR Forecast Model

1 Initialize the states $\mathbf{s}_{[1:T]}$ with $s_t = t$. Initialize $\boldsymbol{\pi}$ by generating a sample from the conditional posterior, Dirichlet$(\mathbf{1}_{(1\times T)}, \alpha)$; see (7). Record the transition matrix $\boldsymbol{N}$ and the counter vector $\boldsymbol{m}$.
2 **for** $t = 2, \ldots, T$ **do**
3 Set $N_{s_{t-1},s_t} = N_{s_{t-1},s_t} - 1$, $N_{s_t,s_{t+1}} = N_{s_t,s_{t+1}} - 1$, and $m_{s_t} = m_{s_t} - 1$. Update the AR parameters $\boldsymbol{\theta}_{s_t}$ according to (5). Set K as the number of unique values in the state set $\mathbf{s}_{[1:T]}$ excluding s_t.
4 Sample a new set of AR model parameters $\boldsymbol{\theta}_{K+1}$ from $G_{\boldsymbol{\theta}}$.
5 Sample s_t from $p(s_t = k|\mathbf{x}_{[1:T]}, \boldsymbol{\pi}, s_{t-1}, s_{t+1}, \boldsymbol{\theta}_k)$ by using (6) for $k = 1, \ldots, K+1$. If $s_t = K+1$, let $K = K+1$, append an empty column and an empty row to the matrix $\boldsymbol{N}$, append a new element 0 to $\boldsymbol{m}$.
6 Set $N_{s_{t-1},s_t} = N_{s_{t-1},s_t} + 1$, $N_{s_t,s_{t+1}} = N_{s_t,s_{t+1}} + 1$ and $m_{s_t} = m_{s_t} + 1$
7 Update the AR parameters $\boldsymbol{\theta}_{s_t}$ according to (5).
8 **end**
9 **for** $k = 1, \ldots, K$ **do**
10 if $m_k = 0$, remove the k-th state from $\boldsymbol{N}$, $\boldsymbol{m}$ and $\boldsymbol{\pi}$, and also remove the AR parameters $\boldsymbol{\theta}_k$. Let $K = K - 1$.
11 **end**
12 Update $\boldsymbol{\pi}$ based on (7).
13 Update $\boldsymbol{p}_i$ for $i = 1, \ldots, K$ based on (8). Generate $\boldsymbol{p}_{K+1}$ by using (2).
14 Repeat Steps 2–13 until convergence, and then record B posterior samples of IMSAR model with $F^{(b)}$ specified by parameters $\{\Theta^{(b)}, \mathbf{P}^{(b)}\}$ for $b = 1, \ldots, B$, where $\Theta^{(b)} \equiv \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \boldsymbol{\theta}_{K+1}\}$ are the AR parameters and $\mathbf{P}^{(b)}$ is a probability transition matrix with the i-th row to be $\boldsymbol{p}_i$ for $i = 1, \ldots, K+1$, where $\boldsymbol{\theta}_{K+1}$ is the average of samples from the prior $G_{\boldsymbol{\theta}}$. And, record the latent states $s_T^{(b)}$ from the corresponding b iteration.

Given the conditional posteriors for the IMSAR model as shown in (5)–(8), we provide a Gibbs sampling procedure described in Algorithm 1 to generate B posterior samples of the forecast model, $F^{(b)} \sim p(F|\mathbf{x}_{[1:T]})$ for $b = 1, \ldots, B$. In Step 1, we initialize the states $\mathbf{s}_{[1:T]}$ and $\boldsymbol{\pi}$. In Steps 2–8, for each $t = 1, \ldots, T$, we update the state s_t and $\boldsymbol{\theta}_{s_t}$ by sampling from the conditional posteriors in (6) and (5). In Steps 9–11, we update the number of active states K. Then, we update $\boldsymbol{\pi}$ and $\boldsymbol{p}_i$ for $i = 1, \ldots, K+1$ in Steps 12–13. Notice that since there is no transition from the inactive or new state $K+1$ occurring in $\mathbf{x}_{[1:T]}$, the transition probability $\boldsymbol{p}_{K+1}$ follows the prior in (2). As all the inactive states share the same prior, the state $K+1$ can be regarded as a group representing inactive states. Thus, we approximate $\boldsymbol{\theta}_{K+1}$ by the average of samples from the prior $G_{\boldsymbol{\theta}}$ according to [36]. The probability of new state is $\alpha/(T+\alpha)$. The procedure is repeated until convergence, and then record $F^{(b)}$ and $s_T^{(b)}$ for $b = 1, \ldots, B$.

In Algorithm 2, at each $F^{(b)}$ with $b = 1, \ldots, B$, we generate m sample paths of $\mathbf{X}_{[T+1:T+\tau]}$ from $p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]}, F^{(b)})$. Then, we can get Bm samples of $\mathbf{x}_{[T+1:T+\tau]}$ from $p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]})$ accounting for both stochastic uncertainty and model estimation error. Since the samples of x_{T+h} with $h = 1, \ldots, \tau$ can be used to estimate the distribution $p(x_{T+h}|\mathbf{x}_{[1:T]})$, we can construct the $(1-\alpha^\star)100\%$ two-sided percentile PI for x_{T+h},

$$\text{PI}(x_{T+h}) \equiv [q_{\alpha^\star/2}(x_{T+h}), q_{1-\alpha^\star/2}(x_{T+h})],$$

quantifying the forecast uncertainty, where $q_\gamma(x_{T+h}) \equiv \inf\{q : F_{X_{T+h}}(q|\mathbf{x}_{[1:T]}) \geq \gamma\}$ for $\gamma = \alpha^\star/2, 1-\alpha^\star/2$, and $F_{X_{T+h}}(\cdot|\mathbf{x}_{[1:T]})$ represents the cumulative distribution function for $p(x_{T+h}|\mathbf{x}_{[1:T]})$. The PI$(x_{T+h})$ can be estimated by using the $\alpha^\star/2$-th and $(1-\alpha^\star/2)$-th order statistics of the Bm scenarios, $\mathbf{x}_{T+h}^{(I)}$ with $I = 1, \ldots, Bm$, in Equation (9).

Algorithm 2: Procedure to Generate Scenarios from the Posterior Predictive Distribution $p(\mathbf{x}_{[T+1:T+\tau]}|\mathbf{x}_{[1:T]})$ and Construct $(1-\alpha^\star)100\%$ Two-sided Percentile PI for x_{T+h} with $h = 1, \ldots, \tau$ Quantifying the Forecasting Uncertainty

1 Specify m and set $I = 0$.
2 **for** $b = 1, \ldots, B$ *and* $i = 1, \ldots, m$ **do**
3 Let $I = I + 1$.
4 **for** $h = 1, \ldots, \tau$ **do**
5 Generate a sample of s_{T+h} from $P^{(b)}_{s_{T+h-1},s_{T+h}}$.
6 Generate a scenario of x_{T+h}, $x_{T+h}^{(I)} = \phi_0^{(b)}(s_{T+h}) + \sum_{j=1}^{p} \phi_j^{(b)}(s_{T+h}) x_{T+h-j} + \sigma^{(b)}(s_{T+h})\epsilon$.
7 **end**
8 **end**
9 Build the posterior predictive distribution $p(x_{T+h}|\mathbf{x}_{[1:T]})$ by using Bm samples of x_{T+h} from Steps 2-8 for $h = 1, \ldots, \tau$.
10 Report a $(1-\alpha^\star)100\%$ two-sided percentile PI for x_{T+h} with $h = 1, \ldots, \tau$,

$$\widehat{\text{PI}}(x_{T+h}) = \left[x_{T+h}^{[\lceil(\alpha^\star/2)Bm\rceil]}, x_{T+h}^{[\lceil(1-\alpha^\star/2)Bm\rceil]}\right] \quad (9)$$

where $x_{T+h}^{[1]} \leq x_{T+h}^{[2]} \leq \ldots \leq x_{T+h}^{[Bm]}$ are the order statistics of all samples $x_{T+h}^{(I)}$ with $I = 1, \ldots, Bm$.

IV. Empirical Study

In this section, we use the year 2006 hourly wind power data from a hypothetical site in Illinois to compare the performance

of our approach with the persistence model [37], the MSAR model [14], the TVQR model [19] and the BELM [22]. The dataset is from the Eastern Wind Integration and Transmission Study provided by NREL [38]. Since the raw data, denoted by $\{x_t^*\}$, are in the range $(0,1)$, by following [27], we take the logit transformation, $x_t = \log(\frac{x_t^*}{1-x_t^*})$, so that the support of x_t is on the whole real line. After predicting x_{t+h}, we transform it back to x_{t+h}^*.

A. Overall Performance for Wind Power Forecasting

To evaluate the behaviors of various approaches, we perform three-step ahead probabilistic prediction by using $L = 1000$ consecutive sequences of hourly wind power. To study the impact of forecast model estimation uncertainty, each sequence consists of historical data with length $T = 100, 500$, denoted by $\mathbf{x}_{[\ell-T+1:\ell]}$ with $\ell = 0, \ldots, L-1$. By following the procedure described in Section III-C, we can generate the scenarios for $x_{\ell+h}$ from the predictive distribution $p(x_{\ell+h}|\mathbf{x}_{[\ell-T+1:\ell]})$ for $h = 1, 2, 3$.

To evaluate the performance of the probabilistic forecast, three metrics are considered: skill score, coverage and sharpness of PIs. Skill score (SC) proposed by [39] gives a comprehensive measurement of a predictive distribution, denoted by f_t, with the real world data x_t. It is defined as

$$\text{SC}(f_t, x_t) = \sum_{j=1}^{J} (\xi^{(\alpha_j)} - \alpha_j)(x_t - q_t^{(\alpha_j)}),$$

where α_j for $j = 1, \ldots, J$ is a sequence of quantile levels, $q_t^{(\alpha_j)}$ is the α_j-th quantile of the predictive distribution f_t, and $\xi^{(\alpha_j)}$ is an indicator variable,

$$\xi^{(\alpha_j)} = \begin{cases} 1 & \text{if } x_t < q_t^{(\alpha_j)} \\ 0 & \text{otherwise} \end{cases}.$$

In the skill score, for a high quantile percentage α, large penalty is taken if $x_t > q_t^{(\alpha)}$, and for a low quantile percentage α, large penalty is taken if $x_t < q_t^{(\alpha)}$. The maximum value of SC is 0 when f_t concentrates at the real world realization x_t. A larger SC denotes better forecasting performance.

In the case studies, we use a sequence of quantiles with $(\alpha_1, \ldots, \alpha_{99})$ evenly distributed on $[0.01, 0.99]$, which represent the global information of the predictive distribution. We record the average SC obtained from L testing sequences of wind power with $h = 1, 2, 3$

$$\overline{\text{SC}}(h) = \frac{1}{L} \sum_{\ell=0}^{L-1} \text{SC}\left(p(x_{\ell+h}|\mathbf{x}_{[\ell-T+1:\ell]}), x_{\ell+h}\right).$$

We compare the performance of IMSAR with MSAR [14], TVQR [19], ELM [22] and the persistence model [37] which has the empirical predictive distribution for $x_{\ell+h}$ specified by $\{x_\ell - x_{\ell-i} + x_{\ell-i-h} : i = 0, \ldots, T-h-1\}$. For MSAR, according to [14], we set the number of states to be $K_0 = 1, 3, 5$. The R package "NHMSAR" provided in [14] is used to train the MSAR model. Our empirical study indicates that the choice of p does not have significant impact on the forecasting performance of IMSAR and MSAR. Thus, we set $p = 1$. To build the posterior predictive distribution and construct the PIs quantifying the forecast uncertainty by following the procedure in Section III-C, we let $B = 100$ and $m = 100$. For TVQR, we use the R package "QRegVCM" developed based on [40]. For BELM, we use the R package "elmNN" to train the extreme learning machines and generate $B = 1000$ bootstrapped samples to quantify both the model estimation uncertainty and stochastic uncertainty.

TABLE I
SKILL SCORE RESULTS OF PREDICTIVE DISTRIBUTIONS

$T = 100$	$h = 1$	$h = 2$	$h = 3$
IMSAR	-1.120	-2.064	-2.983
Persistence	-1.195	-2.357	-3.337
MSAR ($K_0 = 1$)	-1.846	-3.534	-4.962
MSAR ($K_0 = 3$)	-1.385	-2.606	-3.606
MSAR ($K_0 = 5$)	-1.266	-2.424	-3.453
TVQR	-1.313	-2.530	-3.648
BELM	-1.297	-2.385	-3.922
$T = 500$	$h = 1$	$h = 2$	$h = 3$
IMSAR	-0.968	-1.774	-2.763
Persistence	-1.212	-2.542	-3.408
MSAR ($K_0 = 1$)	-1.370	-2.917	-4.053
MSAR ($K_0 = 3$)	-1.096	-2.323	-3.116
MSAR ($K_0 = 5$)	-0.914	-2.187	-2.957
TVQR	-1.127	-2.445	-3.257
BELM	-0.980	-2.034	-3.155

Table I shows the SC results of predictive distributions obtained from different approaches. IMSAR provides the largest SC in all situations. The persistence model works well when the size of historical data is small, but its performance does not improve much when there are more historical data.

Another metric to evaluate the performance of probabilistic forecasting is the coverage of PI for h-step ahead prediction,

$$\text{Coverage}(x_{\ell+h}) = \frac{1}{L} \sum_{\ell=0}^{L-1} \mathbf{1}\left(x_{\ell+h} \in \text{PI}(x_{\ell+h})\right)$$

where $\mathbf{1}(\cdot)$ is an indicator function. Table II shows the coverage of 99%, 95% and 90% percentile two-sided symmetric PIs from different approaches. Since the MSAR doesn't consider the forecast model estimation uncertainty, it underestimates the forecast uncertainty and its PI coverage is much less than the desired confidence level. TVQR also tends to have the under-coverage issue. The PI widths obtained from different approaches are reported in Table III to evaluate the sharpness of probabilistic forecasting. MSAR provides narrower PIs since it underestimate the forecast uncertainty. Notice that a narrower PI doesn't necessarily indicates more accurate forecasting. It should be interpreted together with the coverage.

We also report the posterior mode of the number of active states for IMSAR, denoted as $\hat{K}$. Table IV shows the results obtained by using $L = 1000$ wind power datasets with $T = 100, 500$. Notice that the number of active states depends on the real-world data. As the amount of historical data, T, increases, the number of active states also increases.

We record the mean CPU time of IMSAR to evaluate its computational cost. Algorithms 1 and 2 are implemented in "R" software, and ran on one node of the DRP cluster with two eight-core 2.6 GHz Intel Xeon E5-2650 processors and

TABLE II
COVERAGE OF PIS OBTAINED FROM DIFFERENT APPROACHES

99% PI	$T = 100$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.985	0.974	0.964
	Persistence	0.976	0.968	0.967
	MSAR ($K_0 = 1$)	0.578	0.588	0.531
	MSAR ($K_0 = 3$)	0.879	0.873	0.907
	MSAR ($K_0 = 5$)	0.924	0.923	0.919
	TVQR	0.954	0.922	0.886
	BELM	0.961	0.944	0.929
	$T = 500$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.984	0.981	0.975
	Persistence	0.989	0.982	0.981
	MSAR ($K_0 = 1$)	0.602	0.576	0.552
	MSAR ($K_0 = 3$)	0.853	0.837	0.841
	MSAR ($K_0 = 5$)	0.915	0.910	0.916
	TVQR	0.984	0.947	0.900
	BELM	0.979	0.968	0.957
95% PI	$T = 100$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.953	0.942	0.935
	Persistence	0.940	0.938	0.943
	MSAR ($K_0 = 1$)	0.557	0.554	0.481
	MSAR ($K_0 = 3$)	0.828	0.823	0.844
	MSAR ($K_0 = 5$)	0.905	0.912	0.908
	TVQR	0.927	0.893	0.881
	BELM	0.934	0.923	0.916
	$T = 500$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.959	0.946	0.937
	Persistence	0.951	0.948	0.944
	MSAR ($K_0 = 1$)	0.590	0.574	0.533
	MSAR ($K_0 = 3$)	0.825	0.822	0.840
	MSAR ($K_0 = 5$)	0.896	0.906	0.903
	TVQR	0.924	0.898	0.875
	BELM	0.938	0.925	0.919
90% PI	$T = 100$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.919	0.904	0.867
	Persistence	0.891	0.880	0.886
	MSAR ($K_0 = 1$)	0.504	0.512	0.476
	MSAR ($K_0 = 3$)	0.667	0.701	0.683
	MSAR ($K_0 = 5$)	0.730	0.719	0.727
	TVQR	0.889	0.857	0.828
	BELM	0.885	0.872	0.860
	$T = 500$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.905	0.898	0.882
	Persistence	0.898	0.897	0.895
	MSAR ($K_0 = 1$)	0.529	0.515	0.482
	MSAR ($K_0 = 3$)	0.714	0.718	0.710
	MSAR ($K_0 = 5$)	0.789	0.768	0.789
	TVQR	0.891	0.859	0.826
	BELM	0.894	0.881	0.870

TABLE III
WIDTH OF PIS OBTAINED FROM DIFFERENT APPROACHES

99% PI	$T = 100$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.335	0.474	0.549
	Persistence	0.301	0.502	0.641
	MSAR ($K_0 = 1$)	0.099	0.137	0.165
	MSAR ($K_0 = 3$)	0.151	0.271	0.371
	MSAR ($K_0 = 5$)	0.160	0.300	0.424
	TVQR	0.256	0.370	0.458
	BELM	0.283	0.424	0.570
	$T = 500$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.302	0.451	0.528
	Persistence	0.302	0.484	0.633
	MSAR ($K_0 = 1$)	0.094	0.132	0.160
	MSAR ($K_0 = 3$)	0.148	0.259	0.355
	MSAR ($K_0 = 5$)	0.156	0.281	0.406
	TVQR	0.249	0.352	0.449
	BELM	0.279	0.406	0.558
95% PI	$T = 100$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.221	0.324	0.393
	Persistence	0.214	0.379	0.515
	MSAR ($K_0 = 1$)	0.073	0.114	0.149
	MSAR ($K_0 = 3$)	0.136	0.234	0.291
	MSAR ($K_0 = 5$)	0.141	0.257	0.340
	TVQR	0.206	0.288	0.409
	BELM	0.228	0.365	0.489
	$T = 500$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.203	0.298	0.362
	Persistence	0.204	0.343	0.489
	MSAR ($K_0 = 1$)	0.063	0.097	0.122
	MSAR ($K_0 = 3$)	0.126	0.228	0.284
	MSAR ($K_0 = 5$)	0.132	0.240	0.326
	TVQR	0.196	0.288	0.371
	BELM	0.208	0.349	0.454
90% PI	$T = 100$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.186	0.259	0.331
	Persistence	0.172	0.306	0.416
	MSAR ($K_0 = 1$)	0.063	0.088	0.106
	MSAR ($K_0 = 3$)	0.104	0.195	0.276
	MSAR ($K_0 = 5$)	0.111	0.217	0.307
	TVQR	0.165	0.253	0.276
	BELM	0.179	0.276	0.362
	$T = 500$	$h = 1$	$h = 2$	$h = 3$
	IMSAR	0.164	0.262	0.347
	Persistence	0.181	0.315	0.434
	MSAR ($K_0 = 1$)	0.060	0.084	0.103
	MSAR ($K_0 = 3$)	0.109	0.197	0.275
	MSAR ($K_0 = 5$)	0.114	0.215	0.304
	TVQR	0.165	0.237	0.268
	BELM	0.164	0.261	0.353

128GB of system memory. The CPU time of IMSAR with $T = 100$ historical data is on average 6 seconds with the standard deviation as 0.2 seconds. The CPU time with $T = 500$ is 29 seconds with the standard deviation as 1.6 seconds.

TABLE IV
POSTERIOR MODE OF THE NUMBER OF ACTIVE STATES K FOR IMSAR

$\hat{K}$	1	2	3	4	5	6	7	8	9	≥ 10
$T = 100$	3	386	549	60	2	0	0	0	0	0
$T = 500$	0	0	2	183	319	258	141	70	21	6

B. Robustness Performance

Since there could exist sudden changes in wind power due to the fluctuation of meteorological conditions, we are interested in the robustness performance of various approaches in such situations. For the sudden changes, we follow the definition of wind ramp in [41], that the change of wind power during $\Delta T = 1$ hour is larger than a threshold $P_{val} = 0.25$ on the $[0, 1]$ unit. According to this definition, there are 9 wind ramps during the year of 2006. We compare the performance of all the approaches in forecasting with $T = 100$ historical data.

We report the average skill score over the 9 datasets in Table V. Table VI records the mean of absolute forecast error, where the error is defined as Error $= |\widehat{x}_{T+h} - x_{T+h}|$ for $h = 1, 2, 3$. For IMSAR and MSAR, we use the mean of the predictive distributions as the point forecast $\widehat{x}_{T+h}$. For TVQR

we use the median response, and for persistence model we use $\widehat{x}_{T+h} = x_T$. The results in Tables V and VI demonstrate that IMSAR can provide much better probabilistic forecast and prediction accuracy when there are sudden changes.

TABLE V
SKILL SCORE OF FORECASTING FOR THE DATASETS WITH WIND RAMP

Skill Score	$h=1$	$h=2$	$h=3$
IMSAR	-3.892	-4.767	-5.314
Persistence	-7.921	-9.977	-11.327
MSAR ($K_0=1$)	-4.725	-6.348	-6.748
MSAR ($K_0=3$)	-4.014	-5.868	-6.270
MSAR ($K_0=5$)	-3.788	-5.363	-5.209
TVQR	-3.873	-5.735	-6.437
BELM	-4.195	-5.624	-6.092

TABLE VI
PREDICTION ERROR FOR THE DATASETS WITH WIND RAMP

Error	$h=1$	$h=2$	$h=3$
IMSAR	0.122	0.147	0.164
Persistence	0.233	0.223	0.248
MSAR ($K_0=1$)	0.140	0.177	0.181
MSAR ($K_0=3$)	0.127	0.158	0.168
MSAR ($K_0=5$)	0.124	0.152	0.163
TVQR	0.122	0.155	0.170
BELM	0.129	0.153	0.166

We also use a representative dataset with wind ramp to illustrate the probabilistic forecast obtained by IMSAR. We predict the wind power for hours 4353 to 4355, denoted by x_{T+h} with $h = 1, 2, 3$, using the historical data $\mathbf{x}_{[1:T]}$ from the 4253-th to the 4352-th hour. The wind power data and the prediction intervals are plotted in Fig. 1. The blue solid line represent the real wind power and the red dash lines are the upper and lower bounds of the 95% PI.

Fig. 1. Wind Power Data with Sudden Changes

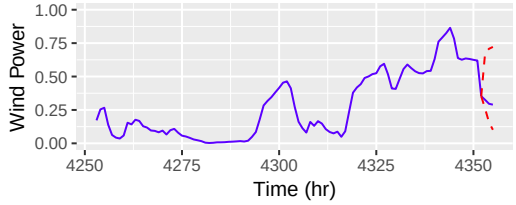
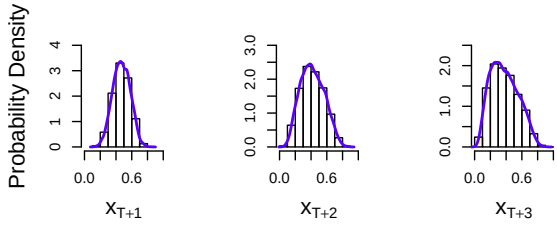


Fig. 2. Posterior Predictive Distributions $p(x_{T+h}|\mathbf{x}_{[1:T]})$ with $h = 1, 2, 3$.



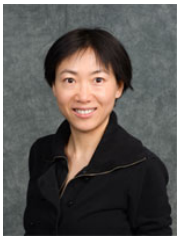
The predictive distributions $p(x_{T+h}|\mathbf{x}_{[1:T]})$ with $h = 1, 2, 3$ are plotted in Fig. 2. As h increases, the prediction uncertainty increases. These plots demonstrate that the prediction distributions are right skewed and have heavy tails. It is risky to make operational decisions without well quantifying the forecast uncertainty. The whole distribution should be considered to provide a reliable guidance on real-time risk management.

V. CONCLUSION

In this paper, we introduce a Bayesian nonparametric framework for short-term wind power probabilistic forecast. It can provide the predictive distribution accounting for both stochastic uncertainty and forecast model estimation uncertainty. Combining with scenario-based stochastic optimization, it allows us to provide reliable operational decisions for smart power grids hedging against wind power forecast uncertainty. The case studies on the wind power data from NREL demonstrate that our approach can provide more accurate, robust and reliable predictive distribution. When the wind power has sudden changes and large fluctuations, the advantages of our approach are more obvious. The potential future research directions include: (1) develop an online predictive distribution updating for wind power; (2) develop an efficient forecast framework for dependent wind power data from many firms.

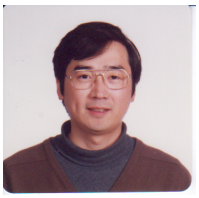
ACKNOWLEDGMENT

Dr. Chen's research was supported in part by NSF grants DMS-1513409 and IIS-1741390.



Wei Xie is an assistant professor in the Department of Mechanical and Industrial Engineering at Northeastern University. She received her Ph.D. in Industrial Engineering and Management Sciences at Northwestern University in 2014. Her research interests focus on computer simulation, data analytics, data-driven stochastic optimization, risk management for complex cyber-physical systems with applications, including smart power grids, supply chains, high-tech manufacturing, and health care. Dr. Xie currently serves as Associate Editor for ACM Transactions on Modeling and Computer Simulation.

Pu Zhang received the B.S. degree in industrial engineering from Shanghai Jiaotong University, China, in 2009 and the M.S. degree in operations research and industrial engineering from the University of Texas at Austin, USA, in 2013. He is currently working toward the Ph.D. degree in industrial and system engineering from Rensselaer Polytechnic Institute. His research interests include stochastic simulation, statistical modeling and data analytics.



Rong Chen received the Ph.D. degree in statistics from Carnegie Mellon University, Pittsburg, PA in 1990. He is a distinguished Professor of Statistics at the Department of Statistics and Biostatistics, Rutgers University. He served as Program Director in the Division of Mathematical Sciences at National Science Foundation from 2005 to 2007. He is an expert in statistical learning of dependent data, Monte Carlo methods statistical and statistical applications with many publications in top statistics, economics and other science journals. From 2013 to 2015 he served as Joint-Editor of Journal of Business & Economic Statistics. Dr. Chen is an elected fellow of American Statistical Association and Institute of Mathematical Statistics and an elected member of International Statistics Institute.

Zhi Zhou received the Ph.D. degree in decision sciences and engineering systems from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2010. Presently, he is a Principal Computational Scientist in the Energy Systems Division at Argonne National Laboratory, Argonne, IL. USA. His research interests include agent-based modeling and simulation, stochastic optimization, statistical forecasting, electricity markets, and renewable energy.

REFERENCES

[1] AWEA, "Awea reliability white paper," 2015.

[2] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1 – 8, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148111002850

[3] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, no. Supplement C, pp. 255 – 270, 2014.

[4] R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang, "Time-adaptive quantile-copula for wind power probabilistic forecasting," *Renewable Energy*, vol. 40, no. 1, pp. 29 – 39, 2012.

[5] A. Botterud, Z. Zhou, J. Wang, R. J. Bessa, H. Keko, J. Sumaili, and V. Miranda, "Wind power trading under uncertainty in lmp markets," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 894–903, May 2012.

[6] P. Sorensen, N. A. Cutululis, A. Vigueras-Rodríguez, H. Madsen, P. Pinson, L. E. Jensen, J. Hjerrild, and M. Donovan, "Modelling of power fluctuations from large offshore wind farms," *Wind Energy*, vol. 11, no. 1, pp. 29–43, 2008.

[7] 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. [Online]. Available: http://dx.doi.org/10.1002/for.1194

[8] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989. [Online]. Available: http://www.jstor.org/stable/1912559

[9] B. E. Hansen, "The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp," *Journal of Applied Econometrics*, vol. 7, no. S1, pp. S61–S82, 1992.

[10] T. H. Goodwin, "Business-cycle analysis with a markov-switching model," *Journal of Business & Economic Statistics*, vol. 11, no. 3, pp. 331–339, 1993.

[11] M. P. Clements and H. Krolzig, "A comparison of the forecast performance of markov-switching and threshold autoregressive models of us gnp," *The Econometrics Journal*, vol. 1, no. 1, pp. 47–75, 1998.

[12] P. Ailliot, V. Monbet, and M. Prevosto, "An autoregressive model with time-varying coefficients for wind fields," *Environmetrics*, vol. 17, no. 2, pp. 107–117, 2006.

[13] P. Pinson, G. Papaefthymiou, B. Klockl, H. Nielsen, and H. Madsen, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2008.

[14] P. Ailliot and V. Monbet, "Markov-switching autoregressive models for wind time series," *Environmental Modelling and Software*, vol. 30, no. Supplement C, pp. 92 – 101, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364815211002222

[15] J. B. Bremnes, "A comparison of a few statistical models for making quantile wind power forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 3–11, 2006. [Online]. Available: http://dx.doi.org/10.1002/we.182

[16] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 95–108, 2006. [Online]. Available: http://dx.doi.org/10.1002/we.180

[17] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, July 2017.

[18] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Optimal prediction intervals of wind power generation," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1166–1174, May 2014.

[19] J. K. Møller, H. A. Nielsen, and H. Madsen, "Time-adaptive quantile regression," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292 – 1303, 2008.

[20] S. A. Pourmousavi Kani and G. Riahy, "A new ann-based methodology for very short-term wind speed prediction using markov chain approach," in *Proceeding of 8th IEEE annual electrical power and energy conference*, 11 2008, pp. 1 – 6.

[21] G. Sideratos and N. D. Hatziargyriou, "Probabilistic wind power forecasting using radial basis function neural networks," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1788–1796, Nov 2012.

[22] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1033–1044, May 2014.

[23] S. Chai, Z. Xu, and W. K. Wong, "Optimal granule-based pis construction for solar irradiance forecast," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3332–3333, July 2016.

[24] J. Zeng and W. Qiao, "Support vector machine-based short-term wind power forecasting," in *2011 IEEE/PES Power Systems Conference and Exposition*, March 2011, pp. 1–8.

[25] J. Wang, J. Sun, and H. Zhang, "Short-term wind power forecasting based on support vector machine," in *2013 5th International Conference on Power Electronics Systems and Applications(PESA)*, Dec 2013, pp. 1–5.

[26] J. Torres, A. García, M. D. Blas, and A. D. Francisco, "Forecast of hourly average wind speed with arma models in navarre (spain)," *Solar Energy*, vol. 79, no. 1, pp. 65 – 77, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X04002877

[27] 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. [Online]. Available: http://dx.doi.org/10.1111/j.1467-9876.2011.01026.x

[28] J. M. Sloughter, T. Gneiting, and A. E. Raftery, "Probabilistic wind speed forecasting using ensembles and bayesian model averaging," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 25–35, 2010. [Online]. Available: https://doi.org/10.1198/jasa.2009.ap08615

[29] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden markov model," in *Advances in neural information processing systems*, 2002, pp. 577–584.

[30] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.

[31] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, no. 196, pp. 209–230, 1973.

[32] J. K. Ghosh and R. V. Ramamoorthi, *Bayesian Nonparametrics*. New York: Springer–Verlag, 2003.

[33] P. D. Hoff, *A First Course in Bayesian Statistical Methods*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[34] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "A sticky hdp-hmm with application to speaker diarization," *Ann. Appl. Stat.*, vol. 5, no. 2A, pp. 1020–1056, 06 2011. [Online]. Available: https://doi.org/10.1214/10-AOAS395

[35] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. New York: Taylor and Francis Group, LLC, 2004.

[36] C. E. Rasmussen, "The infinite gaussian mixture model," in *Advances in Neural Information Processing Systems 12*. MIT Press, 2000, pp. 554–560.

[37] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007. [Online]. Available: http://dx.doi.org/10.1111/j.1467-9868.2007.00587.x

[38] NREL, "Eastern wind integration and transmission study (ewits)," 2006, information at: https://www.nrel.gov/grid/wind-toolkit.html.

[39] P. Pinson, H. A. Nielsen, J. K. Moller, H. Madsen, and G. N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.

[40] X. He, "Quantile curves without crossing," *The American Statistician*, vol. 51, no. 2, pp. 186–192, 1997.

[41] R. Sevlian and R. Rajagopal, "Wind power ramps: Detection and statistics," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.