Local Linear Regression with Adaptive Orthogonal Fitting for the Wind Power Application

Pierre Pinson
Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. (*pp@imm.dtu.dk*)

Henrik Aa. Nielsen
Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. (*han@imm.dtu.dk*)

Henrik Madsen
Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. (*hm@imm.dtu.dk*)

Torben S. Nielsen
ENFOR, Hørsholm, Denmark. (*tsn@enfor.dk*)

Abstract

Short-term forecasting of wind generation requires a model of the function for the conversion of meteorological variables (mainly wind speed) to power production. Such a power curve is nonlinear and bounded, in addition to being nonstationary. Local linear regression is an appealing nonparametric approach for power curve estimation, for which the model coefficients can be tracked with recursive Least Squares (LS) methods. This may lead to an inaccurate estimate of the true power curve, owing to the assumption that a noise component is present on the response variable axis only. Therefore, this assumption is relaxed here, by describing a local linear regression with orthogonal fit. Local linear coefficients are defined as those which minimize a weighted Total Least Squares (TLS) criterion. An adaptive estimation method is introduced in order to accommodate nonstationarity. This has the additional benefit of lowering the computational costs of updating local coefficients every time new observations become available. The estimation method is based on tracking the left-most eigenvector of the augmented covariance matrix. A robustification of the estimation method is also proposed. Simulations on semi-artificial datasets (for which the true power curve is available) underline the properties of the proposed regression and related estimation methods. An important result is the significantly higher ability of local polynomial regression with orthogonal fit to accurately approximate the target regression, even though it may hardly be visible when calculating error criteria against corrupted data.

Key words: wind power; modeling; forecasting; nonparametric regression; local linear regression; total least squares; adaptive estimation; robust estimation.

Corresponding author:
P. Pinson, Informatics and Mathematical Modelling, Technical University of Denmark,
Richard Petersens Plads (bg. 321 - 020), DK-2900 Kgs. Lyngby, Denmark.
Tel: +45 4525 3428, fax: +45 4588 2673, email: pp@imm.dtu.dk, webpage: www.imm.dtu.dk/~pp

1 Introduction

Wind power is a renewable energy that now represents a significant part of the electricity generation mix in a number of European countries. This phenomenon is not only limited to Europe, as the installation of important wind power capacities is also being discussed and planned in rapidly developing countries such as China, India or Brazil. However, large scale integration of wind generation causes difficulties in the management of a power system, owing to its variability and limited predictability. An additional challenge is to conciliate this integration with the on-going deregulation of the European electricity markets. Increasing the value of wind generation through improving the performance of prediction systems is one of the priorities in wind energy research for the coming years (Thor and Weis-Taylor 2002). A status report on wind power forecasting has been published in (Giebel *et al.* 2003).

Short-term predictions of wind power production at a given site, say up to 48-72 hours ahead, are commonly produced by using as input predictions of relevant meteorological variables, provided by a meteorological office. Relevant meteorological variables obviously include wind speed, but also wind direction, air density or temperature. Forecasts of meteorological variables are converted to power forecasts by passing them through the so-called wind farm power curve. From a statistical point of view, the challenge here is to estimate this conversion function, which is nonlinear and bounded, in addition to being nonstationary due to e.g. changes in the site environment, seasonality or climate change.

Local polynomial regression is an appealing nonparametric approach to modeling a wind farm power curve, for which the model coefficients can be adaptively estimated with recursive Least Squares (LS) methods, see e.g. (Nielsen *et al.* 2002, Sanchez 2006). An assumption when applying LS estimation methods is that a noise component is present in the response variable only, i.e. the power output in our case. However, it appears unlikely that the forecasts of meteorological variables used as input do not have an error component. Recently, Pinson *et al.* (2007) showed that the application of local polynomial regression with LS fit may yield somewhat inaccurate estimate of the true conversion function, even though it may still be appropriate for point prediction purposes where error measures are calculated against noisy response data (Jonsson 1994). Our aim here is to describe a method that provides better estimates of the true conversion function, by relaxing the assumption that the noise component is only on the response variable. This work is motivated by the current research efforts on wind power ensemble forecasting (Giebel 2005, and references therein). More particularly, Nielsen *et al.* (2006) concluded that a better understanding of the ability of various estimators to approximate the true regression function that defines a wind farm power curve is needed to improve ensemble predictions of wind generation.

Our approach has been developed with local linear regression as a basis, in order to obtain a nonparametric estimate of the nonlinear power curve. Howver, in contrast to classical LS estimation, the coefficients of the local linear models are then orthogonally fitted by using a Total Least Squares (TLS) criterion. Developments towards the fitting of a linear model with a TLS criterion can be traced back to the works by Golub and Van Loan (1980). This is used here in a nonparametric regression framework to fit local linear models. The aim of this approach is hence similar to that of principal curves, as introduced by Hastie and Stuetzle (1989)

and subsequently discussed by Tibshirani (1992), Verbeek *et al.* (2002), and Einbeck *et al.* (2005), among others: it is to locally minimize the distance between observations and the estimated nonparametric regression function. In order to accommodate nonstationarity, an adaptive estimation method is introduced which, for each local linear model, is based on tracking the left-most eigenvector of the augmented covariance matrix. Adaptivity in time is ensured by exponential forgetting of past observations. Since in practice noise distributions may be skewed and heavy-tailed in addition to not being Gaussian, a robust version of the estimation method is presented for estimates that would be less affected by such noise characteristics.

The paper is structured as follows. The proposed local linear regression with orthogonal fitting is defined in Section 2. The issue of adaptive estimation of the local model coefficients for this regression is then addressed in Section 3. In Section 4, the proposal for robustification of the previously introduced estimation method is described. The properties of the regression and estimation methods are demonstrated in Section 5. Simulations are based on semi-artificial data. They are primarily produced by passing real-world wind speed data through a modelled power curve in order to obtain noise-free power data. Both wind speed and power data are then corrupted to generate realistic datasets of wind speed forecasts and corresponding power measurements. The use of such datasets will allow us to demonstrate that the proposed method is better able to approximate the target regression when noise is present in both wind speed and power variables, since the target regression is indeed available. Concluding remarks in Section 6 end the paper.

2   Definition of the regression

Focus is given to the case for which wind speed is the unique explanatory variable, and thus for which the regression function to be estimated is that for the conversion of wind speed to power production. However, the methodology described hereafter can be generalized to the case of multiple explanatory variables, e.g. wind speed and direction. Let $\{y_i\}$, $i = 1, \ldots, n$, be the time-series of measured power production, and $\{u_i\}$, $i = 1, \ldots, n$, that of wind speed forecasts for the same points in time. Both $\{y_i\}$ and $\{u_i\}$ include a noise component originating from on-site measuring devices and prediction errors, respectively,

$$
\begin{align}
u_i &= u_i^* + \xi_i, & i = 1, \ldots, n \tag{1} \\
y_i &= y_i^* + \varepsilon_i, & i = 1, \ldots, n \tag{2}
\end{align}
$$

where $\{u_i^*\}$ and $\{y_i^*\}$ are seen as the time-series of the true underlying wind speed and power variables, while $\{\xi_i\}$ and $\{\varepsilon_i\}$ are sequences of mutually independent random variables, centred and with finite variance, also independent of u_i^*. It is assumed that a regression model relates the true underlying wind speed and power variables

$$
y_i^* = g(u_i^*), \qquad i = 1, \ldots, n \tag{3}
$$

where g is a nonlinear function to be estimated. Note that g may also be considered as nonstationary.

One recognizes in (1)-(3) a nonparametric regression problem with errors in variables. In contrast to the approaches described in e.g. (Fan and Truong 1993, Carrol *et al.* 1999), the following does not require any form of knowledge on the variance of measurement error in explanatory variables. It is chosen instead to follow the concept of self-consistency, initially introduced by Hastie and Stuetzle (1989) and subsequently generalized by Tarpey and Flury (1996) and Delicado (2001). Such a concept exhibits the statistical property of a smooth curve in the 'middle' of all points that orthogonally project onto it. The local scatter of data hence imposes the nonlinear dependency between noises in explanatory and response variables. Theoretical developments related to self-consistency and regression are available in (Hastie and Stuetzle 1989, Tarpey and Flury 1996). Local linear regression is used here to define g as a smooth curve in a nonparametric framework. It consists of estimating a vector of coefficients for local linear models to be fitted at a certain number of "fitting points" (Cleveland and Devlin 1988), i.e. for some specific wind speed values. Then, the local linear models are fitted with a weighted TLS criterion in order to minimize the Euclidian distance between data points and their orthogonal projections onto the regression curve.

2.1 Local linear models and estimates

Let us focus on a single fitting point $\tilde{u}$, defined on the range of wind speed values. A first-order Taylor expansion of g at $\tilde{u}$ is

$$g(\tilde{u} + \delta u) \ = \ g(\tilde{u}) + \nabla g(\tilde{u})\delta u + \circ(\delta u) \tag{4}$$

with δu being a small deviation from $\tilde{u}$. Assuming that g is sufficiently smooth, such Taylor expansion can serve for local approximation g with linear polynomials. Write

$$\boldsymbol{\phi} \ = \ \boldsymbol{\phi}(\tilde{u}) \ = \ [\phi_0(\tilde{u}) \ \ \phi_1(\tilde{u})]^\top \tag{5}$$

the vector of local model coefficients at $\tilde{u}$. $[.]^\top$ stands for the transposition operator. For obvious dependence of the local model coefficients on the chosen fitting point, and in order to alleviate notations, the dependence on $\tilde{u}$ is omitted in the following, unless absolutely necessary. For a given observation u_i close to $\tilde{u}$, write $[1 \ \ u_i - \tilde{u}]^\top$ the column vector corresponding to the first-order polynomial evaluated at $u_i - \tilde{u}$. Then, if discarding the remaining term $\circ(\delta \mathbf{u})$ in (4), it becomes

$$g(u_i) \ = \ [1 \ \ u_i - \tilde{u}]^\top \boldsymbol{\phi} \tag{6}$$

The coefficients ϕ_0 and ϕ_1 thus correspond to the value and first-order derivative of g at $\tilde{u}$, respectively.

For a dataset of size n, the local linear model of (6) is fitted at $\tilde{u}$ by expressing the model estimates $\hat{\boldsymbol{\phi}}$ as those which minimize

$$\hat{\boldsymbol{\phi}} \ = \ \underset{\boldsymbol{\phi}}{\arg\min} \, S\left(\boldsymbol{\phi}\right) \ = \ \underset{\boldsymbol{\phi}}{\arg\min} \sum_{i=1}^{n} w_i \rho\left(\epsilon_i\right) \tag{7}$$

where ρ is a criterion that defines the loss associated with a model residual ϵ_i , and w_i is a weight

that defines the importance to give to ϵ_i while considering the fitting point $\tilde{u}$. Finally, provided that a local linear model is fitted at a certain number of fitting points, the local polynomial regression for any wind speed value can be obtained by linear-type interpolation of the ϕ_0-coefficients estimated at every fitting point.

2.2 Orthogonal fitting of local linear models

Fitting local linear models with an LS criterion corresponds to the case for which in (7) ρ is a quadratic criterion, i.e. $\rho(\epsilon) = \epsilon^2$, and for which the model residual ϵ_i for time step i is calculated as $\epsilon_i = y_i - [1 \ u_i - \tilde{u}]^\top \phi$. This means that a residual is defined by considering a distance on the response variable axis only (cf. Fig. 1, dashed line). Similarly, w_i is given by a function of the distance between the wind speed observation u_i and the fitting point $\tilde{u}$ (Cleveland and Devlin 1988, Nielsen *et al.* 2000).

In contrast, the curve that passes through the 'middle' of the data points is defined as that which minimizes the Euclidean distance between the data points and the regression line l. Denote by $\tilde{p}$ the localization of the fitting point on l. Such regression line is parameterized by the vector of local coefficients $\phi = [\phi_0 \ \phi_1]^\top$ estimated at $\tilde{u}$. ϕ_0 corresponds to the y-value at $\tilde{u}$, while ϕ_1 relates to the slope of the regression line. For a given data point $p_i = (u_i, y_i)$, the Euclidian distance between p_i and l is given by that between p_i and its orthogonal projection $p_i^\perp$ on l. This corresponds to the dash-dotted lines in Fig. 1. The orthogonal projection $p_i^\perp$ exists and is unique, ensuring the existence and uniqueness of the proposed regression. Introduction to orthogonal fitting of linear models can be found in e.g. (de Groen 1996, Felus 2004).

—— *Here is Fig. 1* ——

Let $d_\perp(p_i, l)$ denote the Euclidean distance between p_i and its orthogonal projection $p_i^\perp$ on l. When aiming at orthogonally fitting the local linear models, the model residual $\epsilon_i^\perp$ related to the wind speed and power observations at time step i is readily given by

$$\epsilon_i^\perp = d_\perp(p_i, l), \qquad i = 1, \ldots, n \tag{8}$$

Then, the objective function to be minimized in order to obtain the estimates of the local coefficients satisfying the orthogonal fit follows the general form of (7),

$$S^\perp(\phi) = \sum_{i=1}^n w_i^\perp \rho\left(\epsilon_i^\perp\right) \tag{9}$$

where ρ is a quadratic criterion, but where the model residuals are defined by (8). Moreover, the definition of the weight $w_i^\perp$ to give to the model residual at time i should be different from that used for LS fitting. Indeed, it appears reasonable here to have $w_i^\perp$ as a function of the distance between $\tilde{p}$ and p_i, but along l. This distance, denoted by $d_l(\tilde{p}, p_i)$, is that between $\tilde{p}$ and $p_i^\perp$. The

weight $w_i^\perp$ can then be obtained e.g. with

$$w_i^\perp = w_i^\perp(l) = \eta\left(\frac{d_l(\tilde{p}, p_i)}{h}\right), \qquad i = 1, \ldots, n \tag{10}$$

where h is the chosen bandwidth for that fitting point, and η is a Kernel function like the tricube function, for instance,

$$\eta \,:\, x \in \mathbb{R}^+ \;\rightarrow\; \eta(x) \in [0, 1], \;\; \eta(x) = \begin{cases} (1 - x^3)^3, & x \in [0, 1] \\ 0 & , \;\; x > 1 \end{cases} \tag{11}$$

Finally, the regression curve is defined at each fitting point by the estimates of the local model coefficients that minimize (9), i.e.

$$\hat{\boldsymbol{\phi}}^\perp = \underset{\boldsymbol{\phi}}{\arg\min}\, S^\perp(\boldsymbol{\phi}) \tag{12}$$

Note that a usual linear model can easily be orthogonally fitted with the methods presented in e.g. (Golub and Van Loan 1980). However, the definition (10) of the local weights means that such weights are actually a function of the model estimates themselves. This yields a complex nonlinear optimization problem which, to our knowledge, has not been treated in the relevant literature.

3 Adaptive estimation of local model coefficients

For the wind power application the target regression g may be seen as nonstationary, though slowly varying. The relation between the wind at a wind farm and the resulting power production evolves with time, due to e.g. ageing of the turbines. In addition, for real-world applications in an online setting, when estimating the local model coefficients one does not want to consider the whole set of available observations every time new observations become available. This calls for the development of a recursive estimation method that permits tracking of the local model coefficients. Another advantage of such a method is to reduce the computational cost of estimation. Hereafter, it is considered that at time n a set of n past observations is available for each time-series, and thus that the dataset grows as time increases. The proposed estimation method follows from the fact that the coefficients of a linear model such as those in (12) can be expressed as a function of the smallest singular value and the related singular vector of the augmented data matrix. For proofs and more details regarding singular value decomposition, refer to (Golub and Van Loan 1980, 1996).

3.1 Formulation of the time-dependent objective function

Denote by $\hat{\phi}_n^\perp$ the estimate of the local model coefficients at time n. The objective function to be minimized for estimating $\hat{\phi}_n^\perp$ is a modified version of that given in (9),

$$\hat{\phi}_n^\perp = \arg\min_{\phi_n} S_n^\perp(\phi_n) = \arg\min_{\phi_n} \sum_{i=1}^n \beta_n(i) w_i^\perp \rho\left(\epsilon_i^\perp\right) \tag{13}$$

where $\epsilon_i^\perp$ and $w_i^\perp$ are the model residual and related weight at time step i, as introduced in (8) and (10). In the following, l_n denotes the regression line defined by ϕ_n. In the above, β_n is a function that permits the exponential forgetting of past observations. This is defined as

$$\beta_n(i) = \begin{cases} \lambda_i^{\mathrm{eff}} \beta_{n-1}(i-1), & 1 \leq i \leq n-1 \\ 1 & , \quad i = n \end{cases} \tag{14}$$

where λ_i^{eff} is the effective forgetting factor, as originally introduced by Nielsen *et al.* (2000). λ_i^{eff} is such that exponential forgetting is applied to an observation if and only if that observation lies in the vicinity of the fitting point $\tilde{u}$. The vicinity is in turn controlled by the bandwidth parameter h. The effective forgetting factor is hereby defined as

$$\lambda_i^{\mathrm{eff}} = 1 - (1-\lambda) w_i^\perp \tag{15}$$

where λ is the user-defined forgetting factor, $0 < \lambda \leq 1$. Note that in the stationary case, one does not want to downweight past observations. Then, λ is set to $\lambda = 1$, resulting in $\beta_n(i) = 1, \ \forall n, i$.

3.2 Updating of the estimates at a given time

An important assumption for using this recursive procedure is that the estimate $\hat{\phi}_{n-1}^\perp$ of the local coefficients at time $n-1$ is the optimal estimate, i.e. that which minimizes the objective function formulated in (13). Also, it is considered that the true regression model g is slowly varying, and thus that the estimates $\hat{\phi}_{n-1}^\perp$ and $\hat{\phi}_n^\perp$ are very similar. Therefore, denoting l_n and l_{n-1} by the estimated regression lines at times n and $n-1$, and by denoting $\tilde{p}_n$ and $\tilde{p}_{n-1}$ as the location of the fitting point on these lines, we have

$$d_{l_n}(\tilde{p}_n, p_i) \simeq d_{l_{n-1}}(\tilde{p}_{n-1}, p_i) \tag{16}$$

at least for the last few time steps i, i.e. those related to model residuals only slightly downweighted by exponential forgetting. Consequently, this translates to

$$w_i^\perp(l_n) \simeq w_i^\perp(l_{n-1}) \tag{17}$$

again for at least a number of i-values, i being inferior though when it approaches n. Such an approximation allows us to use at time n the last regression line for assigning a weight (as defined in (10)) to the new observations to be considered.

Denote by $\mathbf{A}_n \in \mathbb{R}^{n \times 2}$ the data matrix at time n, i.e. the matrix whose i^{th} row contains the vector of observations $[1 \ u_i - \tilde{u}]$ at time i, $i = 1, \ldots, n$. Consequently, the so-called augmented data matrix $\mathbf{A}_n^+ \in \mathbb{R}^{n \times 3}$ is defined as

$$\mathbf{A}_n^+ \;=\; \mathbf{W}_n^{\frac{1}{2}}[\mathbf{A}_n \ y_n] \tag{18}$$

that is, the data matrix to which is added a right column corresponding to the vector of power observations, subsequently multiplied by the square root of the weight matrix $\mathbf{W}_n \in \mathbb{R}^{n \times n}$. $\mathbf{W}_n$ is a diagonal matrix whose i^{th} element on the diagonal is the weight $w_i^\perp$ given to the ith observation.

The recursive estimation method follows from the idea that the SVD of the augmented data matrix $\mathbf{A}_n^+$ is related to the eigenvalue decomposition (EVD) of the augmented inverse covariance matrix $\mathbf{R}_n^+ = {\mathbf{A}_n^+}^\top \mathbf{A}_n^+$. Indeed, if the singular value decomposition of $\mathbf{A}_n^+$ is

$$\mathbf{U}_n^\top \mathbf{A}_n^+ \mathbf{V}_n \;=\; \mathbf{D}_n \tag{19}$$

where $\mathbf{U}_n$ and $\mathbf{V}_n$ are orthogonal matrices of appropriate dimensions, then we readily have

$$\mathbf{V}_n^\top \mathbf{R}_n^+ \mathbf{V}_n \;=\; \mathbf{D}_n^2 \tag{20}$$

This means that the smallest singular value ϱ_n of $\mathbf{A}_n^+$ is equal to the square root of the smallest eigenvalue of $\mathbf{R}_n^+$. And, the right-most singular vector $\mathbf{v}_n$ of $\mathbf{A}_n^+$ corresponds to the right-most eigenvector of $\mathbf{R}_n^+$. Consequently, recursively estimating $\phi_n^\perp$ can be done by tracking the smallest eigenvalue (and related eigenvector) of $\mathbf{R}_n^+$. Several methods are available in the literature for recursive updating of $\mathbf{R}_n^+$ and recursive estimation of its eigenvalues, see for instance (Ljung *et al.* 1978, Yu 1991).

Here, the method employed is based on the updating of the augmented covariance matrix $\mathbf{P}_n^+$. Since the largest eigenvalue ν_n of $\mathbf{P}_n^+$ is equal to the inverse of the smallest eigenvalue of $\mathbf{R}_n^+$, it is clear that

$$\varrho_n \;=\; \nu_n^{-\frac{1}{2}} \tag{21}$$

and consequently, tracking the right-most eigenvector of $\mathbf{R}_n^+$ is equivalent to tracking the left-most eigenvector of $\mathbf{P}_n^+$.

In a first stage, it is necessary to update the augmented covariance matrix $\mathbf{P}_{n-1}^+$ as new observations become available. Write $\mathbf{z}_n$ the augmented vector of observations at time n, centered on $\tilde{u}$,

$$\mathbf{z}_n \;=\; [1 \ u_n - \tilde{u} \ y_n]^\top \tag{22}$$

In parallel, the weight $w_n^\perp(l_n)$ to assign to $\mathbf{z}_n$ is approximated with $w_n^\perp(l_{n-1})$, following (17). The effective forgetting factor λ_n^{eff} is consequently computed using (15).

An update of the augmented covariance matrix $\mathbf{P}_n^+$ at time n can be obtained by using the matrix inversion lemma (see e.g. (Madsen 2006, p. 245))

$$\mathbf{P}_n^+ \;=\; \frac{1}{\lambda_n^{\text{eff}}} \left[\mathbf{P}_{n-1}^+ - \gamma_n \frac{\mathbf{P}_{n-1}^+ \mathbf{z}_n \mathbf{z}_n^\top \mathbf{P}_{n-1}^+}{1 + \gamma_n \mathbf{z}_n^\top \mathbf{P}_{n-1}^+ \mathbf{z}_n} \right] \tag{23}$$

with γ_n defined by

$$\gamma_n = \frac{w_n^{\perp}}{\lambda_n^{\text{eff}}} \tag{24}$$

In a second stage, the power method is used for estimating the largest eigenvalue and the associated eigenvector of $\mathbf{P}_n^+$, as described by Golub and Van Loan (1996, pp. 406-408). Let $\mathbf{v}_n^{(0)}$ be a unit vector, $\mathbf{v}_n^{(0)} \in \mathbb{R}^3$. By iteratively computing $\mathbf{v}_n^{(k)}$, $k > 0$, such that

$$\mathbf{v}_n^{(k)} = \frac{\mathbf{P}_n^+ \mathbf{v}_n^{(k-1)}}{||\mathbf{P}_n^+ \mathbf{v}_n^{(k-1)}||_2} \tag{25}$$

the sequence of vectors $\mathbf{v}_n^{(k)}$ converges to the left-most eigenvector of $\mathbf{P}_n^+$, provided that the related maximum eigenvalue is unique. Denote by $\hat{\mathbf{v}}_n$ and $\hat{\nu}_n$ the estimated largest eigenvector and eigenvalue of $\mathbf{P}_n^+$, respectively. Since it is assumed that the process considered is slowly varying, $\mathbf{v}_n^{(0)} = \hat{\mathbf{v}}_{n-1}$ can serve as an initialization to the power method at time n. Only few steps of the power method should be sufficient to compute $\hat{\mathbf{v}}_n$, due to an expected high similarity between $\mathbf{P}_{n-1}^+$ and $\mathbf{P}_n^+$. In addition, owing to symmetry of $\mathbf{P}_n^+$, it is possible to calculate an estimate of the error bound (Golub and Van Loan 1996, pp. 406-408), i.e. an upper bound on the difference between the largest eigenvalue ν_n of $\mathbf{P}_n^+$ and the kth iterated eigenvalue estimate $\nu_n^{(k)}$. $\nu_n^{(k)}$ is given by

$$\nu_n^{(k)} = \mathbf{v}_n^{(k)\top} \mathbf{P}_n^+ \mathbf{v}_n^{(k)} \tag{26}$$

and the error bound calculation yields

$$|\nu_n - \nu_n^{(k)}| \leq \sqrt{2} \, ||\mathbf{P}_n^+ \mathbf{v}_n^{(k)} - \nu_n^{(k)} \mathbf{v}_n^{(k)}||_2 \tag{27}$$

Therefore, one can iterate on $\mathbf{v}_n^{(k)}$ and $\nu_n^{(k)}$ until a defined error bound is reached.

Finally, since the obtained eigenvector $\mathbf{v}_n$ corresponds the right-most singular vector of the augmented data matrix $\mathbf{A}_n^+$, and following (Golub and Van Loan 1980), the estimate of the local coefficients at time n can be obtained with

$$\hat{\boldsymbol{\phi}}_n^{\perp} = -\frac{1}{\hat{v}_{n,3}} \, [\hat{v}_{n,1} \ \ \hat{v}_{n,2}]^{\top} \tag{28}$$

Note that the updating procedure based on the matrix inversion lemma might prove to be unstable (Liavas and Regalia 1999). In such a case, one may choose to work with $\mathbf{R}_n^+$ instead of $\mathbf{P}_n^+$. When doing so, it will still be necessary to invert $\mathbf{R}_n^+$ at each time step in order to apply the power method to $\mathbf{P}_n^+$.

3.3 Initialization

The initialization part consists of choosing appropriate initial augmented covariance matrix $\mathbf{P}_0^+$ and initial estimate $\hat{\boldsymbol{\phi}}_0^\perp$ of the local model coefficients. A common choice is to take

$$\mathbf{P}_0 = \delta^{-1}\mathbf{I} \tag{29}$$

as an initial inverse augmented covariance matrix, with $\mathbf{I}$ an identity matrix of appropriate size, and δ a small number that is user-defined. δ is usually set at a very small value, reflecting the fact that no information on the covariance of data is available. As a consequence of this lack of information, one also considers an initialization period during which the estimates are not updated (Nielsen *et al.* 2000). This initialization period continues until a point is reached where there are a certain number of occurrences (say 10) of the weight $w_n^\perp$ being larger than a chosen threshold value, e.g. 0.5.

Regarding local model estimates, one can define $\hat{\boldsymbol{\phi}}_0$ as a vector of zeros or as a best guess on the target regression. A best guess can be obtained from the theoretical power curve of the wind farm considered. $\mathbf{v}_0^{(0)}$ is then set to

$$\mathbf{v}_0^{(0)} = \left[\left(\hat{\boldsymbol{\phi}}_0^\perp\right)^\top \quad -1 \right]^\top \tag{30}$$

i.e. so that its last element $v_{0,3}$ is set to $v_{0,3} = -1$, in order to have $\hat{\boldsymbol{\phi}}_n^\perp = \hat{\boldsymbol{\phi}}_{n-1}^\perp$ if $\hat{\mathbf{v}}_n = \hat{\mathbf{v}}_{n-1}$.

4 Robustification of the estimation method

As LS estimates are only optimal (for estimating the target regression) if the noise on the response variable is Gaussian, the orthogonal estimates obtained from the method described above may not be optimal when noise sequences in (1)- (2) deviate from Normal. When considering the modeling of the conversion of wind to power in real-world application, noise distributions may actually be skewed and heavy-tailed in addition to not being Gaussian (Lange and Focken 2006, Pinson 2006). Therefore, in order for the orthogonal estimates to be less affected by such noise characteristics, a proposal for the robustification of the method described in the above Section is given here. The approach is inspired by the M-type estimation methods employed for robustification of LS estimators, initially introduced by Huber (1981) for linear models, and subsequently used in nonparametric regression with LS fit by e.g. (Fan *et al.* 1994, Welsh 1994). It has been shown that this type of robustification approach is suitable for a large range of contaminated distributions (Kelly 1992). The M-type estimator described here directly follows from that developed by Pinson *et al.* (2007) for local polynomial regression with time-varying coefficients. First the use of bounded-influence criteria for robust estimation is presented, followed by the description of resulting changes in the recursive estimation procedure.

4.1 Using a bounded-influence criterion

Robustification is based on downscaling the influence of large residuals which may be considered as suspicious. The quadratic criterion ρ in (9) is replaced by a bounded-influence criterion ρ_m. Bounded influence means that the derivative ψ_m of ρ_m is bounded

$$\psi_m \ : \ \epsilon \in \mathbb{R} \ \rightarrow \ \psi_m(\epsilon) \in [m_{\text{inf}}, m_{\text{sup}}], \quad m_{\text{inf}} < 0, \ m_{\text{sup}} > 0 \tag{31}$$

Also, it is considered that ρ_m is convex and consequently, if denoting by ψ'_m the derivative of ψ_m,

$$\psi'_m \ : \ \epsilon \in \mathbb{R} \ \rightarrow \ \psi'_m(\epsilon) \in [0, M_{\text{sup}}] \tag{32}$$

for almost all ϵ, since ψ'_m cannot be defined for some points if it is a piecewise function. M_{sup} denotes the upper bound on $\psi'_m(\epsilon)$ values.

In order to stay in line with the concept of self-consistency (Tarpey and Flury 1996, Delicado 2001), the objective is still to have local model estimates that minimize a quadratic criterion. ρ_m is selected as the Huber criterion, which combines a quadratic loss in its central part, and a linear one for large $|\epsilon|$-values,

$$\rho_m(\epsilon, c) \ = \ \begin{cases} \frac{\epsilon^2}{2}, & |\epsilon| \leq c \\ c|\epsilon| - \frac{c^2}{2}, & |\epsilon| > c \end{cases} \tag{33}$$

with the c-parameter, referred to as the threshold point, controlling the transition from quadratic to linear. Consequently, the related ψ_m-function is an odd function given by

$$\psi_m(\epsilon, c) \ = \ \rho'_m(\epsilon) \ = \ \begin{cases} \epsilon, & |\epsilon| \leq c \\ c \, \text{sign}(\epsilon), & |\epsilon| > c \end{cases} \tag{34}$$

and its derivative ψ'_m is

$$\psi'_m(\epsilon, c) \ = \ \rho''_m(\epsilon) \ = \ \begin{cases} 1, & |\epsilon| \leq c \\ 0, & |\epsilon| > c \end{cases} \tag{35}$$

The Huber loss function is symmetric and such that $m_{\text{sup}} = -m_{\text{inf}} = c$. The upper bound on the derivative of ψ_m is $M_{\text{sup}} = 1$.

4.2 Changes in the recursive estimation procedure

For the case of local linear regression, defining the M-type estimator related to $\hat{\phi}_n^{\perp}$ necessitates replacing ρ with ρ_m in (9). In addition, the weight $w_i^{\perp}$ associated to the ith observation has to be accounted for. This is because model residuals with a low weight should not be additionally downweighted by the bounded-influence criterion. Consequently, by making the threshold c of the Huber criterion a function of $\left(w_i^{\perp}\right)^{-1/2}$, the objective function can be reformulated so that it instead accounts for the loss related to weighted residuals (Pinson *et al.* 2007). The robust

11

estimates $\hat{\phi}_n^*$ at time n are given by

$$\hat{\phi}_n^{\perp*} = \arg\min_{\phi} S_n^{\perp*}(\phi) = \arg\min_{\phi} \sum_{i=1}^{n} \beta_n^*(i) \rho_m \left(\epsilon_i^{\perp} \sqrt{w_i^{\perp}}, c \right) \tag{36}$$

One notices that in comparison to (13), the weight $w_i^{\perp}$ is no longer used to scale the loss of the ith model residual, but instead to scale the model residual itself before calculating the associated loss.

In addition, β_n^* is still a function that permits an exponential forgetting of past observations in the vicinity of the considered fitting points. However, it is modified here so that only non-suspicious observations are dealt with. It is defined similarly to β_n in (14), except that $\lambda_i^{\text{eff}*}$ the robust effective forgetting factor at time i is instead given by

$$\lambda_i^{\text{eff}*} = 1 - (1 - \lambda) w_i^{\perp} \psi_m' \left(\epsilon_i^{\perp} \sqrt{w_i^{\perp}}, c \right) \tag{37}$$

The robustification of the recursive estimation procedure necessarily implies a change in the updating of the augmented covariance matrix. The updating formula for $\mathbf{P}_{n-1}^{+*}$ is obtained in a similar fashion than that formulated in (23), i.e.

$$\mathbf{P}_n^{+*} = \frac{1}{\lambda_n^{\text{eff}*}} \left[\mathbf{P}_{n-1}^{+*} - \gamma_n \psi_m' \left(\epsilon_n^{\perp} \sqrt{w_n^{\perp}}, c \right) \frac{\mathbf{P}_{n-1}^{+*} \mathbf{z}_n \mathbf{z}_n^{\top} \mathbf{P}_{n-1}^{+*}}{1 + \gamma_n \mathbf{z}_n^{\top} \mathbf{P}_{n-1}^{+*} \mathbf{z}_n} \right] \tag{38}$$

with γ_n given by (24).

The above updating formula means that $\mathbf{P}_{n-1}^{+*}$ is not updated, i.e. $\mathbf{P}_n^{+*} = \mathbf{P}_{n-1}^{+*}$, if the new observations are considered as suspicious. The power method, as well as (28), are finally applied in the same manner in order to obtain the local model coefficients $\hat{\phi}_n^{\perp*}$ at time n.

5 Simulations

Simulation results on semi-artificial datasets are used to highlight the properties of the proposed local linear regression with adaptive orthogonal fitting to model the conversion of wind speed to power. The interest of such semi-artificial datasets is that the true power curve is available. This will allow us to demonstrate the greater ability of the proposed regression with orthogonal fitting (in comparison to classical LS fitting) to approximate the true regression function when noise is present in both wind speed and power variables.

5.1 Semi-artificial data

Semi-artificial means that wind speed measurements from a meteorological mast at the wind farm are used as a basis, but related power values are obtained by transformation through a

modelled power curve. The wind farm considered is located in North Jutland, Denmark. Wind speed measurements consist of hourly averages. It is assumed that such measurements are noise-free. The time-series of wind speed and related power production cover a period of $N = 10000$ hours. They are normalized so that they take values in the unit interval. The model for the nonstationary power curve is introduced in a first stage. The noise that is added to obtain simulated but realistic datasets of wind speed and related power production is then described.

5.1.1 Model for the true power curve

The power curve giving the relation between u_i^* and y_i^*, the true wind speed and power values, is modelled as a nonstationary double exponential function. At time step i $(i = 1, \ldots, N)$, g_i is defined as

$$g_i(u^*) = \exp\left(-\tau_{i,2}\exp\left(-\tau_{i,1}u^*\right)\right) \tag{39}$$

so that the shape of the power curve is controlled by the parameters $\boldsymbol{\tau}_i = [\tau_{i,1} \ \ \tau_{i,2}]^\top$. These parameters are chosen to evolve linearly in time from $\boldsymbol{\tau}_1 = [10 \ 40]^\top$ to $\boldsymbol{\tau}_N = [11 \ 40]^\top$. The resulting nonstationary power curve is illustrated in Fig. 2, by giving its initial and final states. The scatterplot of wind speed and power data over the whole simulation period is depicted in Fig. 3(a). Owing to the distribution of wind speed values, the nonstationary power curve actually has a significantly larger number of data points in its low power part than in its high power part. Note that in real-world test cases, the power curve function may also be a function of other explanatory variables e.g. wind direction.

—— *Here is Fig. 2* ——

5.1.2 Noise sequences and resulting simulated data

In order to obtain simulated, but realistic, wind speed and power data for the wind farm, two different types of noise sequence are envisaged. These noise sequences $\{\omega_i^x\}$ and $\{\zeta_i^x\}$, whose characteristics are a function of a given variable x (i.e. either wind speed or power), are such that:

- $\{\omega_i^x\}$ is an additive Gaussian noise with zero mean and whose standard deviation σ_ω is a function of the level of x, i.e.

$$\omega_i^x \sim \mathcal{N}(0, \sigma_\omega(x)^2), \quad \sigma_\omega(x) = a_\omega + 4x(1-x)b_\omega \tag{40}$$

  Such additive noise simulates a permanent noise in the power measurements. It is assumed that the variance of this noise is directly influenced by the slope of the power curve. This is why a logistic type of function is chosen. On the wind speed axis, this additive noise stands for the amplitude error present in wind speed forecasts. Note that the use of such a logistic type of function is not fully realistic in this case, but it has the interest of increasing the difficulty of estimating the target regression;

- $\{\zeta_i^x\}$ is an impulsive noise of the same form as $\{\omega_i^x\}$, i.e.

$$\zeta_i^x \sim \mathcal{N}(0, \sigma_\zeta(x)^2), \quad \sigma_\zeta(x) = a_\zeta + 4x(1-x)b_\zeta \tag{41}$$

except that this noise is added at random locations characterized by a binary sequence $\{\mathcal{I}_i\}$. The proportion of data corrupted by this impulsive noise is given by π. If considering power measurements, such a noise simulates the presence of gross errors (or even outliers) in the data. They may originate from electronic transmission problems for instance. For the case of wind speed, this type of noise simulates the presence of phase errors in meteorological forecasts, which are less frequent than amplitude errors but of significantly larger magnitude.

Subsequently, the time-series $\{y_i\}$ of simulated power data is obtained by adding these two noises to the noise-free power data $\{y_i^*\}$,

$$y_i = y_i^* + \omega_i^{y^*} + \zeta_i^{y^*}\mathcal{I}_i^{y^*}, \quad i = 1, \ldots, N \tag{42}$$

The time-series $\{u_i\}$ of simulated wind speed data is obtained in a similar way, i.e. with

$$u_i = u_i^* + \omega_i^{u^*} + \zeta_i^{u^*}\mathcal{I}_i^{u^*}, \quad i = 1, \ldots, N \tag{43}$$

For both time-series, data points out of the unit interval are forced to its bounds. The noises in the resulting dataset obviously deviate from being Gaussian.

For the simulations in the present Section, the parameters controlling noise sequences are chosen such that

- wind power: $(a_\omega, b_\omega) = (0.004, 0.9)$ for the additive noise, and $(\pi, a_\zeta, b_\zeta) = (0.2, 0.012, 0.2)$ for the impulsive noise.

- wind speed: $(a_\omega, b_\omega) = (0.005, 0.04)$ for the additive noise, and $(\pi, a_\zeta, b_\zeta) = (0.2, 0.01, 0.15)$ for the impulsive noise.

The resulting dataset is depicted in Fig. 3: Fig. 3(a) shows the noise-free data, while the corrupted ones can be seen from Fig. 3(b). This scatterplot of semi-artificial wind speed and related power data resembles those one would see from real-world data, such as those studied in e.g. (Pinson 2006).

—— *Here is Fig. 3* ——

5.2   Selection of relevant parameters

Remember that $\hat{\phi}_n^\perp$ denotes the local model coefficients obtained with adaptive orthogonal fitting at time n, while $\hat{\phi}_n^{\perp*}$ refers to those obtained with the robust version of the method. By

14

extension, let us use these notations for the related nonparametric and time-varying estimates of the power curve at that same point in time. Both methods are benchmarked against linear regression with adaptive LS fit, as initially described by Nielsen *et al.* (2000) and subsequently used for the wind power application in (Nielsen *et al.* 2002). The power curve estimated with LS fit at time n is denoted by $\hat{\phi}_n$.

For both types of adaptive fitting of local linear models, one has to select or optimize a set of meta-parameters consisting of *(i)* the fitting points on the wind speed range which locate where the local linear models are to be defined and fitted, and *(ii)* the bandwidth value at each fitting point which controls the proportion of local data used for model fitting. The methodology employed for the selection of these meta-parameters is described in the following.

In a first stage, the fitting points $\tilde{u}_{(j)}$ are chosen to be uniformly spread on the unit interval,

$$\tilde{u}_{(j)} \;=\; \frac{j-1}{J-1}, \qquad j = 1, \ldots, J \tag{44}$$

so that only J, the number of these fitting points, has to be selected. It is arbitrarily set to $J = 20$ here, since it has been noticed that when set to a sufficiently large value J did not have qualitative effects on the properties of the various estimators. In a second stage, the bandwidth values $h_{(j)}$ are parameterized from our knowledge of the distribution of wind speed values. Such distribution is right-skewed, with the density of the data being inversely proportional to the level of u. Consequently, $h_{(j)}$ is defined as

$$h_{(j)} \;=\; h_0 + h_1(j-1), \qquad j = 1, \ldots, J \tag{45}$$

with $h_0, h_1 > 0$, so that local model coefficients for high wind speed levels can be estimated from more observations, even if these observations are located further from the fitting point. Another advantage of this parametrization is that instead of determining all the J bandwidth values, only the constant h_0 and scale factor h_1 have to be selected. In general one should have $1/J \ll h_{(j)}$ in order to obtain smooth model estimates.

Then, for each type of estimator, the meta-parameters are determined with one-fold cross-validation: the first 2000 time-steps of both time-series are considered as a buffer training period (for initialization of local model parameters), while the following 2000 time-steps form the cross-validation period. The optimal set of meta-parameters is chosen as that which minimizes a Mean Square Error (MSE) criterion for 1-step ahead forecasts, calculated against the corrupted data (as it would be the case for real-world applications), over the latter period. In practice, it is obtained by trial and error.

Table 1 gathers the optimal sets of meta-parameters for the various estimators determined from the cross-validation procedure. The bandwidth values are significantly larger when employing adaptive orthogonal fitting of the local linear models. This follows from the definition of the weights to be assigned to observations (cf. (10)). Since these are a function of the distance between the fitting point considered and observations, but along the regression curve, distances will generally be of larger magnitude. Hence, in order to have a sufficient number of effective observations for estimating the local coefficients, the bandwidth also needs to have a larger value.

—— *Here is Table 1* ——

In parallel to the selection of the meta-parameters and for the specific case of $\hat{\phi}_n^\perp$ and $\hat{\phi}_n^{\perp*}$, the error bound that controls the number of iterations of the power method every time the augmented covariance matrix is updated is set to 10^{-5}. Finally, the effect of the choice of the threshold point c, which is the last parameter needed for fully defining $\hat{\phi}_n^{\perp*}$, will be studied in the following Paragraph.

5.3   Evaluation of estimated power curves

The remainder of the datasets, consisting of the last 6000 time-steps and referred to as the evaluation set, is used for a 1-step ahead forecasting exercise with the aim of assessing the ability of the various estimators to approximate the true power curve. In addition, since both noise-free and corrupted data are available, this will allow us to comment on the fact that certain error measures, when calculated against corrupted data, may not illuminate the greater ability to approximate the true regression. The 1-step ahead forecasts are evaluated with both a Normalized Mean Absolute Error (NMAE) and a Normalized Root Mean Square Error (NRMSE) criterion. Both error criteria are referred to as normalized since power values are contained in the unit interval. They are thus expressed as a percentage of the nominal power of the wind farm. Even if our aim is clearly to obtain a minimum MSE estimator, the NMAE criterion may provide better information about improving power curve estimation since it would give less weight to large errors related to suspicious data. For a thorough discussion on the choice of error measures for evaluating wind power forecasts, we refer to (Madsen *et al.* 2005). Error criteria are calculated against both noise-free and corrupted data. They will be referred to as $NMAE_t$ (and $NRMSE_t$) for the former case, and as $NMAE_r$ (and $NRMSE_r$) for the latter one.

In addition to the set of meta-parameters given in Table 1 for the case of local linear regression with adaptive orthogonal fitting, one has to choose the threshold parameter c for fully defining $\hat{\phi}_n^{\perp*}$. Here, c is not determined with the same type of cross-validation procedure than that used above (which would be the usual approach for a real-world application), it is instead decided to show the influence of c on the ability of this estimator to approximate the true regression model. This ability is quantified with the $NRMSE_t$ criterion, whose value as a function of c is depicted in Fig. 4. When c equals 1, there is no observation whose influence is downweighted, $\hat{\phi}_n^\perp$ and $\hat{\phi}_n^{\perp*}$ are then equivalent, resulting in the same level of forecast performance. The $NRMSE_t$ criterion decreases as c gets smaller, until a minimum is reached. When further decreasing the value of the threshold point, the $NRMSE_t$ criterion increases again: the fact that there are too many observations whose influence is downweighted negatively impacts the forecast performance of $\hat{\phi}_n^{\perp*}$. A minimum $NRMSE_t$ is reached for $c = 0.11$.

—— *Here is Fig. 4* ——

Table 2 summarizes the 1-step ahead forecast performance when using the three different power curves obtained with the three estimators by gathering the values of all error criteria mentioned

16

above over the evaluation set. The decrease in $NMAE_r$ when going from LS to orthogonal fitting is non-negligible. However, the $NRMSE_r$ criterion is of larger magnitude for $\hat{\phi}_n^\perp$ than for $\hat{\phi}_n$. Meanwhile, it appears that when error criteria are evaluated against the noise-free data (i.e. $NMAE_t$ and $NRMSE_t$), they both have dramatically lower values after orthogonally fitting local linear models. This shows that $\hat{\phi}_n^\perp$ has a higher ability to approximate the underlying regression model. Hoever, this also tells that such higher ability is hard to see when evaluating forecast performance against real-world measurements. Such results nicely fit with the theoretical results in (Jonsson 1994), which indicates that an LS estimator may exhibit higher out-of-sample forecast performance when calculated against noisy data, even though it does not approximate well to the underlying regression model. Here, the $NMAE_r$ measure is actually more informative than the $NRMSE_r$ one, since it makes it possible to see whether the evaluated estimator is more central. This point is also in line with the discussion in (Madsen *et al.* 2005) on error measures in wind power forecasting.

In addition, Table 2 reveals the benefits of the robustification of the estimation procedure used to fit orthogonally to the local linear models. All error measures actually have lower values when focusing on the robust version of the estimator. Primarily, it is clear that the decrease in error criteria calculated against noise-free data is of practical magnitude: the decrease in $NMAE_t$ is of 9.71% while that in $NRMSE_t$ reaches 9.97%. Note that benefits of orthogonal fitting (instead of LS fit) of local linear models are much larger than those from additional robustification.

—— *Here is Table 2* ——

In a last stage, the final power curves obtained with $\hat{\phi}_n$ and $\hat{\phi}_n^{\perp*}$ are compared. They are both depicted in Fig. 5, along with the true power curve at the end of the evaluation set. The curve corresponding to $\hat{\phi}_n^{\perp*}$ provides a closer approximation of the final true power curve. This is mainly true in the zone where the density of data is lower, i.e. for high power values. Using a LS criterion for fitting local linear models yields a tendency to flatten the shape of the estimated power curve. This is not the case when orthogonally fitting these models.

—— *Here is Fig. 5* ——

6   Conclusions

Local linear regression is an appealing approach to modeling the conversion of wind to power, for which the local models are commonly fitted with an LS criterion. Doing so is actually equivalent to making the assumption that a noise component is present in the response variable only. This assumption is not realistic for the wind power forecasting application, when the wind-to-power conversion function is estimated with meteorological forecasts as explanatory variables. Therefore, even if a power curve estimated with LS fit may be considered optimal in a minimum-MSE sense (with the MSE calculated against the noisy data), it does not provide an accurate estimate of the true underlying conversion function. This drawback of the LS-fitting

approach has been clearly illustrated in the present paper.

In order to relax the assumption of the noise component being on the response variable only, the definition of local linear regression with orthogonal fitting has been introduced. This follows the concept of self-consistency, and has an aim similar to that of principal curves, though it is restricted to a nonparametric regression framework: it is to locally minimize the Euclidean distance between the observations and the estimated regression curve. In practice, the estimates of the local model coefficients are defined as those that minimize a TLS criterion. In order to accommodate long-term variations of the local model coefficients, an adaptive estimation method based on tracking of the left-most eigenvector of the augmented covariance matrix $\mathbf{P}_n^+$ has been described. The interest of such a method is that it eliminates the need to compute a singular value decomposition every time a new observation is available, which significantly decreases computational costs. In addition, only the augmented covariance matrix $\mathbf{P}_n^+$ needs to be updated and stored at each time-step. Note that it will be possible in the future to consider other recursive schemes to update $\mathbf{P}_n^+$. More particularly, one may control the information content used to update $\mathbf{P}_n^+$ in the direction of its largest eigenvalue, e.g. with selective forgetting methods (Parkum *et al.* 1992).

The ability of local linear regression with adaptive orthogonal fitting to better estimate the true conversion function of wind to power has been shown from simulations on a semi-artificial dataset, i.e. for which both the true regression model and realistic simulated data were available. When calculating error criteria against the true regression model, the improvement achieved when applying the proposed regression and estimation methods is indeed significant. However, it has been clearly explained that such improvement is difficult to appraise when evaluating estimated regressions against corrupted data. This ability of better approximation of the underlying conversion process is a desirable feature for application to ensemble prediction of wind power. Future works will concentrate on evaluating the benefits of using the proposed regression method for this purpose.

Acknowledgments

References

Carrol R.J., Maca J.D and Ruppert D. 1999. Nonparametric regression in the presence of measurement errors. Biometrika 86: 541-554.

Cleveland W.S. and Devlin S.J. 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83: 596-610.

de Groen P. 1996. An introduction to total least squares. Nieuw Archief voor Wiskunde, Vierde Serie 14: 237-253.

Delicado P. 2001. Another look at principal curves and surfaces. J. Multivariate Anal. 77: 84-116.

Einbeck J., Tutz G. and Evers L. 2005. Local Principal Curves. Stat. Comput. 15: 301-313.

Fan J. and Truong Y.K. 1993. Nonparametric regression with errors in variables. Ann. Stat. 21: 1900-1925.

Fan J., Hu T.C. and Truong Y.K. 1994. Robust non-parametric function estimation. Scand. J. Stat. 21: 433-446.

Felus Y.A. 2004. Application of total least squares for spatial point process analysis. J. Surv. Eng. 130:433-446.

Giebel G., Brownsword R. and Kariniotakis G. 2003. State of the art on wind power prediction. Technical report, Anemos project deliverable report D1.1, available online: http://anemos.cma.fr (last access: 15-05-2007).

Giebel G. 2005. Wind power prediction using ensembles. Risø National Laboratory, Wind Energy Department, Roskilde, Denamrk, Technical Report Risø-R-1527(EN).

Golug G.H. and Van Loan C.F. 1980. An analysis of the total least squares problem. SIAM J. Numer. Anal. 17: 883-893.

Golug G.H. and Van Loan C.F. 1996. Matrix Computations, John Hopkins University Press, Baltimore.

Hastie T. and Stuetzle W. 1989. Principal curves. J. Am. Stat. Assoc. 84:502-516.

Hastie T. and Tibshirani R. 1990. Generalized Additive Models, Chapman & Hall/CRC, London.

Huber P.J. 1981. Robust Statistics, Wiley, New York.

Jonsson B. 1994. Prediction with a linear regression model and errors in a regressor. Int. J. Forecasting 10: 549-555.

Kelly G.E. 1992. Robust regression estimators - the choice of tuning constants. Statistician 41: 303-314.

Lange M. and Focken U. 2006. Physical Approach to Short-term Wind Power Prediction, Springer, Berlin.

Liavas A.P. and Regalia P.A 1999. On the numerical stability and accuracy of the conventional recursive least squares algorithm. IEEE T. Signal Proces. 47: 88-96.

Ljung L., Morf M. and Falconer D. 1978. Fast calculation of gain matrices for recursive estimation schemes. Int. J. Control 27: 1-19.

Madsen H. 2006. Time Series Analysis, Technical University of Denmark, Lyngby.

Madsen H., Pinson P., Kariniotakis G., Nielsen H.Aa. and Nielsen T.S. 2005. Standardizing the performance evaluation of short term wind power prediction models. Wind Eng. 29: 475-489.

Nielsen H.Aa., Nielsen T.S., Joensen A.K., Madsen H. and Holst J. 2000. Tracking time-varying coefficient functions. Int. J. Adapt. Control 14: 813-828.

Nielsen H.Aa., Nielsen T.S., Madsen H. Badger J., Giebel G., Landberg L., Sattler K., Voulund L. and Tøfting J. 2006. From wind ensembles to probabilistic information about future wind power production. Proc. PMAPS 2006, Prob. Meth. App. Pow. Syst., Stockholm, Sweden.

Nielsen T.S., Nielsen H.Aa. and Madsen H. 2002. Prediction of wind power using time-varying coefficient functions. Proc. 15th IFAC World Congress, Barcelona, Spain.

Parkum J.E., Poulsen N.K. and Holst J. 1992. Recursive forgetting algorithms. Int. J. Control 55: 109-128.

Pinson P. 2006. Estimation of the uncertainty in wind power forecasting. Ph.D. thesis, Ecole des Mines de Paris, Paris, France.

Pinson P. Nielsen H.Aa. and Madsen H. 2007. Robust estimation of time-varying coefficient functions – Application to the modeling of wind power production. Technical University of Denmark, Informatics and Mathematical Modeling, Technical report IMM-2007-09.

Sanchez I. 2006. Short-term prediction of wind energy production. Int. J. Forecasting 22: 43-56.

Tarpey T. and Flury B. 1996. Self-consistency: a fundamental concept in statistics. Stat. Sci. 11: 229-243.

Thor S.-E. and Weis-Taylor P. 2002. Long-term research and development needs for wind energy for the time frame 2000-2020. Wind Energ. 5:73-75.

Tibshirani R. 1992. Principal curves revisited. Stat. Comput. 2: 183-190.

Verbeek J.J., Vlassis N. and Kröse B. 2002. A k-segments algorithm for finding principal curves. Pattern Recogn. Lett. 23: 1009-1017.

Welsh A.H. 1994. Robust estimation of smooth regression and spread functions and their derivatives. Stat. Sinica 6: 347-366.

Yu K.B. 1991. Recursive updating of the eigenvalue decomposition of a covariance matrix. IEEE T. Signal Proces. 39: 1136-1145.
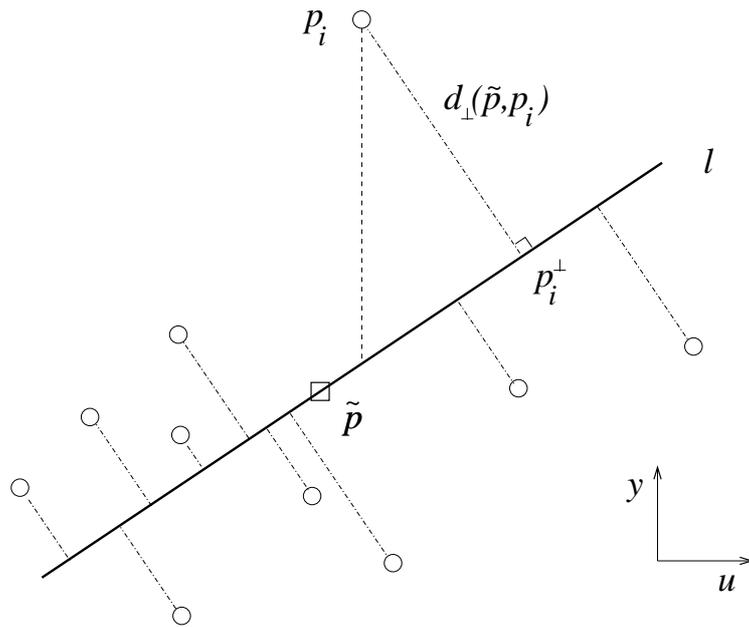
Fig. 1. Fitting of a local linear model with LS and TLS criteria. In the former case, the distance to be minimized is measured along the y-axis (dashed line), while for the latter, this distance is that between the observation p_i and its orthogonal projection $p_i^{\perp}$ on the curve (dash-dotted line). The fitting point $\tilde{p}$ is represented by a square while the data points are represented by circles. For both types of fitting, the location of $\tilde{p}$ on the u-axis is user-defined, while its position on the y-axis results from the fitting approach employed.
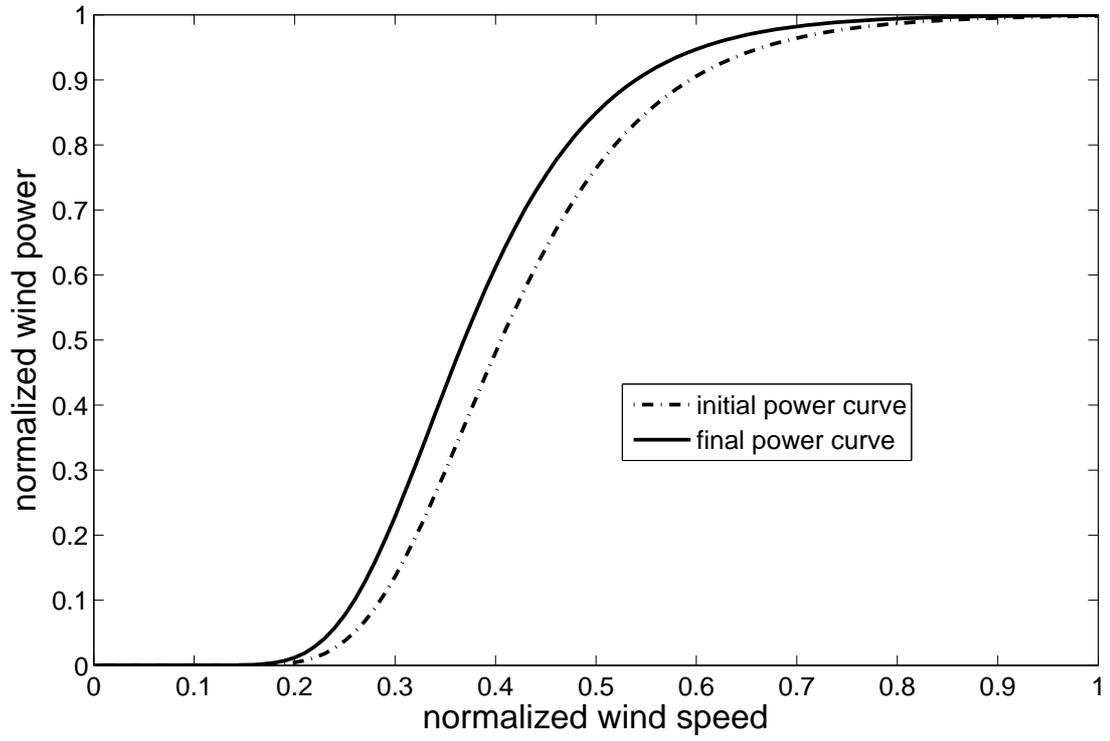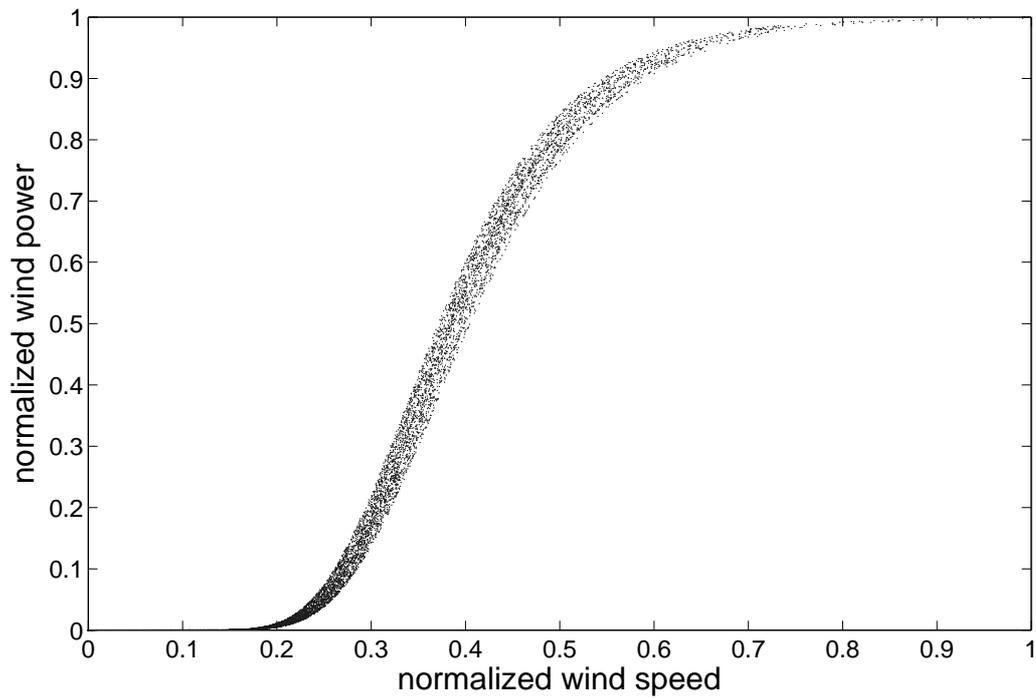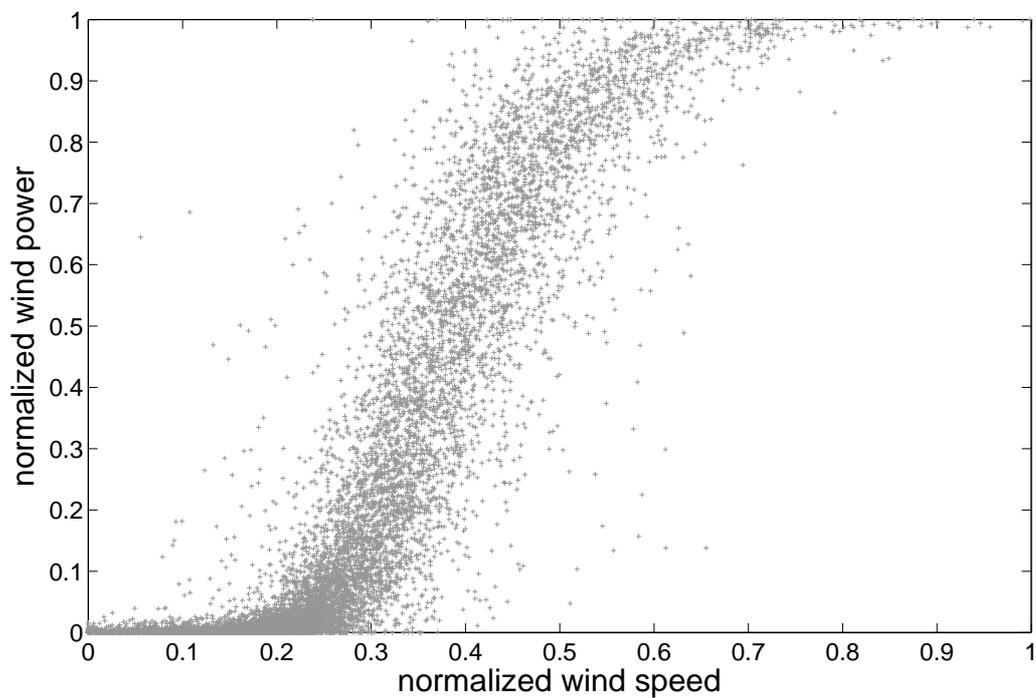
Fig. 2. Model for the true power curve. This model consists in a double exponential function, whose parameters $\boldsymbol{\tau}_i$ linearly vary (in time) from $\boldsymbol{\tau}_1 = [10\ 40]^\top$ to $\boldsymbol{\tau}_N = [11\ 40]^\top$ over the dataset.

(a) Noise-free nonstationary power curve

(b) Corrupted nonstationary power curve

Fig. 3. Simulated data. Both wind speed and power data defining the true power curve are corrupted with independent noise sequences. The noise sequences include additive and impulsive components.
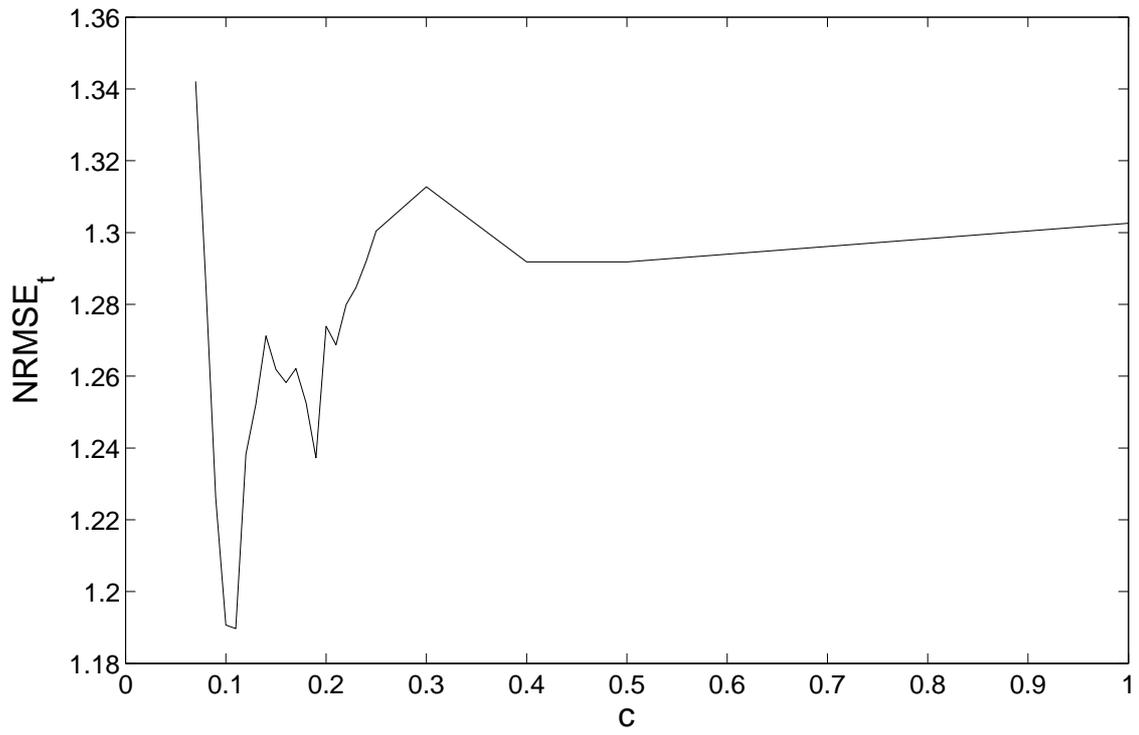
Fig. 4. Evolution of the NRMSE$_t$ criterion as a function of the value of the threshold parameter c for $\hat{\phi}_n^{\perp*}$. The minimum NRMSE$_t$ is attained for $c = 0.11$.
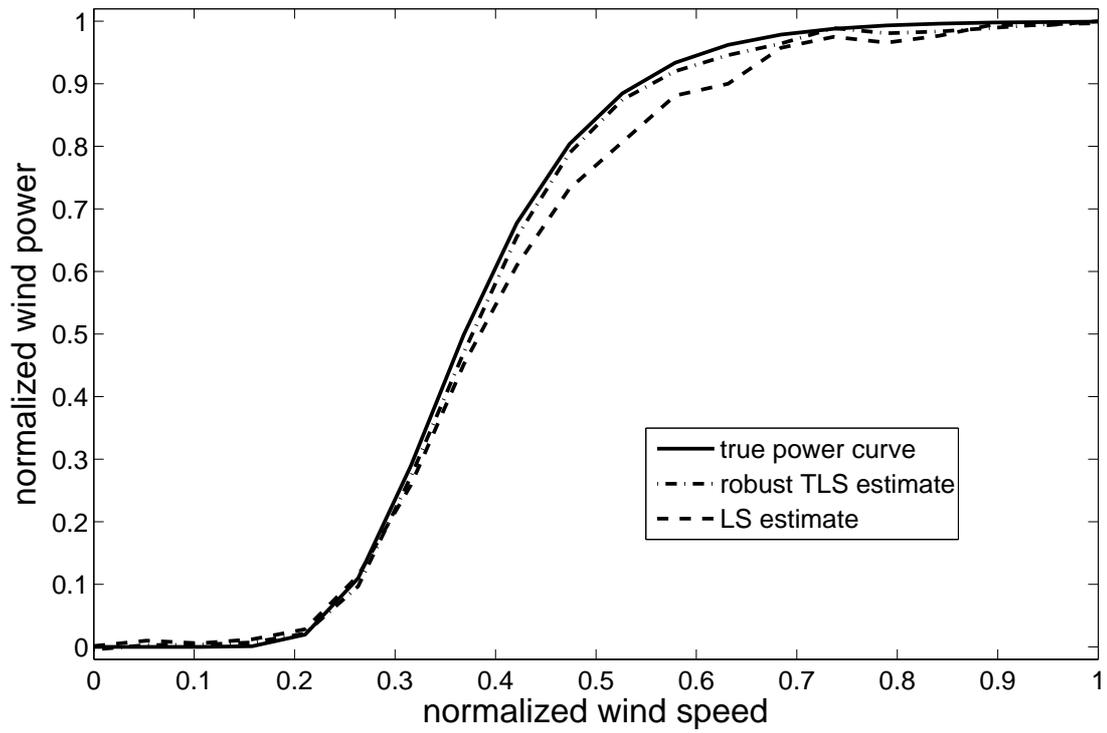
Fig. 5. Comparison of the final power curves obtained with $\hat{\phi}_n$ and $\hat{\phi}_n^{\perp*}$, that is, from local linear regression whose coefficients are either fitted with an LS or a TLS criterion. One can visually appraise the greater ability of the former estimation method to yield a better estimate of the target regression.

Table 1. Optimal tuning parameters when fitting local linear regression with LS and TLS criteria. They are obtained from a one-fold cross-validation procedure.

	h_0	h_1	λ
LS fitting	0.024	1.5	0.987
TLS fitting	0.05	3.5	0.994

Table 2. Minimum values of the NRMSE$_r$ criterion and related values of the other evaluation criteria for the various estimators. Error criteria are expressed as a percentage of the nominal capacity of the wind farm.

	$\hat{\phi}_n$	$\hat{\phi}_n^{\perp}$	$\hat{\phi}_n^{\perp *}$
NMAE$_r$ [%]	7.1453	6.9398	6.8874
NMAE$_t$ [%]	2.4677	1.0805	0.9756
NRMSE$_r$ [%]	11.4801	11.5909	11.5713
NRMSE$_t$ [%]	3.0519	1.3215	1.1897