

Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction

Niya Chen, Zheng Qian, Ian T. Nabney, and Xiaofeng Meng

Abstract—Since wind at the earth's surface has an intrinsically complex and stochastic nature, accurate wind power forecasts are necessary for the safe and economic use of wind energy. In this paper, we investigated a combination of numeric and probabilistic models: a Gaussian process (GP) combined with a numerical weather prediction (NWP) model was applied to wind-power forecasting up to one day ahead. First, the wind-speed data from NWP was corrected by a GP, then, as there is always a defined limit on power generated in a wind turbine due to the turbine controlling strategy, wind power forecasts were realized by modeling the relationship between the corrected wind speed and power output using a censored GP. To validate the proposed approach, three real-world datasets were used for model training and testing. The empirical results were compared with several classical wind forecast models, and based on the mean absolute error (MAE), the proposed model provides around 9% to 14% improvement in forecasting accuracy compared to an artificial neural network (ANN) model, and nearly 17% improvement on a third dataset which is from a newly-built wind farm for which there is a limited amount of training data.

Index Terms—Censored data, Gaussian process, numerical weather prediction, wind power forecasting.

I. INTRODUCTION

As a green and renewable energy resource, the utilization of wind energy has been growing rapidly around the world [1]. In China, the current total capacity of wind farms is approximately 75.6 GW, with a growth rate of 21.2% in 2012 [2]. However, the intrinsically variable and uncontrollable characteristics of wind pose several operational challenges. Thus wind power prediction is an essential process for the maintenance of wind power units and energy reserve scheduling [3], [4].

Accurate short-term wind power forecasts with a prediction horizon from one hour to several days are critical to optimize the scheduling of wind farm maintenance and reserve electricity generation: these have an impact on grid reliability and market-based ancillary service costs. Broadly speaking, there are three

approaches for short-term wind power forecasting: physical numerical weather prediction (NWP) models, statistical models based purely on historical data, and statistical models with NWP data as additional exogenous inputs. Physical models have advantages over longer horizons (from several hours to dozens of hours), because they include (3-D) spatial and temporal factors in a full fluid-dynamics model of the atmosphere. However, such models have limitations, such as the limited observation set for calibration (a serious issue given the extremely large number of variables in these models), the relatively limited spatial resolution possible over such a wide area (typically the whole earth), and the impossibility of accounting for local topography [5]. Statistical models which use only historical wind speed and power data and do not include any explicit model of the physical processes have been developed based on single models or hybrids of several methods, such as autoregressive integrated moving average (ARIMA) [6], [7], Kalman filters [8], artificial neural networks [9], [10], support vector machines [11], [12] and Bayesian theory [13]. Since wind has an intrinsic non-stationary character, these purely statistical models are effective only for very short-term forecasts (about 1–4 hours ahead), and generally present very inaccurate predictions as the forecast horizon grows [14]. To overcome these limitations and obtain longer forecast horizons, some authors have combined both types of model, by using NWP data from a physical model as inputs to a statistical model [15], [16].

Two forecasting systems based on artificial neural networks using NWP data and measured power generation from supervisory control and data acquisition (SCADA) systems as inputs for short-term wind power prediction (over a horizon of 72 hours) are presented in [17]: they can be used for electricity market bid offers and wind-farm maintenance tasks. The average RMS errors of the two presented forecast systems vary from 14% for the 12–24 hour horizon to 19.7% for the 48–72 hour horizon. Another forecast model combined with NWP data is proposed by Vaccaro [18], in which one-day-ahead wind power forecasting is based on information amalgamation from a local atmospheric model and measured data. De Giorgi [19] developed a series of forecast systems by combining Elman and MLP neural networks, and predicted power production of a wind farm with three wind turbines for 5 time horizons: 1, 3, 6, 12 and 24 hours. The normalized absolute average error of the presented systems vary from 5.67% to 15.50% depending on forecast time and different combinations of networks.

Recently, as an effective nonlinear prediction method, Gaussian processes (GP) have been applied in many domains, both in regression [20] and classification [21], including wind energy prediction. Jiang and Dong focused on very short term

Manuscript received January 07, 2013; revised May 25, 2013 and August 22, 2013; accepted September 04, 2013. Date of publication October 02, 2013; date of current version February 14, 2014. This work was supported in part by the National Fund for Creative Groups of China (Grant No. 61121003) and in part by a BUAA Scholarship, which funded N. Chen's visit to Aston University. Paper no. TPWRS-00016-2013.

N. Chen, Z. Qian, and X. Meng are with Beihang University, Beijing, China. I. T. Nabney is with the Non-linearity and Complexity Research Group, Aston University, Birmingham, U.K. (e-mail: i.t.nabney@aston.ac.uk).

Digital Object Identifier 10.1109/TPWRS.2013.2282366

(< 30 min) wind-speed prediction using GPs [22]. They evaluated their model on real-world datasets, and found that the GP performs better than ARMA and the Mycielski algorithms [23].

In this paper, wind-farm datasets including NWP results and measured data from SCADA system are analyzed and applied to wind power forecasting over a horizon of up to one day, using a Gaussian Process. The main innovations in this paper are:

- 1) The predicted wind speed from an NWP model is *corrected* using a GP with predicted wind speed, wind direction, temperature, humidity and pressure as inputs. The simplification of the regression task (by modeling a correction to the NWP data) helps to improve performance compared with earlier methods.
- 2) Automatic relevance determination (ARD) is used for feature selection in order to improve generalization performance.
- 3) A *censored* Gaussian process (CGP) method is applied to build the relationship between corrected wind speed and wind power, as wind power generation is censored due to the controlling strategy of the wind turbine.
- 4) A subset of high-wind speed data is treated separately, because of its different characteristics as shown on analysis of the initial models.
- 5) Historical wind-speed data from a SCADA system is used as an additional input to the forecasting model for 1–4 hours-ahead prediction, since we have proved this to be effective in this range of time horizons.
- 6) The influence on forecast accuracy of training set history is investigated, and empirical results show that the GP method is able to obtain better prediction performance when there is limited data for model training than other techniques.

Section II describes the NWP model used to provide daily forecasts of meteorological variables. Section III defines the Gaussian process, ARD and censored GP methods applied to build all the regression models. Section IV describes the whole detailed modeling process from NWP data to forecast wind power results. Simulation results are presented and analyzed in Section V. Finally, Section VI includes the overall conclusions.

II. NWP MODEL

Numerical weather prediction uses hydrodynamic and thermodynamic models of the atmosphere to predict weather based on certain initial-value and boundary conditions. Though there were attempts since the 1920s, reliable numerical weather prediction was not achieved until the development of powerful computers. Nowadays, a number of global and regional NWP forecast models have been developed. In our research, we use the weather research and forecasting (WRF) NWP model. WRF was created through a partnership that includes National Center for Atmospheric Research (NCAR), National Oceanic and Atmospheric Administration (NOAA), and more than 150 other organizations and universities in the United States and internationally, and is released as a free model for public use [24].

In general, short-term wind-power forecasting needs predictions from an NWP model with high spatial resolution. In our case, the WRF model with the advanced research WRF (ARW) core is used. First, the accurate geographical position and hub

height of one wind turbine or wind mast in a wind farm is chosen as a reference point, then a 3-layer nested grid is built centered around this point. Finally, using the initial data provided by Global Forecast System (GFS) model, the NWP data can be abstracted from the corresponding point in the 3rd layer. The NWP data used in our paper is available at 20:00 GMT every day. The data including wind speed and direction, temperature, humidity and pressure, is provided at an interval of 10 min for the following 72 hours.

The Chinese government imposes very specific demands on wind-power forecasting in the energy sector, which include the following conditions: the forecast error of 1–4 hour ahead should be less than 10% of wind turbine's installed capacity; the 5–24 hour-ahead forecasting error should be below 20% of installed capacity. All the forecast errors contained in this paper are calculated using hourly data. Therefore, we focus on forecasting for a horizon from 1 to 24 hours for wind power, based on hourly sampled NWP data.

III. GAUSSIAN PROCESS METHODS

Recently, there has been much activity concerning the application of Gaussian process to machine learning tasks. Because a systematic and detailed explanation of Gaussian process regression and ARD can be found in Rasmussen's book [25], and the model of *censored* GP was proposed by Groot [26], here we only provide a brief description of these methods.

A. Standard Gaussian Process

A Gaussian process $f(x)$ can be completely specified by its mean function and covariance function, written as $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$, where the mean function $m(x)$ and covariance function $k(x, x')$ are defined as

$$\begin{aligned} m(x) &= E[f(x)] \\ k(x, x') &= E[(f(x) - m(x))(f(x') - m(x'))]. \end{aligned} \quad (1)$$

Usually, the mean function is assumed to be zero and the target variable is normalized to have zero mean.

Consider the GP in a classic regression problem: assume we have a training set \mathcal{D} of n observations, $\mathcal{D} = \{(x_i, y_i) | i = 1, \dots, n\}$, where x denotes an input vector of dimension D and y denotes a scalar output. As we have n cases of x , the whole input data is represented by a $D \times n$ matrix, and with targets collected in a vector y , we can write $\mathcal{D} = (X, y)$. Therefore, the key point of the regression problem is to model the relationship between inputs and targets, that is to build a function to satisfy

$$y_i = f(x_i) + \epsilon_i \quad (2)$$

where the observed values y differ from the function values $f(x)$ by additive noise ϵ , which is assumed to be an independent, identically distributed Gaussian distribution with zero mean and variance σ_n^2 , i.e., $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. Note that y is a linear combination of Gaussian variables and hence is itself Gaussian [27]. The prior on y becomes

$$\begin{aligned} E[y] &= E[f + \epsilon] = 0 \\ \text{cov}[y] &= K(X, X) + \sigma_n^2 I \end{aligned} \quad (3)$$

where K is a matrix with elements $K_{ij} = k(x_i, x_j)$, which is also known as the kernel function.

Given a training set $\mathcal{D} = (X, y)$, our goal is to make predictions of the target variable f_* for a new input x_* . Since we already have $p(y|X, k) = \mathcal{N}(0, K + \sigma_n^2 I)$, the distribution with new input can be written as

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & k(X, x_*) \\ k(x_*, X) & k(x_*, x_*) \end{bmatrix}\right) \quad (4)$$

where $k(X, x_*) = k(x_*, X)^T = [k(x_1, x_*), \dots, k(x_n, x_*)]$, which we will abbreviate as k_* . Then according to the principle of joint Gaussian distributions, the prediction result for the target is given by

$$\begin{aligned} \bar{f}_* &= k_*^T (K + \sigma_n^2 I)^{-1} y \\ V[f_*] &= k(x_*, x_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_* \end{aligned} \quad (5)$$

Now, the whole regression model based on Gaussian process is completed.

B. Automatic Relevance Determination

In practice, rather than fixing the covariance function for a Gaussian process model, we may prefer to use a parametric function and infer the parameter values from the observed data. This learning process, which is usually called “learning the hyperparameters”, can be accomplished by maximizing the log likelihood function

$$\ln p(y|\theta) = -\frac{1}{2} \ln |K| - \frac{1}{2} y^T K^{-1} y - \frac{n}{2} \ln(2\pi). \quad (6)$$

This function is maximized using a standard nonlinear gradient-based optimization algorithm. A detailed description can be found in [25].

As the maximum likelihood can be used to determine the parameter values in the Gaussian process, we can extend this technique by incorporating a separate lengthscale parameter for each input variable, and hence the relative importance of different inputs can be inferred from the observed data, which is known as ARD. For example, the covariance function used in our model is the rational quadratic

$$k(x, x') = \theta_0 \left(1 + \sum_{i=1}^D l_i (x_i - x'_i)^2\right)^{-v} + b \quad (7)$$

where b represents a bias, and all the hyperparameters can be contained in a vector $\theta = (\theta_0, L, b)^T$, where L denotes the vector of all the lengthscale parameters $L = \{l_1, \dots, l_D\}$. The hyperparameters $\{l_1, \dots, l_D\}$ are used to implement ARD: as l_i becomes larger, the function becomes relatively more sensitive to the corresponding input variable x_i . Therefore, the importance of each input variable is revealed, and inputs with a small value of l_i can be discarded.

C. Censored Gaussian Process

As a result of the wind-turbine control strategy, there is always a defined upper limit to the power generation of each turbine type. Denote the upper limit of power generation by C , the

unrestricted power output by y^* (a *latent* variable since it is not observed directly), and the observed output of wind turbine as y , then with the physical non-negative property of power also considered, we have

$$y = \begin{cases} 0 & \text{if } y^* \leq 0 \\ y^* & \text{if } 0 < y^* < C \\ C & \text{if } y^* \geq C. \end{cases} \quad (8)$$

In statistical terminology, the true values are “censored” in that they are not observed directly but instead a different value is substituted. Exploratory analysis of the data shows that 1.1% percent of the data is approximately (within 1% of) the upper limit C and a further 5.4% percent is within 5% of the upper limit (and is thus in the range where the noise distribution of the unrestricted prediction overlaps significantly with the censored range). Thus it is important to account for this constraint in the model itself rather than simply pre-process the predictions of a “standard” regression model by thresholding them at C .

Following Groot’s paper [26], we assume that the latent values $y^* = f(x)$ can be realized by a Gaussian process regression model, and use $\phi(\cdot)$, $\Phi(\cdot)$ to denote the standard normal density and cdf respectively, then the likelihood can be expressed as a mixture of Gaussian and probit likelihood terms

$$\begin{aligned} L &= \prod_{i=1}^n p(y_i | f_i) \\ &= \prod_{y_i=0} \left[1 - \Phi\left(\frac{f_i}{\sigma}\right)\right] \\ &\quad \times \prod_{0 < y_i < C} \left[\frac{1}{\sigma} \phi\left(\frac{y_i - f_i}{\sigma}\right)\right] \prod_{y_i=C} \left[\Phi\left(\frac{f_i - C}{\sigma}\right)\right]. \end{aligned} \quad (9)$$

By Bayes’ theorem, the posterior distribution over the latent variables is the product of the prior and the likelihood

$$p(f|X, y) = \frac{1}{Z} p(f|X) \prod_{i=1}^n p(y_i | f_i) \quad (10)$$

where $Z = p(y|X)$ is a normalization term called the marginal likelihood. However, this posterior is analytically intractable because of the form of the likelihood in (9): thus one has to use approximate techniques for computing it. Here we use an expectation propagation (EP) method: assume that the likelihood of latent variable f_i is $p(y_i | f_i) \simeq t_i(f_i | \tilde{Z}_i, \tilde{\mu}_i, \tilde{\sigma}_i^2)$, and abbreviate it as t_i . Then the posterior $p(f|\mathcal{D})$ is approximated by $q(f|\mathcal{D})$ where

$$\begin{aligned} q(f|\mathcal{D}) &= \frac{1}{Z_{EP}} p(f) \prod_{i=1}^n t_i(f_i | \tilde{Z}_i, \tilde{\mu}_i, \tilde{\sigma}_i^2) \\ &= \frac{1}{Z_{EP}} p(f) \mathcal{N}(\tilde{\mu}, \tilde{\Sigma}) \prod_{i=1}^n \tilde{Z}_i \end{aligned} \quad (11)$$

where Z_{EP} is the normalization term. Then we update the t_i approximations sequentially by the EP algorithm, finally the approximation to the posterior is computed, and the

predictive distribution $q(f_*)$ of input x_* can be written as $q(f_*) = \mathcal{N}(\mu_*, \sigma_*^2)$, where

$$\begin{aligned}\mu_* &= k_*^T (K + \tilde{\Sigma})^{-1} \tilde{\mu} \\ \sigma_*^2 &= k(x_*, x_*) - k_*^T (K + \tilde{\Sigma})^{-1} k_*\end{aligned}\quad (12)$$

where $k_* = [k(x_1, x_*), \dots, k(x_n, x_*)]$, as defined in Section III-A.

IV. MODELING PROCESS

The forecasting framework proposed in this paper uses GP models and three additional features: 1) ARD is used to select model inputs; 2) predicted wind speed from the NWP model is corrected before modelling the mapping from wind speed to wind power; 3) detailed adjustments were applied to improve forecast accuracy, such as including historical data and building a separate model for high wind speed.

The NWP data usually includes several kinds of meteorological information: wind speed, wind direction, temperature, air pressure, and humidity. It is clear that wind power mainly depends on the actual wind speed, however, we do not know for sure if any other variable plays an important role, too. The ARD toolbox in Netlab [28] was used to determine the appropriate inputs for each model mentioned in this paper.

Since wind power generation is mainly determined by wind speed, the most important practical task for companies operating wind farms is to get accurate wind speed prediction. There are two ways to obtain wind power from NWP data: directly learning the model between NWP data and wind power using a censored GP; or correcting the error in NWP wind speed prediction first and then using this corrected wind speed to replace original NWP speed for forecasting power generation. The second method is based on the belief, underpinned by a large body of empirical analysis, that there are some systematic and stochastic biases present in the original NWP speed forecasts—the detailed empirical analysis is illustrated in Section V-B. We denote the first way of modeling wind power as GP-Direct, and the second as GP-CSpeed (meaning based on corrected speed).

A schematic of the whole modeling process of the proposed GP-CSpeed model, including ARD inputs selection and NWP speed correction, is shown in Fig. 1. Firstly, an ARD model (ARD I) is applied to determine which features in NWP are most relevant to predict measured wind speed, then the corrected wind speed is obtained using a GP model with the selected inputs. Second, the wind speed in NWP data is replaced with the corrected speed from the first step, and another ARD model (ARD II) is used to choose features that are relevant to power generation. Finally, with the data of selected features used as inputs, a censored GP-based model is built for forecasting wind power.

Fig. 2 shows the detailed structure of the proposed correction process, which obtains corrected wind speed from selected NWP variables as shown in the dashed line box in Fig. 1. Certain constraints are used to improve the accuracy of modeling. Since the character of wind changes with the diurnal cycle, the

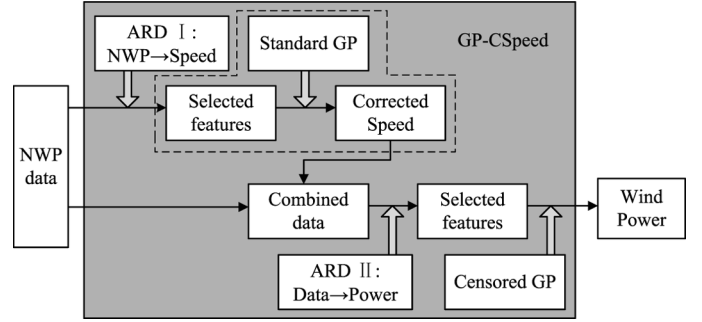


Fig. 1. Modeling process of the proposed GP-CSpeed model.

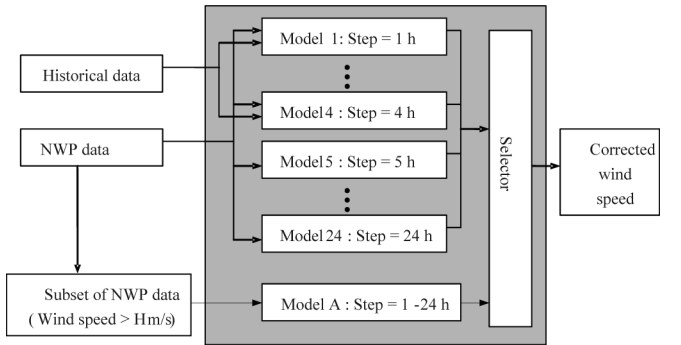


Fig. 2. Structure of correction process.

prediction error of the NWP model shows different properties depending on the forecast horizon (1–24 hours). Therefore, we built a separate model for each of the 24 hourly forecast horizons and the training dataset was divided into 24 subsets. Secondly, as mentioned before, historical data of measured wind speed turns out to be useful for very short-term forecasting, 1–4 hours ahead for our datasets according to empirical analysis (refer to Section V-B), and is therefore included as an input for models 1–4 only (i.e., 1–4 hours forecast horizon). Thirdly, the prediction error of the NWP model is larger and more related to humidity when the predicted wind speed is high, thus we developed a separate GP model (labeled Model A in the figure) specifically to improve the correction accuracy when the predicted wind speed is larger than H m/s, where H denotes the threshold value of high speed and depends on the location of the wind farm.

As shown in Fig. 2, models 1–24 give results of correction for 1–24 hours-ahead predicted wind speed, while model A gives corrected wind speed for high speeds. The reason for just building one model for high wind speed, is that such data only accounts for about 5% of the whole dataset, and thus there is not enough data for training several high-speed models separately. The “selector” component means that when the original predicted wind speed in the NWP data is larger than H m/s, then the corresponding corrected speed in the result of model 1–24 is replaced by the output of model A.

V. EXPERIMENTAL VALIDATION

Three real-world datasets based on wind farms were used in this paper to evaluate our approach. The first one is from a wind farm in Gansu province (denoted as Farm-G), which is

TABLE I
WIND TURBINE PROPERTIES

Location	Installed Capacity (kW)	Height (m)	Number
Farm-G1	850	55	58
Farm-G2	1500	65	33
Farm-J	1500	80	67
Farm-R	1500	80	48

located in the windy western part of China. The second, denoted by Farm-J, is from Jiangsu province, a coastal area located in southern China. These two farms are about 2400 km apart, and therefore the weather conditions are independent from each other. The third one is a recently built wind farm nearby Farm-J, denoted by Farm-R.

As shown in Table I, Farm-G is a large wind farm with 91 wind turbines, which are divided into two groups: the installed capacity of one kind is 850 kW, and the other is 1500 kW. Actually, the NWP data of Farm-G were obtained from two different models separately developed for these two groups, because of the difference in turbine height. Correspondingly, all the prediction models established in this paper are developed separately for different wind turbine groups. Datasets from Farm-G and Farm-J have a time scale from April 2010 to April 2013: we have used a whole year from April 1, 2010 to April 1, 2011 as a training set, and the remainder as a test set. The newly built Farm-R only has 2 and a half months' data, and is specially modeled in Section V-D, to illustrate the performance of the proposed GP-CSpeed model on a small dataset.

A. Forecasting Accuracy Evaluation

Several criteria were used to evaluate the accuracy of the proposed approach. This accuracy is computed as a function of the measured wind speed or power. Four error measures were employed for model evaluation and model comparison: the root mean square error (RMSE), the normalized root mean square error (NRMSE), the mean absolute error (MAE), and normalized mean absolute percentage error (NMAPE). The error measures are defined as follows:

$$e_t = y_t - \hat{y}_t \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (14)$$

$$NRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \times \frac{100}{C} \quad (15)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (16)$$

$$NMAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{C} \right| \times 100 \quad (17)$$

where y_t represents the actual observation value at time t , \hat{y}_t represents the forecast value for the same period, n is the number of forecasts, and the installed capacity of wind farm is denoted by C .

TABLE II
ARD RESULTS ON TWO FARMS

Location	Wind speed	Wind direction	Temperature	Air pressure	Humidity
Farm-G1	1.5696	0.0002	3.4872	9.6381	2.5239
Farm-G2	0.3882	0.0463	2.2647	3.5917	0.9233
Farm-J	0.1203	0.0014	3.2269	0.0383	0.3569

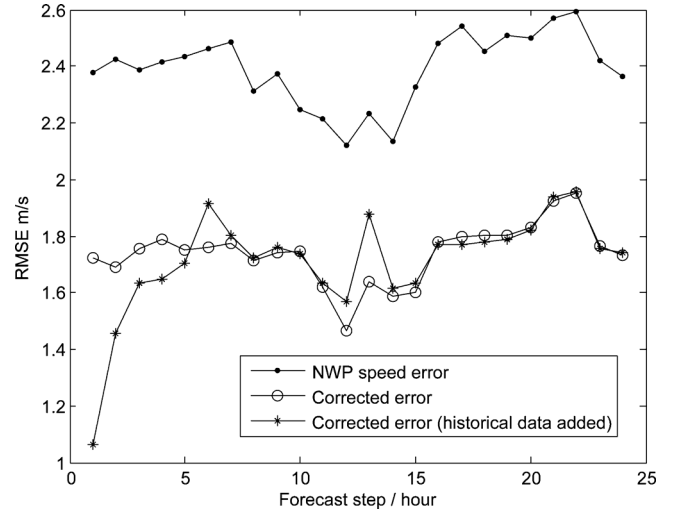


Fig. 3. Comparison of forecast error of 3 models.

The NMAPE is used for wind power as a requirement of the Chinese government, and it is also a suitable expression to interpret the quality of forecast, since without normalization, the same value of $|e|$ may imply very different performance according to different wind farms with variant installed capacity.

As specified requirements are made separately for 1–4 hours ($NMAPE \leq 10\%$) and 5–24 hours ($NMAPE \leq 20\%$) forecast, we define an extra criterion to determine the prediction performance

$$P_x = \frac{1}{m} |\{ |e_i| \leq x \cdot C \}| \times 100\% \quad (18)$$

where m denotes the size of the corresponding test dataset, and x denotes the threshold NMAPE value, i.e., $x = 0.1$ for 1–4 hours forecast and $x = 0.2$ for 5–24 hours forecast. In this way the percentage of points which meet the requirements can be calculated.

B. Effectiveness of Proposed Model

First of all, ARD was applied to determine which NWP variables should be included as inputs to the speed correction model. Using the difference between measured wind speed and NWP predicted speed as the target variable, the relevance values are listed in Table II.

We note that wind speed, temperature and humidity impact the prediction accuracy of 3 groups of wind turbines in 2 wind farms. For Farm-G, air pressure plays a much more important role, which is reasonable considering its high altitude. Therefore, we choose these three features as inputs to the GP correction process for Farm-J, and add air pressure for Farm-G. By the same selection process, ARD is applied to choose features

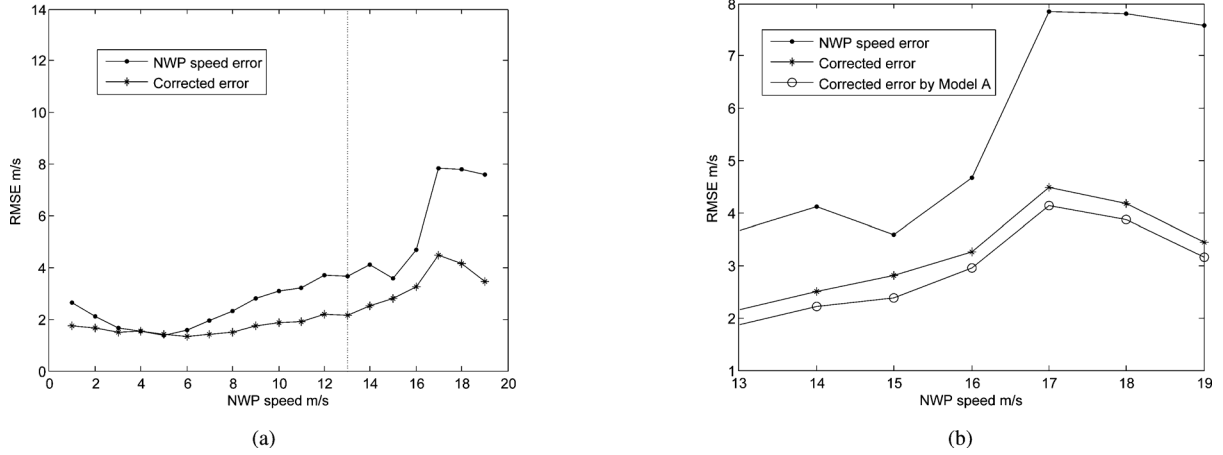


Fig. 4. Influence of building additional model for high speed NWP data. (a) Standard correction model. (b) Comparison with additional high-speed model.

TABLE III
ARD RESULTS ON HIGH-SPEED SUBSET

Location	Wind speed	Wind direction	Temperature	Air pressure	Humidity
Farm-G2	3.3159	0.0002	18.183	12.436	26.099

that are relevant to wind power: corrected speed, temperature, air pressure and humidity for Farm-G, corrected speed, wind direction and temperature for Farm-J.

In order to illustrate the effectiveness of the whole modeling process, we show the detailed results for a single wind turbine from Farm-J. A comparison of the original NWP wind speed error and the results of two correction models (with or without historical data) is displayed in Fig. 3. It can be clearly observed that both corrected results are much better than the original NWP data, and for the short-horizon forecasts, the addition of historical data successfully reduced prediction error, but introduced a small amount of additional error (probably due to over-fitting) as the forecast horizon grows. Therefore, we chose to add historical data for 1–4 hours ahead wind speed correction only.

In order to display and analyze the performance of the correction models, we first divided the dataset by NWP-predicted wind speed \hat{y}_t , that is $\mathcal{D}_s = \{s-1 \leq \hat{y}_t < s, s = 1, 2, \dots\}$, then calculated the RMSE of each subset \mathcal{D}_s , and plotted the RMSE value against s . As shown in Fig. 4(a), the performance of the correction models 1–24 is consistently good when $s \leq 10$: however, for higher wind speed the corrected results become worse. On further analysis, it was noticed that humidity is influential as wind speed increases, especially for $s > 13$, as shown by Table III (compare with Table II).

As a result, the high-speed subset is selected with a threshold $H = 13$ m/s for Farm-J, and special correction model A can be built with predicted wind speed, temperature and humidity as inputs. By the same analysis process, the threshold value is determined to be $H = 12$ m/s for Farm-G1, and $H = 13$ m/s for Farm-G2. The correction results for the high-speed subset are shown in Fig. 4(b), which illustrates the effectiveness of building model A for high speed data.

TABLE IV
SPEED FORECAST ERROR OF A TURBINE IN FARM-J

Data	RMSE(m/s)	MAE(m/s)	MAE Improvement
NWP-speed	2.4160	1.8663	-
Corrected speed	1.6405	1.2399	33.56%

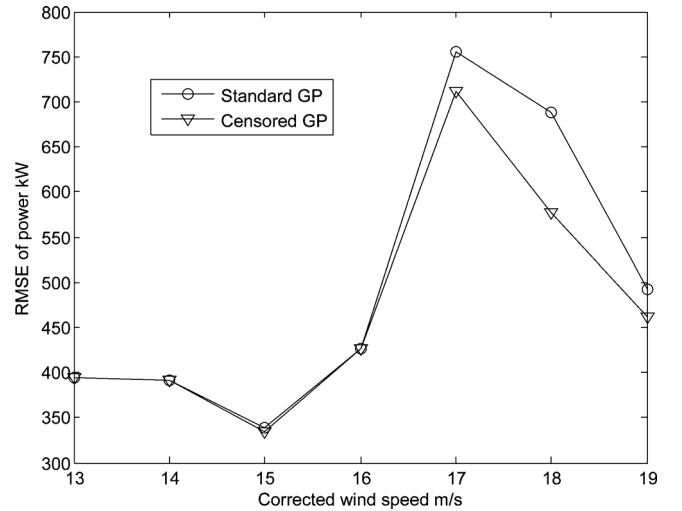


Fig. 5. Power prediction accuracy for normal and censored GP.

With the whole correction process, the corrected wind speed has a much better test-set accuracy than the original NWP wind-speed predictions, as shown in Table IV.

Since the generated wind power is limited to the installed capacity of a wind turbine by a controller, a censored GP is used in this paper to build the wind power forecasting model. The RMSE of predicted wind power from both censored and standard GP models is calculated and the predictions with a high corrected wind speed s is plotted in Fig. 5, because this is where the generated wind power might be censored.

The standard GP results include a threshold on the GP output at capacity, but the GP can predict impossible values, whereas the censored GP builds the constraint into the parameter

learning process. From the figure, we can clearly see that the censored and standard GPs perform equally well at medium speed, but as the wind speed grows, the difference becomes more obvious and the censored GP is more accurate. The improvement in accuracy at very high predicted wind speeds is because at those speeds (provided the speed prediction is reasonably accurate), the turbine output is always installed capacity and hence is easier to predict.

C. Comparative Results

In this section we present the results of our power-prediction framework and some benchmarks: the persistence model, ARIMA method, and a multi-layer perceptron (MLP) neural network. The persistence method simply uses the current value as the forecast, which means that at time t , the prediction $\hat{y}_{t+1} = \hat{y}_{t+2} = \dots = \hat{y}_{t+24} = y_t$. Since sometimes the wind power data may be invalid because of a turbine fault (invalid data was excluded from the dataset at the pre-processing stage), we use the current wind speed data in this model, and calculate the wind power with the wind turbine curve. The ARIMA method is a classical time series based model that has been frequently applied to wind forecasting [7], and is therefore picked as a benchmark.

MLP networks have been applied to short-term wind power forecasting before, and have achieved a much better performance than the persistence method [29], hence an MLP-based model which builds a relationship between selected NWP features (using ARD) and wind power is chosen for comparison. Historical data is also added as input to MLP model for 1–4 hour ahead forecast. A 9-neuron hidden layer was chosen for the MLP model, based on empirical cross-validation results (model comparison on a validation set).

An example of one-day-ahead wind power forecasting for Farm-G2 is illustrated in Fig. 6. The forecast error is reasonably small for very-short term forecasting (1–4 hours ahead). The figure clearly shows how the forecast error increases with the forecast horizon, but the proposed GP-CSpeed model still captures the shape of the actual power time series, and has better performance than the MLP model. The results of applying the proposed model to the long test datasets are shown in Tables V–VII.

As shown in Tables V–VII: the proposed GP-CSpeed model has better performance than the other models. In terms of MAE, the improvement of accuracy compared to persistence method is 17.74%, 30.34% and 44.21% for 3 datasets. If comparing to MLP model, the improvement is 14.26% for dataset of Farm-G1, 9.52% for Farm-G2, and 11.05% for Farm-J.

Please note: for both wind turbine groups in Farm-G, the proposed GP-CSpeed model succeed in improving forecast performance by all the criteria; but for Farm-J, though with a lower average forecast error, the GP-CSpeed model is a little bit worse than MLP model on 1–4 hour forecasting (criterion $P_{0.1}$ in Table VII). Therefore, we have analyzed the forecast results of Farm-J more deeply by plotting the Diebold-Mariano (DM) test [30] result for each forecast horizon (1–24 hour) in Fig. 7. DM test was designed for comparing predictive accuracy of different models, where larger DM test result (absolute value) shows larger difference.

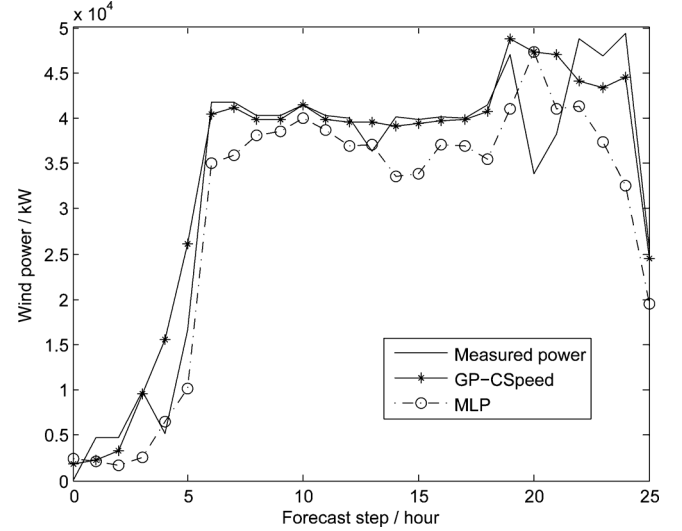


Fig. 6. One-day-ahead wind power forecasting.

TABLE V
POWER FORECAST ERROR OF FARM-G1 (NOMINAL POWER: 49.3 MW)

Model	RMSE (MW)	NRMSE	MAE (MW)	NMAPE	Improve- $P_{0.1}$ ment ¹	$P_{0.1}$ 1-4h	$P_{0.2}$ 5-24h
Persistence	9.61	19.5%	6.43	13.04%	—	69.96%	75.02%
ARIMA	9.55	19.37%	6.36	12.90%	1.02%	70.71%	75.36%
MLP	8.65	17.54%	6.17	12.52%	3.96%	72.09%	78.24%
GP-Direct	8.61	17.45%	6.13	12.43%	4.65%	71.84%	78.60%
GP-CSpeed	7.63	15.48%	5.29	10.72%	17.74%	75.97%	84.26%

1. Improvement is calculated relative to the persistence method.

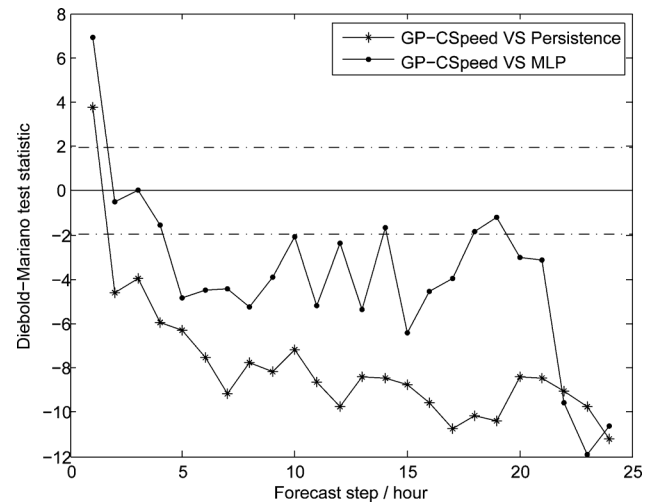


Fig. 7. DM test results of Farm-J.

As we can see from the figure, the forecast accuracy of the proposed GP-CSpeed model is the worst for 1-hour-ahead forecast (DM test statistic value $S > 0$), but the best at all the other points ($S < 0$) and shows a significant improvement as the forecast horizon grows, as most of the DM statistic values exceed the 5% significant level, i.e., $|S| > 1.96$. The overall DM statistic value of Farm-J is -20.40 for GP-CSpeed model comparing to MLP model, thus shows a significant improvement on forecast accuracy.

TABLE VI
POWER FORECAST ERROR OF FARM-G2 (NOMINAL POWER: 49.5 MW)

Model	RMSE (MW)	NRMSE	MAE (MW)	NMAPE	Improve- $P_{0.1}$ ment ²	$P_{0.2}$ 1-4h	$P_{0.2}$ 5-24h
Persistence	10.88	21.99%	7.90	15.97%	—	65.13%	65.74%
ARIMA	10.74	21.71%	7.75	15.66%	1.91%	66.26%	66.74%
MLP	8.46	17.08%	6.09	12.30%	22.99%	68.77%	78.71%
GP-Direct	8.43	17.04%	6.06	12.24%	23.35%	69.03%	78.82%
GP-CSpeed	7.90	15.96%	5.51	11.12%	30.34%	72.16%	82.19%

2. Same as I.

TABLE VII
POWER FORECAST ERROR OF FARM-J (NOMINAL POWER: 100.5 MW)

Model	RMSE (MW)	NRMSE	MAE (MW)	NMAPE	Improve- $P_{0.1}$ ment ³	$P_{0.2}$ 1-4h	$P_{0.2}$ 5-24h
Persistence	26.34	26.21%	18.22	18.13%	—	64.73%	63.39%
ARIMA	26.21	26.08%	17.94	17.85%	1.54%	67.77%	64.27%
MLP	17.32	17.24%	12.04	11.98%	33.94%	71.89%	79.87%
GP-Direct	17.48	17.39%	12.20	12.14%	33.06%	70.66%	79.31%
GP-CSpeed	15.63	15.55%	10.71	10.66%	41.21%	69.61%	83.64%

3. Same as I.

TABLE VIII
AVERAGE COMPUTATIONAL COST

Model	MLP	GP-Direct	GP-CSpeed
Time	13.27s	83.64s	83.17s

All the models were run on a single computer (Intel Core 2 Duo CPU, 2.93 GHz, 2 GB memory): the mean computational cost for each model is listed in Table VIII.

D. Effectiveness of Proposed Model for Small Dataset

Since wind power is seen as one of the most important carbon-free energy generation methods, many new wind farms are being built [2]. During their initial operation it is hard to obtain a significant quantity of historical data to build accurate forecasting models. In this section, we investigate the effectiveness of our GP method when there is little data. First of all, to analyze the influence of training set size, based on Farm-G2 dataset, various forecast results are obtained by changing the length of the training set, while test set is still 1 year long (from April 1, 2011 to April 1, 2012). The forecasting accuracy of proposed GP-CSpeed model is compared to MLP model in Fig. 8.

As shown in Fig. 8, the prediction error of the GP-CSpeed model is already quite low when there is only one month of training set data. In addition, as the training set grows, the GP-CSpeed model achieves a stable high performance quickly, while the accuracy of the MLP model is still relatively poor.

Then, to validate the performance of the proposed model in a real-world application, a dataset from a newly built wind farm was used to build a forecast model and obtain prediction results. The newly built wind farm we focus on is located in the coastal area of China where wind energy is abundant, and there are plans to construct more wind farms in the region in the near future. The data, including 2 and a half months' NWP and SCADA data, is divided into a 45-day training set and a 1-month test set. Note that one advantage of splitting the power prediction problem into separate wind-forecasting and power-prediction

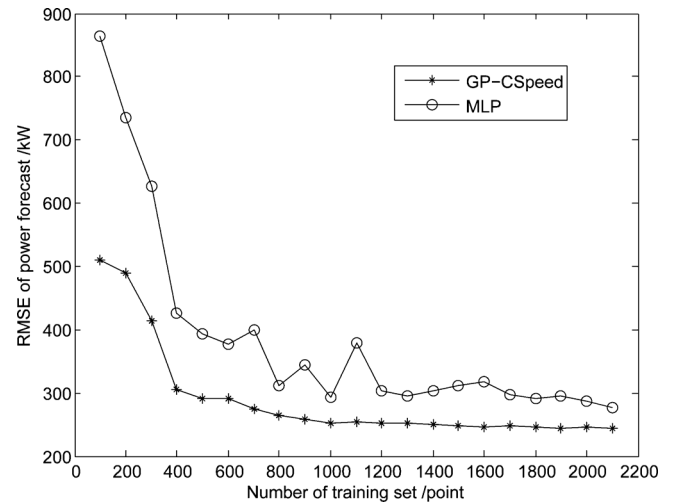


Fig. 8. Forecasting accuracy influenced by length of training set.

TABLE IX
WIND POWER FORECAST ERROR OF FARM-R (NOMINAL POWER: 72 MW)

Model	RMSE (MW)	NRMSE	MAE (MW)	NMAPE	Improve- $P_{0.1}$ ment ⁴	$P_{0.2}$ 1-4h	$P_{0.2}$ 5-24h
Persistence	13.19	18.32%	8.03	11.15%	—	72.88%	85.86%
MLP	9.96	13.84%	6.56	9.12%	18.22%	66.95%	91.92%
GP-CSpeed	8.42	11.69%	5.47	7.60%	31.85%	73.73%	92.42%

4. Same as I.

tasks is that the model mapping wind speed to power output can be reused on new wind farms provided the turbine is of a known type. The forecasting results are contained in Table IX.

As shown in Table IX, despite of the limited quantity of data, the proposed GP-CSpeed model still performs an acceptable accuracy of forecast. This advantage is especially obvious compared with the MLP model with a 16.67% improvement.

VI. CONCLUSION

Short-term wind power forecasting, which strongly impacts the safety and economics of the electricity grid, is an important and challenging task, considering the uncontrollable and stochastic nature of wind. Current research in this domain based on statistical models can be divided into two parts: forecasts several hours ahead using only historical data; predictions dozens of hours ahead based on NWP data.

In this paper, we investigated a combination of numeric and probabilistic models: a GP combined with an NWP model was applied to one-day-ahead wind power forecasting. Certain methods were employed to improve the forecast accuracy: predicted wind speed is firstly corrected by GP before it is used to forecast wind power; as there is a defined limit on power generation of wind turbine, a censored GP is applied to build speed-power model; ARD is used to choose effective NWP variables as inputs of each model; for 1–4 hour ahead forecasts, historical data is added into modeling process; and a high wind-speed subset is treated separately by building a single forecast model.

The simulation results shows that, compared to an MLP model, the proposed model has around 9% to 14% improvement of accuracy for the regular large datasets of Farm-G

and Farm-J, and hence the effectiveness and performance of the GP-CSpeed model is proved. Furthermore, the proposed GP-CSpeed model is especially effective when there is limited training data, as it achieves an significant 16.67% improvement for dataset from newly built Farm-R.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for the comments which have materially improved this paper.

REFERENCES

- [1] J. K. Kaldellis and D. Zafirakis, "The wind energy evolution: A short review of a long history," *Renew. Energy*, vol. 36, no. 7, pp. 1887–1901, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S096014811000085>.
- [2] G. W. E. Council, "Global wind statistics 2012," in *Global Wind Report*, 2012, pp. 1–4.
- [3] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Freitas, "A review on the young history of the wind power short-term prediction," *Renew. Sustain. Energy Rev.*, vol. 12, no. 6, pp. 1725–1744, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032107000354>.
- [4] M. Lei, L. Shiyang, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032108000282>.
- [5] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *Proc. North Amer. Power Symp. (NAPS)*, 2010, pp. 1–8.
- [6] H. Liu, E. Erdem, and J. Shi, "Comprehensive evaluation of ARMA arch(m) approaches for modeling the mean and volatility of wind speed," *Appl. Energy*, vol. 88, no. 3, pp. 724–732, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261910003934>.
- [7] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renew. Energy*, vol. 34, no. 5, pp. 1388–1393, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0960148108003327>.
- [8] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos, "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering," *J. Wind Eng. Ind. Aerodynam.*, vol. 96, no. 12, pp. 2348–2362, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167610508001074>.
- [9] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Appl. Energy*, vol. 87, no. 7, pp. 2313–2320, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261909005364>.
- [10] Y.-Y. Hong, H.-L. Chang, and C.-S. Chiu, "Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs," *Energy*, vol. 35, no. 9, pp. 3870–3876, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360544210003117>.
- [11] S. Salcedo-Sanz, E. G. Ortiz-García, Á. M. Pérez-Bellido, A. Portilla-Figueras, and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4052–4057, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417410010249>.
- [12] J. Zhou, J. Shi, and G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 52, no. 4, pp. 1990–1998, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0196890410005078>.
- [13] Y. Jiang, Z. Song, and A. Kusiak, "Very short-term wind speed forecasting with Bayesian structural break model," *Renew. Energy*, vol. 50, no. 0, pp. 637–647, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0960148112004764>.
- [14] M. G. De Giorgi, A. Ficarella, and M. Tarantino, "Error analysis of short term wind power prediction models," *Appl. Energy*, vol. 88, no. 4, pp. 1298–1311, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S030626191000437X>.
- [15] S. Salcedo-Sanz, Á. M. Pérez-Bellido, E. G. Ortiz-García, A. Portilla-Figueras, L. Prieto, and D. Paredes, "Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction," *Renew. Energy*, vol. 34, no. 6, pp. 1451–1457, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S096014810800390X>.
- [16] S. Al-Yahyai, Y. Charabi, and A. Gastli, "Review of the use of numerical weather prediction (NWP) models for wind energy assessment," *Renew. Sustain. Energy Rev.*, vol. 14, no. 9, pp. 3192–3198, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032110001814>.
- [17] I. J. Ramirez-Rosado, L. A. Fernandez-Jimenez, C. Monteiro, J. Sousa, and R. Bessa, "Comparison of two new short-term wind-power forecasting systems," *Renew. Energy*, vol. 34, no. 7, pp. 1848–1854, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0960148108004321>.
- [18] A. Vaccaro, P. Mercogliano, P. Schiano, and D. Villacci, "An adaptive framework based on multi-model data fusion for one-day-ahead wind power forecasting," *Elect. Power Syst. Res.*, vol. 81, no. 3, pp. 775–782, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378779610002816>.
- [19] M. G. De Giorgi, A. Ficarella, and M. Tarantino, "Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods," *Energy*, vol. 36, no. 7, pp. 3968–3978, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360544211003215>.
- [20] G. Salimi-Khorshidi, T. E. Nichols, S. M. Smith, and M. W. Woolrich, "Using Gaussian-process regression for meta-analytic neuroimaging inference based on sparse observations," *IEEE Trans. Med. Imag.*, vol. 30, no. 7, pp. 1401–1416, 2011.
- [21] Y. Bazi and F. Melgani, "Gaussian process approach to remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 186–197, 2010.
- [22] X. Jiang, B. Dong, L. Xie, and L. Sweeney, H. Coelho, R. Studer, and M. Wooldridge, Eds., "Adaptive Gaussian process for short-term wind speed forecasting," in *Proc. 19th European Conf. Artificial Intelligence (ECAI 2010)*, Lisbon, Portugal, Aug. 16–20, 2010, vol. 215, ser. Frontiers in Artificial Intelligence and Applications, pp. 661–666, IOS Press.
- [23] F. O. Hocaoglu, M. Fidan, and O. N. Gerek, "Mycielski approach for wind speed prediction," *Energy Convers. Manage.*, vol. 50, no. 6, pp. 1436–1443, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0196890409000788>.
- [24] D. Hosansky, "Weather forecast accuracy gets boost with new computer model," *UCAR*, 2006. [Online]. Available: <http://www.ucar.edu/news/releases/2006/wrf.shtml>.
- [25] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [26] P. L. G. Perry, "Gaussian process regression with censored data using expectation propagation," in *Proc. 6th Eur. Workshop Probabilistic Graphical Models*, 2012.
- [27] C. Chatfield and A. J. Collins, *Introduction to Multivariate Analysis*. London, U.K.: Chapman and Hall, 1980.
- [28] I. T. Nabney, *Netlab: Algorithms for Pattern Recognition*. New York, NY, USA: Springer, 2004.
- [29] N. Amjadi, F. Keynia, and H. Zareipour, "Short-term wind power forecasting using ridgelet neural network," *Elect. Power Syst. Res.*, vol. 81, no. 12, pp. 2099–2107, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378779611001945>.
- [30] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Business Econ. Statist.*, vol. 13, no. 3, pp. 253–263, 1995. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599>.

Niya Chen was born in Anhui, China, in 1987. She received the B.Sc. degree in electrical engineering in 2007 from Beihang University, Beijing, China, where she is currently pursuing the Ph.D. degree.

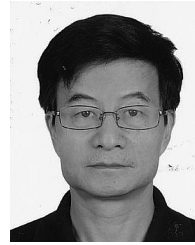
Her research interests include renewable energy sources and artificial intelligence techniques.





Zheng Qian was born in Zhengzhou, China, in 1973. He received the Ph.D. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2000.

He was a Postdoctoral Fellow with Tsinghua University, Beijing, China. Since 2002, he has been a Professor of instrument science and technology with Beihang University, Beijing. His research interests include advanced sensor technology, high voltage test, application of renewable energy, etc.



Xiaofeng Meng received the Ph.D. degree from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1995.

He became a Professor at Beijing University of Aeronautics and Astronautics in 1999. His research interests include dynamic test, automatic test and fault analysis, application of renewable energy, etc.



Ian T. Nabney received the B.A. degree in mathematics from Oxford University, Oxford, U.K., in 1985, and the Master of Advanced Study and Ph.D. degree in mathematics from Cambridge University, Cambridge, U.K., in 1986 and 1989, respectively.

Between 1989 and 1994, he worked as a researcher at Logica Cambridge Ltd. He joined Aston University as a lecturer in 1995 and was promoted to Professor in 2004. His research interests span both theory and applications of principled methods of pattern analysis (particularly using Bayesian techniques), time series analysis and data visualization. He was the system architect and lead developer of the Netlab pattern analysis toolbox, which has been downloaded more than 40 000 times. He is the chair of the Natural Computing Applications Forum and was awarded a Young Researcher Prize by Pfizer in 2001.