Wind Energy

WIND ENERGY
Wind Energ. 2016; **19**:2321–2336
Published online 27 April 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/we.1985

RESEARCH ARTICLE

Adaptive robust polynomial regression for power curve modeling with application to wind power forecasting

Man Xu[1], Pierre Pinson[2], Zongxiang Lu[1], Ying Qiao[1] and Yong Min[1]

[1] Department of Electrical Engineering, Tsinghua University, Beijing, China
[2] Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

ABSTRACT

Wind farm power curve modeling, which characterizes the relationship between meteorological variables and power production, is a crucial procedure for wind power forecasting. In many cases, power curve modeling is more impacted by the limited quality of input data rather than the stochastic nature of the energy conversion process. Such nature may be due the varying wind conditions, aging and state of the turbines, etc. And, an equivalent steady-state power curve, estimated under normal operating conditions with the intention to filter abnormal data, is not sufficient to solve the problem because of the lack of time adaptivity. In this paper, a refined local polynomial regression algorithm is proposed to yield an adaptive robust model of the time-varying scattered power curve for forecasting applications. The time adaptivity of the algorithm is considered with a new data-driven bandwidth selection method, which is a combination of pilot estimation based on blockwise least-squares parabolic fitting and the probability integral transform. The regression model is then extended to a more robust one, in which a new dynamic forgetting factor is defined to make the estimator forget the out-of-date data swiftly and also achieve a better trade-off between robustness against noisy data and time adaptivity. A case study based on a real-world dataset validates the properties of the proposed regression method. Results show that the new method could flexibly respond to abnormal data at different lead times and has better performance than common benchmarks for short-term forecasting. Copyright © 2016 John Wiley & Sons, Ltd.

KEYWORDS
wind power modeling; forecasting; local polynomial regression; recursive least squares; adaptive estimation; robust estimation

Correspondence
Man Xu, Department of Electrical Engineering, Tsinghua University, Beijing, China.
E-mail: m-xu10@mails.tsinghua.edu.cn

Received 30 September 2015; Revised 10 March 2016; Accepted 10 March 2016

NOMENCLATURE

The main notations used in the paper are stated next for quick reference. Other symbols are defined as required.

y_i	Response variable at time i, $i = 1, 2, \ldots, N$
$\mathbf{x}_i$	The vector of explanatory variables at time i
x_i^k	The k th dimension of $\mathbf{x}_i$, $k = 1, 2, \ldots, D$
θ	Smooth function of multi-input-single-output system
$\mathbf{x}_{(j)}$	The j th fitting point of θ, $j = 1, 2, \ldots, J$
$\phi_{n,(j)}$	The local coefficients for a given $\mathbf{x}_{(j)}$ at time n
$\beta_{n,(j)}$	The exponential forgetting function of past observations for a given $\mathbf{x}_{(j)}$ at time n
λ	Forgetting factor
$w_{i,(j)}$	The weight of $\mathbf{x}_i$ to $\mathbf{x}_{(j)}$
$\overline{\lambda}_{i,(j)}$	The effective forgetting factor of $\mathbf{x}_i$ for a given $\mathbf{x}_{(j)}$
$h_{(j)}^k$	The bandwidth for $x_{(j)}^k$, the k th dimension of $\mathbf{x}_{(j)}$

W	The tricube kernel function
$e_{n,(j)}$	Residual of y_n
M	The number of past records used for obtaining the optimal bandwidth
V	The variance of the smoother
B	The bias of the smoother
σ^2	The conditional variance function
f	The marginal density of the explanatory variable
N_S	The number of blocks for the block-wise parabolic fitting
F_{CDF}	The cumulative distribution function
ψ	The derivative of the loss function
$\mathbf{c}$	$\mathbf{c} = [c_{\text{inf}}, c_{\text{sup}}]$, inferior and superior threshold points of the M-type loss function
α	The proportion of residuals considered as suspicious
a_e, b_e, c_e	Parameters of dynamic forgetting factor
q	Forecasting lead time

1. INTRODUCTION

Many studies on short-term wind power forecasting with lead times from a few hours to a few days ahead can be found in the literature with focus on both point and probabilistic forecasting.[1–4] As input to operational problems, short-term wind power forecasting has great value in dealing with the challenges stemming from the participation of wind energy in electricity markets and quantifying necessary power systems' reserves,[5–7] among others. Some studies have quantified the benefits of improving forecasting accuracy for better decision-making in power system operation.[8–10] For example, it has been shown in[8] that inaccurate forecasting may lead to high operating cost and increased wind power curtailment for autonomous power systems. Ref.[9] compares the market price based on five different forecast scenarios and states that the value of the short-term forecasting in electricity markets depends significantly on its accuracy.

Most forecasting tools involve a conversion procedure from meteorological forecasts to wind power, i.e. power curve modeling. Forecasting accuracy is strongly dependent on the quality of power curve modeling, the typical contribution of which to the forecasting uncertainty could be 10% to 15% of the Root Mean Square Error (RMSE).[11]

The conversion function described by the power curve is basically nonlinear and non-stationary because of the fluctuating and stochastic nature of the wind resource, and it also comprises a noise component which represents all the unavailable microscopic interactions. Moreover, empirical power curves display much more scattered data because of the limited quality of wind power data used for modeling and forecasting, which is common for a substantial number of wind farms throughout the world. Because these wind farms are located in remote and harsh areas where the communication and maintenance conditions are not satisfactory, and interferences during measurement, transmission and conversion of the raw data make the data quality lower than expected. Another reason is the lack of on-site information related to real-time operation, e.g. maintenance plans, wind turbine operation status and special wind power operation schedules. This leads to unforeseen changing relationship between the variables involved, eventually contributing to the scattered and non-stationary characteristics of the power curve greatly.

Consequently, it is crucial for forecast accuracy improvement to make the wind power conversion function adaptive and robust. Raw data preprocessing before training is one possible way to handle data quality related problems in power curve modeling.[12,13] Those methods could mitigate the effect of abnormal data to some extent, but they do not help capturing the time-varying and scattered properties of the power curve essentially. An alternative approach is to perform power curve modeling with adaptive algorithms. A variety of different approaches have been put forward to model power curves, motivated by monitoring and forecasting, aiming at describing their nonlinear relationship between meteorological variables and power generation.[14,15] Some popular approaches, like the data-mining ones,[16–18] try to model an equivalent steady state power curve under normal operating conditions, with the intension to filter abnormal data, which lack time-adaptivity for on-site use. Very little literature exists, however, regarding how to track non-stationary power curves when a considerable proportion of the data points are corrupted because of the reasons introduced in the above.

Local polynomial regression (LPR) has nice nonparametric features and is considered as an appealing approach for wind power curve modeling and forecasting.[19–21] The time adaptivity of LPR can be obtained from a combination of recursive least squares (RLS) and exponential forgetting of old measurements.[22] The value of the forgetting factor of RLS-LPR reflects the modeling ability to capture the changing parameters and the requirement to reduce wind power forecasting errors introduced by the inaccurate power curve model based on outdated information. An appropriate local bandwidth is essential for the regression model to attain a good bias-variance balance on wind power data with time-varying characteristics.[23] In this paper, we use the RLS-LPR model of[22] as the reference model and refer to it as the basic RLS-LPR model.

A new RLS-LPR based approach is developed in this paper, so as to yield a reliable model of the nonlinear power curve for forecasting applications. We refine the basic RLS-LPR to a highly adaptive model that can achieve both robustness against noisy data and time adaptivity to the non-stationary relationship into consideration. And, a new data-driven bandwidth selection scheme is proposed to adapt to the density of the data used as input. Then, the basic model is robustified based on an M-type estimator, i.e. the robust estimator that uses the Huber loss function.[24] A new dynamic forgetting factor is defined to make the estimator forget the outdated data swiftly in order to capture the most recent relationship in the dataset. Power curve modeling could be influenced by the uncertainty in both response and explanatory variables. The enhanced RLS-LPR approach mainly deals with the low data quality problem related to wind power (i.e. the response variable). Like most robust approaches,[25] the explanatory variables are assumed to be noise free.

The paper is structured as follows. The key problems we aim to deal with and the basic RLS-LPR regression model are described in Section 2. In order to get better performance in terms of both time adaptivity and robustness, several improvements to the LPR model are made in Section 3, including a new bandwidth selection method, a robust M-type estimator proposal and a dynamic tracking of the forgetting factor. A case study based on real-world data is considered in Section 4 to verify the performance of the enhanced model. Finally, conclusions and perspectives regarding future works are gathered in Section 5.

2. LOCAL POLYNOMIAL REGRESSION FOR POWER CURVE MODELING

2.1. The impact of highly corrupted wind power data on power curve modeling

Figure 1 gives some representative situations with wind power samples and adaptively updated power curves to illustrate the modeling problem brought by the potential limited quality of wind power observations. Figure 1(a) is based on actual wind speed data measured by a wind mast and artificial wind power data. The method in[20] is applied to simulate noise-free wind power data. The generated wind power data, modeled by a double exponential function, are then combined with Gaussian noise series with given standard deviations. From this figure one can see that the data points are less scattered than those in Figure 1(b). In this case, even an averaging model that goes through the center of the data points could provide a good result without delicate consideration about the time-varying and noisy features of the dataset.

In Figure 1(b), the wind speed data are the same as in Figure 1(a), while the wind power data are real-world measurements by the SCADA system of the wind farm. The conversion from wind speed to power varies dramatically from time to time, and the data points are more scattered with much noise. On the one hand, one can see that a model estimated only on the normal sample points will perform poorly when used for the abnormal points, such as Period 2 in Figure 1(b), caused by the practical reasons introduced in Section 1. On the other hand, severe mistakes could happen if no robust estimation approach is used and the model estimation accounts for all the dispersive points. Thus the power curve model needs to be enhanced by combining time adaptivity and robust estimation, in order to deal with the non-stationary and scattered properties of the power curve.

The data used here for the illustration is not related to the case study in Section 4. For the training process of the power curve estimation with application to wind power forecasting, the wind speed data could be from wind measurements or

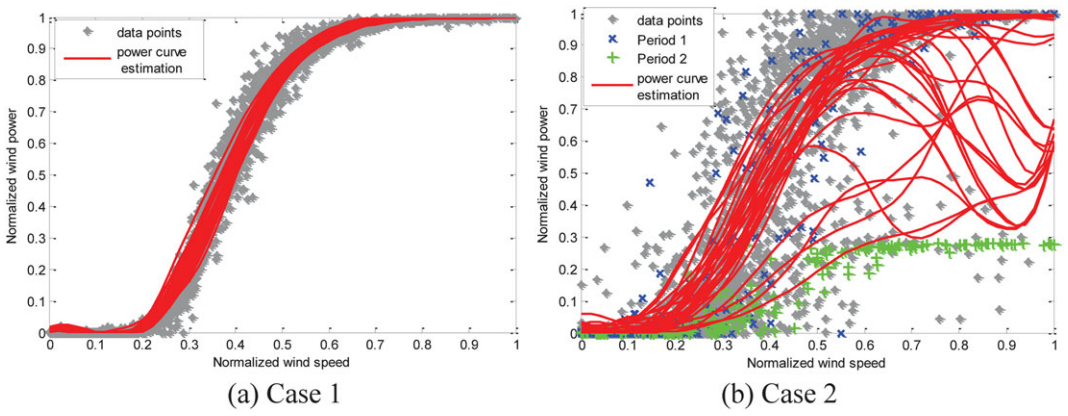

Figure 1. Influence of the dataset quality on power curve modeling. Note: Case 1 refers to the dataset with less noise and simple non-stationary relationship and Case 2 refers to the highly corrupted dataset. For case 2, period 1 and period 2 have occurred successively.

historical NWP results.[26] In this paper, we only use NWP as the training data. For cases with other types of training data, our approach could be also applied especially when the wind power data are highly corrupted.

2.2. Basic estimating model and its recursive formulation

This section is devoted to the framework of RLS-LPR, which can be equally used with meteorological measurements and forecasts as input, with application to power curve monitoring or forecasting. The basic model is outlined in a compact way to make the paper self-content.

In a general manner, we have a set of multi-input-single-output observations: $[\mathbf{x}_i, y_i]$, $i = 1, 2, \ldots, N$, $\mathbf{x}_i = \left[x_i^1, \ldots, x_i^k, \ldots, x_i^D\right]^{\mathrm{T}}$. The response variable follows a general functional of the explanatory variables:

$$y_i = \theta(\mathbf{x}_i) + \varepsilon_i, \tag{1}$$

where a model for θ fitted in a least-squares sense specifies the conditional mean of y_i given $\mathbf{x}_i$, while ε_i is the white noise. As for the modeling of a power curve with a forecasting application in mind, y_i is the power measurement at time i and $\mathbf{x}_i$ is the corresponding vector of meteorological forecasts, such as wind speed and wind direction, which are the drivers of the power generation process.

At time step n ($n \leq N$), the model in 1 is approximated by local polynomial fitting for some distinct values of the space spanned by $\mathbf{x}_i$, $i = 1, 2, \ldots, n$. Such values are the fitting points for θ, denoted by $\mathbf{x}_{(j)}, j = 1, \ldots, J$. For a single fitting point, the local linear model is formulated as

$$y_i = \mathbf{p}(\mathbf{x}_i)^{\mathrm{T}} \boldsymbol{\phi}_{n,(j)} + \varepsilon_i, \tag{2}$$

where $\mathbf{p}(\mathbf{x}_i)$ is for the polynomial terms evaluated at $\mathbf{x}_i$, $\boldsymbol{\phi}_{n,(j)}$ are the local coefficients and ε_i is a white noise. Equation 2 is fitted using weighted least squares, i.e.

$$\widehat{\boldsymbol{\phi}}_{n,(j)} = \underset{\boldsymbol{\phi}_{n,(j)}}{\arg\min} \sum_{i=1}^{n} \beta_{n,(j)}(i) w_{i,(j)} \rho\left(y_i - \mathbf{p}(\mathbf{x}_i)^{\mathrm{T}} \boldsymbol{\phi}_{n,(j)}\right), \tag{3}$$

where the loss function ρ is commonly chosen as quadratic, in line with the idea of minimizing the RMSE of the resulting forecasts. The exponential forgetting function $\beta_{n,(j)}$ reads

$$\beta_{n,(j)}(i) = \begin{cases} \overline{\lambda}_{i,(j)} \beta_{n-1,(j)}(i), & 1 \leq i \leq n-1 \\ 1, & i = n \end{cases}, \tag{4}$$

and the effective forgetting factor,

$$\overline{\lambda}_{i,(j)} = 1 - (1 - \lambda) w_{i,(j)}, \; 0 < \lambda \leq 1, \tag{5}$$

which insures that past samples are down-weighted only when new information becomes available. The weight $w_{i,(j)}$ of $\mathbf{x}_i$ is allocated by the product kernel,

$$w_{i,(j)} = \prod_{k=1}^{D} W\left(\frac{\left|x_i^k - x_{(j)}^k\right|}{h_{(j)}^k}\right). \tag{6}$$

$W(\cdot)$ writes

$$W(u) = \begin{cases} \left(1 - u^3\right)^3, & u \in [0, 1) \\ 0, & u \in [1, +\infty) \end{cases}, \tag{7}$$

such that $W: R_0 \rightarrow [0, 1]$. $h_{(j)}^k$ indicates the bandwidth for the k th dimension of a particular fitting point $\mathbf{x}_{(j)}$, which controls the regression vicinity of $\mathbf{x}_{(j)}$.

Hereafter $\theta_{(j)} = \theta(\mathbf{x}_{(j)})$ is estimated as

$$\widehat{\theta}_{(j)} = \mathbf{p}\left(\mathbf{x}_{(j)}\right)^{\mathrm{T}} \widehat{\phi}_{n,(j)}. \tag{8}$$

And for a given $\mathbf{x}_i$, $\widehat{\theta}(\mathbf{x}_i)$ is obtained by interpolation through the fitted estimates $\widehat{\theta}_{(1)}, \ldots, \widehat{\theta}_{(J)}$.

The recursive estimation of $\phi_{n,(j)}$ is summarized in[22] as following,

$$e_{n,(j)} = y_n - \mathbf{p}(\mathbf{x}_n)^{\mathrm{T}} \widehat{\phi}_{n-1,(j)}, \tag{9}$$

$$\mathbf{R}_{n,(j)} = \overline{\lambda}_{i,(j)} \mathbf{R}_{n-1,(j)} + w_{n,(j)} \mathbf{p}(\mathbf{x}_n) \mathbf{p}^{\mathrm{T}}(\mathbf{x}_n), \tag{10}$$

$$\widehat{\phi}_{n,(j)} = \widehat{\phi}_{n-1,(j)} + w_{n,(j)} \left(\mathbf{R}_{n,(j)}\right)^{-1} \mathbf{p}(\mathbf{x}_n), \tag{11}$$

where $e_{n,(j)}$ is the regression residual of y_n given $\widehat{\phi}_{n-1,(j)}$, and $\mathbf{R}_{n,(j)}$ is for the local covariance at the fitting point of interest.

3. IMPROVEMENTS TO LOCAL POLYNOMIAL REGRESSION

To deal with highly corrupted wind power measurement data, additional features are introduced in this section to improve the time-adaptivity and robustness of the basic model. As for the application to wind power modeling, $h^k_{(j)}$ in 6 is mostly determined according to the non-uniform distribution of samples or the practical implementation requirements.[22,27] To make $h^k_{(j)}$ adaptive to both the time-varying curve shape and data density, a new data driven bandwidth selection method is discussed in details in section 3.1. Since the residual distribution from power curve modeling is skewed and heavy-tailed in practice,[28] an M-type estimator is applied to down-weight the influence of the largely scattered points and improve the robustness of the LS estimator in Section 3.2. The approach follows directly from that developed in,[20] derived from the M-type estimation methods in[24] and.[29] Finally, a new dynamic tracking method for the effective forgetting factor is put forward in Section 3.3, motivated by the fact that the behavior of wind power generation can shift quickly because of the changing weather conditions or special operating strategies of the wind farm.

3.1. Optimal bandwidth selection

Since $\theta(\cdot)$ in 1 may vary considerably over time on a corrupted dataset, it is reasonable to consider that the nonlinear power curve shape may be changing when deciding on the bandwidth. The basic idea of our data driven approach is to approximate unknown quantities in the asymptotic formula of the optimal bandwidth in[30] by a pilot estimator based on minimizing the Mean Square Error over the dataset. The blockwise least-squares parabolic fitting is chosen as the pilot estimator in this paper, which is known to have good performance for uniformly sampled data.[31]

As a simplified version of that in equation 1, the regression form between a single dimension of the explanatory variable and response variable is given by:

$$y_m = \theta(x_m) + \varepsilon_m, \tag{12}$$

where $[x_m, y_m]$, $m = n - M, \ldots, n - 1$ are the past M observations and there should be a strong correlation between x_m and y_m to make ε_m close to the Gaussian noise, which lays the basis for further bandwidth selection. For wind power modeling, tests have shown that this method can be readily employed with the orthogonal components of wind (i.e. u and v), but not directly to a polar variable like wind direction. For simplicity of description, the k-index representing the dimension of the explanatory variable is omitted in this section when there is no ambiguity.

The asymptotic optimal bandwidth can by calculated with

$$\widehat{h} = \arg\min_{h} \left\{ M^{-1} h^{-1} \int V(x)\mathrm{d}x + h^4 \int B^2(x)\mathrm{d}x \right\}, \tag{13}$$

where $V(x)$ and $B(x)$ are the variance and bias of the estimator respectively. And we then obtain the optimal bandwidth at each x,

$$h(x) = M^{-1/5} \left[\frac{V(x)}{4B^2(x)} \right]^{1/5}. \tag{14}$$

One can refer to[32] for the specific formulation of $V(x)$ and $B(x)$, in which the unknown quantities, such as the conditional variance $\sigma^2(x)$, derivatives of $\theta(x)$ and marginal density $f(x)$, are estimated in the following pilot procedure.

The pilot estimation is there to estimate the unknown quantities by doing low degree polynomial fitting in each block divided over coordinate axis of the explanatory variable. The key point is choosing the proper number of blocks. The number is recommended to be 3 for computational efficiency,[31] which is in accordance with the shape of the power curve.

The pilot procedure translates to:

1) Partition the design regression interval $[\mu, v]$ of X into N_S blocks:

$$S_l = \left[\mu + \frac{(l-1)(v-\mu)}{N_S}, \mu + \frac{l(v-\mu)}{N_S}\right], \quad l = 1, 2, \ldots, N_S. \tag{15}$$

2) S_l can be split into the left part $n_{L,l} = \sum_{m=n-M}^{n-1} I(L_l \le x_m < c_l)$ and the right part $n_{R,l} = \sum_{m=n-m}^{n-1} I(c_l \le x_m < R_l)$, where the block center $c_l = (R_l + L_l)/2$, the block radius $r_l = (R_l - L_l)/2$ and $I(\cdot)$ is the indicator function. Loop over all S_l, $l = 1, 2, \ldots, N_S$ and:
   a. Use least-squares parabolic fitting in S_l to obtain the estimates of $\theta_l(x)$, $\theta_l'(x)$ and $\sigma^2(x)$.
   b. Estimate f_l and its derivative f_l' by:

$$\widehat{f}_l = \frac{n_{L,l} + n_{R,l}}{2Mr_l} \tag{16}$$

$$\widehat{f'}_l = \frac{1}{r_l}\frac{n_{L,l} - n_{R,l}}{Mr_l} \tag{17}$$

   c. Calculate $\widehat{V}(c_l)$ and $\widehat{B}(c_l)$ based on the above estimating results, which are substitutes for the unknown quantities.
   d. Calculate $h(c_l)$ using equation 14.
3) Obtain the local bandwidth function $h(x)$ by smoothing the step function with intervals valued by $h(c_l)$ based on the kernel function in equation 7, using r_l as the smoothing bandwidth to make sure that it is most accurate at the block centers.

For the clearly non-uniformly distributed samples, a uniformization method based on the probability integral transform is applied before the above pilot procedure in order to keep the selected bandwidths adaptive to the data density with high relevance. F_{CDF} is the cumulative distribution function and the probability integral transform insures that $\widetilde{X} = F_{CDF}(X)$ is uniformly distributed on the unit interval.[33] $F_{CDF}(X)$ can be estimated by user-defined fitting models based on x_m. The bandwidth selection procedure is carried out based on the transformed datasets $[\widetilde{x}_m, y_m]$ to obtain the local optimal bandwidth function $\widetilde{h}(\widetilde{x})$.

The optimal bandwidth selection for $x_{(j)}$ is summarized in Figure 2. First, $x_{(j)}$ is transformed to $\widetilde{x}_{(j)} = F_{CDF}\left(x_{(j)}\right)$. Then we have $\widetilde{h}_{(j)} = \widetilde{h}\left(\widetilde{x}_{(j)}\right)$ and the corresponding interval along $\widetilde{X}$ axis, i.e. $[\widetilde{x}_{(j),a}, \widetilde{x}_{(j),b}]$, where $\widetilde{x}_{(j),a} = \widetilde{x}_{(j)} - \widetilde{h}_{(j)}, \widetilde{x}_{(j),b} = \widetilde{x}_{(j)} + \widetilde{h}_{(j)}$. We finally transform the boundaries back to $x_{(j),a} = F_{CDF}^{-1}\left(\widetilde{x}_{(j),a}\right)$, $x_{(j),b} = F_{CDF}^{-1}\left(\widetilde{x}_{(j),b}\right)$, and set the bandwidth $h_{(j)} = \max\{|x_{(j)} - x_{(j),a}|, |x_{(j)} - x_{(j),b}|\}$.

3.2. Local robustification with an M-type estimator

The basic idea is to reduce the impact of the residuals in equation 9 that are determined to be suspicious, by using a bounded influence quadratic loss function different from that in equation 3.

ρ is selected as a Huber loss function with the relaxed symmetric influence bound. Taking the asymmetric distribution of the residuals into consideration, we have:

$$\rho(e; \mathbf{c}) = \begin{cases} c_{\inf} e - \dfrac{c_{\inf}^2}{2}, & e < c_{\inf} \\ \dfrac{e^2}{2}, & e \in \left[c_{\inf}, c_{\sup}\right] \\ c_{\sup} e - \dfrac{c_{\sup}^2}{2}, & e > c_{\sup} \end{cases}, \tag{18}$$

DOI: 10.1002/we

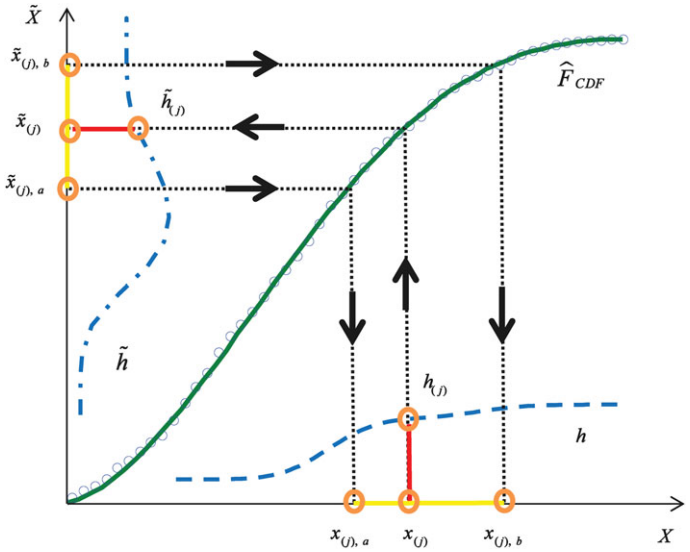

Figure 2. Illustration of the optimal bandwidth selection for $x_{(j)}$. Note: The horizontal axis is the original variable X, and the vertical axis is the transformed variable $\widetilde{X}$. The green line is the fitted $\widehat{F}_{CDF}$ based on the sampled cumulative probability values of X, which are marked by the small blue circles. The dotted blue lines are the sketches of the optimal bandwidth h, $\widetilde{h}$ for X, $\widetilde{X}$, respectively. The red lines denote the specific optimal bandwidths $h_{(j)}$, $\widetilde{h}_{(j)}$ for $x_{(j)}$, $\widetilde{x}_{(j)}$, respectively, and the yellow lines are the corresponding intervals $[\widetilde{x}_{(j),a}, \widetilde{x}_{(j),b}]$, $[x_{(j),a}, x_{(j),b}]$ along the axis. The bold arrows (upward and to the left) denote the probability integral transform from X to $\widetilde{X}$; The bold arrows (to the right and downward) denote the back transform from $\widetilde{X}$ to X.

where $\mathbf{c} = [c_{\text{inf}}, c_{\text{sup}}]$ are the inferior and superior threshold points of ρ. ρ has a quadratic loss in the central part located between c_{inf} and c_{sup}, and turns to be linear for the out-of-bound part.

The derivative of ρ is denoted by ψ, which consequently writes:

$$\psi(e;\mathbf{c}) = \begin{cases} c_{\text{inf}}, & e < c_{\text{inf}} \\ e, & e \in \left[c_{\text{inf}}, c_{\text{sup}}\right] \\ c_{\text{sup}}, & e > c_{\text{sup}} \end{cases}, \tag{19}$$

and the derivative of ψ is such that:

$$\psi'(e;\mathbf{c}) = \begin{cases} 1, & e \in \left[c_{\text{inf}}, c_{\text{sup}}\right] \\ 0, & e < c_{\text{inf}} \text{ or } e > c_{\text{sup}} \end{cases}. \tag{20}$$

Thus, the negative residuals smaller than c_{inf} and positive residuals bigger than c_{sup} are downweighted when updating the model.

Additionally, the weighted residual $e_{n,(j)}\sqrt{w_{n,(j)}}$ substitutes the residual $e_{n,(j)}$ calculated by equation 9 to make sure that the large residuals are not downweighted a second time with the loss function in its linear part. The robust estimator is reformulated as:

$$\widehat{\phi}_{n,(j)} = \underset{\phi_{n,(j)}}{\arg\min} \sum_{i=1}^{n} \beta_{n,(j)} \rho\left(\left(y_i - \mathbf{p}(\mathbf{x}_i)^{\mathrm{T}} \phi_{n,(j)}\right)\sqrt{w_{i,(j)}}, \mathbf{c}(\alpha)\right), \tag{21}$$

where the threshold points $\mathbf{c}(\alpha) = [c_{\text{inf}}(\alpha), c_{\text{sup}}(\alpha)]$ are determined adaptively based on an empirical estimation of the residual distribution, as the quantiles with proportion $\alpha/2$ and $1-\alpha/2$ of the distribution. Consequently the recursive updating formula in equations 10 and 11 are changed correspondingly to:

$$\mathbf{R}_{n,(j)} = \overline{\lambda}_{n,(j)} \mathbf{R}_{n-1,(j)} + \psi'\left(e_{n,(j)}\sqrt{w_{n,(j)}}; \mathbf{c}(\alpha)\right)\mathbf{p}(\mathbf{x}_n)\mathbf{p}^{\mathrm{T}}(\mathbf{x}_n), \tag{22}$$

$$\widehat{\phi}_{n,(j)} = \widehat{\phi}_{n-1,(j)} + \psi\left(e_{n,(j)}\sqrt{w_{n,(j)}}; \mathbf{c}(\alpha)\right)\left(\mathbf{R}_{n,(j)}\right)^{-1}\mathbf{p}(\mathbf{x}_n), \tag{23}$$

for $w_{n,(j)} > 0$. The local polynomial estimation $\widehat{\theta}_{(j)}$ at fitting point $\mathbf{x}_{(j)}$ is obtained as in equation 8.

3.3. Dynamic tracking of the effective forgetting factor

The effective forgetting factor $\overline{\lambda}_{n,(j)}$ in equation 4 controls the influence of old observations on the current LPR modeling, allowing for adaptivity in time. For really noisy and non-stationary wind power data, however, the relationship between explanatory and response variables can change quickly because of the changing weather conditions or special operating strategies of the wind farm. Therefore, it is necessary to set $\overline{\lambda}_{n,(j)}$ in a more dynamic way to forget the outdated data swiftly while tracking relevant changes effectively. Inspired by the dynamic definition for a forgetting factor in[34,35] and,[36] we propose a new version of a dynamic effective forgetting factor working for our RLS-LPR model, which is defined as:

$$\overline{\lambda}_{n,(j)} = \begin{cases} 0.995 - \dfrac{b_e}{1+\exp\left[-c_e\left(\left|e_{n,(j)}\sqrt{w_{n,(j)}}\right| - a_e\right)\right]}, & e_{n,(j)}\sqrt{w_{n,(j)}} \in \mathbf{c}(\alpha) \\ 1, & \text{otherwise} \end{cases}. \tag{24}$$

Such definition is consistent with that in equation 5, which is such that old observations are not downweighted as long as non-suspicious new information is not available. The parameter a_e controls when the forgetting factor begins to be reduced according to the weighted residual. The parameter b_e insures the minimum value of $\overline{\lambda}_{n,(j)}$ and c_e determines the reduction gradient. The working space of the dynamic forgetting factor is given by the threshold vector $\mathbf{c}(\alpha)$ in order to make sure that the time adaptivity function of $\overline{\lambda}_{n,(j)}$ is not contradictory with the robust model scaled by α.

4. APPLICATION RESULTS

A case study based on practical wind power data is used to demonstrate the proposed model. The basic model introduced in Section 2.2, with global fixed bandwidth, denoted by $\widehat{\theta}_{GF}$, is our benchmark model. The purpose is to model the conversion from meteorological forecasts to power generation and testing its performance on short-term wind power forecasting. Several tests have been carried out based on the orthogonal components of wind speed and wind direction. For those tests, however, there is no significant difference in the power curve shape and forecasting accuracy at lead times that are more than 2 hours. Consequently, the case study presented below consists in the sole conversion from wind speed to power. Wind direction should certainly affect the power curve because of the wake effects in the wind farm. And it could be investigated more in the future for other applications, e.g. power curve modeling for very short-term forecasting or based on a dataset with less noise. The conversion model in equation 1 is reformulated as:

$$y_{i+q} = \theta_{i,q}(x_{i+q}) + \varepsilon_{i+q}, \tag{25}$$

where i is the model issued time and q is the lead time. $\mathbf{x}_i,\ i=1,\ldots,N$ are reduced to a single explanatory variable $x_i,\ i=1,\ldots,N$.

4.1. Data description

The wind power dataset used here is for a wind farm in western China comprising 32 wind turbines and a nominal capacity of 96 MW. The wind farm is located at the Gobi desert with harsh natural environment. The NWP data are from ECMWF, for a grid node close to the wind farm, with lead times ranging from 1 to 24 h ahead. Meteorological forecasts at 100 m above ground level have been chosen. An hourly averaging has been made to the wind power data to match it with the hourly meteorological data. NWP data and power data are normalized by their maximum values and available for one year. It is important to note that the on-site information of the wind farm is not available, including the maintenance plan, wind turbine operation status and wind power curtailment schedule. The forecasting system, as well as the online power curve modeling, must be able to handle the raw data without an automatic data correction because of limited information.

A 10 day episode of wind power and NWP wind speed time series is depicted in Figure 3, from which one can see the examples of special periods in power generation, marked by dotted boxes. Such data are probably caused by the practical reasons introduced in Section 1, and should be accommodated for online modeling.

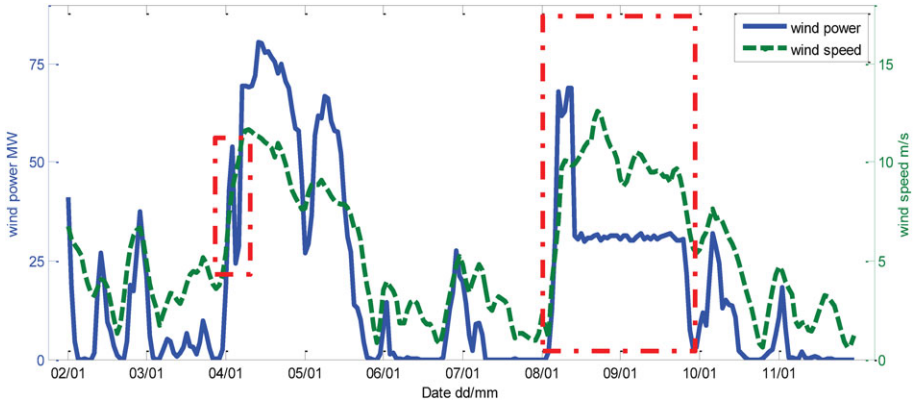

Figure 3. Time series of wind power and wind speed from NWP at a wind farm in China during a 10 day episode.

4.2. Design of the case study

For initializing the recursive process of $\widehat{\phi}_{GF}$, the recursive matrix $\mathbf{R}_{0,(j)}$, $j \in [1, J]$ is valued as a diagonal matrix $diag(\gamma)$, where γ is a small positive number. And $\widehat{\phi}_{0,(j)}$ is chosen as a vector of zeroes. For all the estimators, the fitting points $x_{(j)}$ are set uniformly as:

$$x_{(j)} = \frac{j-1}{J-1}, \; j = 1, \ldots, J. \tag{26}$$

The number of fitting points J is chosen arbitrarily as 20 since the influence of different J on the estimators is quite limited when J is large enough.

Other parameters, set by cross validation or expert judgment, are introduced in details in Section 4.3 and 4.4. The data of both wind power and NWP are divided into three groups: the first three-month data for the initialization of the estimators, the following three-month data for the cross validation of the parameters, and the last six-month data for the evaluation and comparison of the estimators.

4.3. Bandwidth selection

To have a deeper look at the proposed optimal bandwidth selection (OBS) method, other three methods are compared to $\widehat{\theta}_{OBS}$ based on the RLS-LPR model introduced in Section 2.2:

(1) Estimator with global fixed bandwidth $\widehat{\theta}_{GF}$. For all the fitting points, the bandwidth $h_{(j)}, j = 1, \ldots, J$ are fixed with the same value h_{GF}, which is determined by trial and error. This method is very simple and can achieve acceptable performance for on-site use;
(2) Estimator with nearest neighbor bandwidth $\widehat{\theta}_{NN}$. This method is also simple in nature and quite appealing to use when the data points are not uniform. The basic rule is to order the explanatory variable x_i, $i = 1, \ldots, n$ by $x_{(p_1)} \le x_{(p_2)} \le \ldots \le x_{(p_n)}$ and choose $p_s \in p_1, \ldots, p_n$ with $h(x_{(j)}) = |x_{p_s} - x_{(j)}|$ as the bandwidth. In this study, the parameter s is replaced by $s = f_s \cdot n, f_s \in (0, 1)$ and f_s is determined by trial and error;
(3) Estimator with the local bandwidth selection method $\widehat{\theta}_{LBS}$. Apply the optimal bandwidth selection method introduced in Section 3.1 directly to the non-uniform x_i, $i = 1, \ldots, n$ without the probability integral transform.

For $\widehat{\theta}_{LBS}$ and $\widehat{\theta}_{OBS}$, the $V(x)$ and $B(x)$ in equation 14 are approximated by those of local linear smoother, i.e.

$$n^{-1}h^{-1}V(x) = n^{-1}h^{-1}\int K^2(u)du\frac{\sigma^2(x)}{f(x)}, \tag{27}$$

$$h^2 B(x) = \frac{h^2}{2}Q''(x)\int u^2 K(u)du, \tag{28}$$

where $K(u) = (1-u^3)^3 I(|u| \le 1)$. N_S is chosen as 3 and M is set to be 2500, which is sufficiently large for the estimation.

Table I summarizes the 12 h ahead forecasting performance of the models. The value of the basic forgetting factor λ in equation 5 is chosen as the same for all the estimators and determined by cross validation for $\widehat{\theta}_{GF}$. All other three estimators

Table I. Forecasting error evaluation criteria for $\widehat{\theta}_{GF}$, $\widehat{\theta}_{NN}$, $\widehat{\theta}_{LBS}$ and $\widehat{\theta}_{OBS}$, normalized by the rated wind power capacity. ME, MAE and RMSE are calculated based on the evaluation set. $RMSE_L$, $RMSE_M$ and $RMSE_H$ are calculated based on data points with wind speed belongs to [0, cut-in speed], [cut-in speed, rated speed] and [rated speed, max speed] respectively. $q = 12$ h. $\lambda = 0.976$.

	$\widehat{\theta}_{GF}$	$\widehat{\theta}_{NN}$	$\widehat{\theta}_{LBS}$	$\widehat{\theta}_{OBS}$
ME %	0.12	0.06	0.19	−0.01
MAE %	12.23	12.04	12.19	12.01
RMSE %	18.17	18.03	18.11	17.92
$RMSE_L$ %	9.47	9.40	9.51	9.34
$RMSE_M$ %	20.20	19.89	20.04	19.82
$RMSE_H$ %	26.06	25.43	25.84	25.22
Key parameters	$h_{GF} = 5m/s$	$f_s = 0.75$	$N_S = 3$, $M = 2500$	$N_S = 3$, $M = 2500$

outperform $\widehat{\theta}_{GF}$ for most error criteria, showing the advantage of using an adaptive bandwidth instead of a fixed one. $\widehat{\theta}_{OBS}$ encompasses the features of $\widehat{\theta}_{NN}$ and $\widehat{\theta}_{LBS}$ in the sense of respecting the data density, as well as responding to the changes in the power curve shape and, thus has better performance than its competitors.

Table I also gives the error criteria calculated on the evaluation set for different levels of wind speed. It can be seen that the advantage of $\widehat{\theta}_{OBS}$ is more obvious at high wind speed interval, especially compared to $\widehat{\theta}_{GF}$ and $\widehat{\theta}_{LBS}$.

The bandwidth selection for all the fitting points at three randomly chosen model updating time is shown in Figure 4(a). The resulting power curve estimates for each updating time are shown in Figure 4(b), (c), (d). One can see that $\widehat{\theta}_{OBS}$ has the same trend in the bandwidth and power curve estimations as $\widehat{\theta}_{NN}$, while $\widehat{\theta}_{LBS}$ and $\widehat{\theta}_{GF}$ have the similar bandwidths and curve estimations. Without the probability integral transform procedure, LBS method has narrow bandwidths for data points of high wind speed and low data density, and finally the similar performance to $\widehat{\theta}_{GF}$ in error criteria.

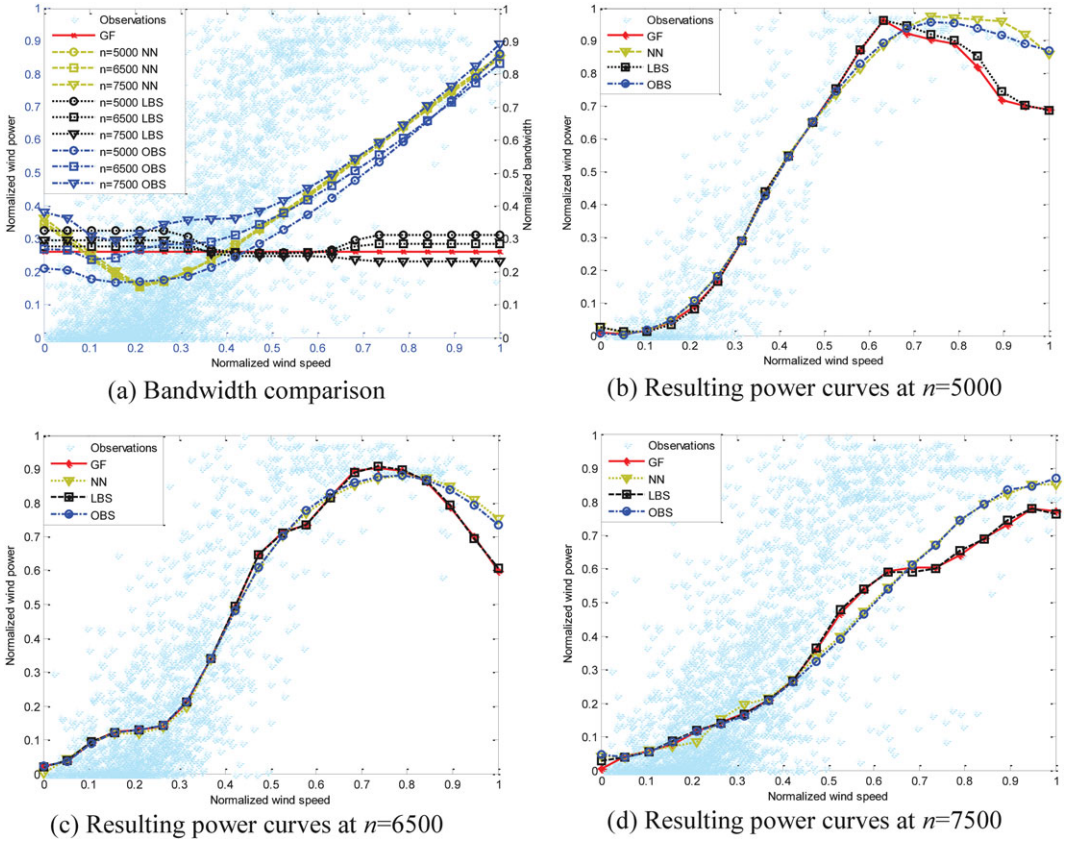

Figure 4. Results of different bandwidth selection methods at model updating time $n = 5000, 6500, 7500$ when forecasting lead time $q = 12$ h. Note: (a) gives the local bandwidth results of different methods with observations in the evaluation dataset. The lines refer to the right axis. (b), (c) and (d) give the resulting power curves with observations in the evaluation dataset at each updating time.

Wind Energ. 2016; **19**:2321–2336 © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/we

As for the comparison between $\widehat{\theta}_{OBS}$ and $\widehat{\theta}_{NN}$, there could be a few changes in the distribution of wind power data for one wind farm during a few months, but the shape of the online modeling power curve could vary distinctly from time to time because of noisy data. This is corresponding to the fact in Figure 4(a) that the three bandwidth curves of $\widehat{\theta}_{NN}$ almost remain the same while a non-negligible time-varying characteristic can be seen for OBS bandwidths at different updating time. This is the key for $\widehat{\theta}_{OBS}$'s advantage over $\widehat{\theta}_{NN}$ in the sense of time adaptivity to the changing power curve shape. Another benefit of OBS is that it is simple to choose N_S and M. This saves the efforts of trial and error, and is quite valuable for practical implementation requirements.

4.4. Robust LPR model with dynamic effective forgetting factor

To validate the properties of the proposed dynamic effective forgetting factor, the boundary weighted forgetting factor, as is introduced in,[20] is applied for comparison,

$$\overline{\lambda}_{n,(j)} = 1 - (1-\lambda)\psi'\left(e_{n,(j)}\sqrt{w_{n,(j)}}; \mathbf{c}(\alpha)\right). \tag{29}$$

For which, the user defined parameter λ should be no less than 0.95 to insure the vicinity is big enough to support the LPR.

Let $\widehat{\theta}^{\dagger}_{OBS}$ denote the robust version of $\widehat{\theta}_{OBS}$ with effective forgetting factor defined by equation 29, while $\widehat{\theta}^{\dagger}_{D,OBS}$ refers to the local robust estimator with the dynamic forgetting factor given by equation 24. The basic forgetting factor λ for $\widehat{\theta}^{\dagger}_{OBS}$ is chosen as the same value as that of $\widehat{\theta}_{OBS}$ and $\widehat{\theta}_{GF}$. The proportion parameter α is determined by cross validation for $\widehat{\theta}^{\dagger}_{OBS}$ and $\widehat{\theta}^{\dagger}_{D,OBS}$. Additional parameters of the dynamic forgetting factor are chosen by expert knowledge, as $a_e = 0.3$, $b_e = 0.4995$ and $c_e = 30$. $a_e = 0.3$ is such that the forgetting factor begins to reduce when the absolute value of the weighted residual exceeds 0.3. $b_e = 0.4995$ insures that the forgetting factor is no less than 0.5. And $c_e = 30$ gives a quick reduction speed to capture the relevant changes in the power curve in time.

Power curve modeling based on different estimators is tested for lead times from 2 to 24 h in this study. Figure 5 summarizes the performance of those five estimators with different forecasting lead times. For most lead times, the decrease in RMSE is non-negligible when going from the basic model to the robust LPR models. Figure 6 shows the evolution of the optimal value of key parameters (λ for $\widehat{\theta}_{GF}$, $\widehat{\theta}_{OBS}$, and $\widehat{\theta}^{\dagger}_{OBS}$, α for $\widehat{\theta}^{\dagger}_{OBS}$ and $\widehat{\theta}^{\dagger}_{D,OBS}$). For short lead times, i.e. a few hours ahead, it is crucial for the estimator to capture the current operation status of the wind farm. The minor values of α and λ insure that less data points are considered as suspicious, and a narrow sliding window is used for the regression to achieve higher time adaptivity to the changing non-stationary features of the dataset. With growing lead times, the current changes in the power curve could act as disturbance for the conversion model from wind speed to power in future steps. Thus it is important for the model to remain robust when dealing with the temporary dispersive data points. This results in the growing λ and α to obtain a strong robust model.

In Figure 5, it also can be seen that the benefits from employing a dynamic forgetting factor is related to the lead time and the predominance of $\widehat{\theta}^{\dagger}_{D,OBS}$ increases with shorter lead times. With the same value of α, the dynamic forgetting factor gives $\widehat{\theta}^{\dagger}_{D,OBS}$ more flexibility to track the non-stationary conversion from wind speed to power.

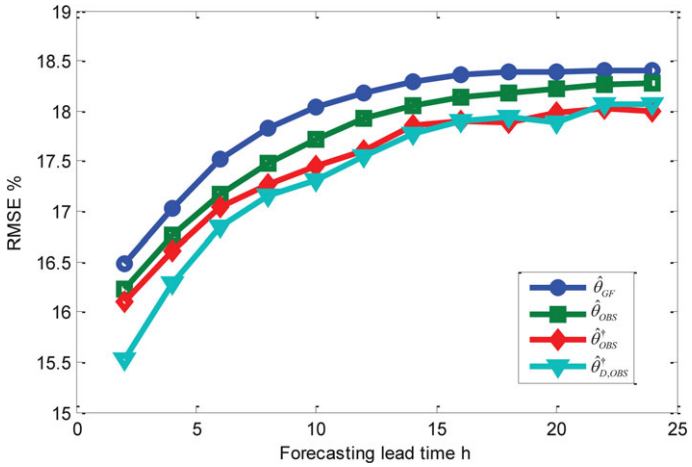

Figure 5. RMSE for different lead times and power curve models, normalized by the rated wind power capacity.

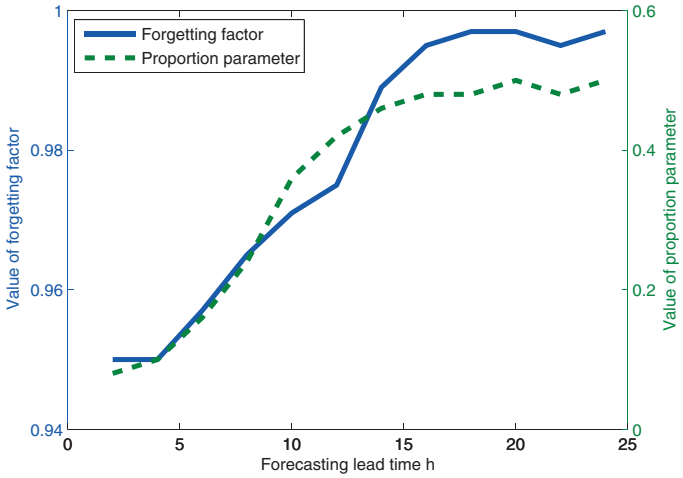

Figure 6. Optimal value of λ and α for different lead times.

To give a further explanation of the benefits from the proposed $\widehat{\theta}^{\dagger}_{D,OBS}$, an extra dataset consisting of 10 outliers, as the marked crosses shown in Figure 7, is added to the modeling test of $\widehat{\theta}^{\dagger}_{D,OBS}$, with the basic model $\widehat{\theta}_{GF}$ as the reference. The outliers have been chosen directly from the scattered plot of additional wind speed and power data. Such outliers are used to analyze the situations where the relationship between wind speed and power deviates from that of the standard power curve significantly and continuously for hours.

The last recursive step of the evaluation set is mark by t_0 as the initial state and the power curve updated time at each added point is marked by t_i, $i = 1, 2, \ldots, 10$. We define L_G as the distance between the power curve and a group of data points, as is illustrated in Figure 7. L_G is calculated with

$$L_G = \left\| \widehat{\theta}_{t_i} - T_G \right\| = \sqrt{\sum_{i=1}^{G} \left(\widehat{\theta}_{t_i}(x_{t_i}) - y_{t_i} \right)^2}, \tag{30}$$

where $\widehat{\theta}_{t_i}$ is the power curve estimation at t_i; T_G is a group of data points, $T_G : \left[x_{t_i}, y_{t_i}\right], i = 1, 2, \ldots, G$, and G is the total number of data points in T_G. For the outlier test, T_G consists of the chosen outliers, i.e. $\left[x_{t_i}, y_{t_i}\right], i = 1, 2, \ldots, 10$. We use the changes in L_G from time step t_0 to t_{10} to illustrate the responses of $\widehat{\theta}_{t_1}, \ldots, \widehat{\theta}_{t_{10}}$ to the outliers.

Figure 8 shows the influence of the outliers on L_G at different lead times. $q = 2$ h and $q = 12$ h are chosen as two typical lead times to address the features of $\widehat{\theta}^{\dagger}_{D,OBS}$. When $q = 2$ h, $\widehat{\theta}^{\dagger}_{D,OBS}$ has a clear advantage in responding to the test points

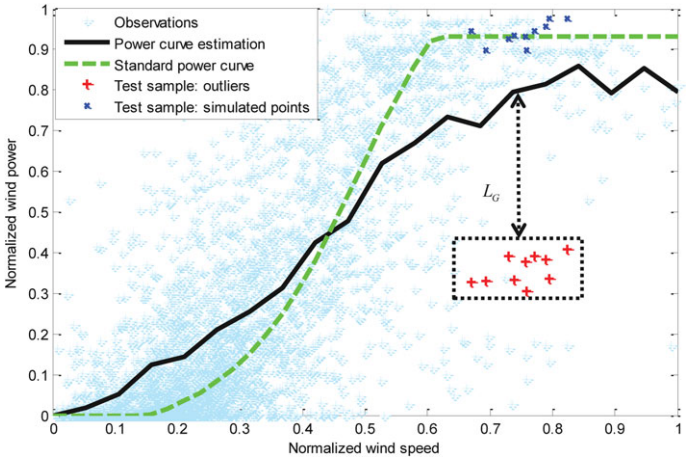

Figure 7. The added test samples and the illustration of the distance L_G between the group of outliers and the estimating power curve.

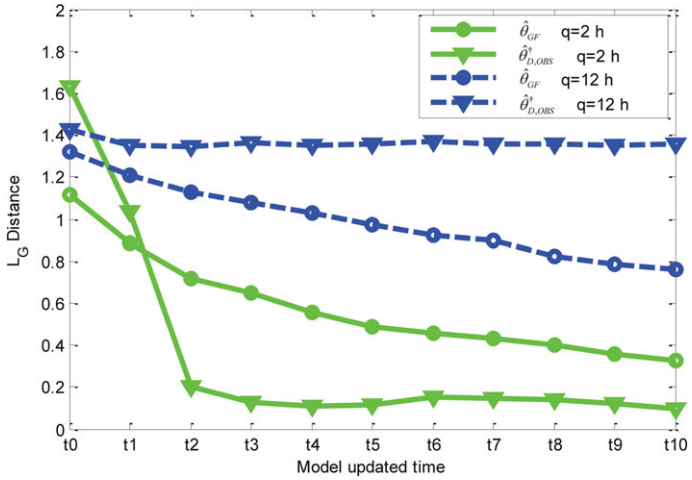

Figure 8. Comparison in the changing L_G for different estimators at $q = 2$ h and $q = 12$ h.

rapidly and accurately because of the dynamic nature of our effective forgetting factor. When $q = 12$ h, the robustness of $\widehat{\theta}^{\dagger}_{D,OBS}$ is strengthened with $\alpha = 0.42$ and the modeled power curves remain almost unaffected by the outliers. One can see from Figure 8 that $\widehat{\theta}_{GF}$ could also achieve the tradeoff between robustness and time adaptivity to some extent by adjusting λ. The dynamic forgetting factor enables $\widehat{\theta}^{\dagger}_{D,OBS}$ to make such a trade-off more flexible, and this is the key for its better performance at different lead times.

To illustrate how the dynamic forgetting factor contributes to power forecasting, we add another group of test data to demonstrate the normal operating conditions of the wind farm, marked by t_i, $i = 11, 12, \ldots, 20$. These data, depicted by the dark stars in Figure 7, are obtained from the same wind speed series of the outliers and the simulated wind power based on the standard power curve combined with a Gaussian noise.

The forecasts with $\widehat{\theta}_{GF}$ and with $\widehat{\theta}^{\dagger}_{D,OBS}$ are shown in Figure 9. In Figure 9(a), the forecasting results at t_3,...., t_{12} are with the power curve estimation updated with the outliers. One can see that forecasting series with $\widehat{\theta}^{\dagger}_{D,OBS}$ could track the test samples much more in a more timely manner than $\widehat{\theta}_{GF}$. While in Figure 9(b), the forecasting series with $\widehat{\theta}^{\dagger}_{D,OBS}$ at t_{13},...., t_{20} is almost unaffected by the outliers, compared to that with $\widehat{\theta}_{GF}$. Such results validate that the enhanced model could make a better forecasting at different lead times.

4.5. Statistical significance test

In order to validate the consistency of $\widehat{\theta}^{\dagger}_{D,OBS}$'s advantage in RMSE in Figure 5, statistical significance tests between the basic model and the enhanced model are shown in this section based on the Diebold–Mariano (DM) test.[37] The basic idea is to judge whether the differential between the loss functions of the compared error is significantly different from zero, or

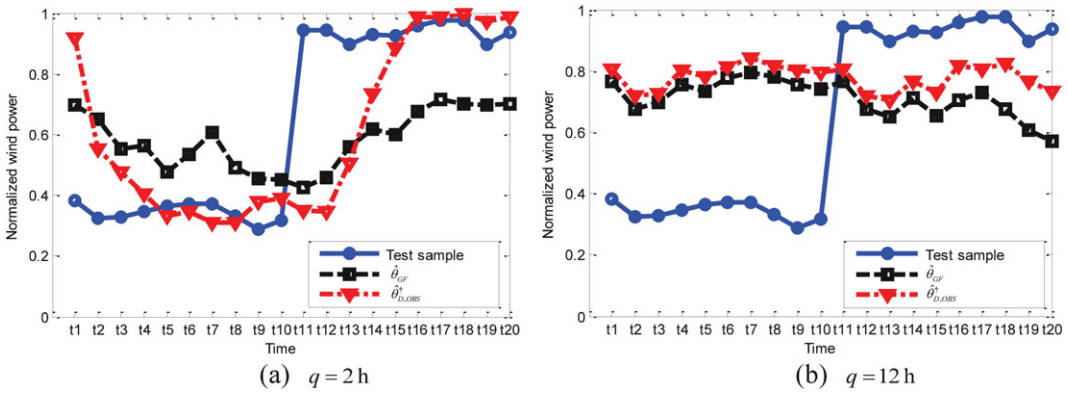

Figure 9. Forecasting time series of the test samples with the basic model and the enhanced RLS-LPR.

Table II. *p* Value of the DM test for the forecasting error comparison between $\widehat{\theta}_{GF}$ and $\widehat{\theta}_{OBS}$, $\widehat{\theta}_{GF}$ and $\widehat{\theta}^{\dagger}_{OBS}$, and $\widehat{\theta}_{GF}$ and $\widehat{\theta}^{\dagger}_{D,OBS}$ at different lead times. The value of the DM test statistic is positive for all the following cases. '*' and '**' denote that the result satisfies 0.05 and 0.1 confidence level respectively.

	2 h	4 h	6 h	10 h	12 h	14 h	20 h
$\widehat{\theta}_{GF}, \widehat{\theta}_{OBS}$	0.1213	0.0946**	0.0634**	0.0831**	0.1016**	0.3928	0.5150
$\widehat{\theta}_{GF}, \widehat{\theta}^{\dagger}_{OBS}$	0.1049**	0.088**	0.0517*	0.0269*	0.0397*	0.0674**	0.1793
$\widehat{\theta}_{GF}, \widehat{\theta}^{\dagger}_{D,OBS}$	0.0007*	0.0077*	0.0303*	0.0175*	0.0314*	0.0486*	0.0593**

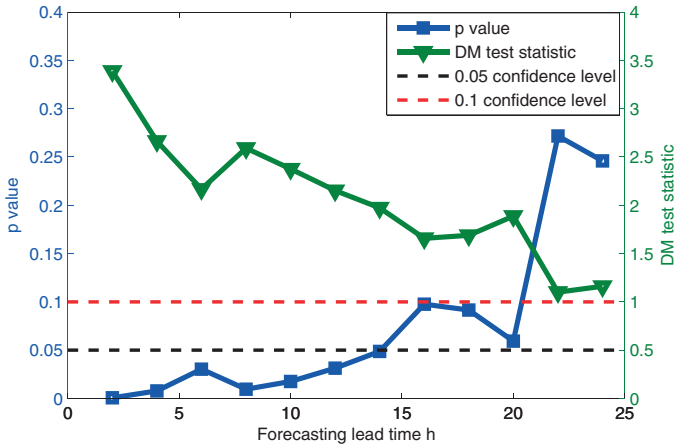

Figure 10. Results of the DM test between $\widehat{\theta}_{GF}$ and $\widehat{\theta}^{\dagger}_{D,OBS}$ at different lead times.

not. In the null hypothesis of equal forecasting error, the DM statistic, the specific formulation of which can be found in,[37] tends to have standard Gaussian distribution.

To help illustrate the benefit of our three proposed contributions, the DM test has been conducted between $\widehat{\theta}_{OBS}$, $\widehat{\theta}^{\dagger}_{OBS}$, $\widehat{\theta}^{\dagger}_{D,OBS}$ and $\widehat{\theta}_{GF}$. Table II gives the test results at typical lead times. From $\widehat{\theta}_{OBS}$, $\widehat{\theta}^{\dagger}_{OBS}$ to $\widehat{\theta}^{\dagger}_{D,OBS}$, the observed difference between the compared models becomes more and more significant at each lead time. To have a clear look at the improvement of all three contributions together, Figure 10 shows the test results between $\widehat{\theta}_{GF}$ and $\widehat{\theta}^{\dagger}_{D,OBS}$. One can see that the DM statistic stays positive for all forecasting lead times, reflecting the advantage of the enhanced model to some extent. The significant difference in forecasting accuracy, indicated by the p value, may be divided into three groups:

1) For lead times that are no more than 14 h, the p value remains below 0.05, which is meant that $\widehat{\theta}^{\dagger}_{D,OBS}$ has better forecasting accuracy than $\widehat{\theta}_{GF}$ at 0.05 significant level.
2) For lead times 16 to 20 h, the advantage of $\widehat{\theta}^{\dagger}_{D,OBS}$ can be validated at 0.1 significant level.
3) For lead times that are longer than 20 h, the p value is bigger than 0.1. Thus the difference in RMSE cannot be concluded to be significant.

5. CONCLUSION

Wind power curve modeling is challenged by the limited quality of input data under practical circumstances. And, the key issue here is how to consider both time adaptivity and robustness towards the non-stationary and scattered data when estimating the wind-to-power conversion model.

An adaptive and robust RLS-LPR model is proposed in this paper, which uses an optimal bandwidth selection scheme and a local robust M-type estimator combined with dynamic tracking of the effective forgetting factor. The power curve modeling for short-term wind power forecasting on a practical case study allows illustrating the benefits of the adaptive robust model in a step-by-step fashion, by going from the basic LPR model to the various refined LPR models. It has been shown that the new data-driven bandwidth selection scheme enables the local bandwidth of the LPR model to be adaptive to both data density and to changes in the power curve shape. Also, it has been clearly explained how the proposed model

Wind Energ. 2016; **19**:2321–2336 © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/we

responds adaptively at different lead times when dealing with abnormal data points. The model could obtain better performance for most lead times and the predominance increases with shorter lead times.

The modeling details are specific to the dataset used here, but the techniques and methodology are highly generalizable for wind power modeling with different forecasting lead times and data quality. The model has great value for online applications. It makes use of past data for current estimation to track the changing state of the power curve with an apparently high degree of flexibility. Also, compared to other power curve modeling approaches, our method is simple in nature with a transparent model structure and parameters of definite physical meanings that are more feasible to adjust. Practically, the model is suitable for real-time operation because of the recursive estimation procedure, which limits the amount of computation required per new sample as the number of samples increases.

Future work should be carried out with two focuses. One is to take other available data and variables into consideration, not only the wind direction but the wind mast and neighboring wind farm records as well, to make the model feasible for more applications, like wind power monitoring. The other focus could be given to the modeling adaptivity in terms of setting more parameters in a data-driven way. Considering the computational cost, fast and simple data driven approaches are our target, like the optimal bandwidth selection method in this paper.

ACKNOWLEDGEMENT

The authors acknowledge ECMWF for providing the meteorological forecasts, as well as Gansu Electric Power Corporation for providing the wind power data. We also thank Huajie Ding for some innovative ideas. This work was partly funded by National Natural Science Foundation of China (51190101) and National Key Technologies R&D Program (2013BAA01B03). Besides, Pierre Pinson was partly supported by the Danish Council for Strategic Research through the project CITIES (DSF-1305-00027B) and '5 s'—Future Electricity Markets (no. 12-132636/DSF). The authors are grateful to the editor and reviewers, whose efforts allowed to enhance the paper.

REFERENCES

1. Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. *The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview* (2nd edn). ANEMOS.plus: Roskilde, 2011.
2. Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. *Renewable Energy* 2012; **37**: 1–8. DOI: 10.1016/j.renene.2011.05.033.
3. Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews* 2014; **32**: 255–270. DOI: 10.1016/j.rser.2014.01.033.
4. Yoder M, Hering AS, Navidi WC, Larson K. Short-term forecasting of categorical changes in wind power with Markov chain models. *Wind Energy* 2014; **17**: 1425–1439. DOI: 10.1002/we.1641.
5. Morales JM, Conejo AJ, Madsen H, Pinson P, Zugno M. *Integrating Renewables in Electricity Markets: Operational Problems*. Springer Science + Business Media: New York, 2014.
6. Matos MA, Bessa RJ. Setting the operating reserve using probabilistic wind power forecasts. *Power Systems, IEEE Transactions on* 2011; **26**: 594–603. DOI: 10.1109/TPWRS.2010.2065818.
7. Wang J, Botterud A, Bessa R *et al.* Wind power forecasting uncertainty and unit commitment. *Applied Energy* 2011; **88**: 4014–4023. DOI: 10.1016/j.apenergy.2011.04.011.
8. Tsikalakis AG, Hatziargyriou ND, Katsigiannis YA, Georgilakis PS. Impact of wind power forecasting error bias on the economic operation of autonomous power systems. *Wind Energy* 2009; **12**: 315–331. DOI: 10.1002/we.294.
9. Barthelmie RJ, Murray F, Pryor SC. The economic benefit of short-term forecasting for wind energy in the UK electricity market. *Energy Policy* 2008; **36**: 1687–1696. DOI: 10.1016/j.enpol.2008.01.027.
10. Usaola J, Ravelo O, Gonz A, Lez G, Soto F, Davila MTC, Diaz-Guerra BEN. Benefits for wind energy in electricity markets from using short term wind power prediction tools; a simulation study. *Wind Engineering* 2004; **28**: 119–127.
11. Paiva LT, Rodrigues CV, Palma JMLM. Determining wind turbine power curves based on operating conditions. *Wind Energy* 2013; **17**: 1563–1575. DOI: 10.1002/we.1651.
12. Wang Y, Infield DG, Stephen B, Galloway SJ. Copula-based model for wind turbine power curve outlier rejection. *Wind Energy* 2013; **17**: 1677–1688. DOI: 10.1002/we.1661.
13. Zheng L, Hu W, Min Y. Raw wind data preprocessing: a data-mining approach. *Sustainable Energy, IEEE Transactions on* 2015; **6**: 11–19. DOI: 10.1109/TSTE.2014.2355837.

14. Carrillo C, Obando Montaño AF, Cidrás J, Díaz-Dorado E. Review of power curve modelling for windturbines. *Renewable and Sustainable Energy Reviews* 2013; **21**: 572–581. DOI: 10.1016/j.rser.2013.01.012.
15. Lydia M, Kumar SS, Selvakumar AI, Prem Kumar GE. A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews* 2014; **30**: 452–460. DOI: 10.1016/j.rser.2013.10.030.
16. Marvuglia A, Messineo A. Monitoring of wind farms' power curves using machine learning techniques. *Applied Energy* 2012; **98**: 574–583. DOI: 10.1016/j.apenergy.2012.04.037.
17. Yampikulsakul N, Byon E, Huang S, Sheng S, You M. Condition monitoring of wind power system with nonparametric regression analysis. *IEEE Transactions on Energy Conversion* 2014; **29**: 288–299. DOI: 10.1109/TEC.2013.2295301.
18. Kusiak A, Zheng H, Song Z. On-line monitoring of power curves. *Renewable Energy* 2009; **34**: 1487–1493. DOI: 10.1016/j.renene.2008.10.022.
19. Nielsen TS, Nielsen HA, Madsen H. Prediction of wind power using time-varying coefficient functions. In *Proceedings of the XV IFAC World Congress*, 2002.
20. Pinson P, Nielsen HA, Madsen H. Robust estimation of time-varying coefficient functions—application to the modeling of wind power production. Technical University of Denmark, Tech. Rep, 2007.
21. Sánchez I. Short-term prediction of wind energy production. *International Journal of Forecasting* 2006; **22**: 43–56. DOI: 10.1016/j.ijforecast.2005.05.003.
22. Nielsen HA, Nielsen TS, Joensen AK, Madsen H, Holst J. Tracking time-varying coefficient-functions. *International Journal of Adaptive Control and Signal Processing* 2000; **14**: 813–828.
23. Fan J, Gijbels I. Variable bandwidth and local linear regression smoothers. *The Annals of Statistics* 1992; **20**: 2008–2036.
24. Huber PJ. *Robust Statistics*. Wiley: New York, 1981.
25. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons: New York, 2011.
26. Cutler NJ, Outhred HR, MacGill IF. Using nacelle-based wind speed observations to improve power curve modeling for wind power forecasting. *Wind Energy* 2012; **15**: 245–258. DOI: 10.1002/we.465.
27. Shokrzadeh S, Jafari Jozani M, Bibeau E. Wind turbine power curve modeling using advanced parametric and nonparametric methods. *IEEE Transactions on Sustainable Energy* 2014; **5**: 1262–1269. DOI: 10.1109/TSTE.2014.2345059.
28. Pinson P. Estimation of the uncertainty in wind power forecasting. *Centre Energétique et Procédés–Ecole des Mines de* 2006.
29. Fan J, Hu T-C, Truong YK. Robust non-parametric function estimation. *Scandinavian Journal of Statistics* 1994: 433–446. Online ISSN: 1467–9469.
30. Fan J, Gijbels I. *Local Polynomial Modelling and its Applications*. CRC Press, 1996.
31. Härdle W, Marron JS. Fast and simple scatterplot smoothing. *Computational Statistics & Data Analysis* 1995; **20**: 1–17. DOI: 10.1016/0167-9473(94)00031-D.
32. Fan J. Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 1992; **87**: 998–1004. DOI: 10.1080/01621459.1992.10476255.
33. Fisher RA. *Statistical Methods for Research Workers, the fifth edition*. Oliver and Boyd: Edinburgh, 1934.
34. Paleologu C, Benesty J, Member S, Ciochin S. Least-squares algorithm for system identification. *IEEE Signal Processing Letters* 2008; **15**: 597–600. DOI: 10.1109/LSP.2008.2001559.
35. Leung SH, So CF. Gradient-based variable forgetting factor RLS algorithm in time-varying environments. *IEEE Transactions on Signal Processing* 2005; **53**: 3141–3150. DOI: 10.1109/TSP.2005.851110.
36. Dowell J, Pinson P. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid* 2015; **7**: 763–770. DOI: 10.1109/TSG.2015.2424078.
37. Diebold FX, Mariano RS. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 1995; **13**: 253–263. DOI: 10.1080/07350015.1995.10524599.