WIND ENERGY
Wind Energ. 2006; **9**:95–108
Published online 19 December 2005 in Wiley Interscience
(www.interscience.wiley.com) DOI: 10.1002/we.180

Research Article

Using Quantile Regression to Extend an Existing Wind Power Forecasting System with Probabilistic Forecasts

Henrik Aalborg Nielsen*, Henrik Madsen and Torben Skov Nielsen, Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

Key words:
wind power forecasting;
uncertainty;
quantile regression;
additive model

For operational planning it is important to provide information about the situation-dependent uncertainty of a wind power forecast. Factors which influence the uncertainty of a wind power forecast include the predictability of the actual meteorological situation, the level of the predicted wind speed (due to the non-linearity of the power curve) and the forecast horizon. With respect to the predictability of the actual meteorological situation a number of explanatory variables are considered, some inspired by the literature. The article contains an overview of related work within the field. An existing wind power forecasting system (Zephyr/WPPT) is considered and it is shown how analysis of the forecast error can be used to build a model of the quantiles of the forecast error. Only explanatory variables or indices which are predictable are considered, whereby the model obtained can be used for providing situation-dependent information regarding the uncertainty. Finally, the article contains directions enabling the reader to replicate the methods and thereby extend other forecast systems with situation-dependent information on uncertainty. Copyright © 2005 John Wiley & Sons, Ltd.

Received 22 November 2004; Revised 22 September 2005; Accepted 1 October 2005

Introduction

In recent years a growing interest in information about the uncertainty of wind power forecasts in different weather situations has emerged. Based on wind speed measurements and standard meteorological forecasts, Bremnes[1] estimates the power curve of a small wind farm and then models the relation between the actual and forecasted wind speed with respect to both the mean and the covariance. Considering the same wind farm, Bremnes[2] uses local linear quantile regression to obtain a probabilistic model based on meteorological forecasts and on observations of power. Pinson and Kariniotakis[3–5] use consecutive forecasts and, based on these, define a quantity called the 'meteo-risk index'. This quantity measures the agreement between the consecutive forecasts and is used to predict the uncertainty of the wind power forecast. Lange and Heinemann[6] identify relations between typical weather situations and the magnitude of the forecast error. In a research project carried out with Eltra (the transmission system operator (TSO) in western Denmark), Nielsen and Madsen[7] have developed a stochastic model of the forecast errors when using WPPT Version 2.[8] The model describes the variance and correlation within and between the daily forecasts.

Over the last decade, much effort has been spent on developing wind power forecasting systems supplying a point forecast of the wind power production of a farm or a region. For this reason it is desirable to be able to extend existing point forecast systems to probabilistic forecast systems. In this article we consider forecast errors from an existing system (WPPT Version 4)[9,10] and use linear quantile regression[11] together with spline

*Correspondence to: H. A. Nielsen, Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark
E-mail: han@imm.dtu.dk
Contract/grant sponsor: European Commission; contract/grant number: ENK5-CT-2002-00665.

bases[12] in order to obtain a model for the 25% and the 75% quantiles of the forecast errors. The methods are easily applicable and can be applied to any forecasting system using the free software called 'R' (http://www.r-project.org) with the add-on package 'quantreg', which can be downloaded from the same homepage. The meteorological forecasts used by WPPT in the particular set-up considered are the wind speed and direction 10 m above ground level (10 m a.g.l.) from DMI-HIRLAM,[13] and we consider a few other forecasted variables from DMI-HIRLAM as candidates for the quantile model. Furthermore, for each of these variables we consider risk indices inspired by the meteo-risk index mentioned above. It was decided to focus on the horizons relevant for reporting to NordPool (http://www.nordpool.com). Considering timing and calculation times, we have therefore focused on the 06Z DMI-HIRLAM forecast for horizons 18–42 h. However, in order to be able to calculate the risk indices for each of the meteorological variables, we consider only 18–36 h (see the beginning of Risk Indices of Meteorological Variables).

The outline of the article is as follows. The data, including training and test periods, are described in section two. The methods used in the article, i.e. quantile regression and parametric additive models, are briefly described in section three. Also in section three it is described why we have chosen to use additive models. Sections four and five describe the model building process and the evaluation on test data respectively. Finally, in section six we give conclusions. In the Appendix it is outlined how the models can be fitted and visualized and how forecasts can be produced using 'R'.

Data

The data used in this study consist of the following.

- 15 min power averages from the Tunø Knob offshore wind farm consisting of 10 Vestas V39 turbines (500 kW nominal); location 55° 58′ 08″ N, 10° 21′ 10″ E.
- Forecasts of the wind power production of the farm based on WPPT Version 4; time step 15 min, with a maximum horizon of 48 h.
- Meteorological forecasts of air density, friction velocity, 10 m wind speed and direction from DMI-HIRLAM;[13] time step 60 min, with a maximum horizon of 48 h.
- Data from 1 January to 31 May 2003 are used for developing and training the models. Data from 1 June to 31 October, 2003 are used to test the forecast results. Measurements of power production are available back to 1 July 1999; these are used as a climatological reference.

Interpolation is used to obtain meteorological forecasts for all time points at which power forecasts and observations are available. It is noted that on 2 September 2003 a model change was introduced into DMI-HIRLAM which is expected to have a large influence on the forecasted 10 m wind. Over time, WPPT will adapt to this change and will use the meteorological forecasts in an optimal way within the framework of the system. However, the distributional properties of the error may change permanently.

Figure 1 shows the observed power plotted against the forecasted power for the training and test data and for the data split at 2 September 2003. The plots of the training and test data differ qualitatively. In particular, the saturation at high levels of production occurs much more frequently in the training data than in the test data. The other plots in the figure indicate that this could be related to the change in DMI-HIRLAM on 2 September 2003. For this reason the evaluation on the test data will be performed on the total test set and on the test set split on the date just mentioned.

Methods

Quantile Regression

Considering a random variable Y, the median is the most well-known quantile and is characterized as the value $Q(1/2)$ for which the probabilities of obtaining values of Y above or below $Q(1/2)$ both equal 1/2. Generally,

DOI: 10.1002/we

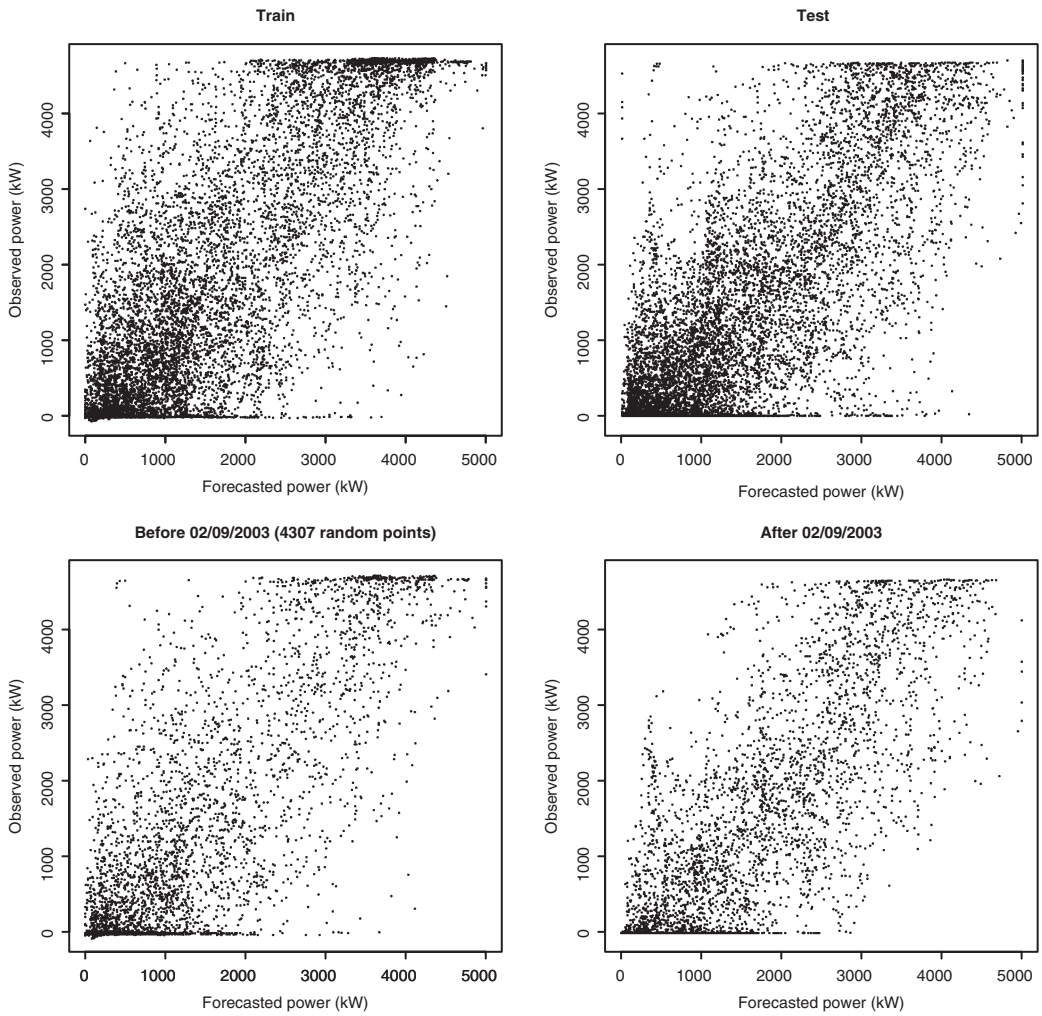

Figure 1. Observed versus forecasted power (horizons 18–36 h since 06Z) for the training and test data (top row) and before and after a model change in DMI-HIRLAM which is expected to have a large influence on the forecasted 10 m wind (bottom row). Note that the lengths of the training and test periods are both 5 months. Also, the two plots in the bottom row contain the same number of points; to achieve this, a random subset of the points before the model change is selected

$Q(\tau)$ is defined as the value for which the probability of obtaining values of Y below $Q(\tau)$ is τ. In quantile regression,[11,14] $Q(\tau)$, $0 < \tau < 1$, is expressed as a linear combination of some known regressors and unknown coefficients, exactly as the mean is modelled in (multiple) linear regression. Thus the τ-quantile is modelled as

$$Q(\tau) = \beta_0(\tau) + \beta_1(\tau)x_1 + \ldots + \beta_P(\tau)x_P \tag{1}$$

where $x.$ are the p known regressors, also called explanatory variables, and $\beta.(\tau)$ are unknown coefficients, depending on τ, to be determined from observations $(y_i, x_{i,1}, \ldots, x_{i,p})$, $i = 1, \ldots, N$.

Given the check function

$$\rho_\tau(e) = \begin{cases} \tau e, & e \geq 0 \\ (\tau - 1)e, & e < 0 \end{cases} \tag{2}$$

DOI: 10.1002/we

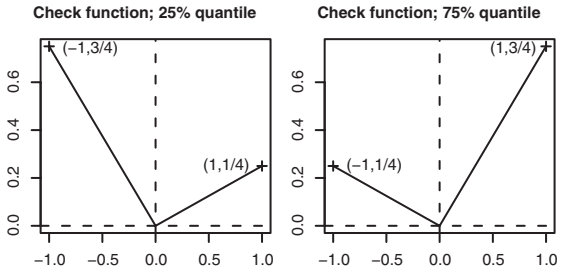

Figure 2. The check function $\rho_\tau(e)$ *for* $\tau = 0.25$ *(left) and 0.75 (right)*

the sample τ-quantile can be found by minimizing $\sum_{i=1}^{N} \rho_\tau(y_i - q)$ with respect to q (Reference 15, p. 417). Figure 2 shows the check function for two values of τ. Replacing q with the right-hand side of (1) leads to the estimates

$$\hat{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \rho_\tau[y_i - (\beta_0 + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p})] \tag{3}$$

where $\beta(\tau)$ is a vector containing the unknown coefficients. The estimates can be obtained by use of linear programming techniques.[14] Here we have used the add-on package 'quantreg' for 'R' (see section one and Appendix). It is noted that, if the check function is replaced with squared loss, i.e. if $\rho_\tau(e) = e^2$, then equation (3) leads to least squares estimates.

Parametric Additive Quantile Models

To simplify the discussion, we start by considering models for the mean of a random variable and later we consider models for the quantiles. Generally, when the dependence of y on $x_1, \ldots, x_p$ is not known, a very general model is

$$y = g(x_1, \ldots, x_p) + \varepsilon \tag{4}$$

where g is an unknown function and ε represents independent identically distributed errors with mean zero and variance σ^2. In principle, it is possible to estimate g, e.g. by use of local regression.[16] However, when investigating many explanatory variables, i.e. more than two or three, the *curse of dimensionality*[17] makes practical application of (4) problematic (Reference 18, pp. 83–84). To circumvent this problem, additive models[18] are used in this article. Models of this type can be expressed as

$$y = \alpha + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p) + \varepsilon \tag{5}$$

The constant α and the *functions* $f_\cdot(\cdot)$ can be estimated based on data using non-parametric methods together with backfitting.[18] However, note that, unless the levels of functions are restricted, the estimates are non-unique, e.g. a constant can be added to one function and subtracted from another. Hastie and Tibshirani[18] impose the restriction that each of the function estimates has zero mean over the data.

As described by Hastie and Tibshirani[18] (Section 9.3), each of the functions can be approximated by linear combinations of known basis functions of the corresponding explanatory variable, i.e.

$$f_j(x_j) = \sum_{k=1}^{n_j} b_{jk}(x_j)\theta_{jk} \tag{6}$$

where $b_{j\cdot}(x_j)$ are the basis functions and $\theta_{j\cdot}$ are unknown coefficients. The resulting model, i.e. the model consisting of (5) and (6), is a linear regression model. However, the price paid for the simplicity obtained by using

DOI: 10.1002/we

the approximation (6) is that the resulting estimates of the functions generally have larger bias than those based on non-parametric approximations and backfitting (Reference 18, Section 9.3).

To obtain unique estimates, a restriction must be imposed on (6) and the resulting basis functions derived. If e.g. it is required that $f_j(0) = 0$, it follows from (6) that $\theta_{j1} = -\sum_{k=2}^{n_j} \theta_{jk} b_{jk}(0)/b_{j1}(0)$. Plugging this θ_{j1} into (6) results in the expression

$$f_j(x_j) = \sum_{k=2}^{n_j} \left(b_{jk}(x_j) - \frac{b_{jk}(0)}{b_{j1}(0)} b_{j1}(x_j) \right) \theta_{jk} \tag{7}$$

where the term in front of the coefficients θ_{jk} defines the $n_j - 1$ new basis functions. Note that the basis function to be eliminated can be chosen with some freedom as long as it is non-zero for the value at which the function f_j is zero. Likewise, if some of the functions in (5) are known to be periodic, this restriction can be imposed on the basis functions. Using e.g. cubic B-spline basis functions,[12] the functions in (5) have continuous derivatives up to order two. This property should also be imposed when constructing the periodic basis. The resulting model is a linear regression model for which the least squares estimates of α and θ_{jk} are unique.

Comparing with (1), it is seen that this can be generalized to quantile regression by modelling $Q(\tau)$ as

$$Q(\tau) = \alpha(\tau) + \sum_{j=1}^{p} f_j(x_j; \tau) = \alpha(\tau) + \sum_{j=1}^{p} \sum_{k=1}^{n_k} b_{jk}(x_j) \theta_{jk}(\tau) \tag{8}$$

with the basis functions constructed under appropriate restrictions on $f_j(\cdot)$, $j = 1, \ldots, p$, as outlined above.

In this article, focus will be on the 25% and 75% quantiles. As mentioned above, the levels of the functions are arbitrary, e.g. $f_j(0) = 0$. Hence the effect of x_j should be quantified by plotting the sum of the corresponding estimated function and the estimated intercept. To centre the plots around zero, we subtract the average of the intercepts estimated for the 25% and 75% quantiles. Thus the effect of x_j is quantified by plotting

$$\hat{f}_j(x_j; \tau) + \hat{\alpha}(\tau) - \frac{\hat{\alpha}(0.25) + \hat{\alpha}(0.75)}{2}$$

for $\tau = 0.25, 0.75$. Otherwise, differences in $\hat{\alpha}(0.25)$ and $\hat{\alpha}(0.75)$ may cause apparent crossings of the 25% and 75% quantiles. The 'hat' denotes estimated values.

Building the Quantile Model

In this section, models for the 25% and 75% quantiles are developed. First a model considering the explanatory variables

`pow.fc`	forecasted power from WPPT (kW)
`horizon`	number of hours since 06Z (h)
`ad`	forecasted air density from DMI-HIRLAM (g m^{-3})
`fv`	forecasted friction velocity from DMI-HIRLAM (m s^{-1})
`wd10m`	forecasted wind direction 10 m a.g.l. from DMI-HIRLAM (°)
`ws10m`	forecasted wind speed 10 m a.g.l. from DMI-HIRLAM (m s^{-1})

is developed. Hereafter, risk indices based on the meteorological forecast variables are considered.

Basic Model

Owing to the non-linearity of the power curve, it is natural to require that the forecasted power production (`pow.fc`) is included in the quantile model of the forecast error. Figure 3 shows all pairwise scatter plots of

DOI: 10.1002/we

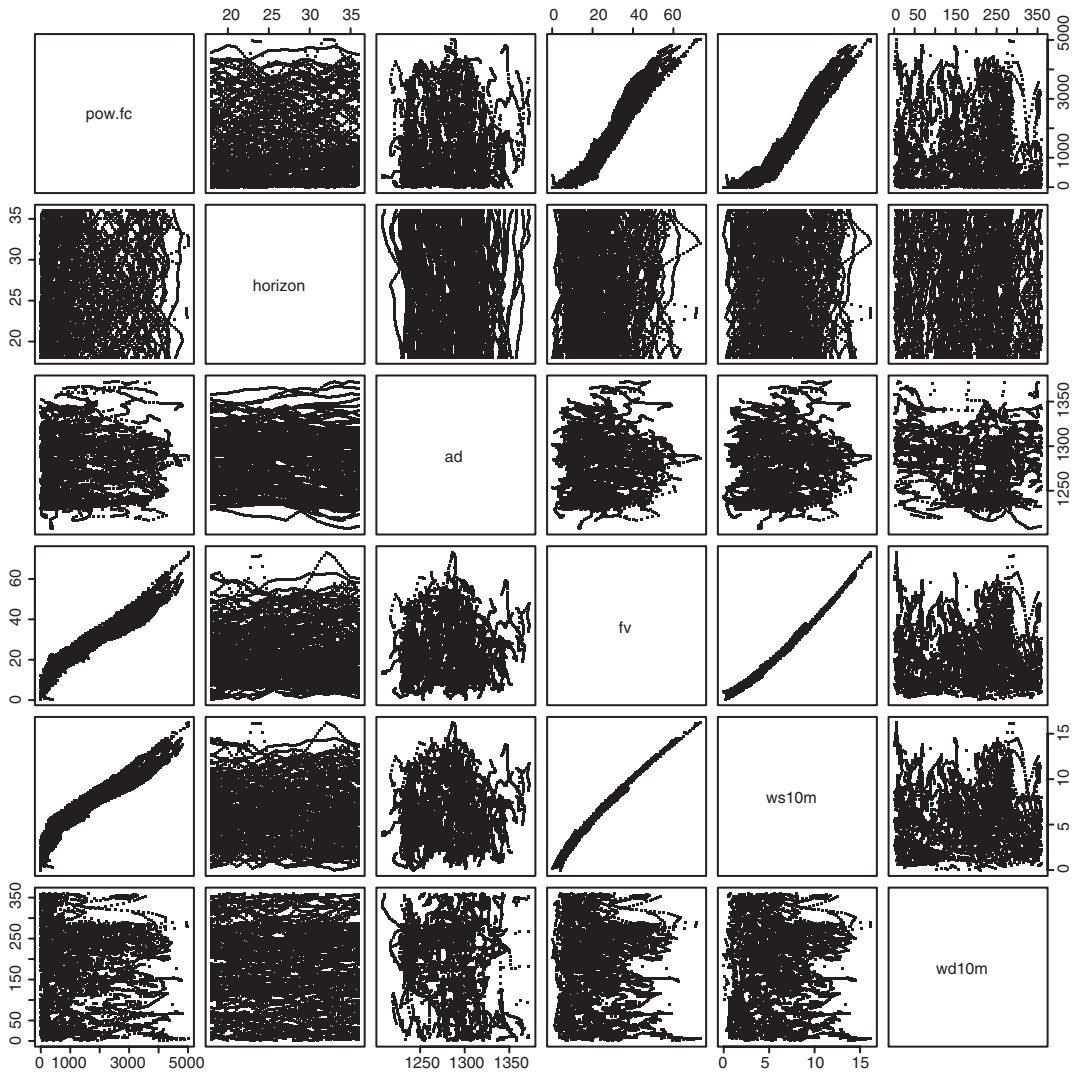

Figure 3. All pairwise scatter plots of the explanatory variables (training data)

the potential explanatory variables. Owing to the close relations between some of the variables (`pow.fc`, `fv` and `ws10m`), it is seen that, with the requirement just stated, the friction velocity (`fv`) and the 10 m wind speed (`ws10m`) cannot be included in the model.

For each of the remaining explanatory variables (`pow.fc`, `horizon`, `ad` and `wd10m`) a spline basis with 10 degrees of freedom is constructed.[12] For the wind direction (`wd10m`) a periodic cubic spline basis with equidistant knots is used. The periodic basis is constructed so that it integrates to zero over the period (360°). For the non-periodic variables, natural spline bases without intercepts are used; this implies that the functions are restricted to be zero at the lower boundary knot. The boundary knots are placed at the limits of the data and the internal knots are placed according to the quantiles of the individual explanatory variables. In this way the model allows for more flexibility where the observations are relatively dense. Note that for prediction it is important to use the same actual knots. Since none of the bases allows for a free intercept, this is handled by an intercept in the model. The intercept is expected to vary with the quantile considered.

DOI: 10.1002/we

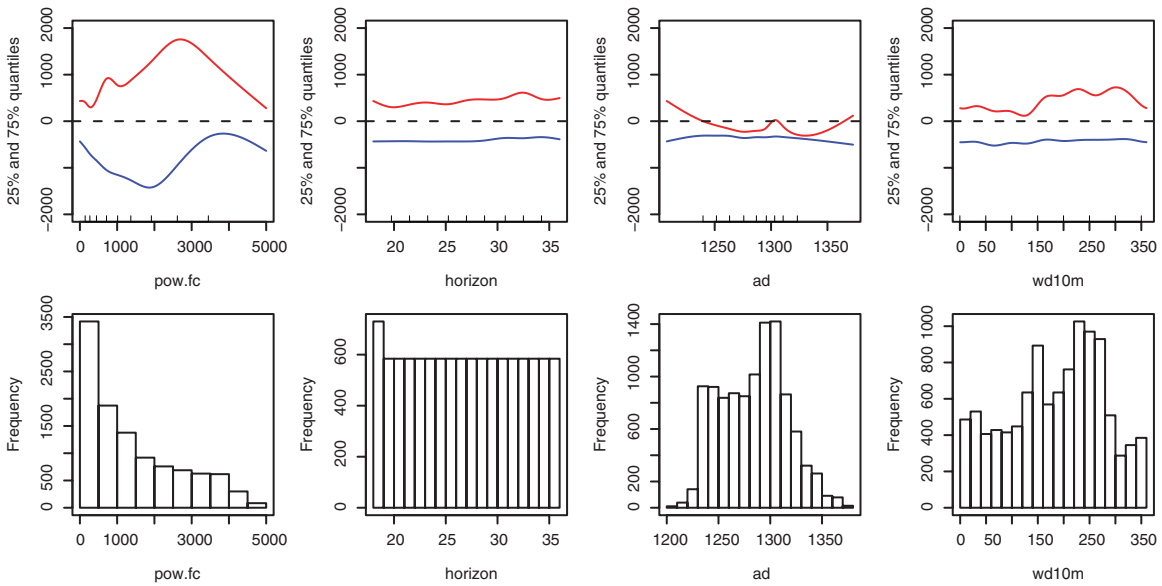

Figure 4. Estimated 25% (lower) and 75% (upper) quantiles (top row) together with histograms (bottom row) of the explanatory variables. The internal markings on the horizontal axis in the top row of plots indicate the placement of knots

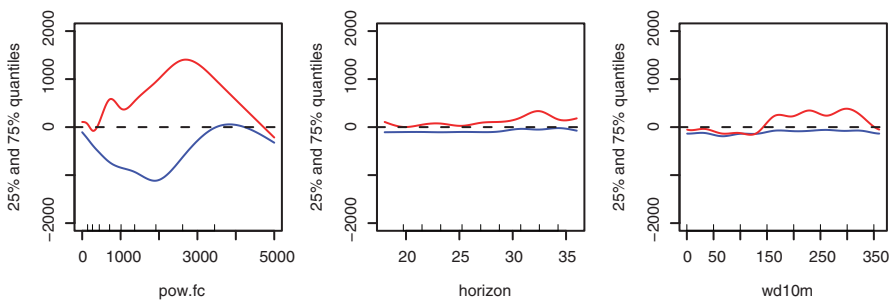

Figure 5. Estimated 25% (lower) and 75% (upper) quantiles when excluding the forecasted air density from the model. The internal markings on the horizontal axis of the plots indicate the placement of knots

The resulting model for each of the quantiles (25% and 75%) is depicted in Figure 4, which shows the effect of each variable. For each pair of estimates the difference in estimated intercepts is visible for the minimal value of the explanatory variable. It is seen that the effect of horizon is small (almost flat curves) and there is some increased uncertainty for westerly winds. The dependence on air density seems to be minor. Overall, the most important explanatory variable is the forecasted power. For the training data, crossings of the 25% and 75% quantiles occur in 111 out of 10,658 cases.

The estimates describing the dependence on the air density do not seem to have any reasonable interpretation and the differences for low and high densities are supported by very few data points. For this reason it is decided to exclude it from the model. The number of crossings decreases to 46 for the resulting model. The estimates are depicted in Figure 5. Since the curves are close when `pow.fc` is zero, it is seen that the intercepts of the 25% and 75% quantile models are very close.

DOI: 10.1002/we

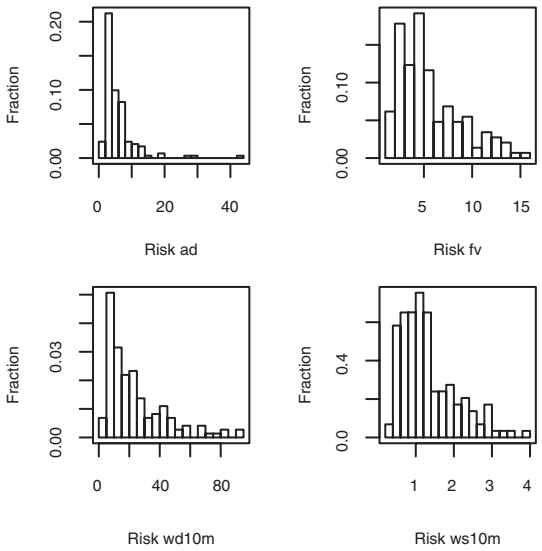

Figure 6. Histograms of risk indices (training data)

Risk Indices of Meteorological Variables

The European Centre for Medium-range Weather Forecasts (ECMWF, http://www.ecmwf.int), which runs the global model supplying boundary conditions for DMI-HIRLAM, performs data assimilation based on 12 h intervals (00Z and 12Z). For the assessment of the forecast risk the two DMI-HIRLAM forecasts which are based on the two latest global data assimilations are compared. Since the primary interest is in the 06Z forecast (please refer to the Introduction), this is compared with the preceding 18Z forecast. Since the maximum forecast horizon of DMI-HIRLAM is 48 h, risk indices can only be calculated for the 06Z forecast up to a horizon of 36 h.

Following Pinson and Kariniotakis,[5] the differences in the two forecasts are squared and summed over the entire range of horizons for a particular 06Z forecast. The square root of this number is used as the risk index of each variable, corresponding to each 06Z meteorological forecast. Figure 6 shows histograms of the risk indices. It is seen that the risk indices generally only show a few high values and therefore it will only be possible to detect simple relationships; for this reason it is decided only to investigate linear relationships between the quantiles and the risk indices.

When adding the risk indices one at a time to the model shown in Figure 5, i.e. without air density, the results shown in Figure 7 are obtained. Generally the risk indices seem to be of minor importance for the quantiles and it is chosen to use only the one with the clearest signal, i.e. `fv`. Figure 8 shows the estimates obtained for this model. It is seen that the effect of the risk index is comparable to the effect of the horizon. For the training data the number of crossings of the 25% and 75% quantiles decreases to 39 for this model.

Evaluation on Test Data

The following models are fitted to the training data and evaluated on the test data:

- *basic*—a model using only the forecasted power productions as explanatory variables;
- *full*—the model corresponding to Figure 4;
- *w/o density*—the model corresponding to Figure 5;
- *incl. risk*—the model corresponding to Figure 8.

DOI: 10.1002/we

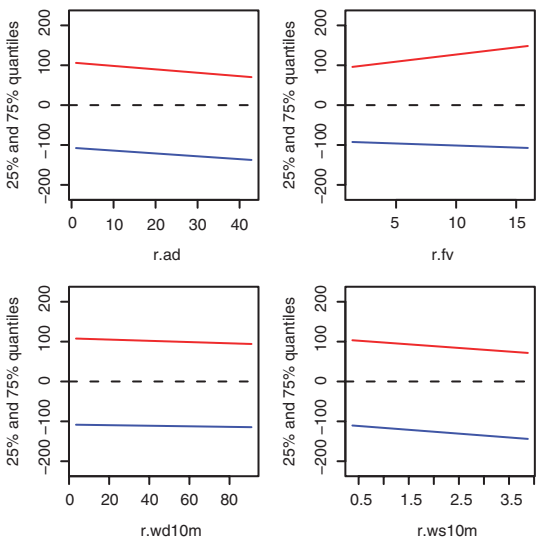

Figure 7. Estimates (lower, 25% quantile; upper, 75% quantile) of the dependence on risk indices when requiring the dependence to be linear and adding the risk indices one at a time to the model depicted in Figure 5

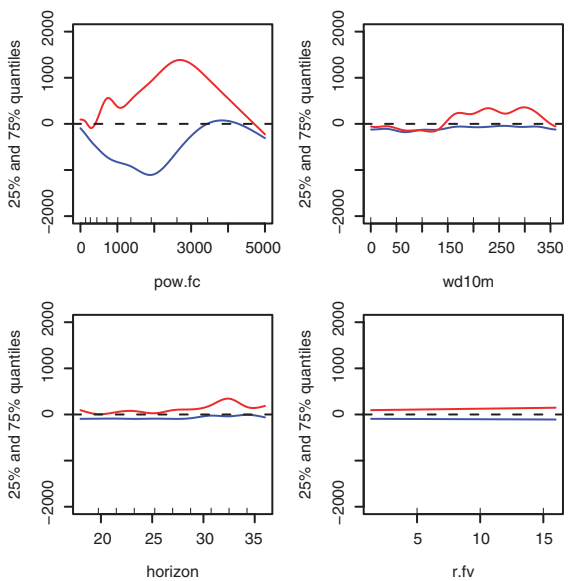

Figure 8. Estimates (lower, 25% quantile; upper, 75% quantile) in the model consisting of the model depicted in Figure 5 with the risk index of `fv` *added (bottom right)*

The number of crossings on the test data ranges from 60 to 82 of 11,168 cases. In the case of crossing of the two quantiles, these have been set to their common average. The actual frequencies with which the prediction error is below the 25% quantile or above the 75% quantile in the test data are listed in Table I. A marked difference is seen when splitting the data according to the date at which a presumably important change was introduced into DMI-HIRLAM. For this reason we focus on the part of the test period up to 2 September 2003.

DOI: 10.1002/we

Table I. Observed frequencies (test data) below the forecasted 25% quantile and above the forecasted 75% quantile for the four models considered. Values (%) are given both for the full test set and for the test set split into two parts based on a presumably important model change in DMI-HIRLAM (see section two)

	Basic	Full	W/o density	Incl. risk
All test data				
Above	18	13	17	17
Centre	61	58	53	52
Below	21	29	30	31
Test data before 2/9/2003				
Above	19	10	19	19
Centre	62	64	53	52
Below	19	26	28	29
Test data after 2/9/2003				
Above	16	16	15	14
Centre	60	50	50	53
Below	24	34	35	33

The model termed 'full' seems to result in some problems for the forecasted 75% quantile, since this forecast is exceeded in only 10% of the cases. For the model termed 'basic', the forecast intervals seem to be symmetric but too wide. The two remaining models both perform well (52%–53% change to be between the forecasted quantiles), but the forecast intervals seems to be shifted upwards corresponding to approximately 5%. Random variation may account for some of these differences, but such variations are difficult to quantify owing to the inherent and presumably complicated correlation of the data.

Given quantiles which are correct in a probabilistic sense, the quality of these depends on (i) the ability to distinguish between situations with low and high uncertainty and (ii) the sharpness of the distributions. Here the sharpness is measured as the inter quartile range (IQR), i.e. the difference between the forecasted 75% and 25% quantiles.

Qualitatively, (i) is fulfilled if both low and high values of the IQR occur, and with respect to (ii) the IQR should be smaller than the IQR obtained from historic production data. These aspects are addressed in Figure 9. Results are shown for three models, where the basic model is included for reference, although it is not very precise with respect to the observed frequencies. It is seen that the basic model differs from the other two models. Also, since the plots for the other two models do not differ markedly, it does not seem very important to include the particular risk index.

The relatively large difference between the 5% and 95% quantiles of the IQR indicates high variability, and for probabilistic correct quantiles this can be interpreted as the fulfilment of (i). Furthermore, it is seen that in many situations the IQR is significantly smaller than the IQR of the historic power productions, i.e. the forecast is sharp compared with historic data.

Discussion and Conclusions

We have proposed a method for building models of e.g. the 25% and 75% quantiles of the forecast errors from existing wind power forecast systems. Such models can be used together with existing systems and have the potential of providing situation-specific information about the uncertainty of a particular forecast.

The quantiles are modelled as a sum of non-linear smooth functions of variables forecasted by the meteorological model or variables derived from such forecasts. The additive model structure is used since it allows

DOI: 10.1002/we

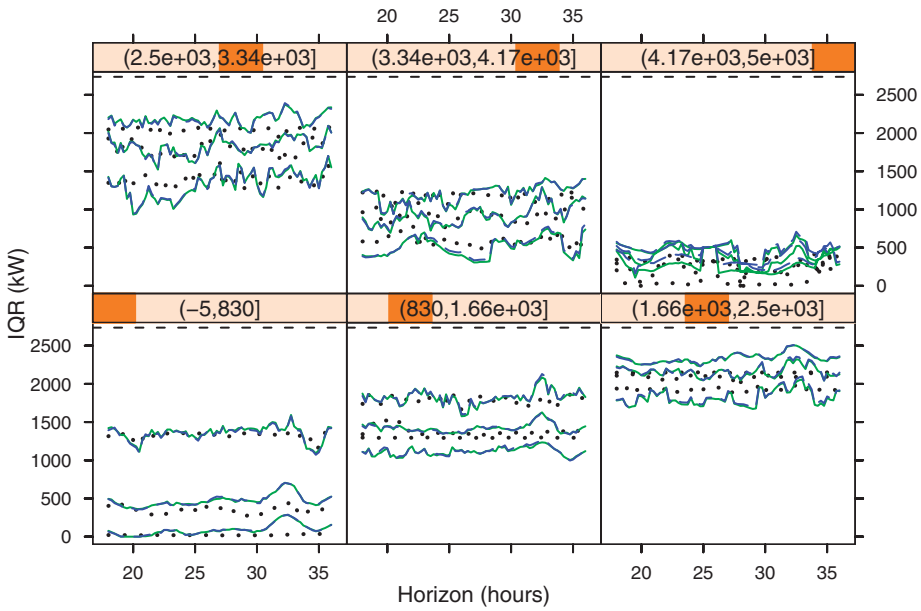

Figure 9. Quantiles (5%, 50%, 95%) of the inter-quantile range (IQR). Models: basic (dotted), w/o density (full), incl. risk (dashed). The grouping variable is the power production (kW) as forecasted by WPPT and split into six intervals of equal length. The broken horizontal line at 2736 kW indicates the IQR of actual power productions between 1 July 1999 and 1 June 2003

for the inclusion of more explanatory variables than more general non-parametric models. Furthermore, additive models are relatively easy to visualize and interpret. Using spline bases to approximate each of the smooth functions as a linear combination of basis functions depending only on known quantities permits the use of existing linear quantile regression software to fit the models.

The software used is 'R' together with the add-on package 'quantreg', which can both be freely downloaded from http://www.r-project.org. An example R-script is included in the Appendix 'R' could be used to easily extend a given wind power forecast system, and it is even possible to embed 'R' into other software products.

With respect to the analysis of the specific data it is noted that the risk indices, which all are inspired by Pinson and Kariniotakis,[5] do not seem to have very much influence on the 25% and 75% quantiles. However, it is noted that Pinson and Kariniotakis[5] consider horizons ranging from 0 to 24 h, whereas we consider horizons ranging from 18 to 36 h. Note that the horizons mentioned do not take into account the calculation time of HIRLAM. Not surprisingly, the most influential variable is the forecasted wind power production. Furthermore, the results show increased uncertainty for westerly winds. Also, the effect of the horizon on the quantiles is minor.

Bremnes[2] shows that the optimal quantile to use depends on the actual prices in the market. This will require a range of quantile models to be applied in parallel. Using models with several predictors and spline bases as suggested in this article is likely to result in crossing of some of these quantiles. Ideally, the coefficients estimated should be constrained in order to avoid crossings. However, we are not aware of software which can handle this easily. In the situation just outlined, it is probably sufficient to use the quantile model indicated by the prices in the market and disregard the fact that this quantile model may cross some other quantile models.

As just outlined, quantile regression is characterized by estimating separate models for each quantile. As a consequence, crossing of quantiles may occur; indeed, in the analysis of the data here a few crossings of the 25% and 75% quantiles occurred. In practice this probably indicates low uncertainty and is therefore of less practical importance. It is, however, undesirable from a theoretical point of view. One solution would be to

DOI: 10.1002/we

start with the median (50% quantile) and find solutions to successive lower and higher quantiles under the restriction that the quantiles do not cross. The restriction should be valid for all possible values of the explanatory variables. Considering the data at hand, this could be approximated by considering all observations, i.e. the number of restrictions will be high. We believe that methods based on approximations of the full distribution should also be investigated, since this will automatically supply non-crossing quantiles.

As mentioned above, we chose the additive model structure since it allows for inclusion of many explanatory variables. However, the additive model structure may lead to unrealistic quantiles, i.e. quantiles above or below the capacity range. For the additive model displayed in Figure 5, the lower and upper bound is broken in approximately 500 out of 11,000 cases. As expected, very similar results are obtained when including the risk index, i.e. for the model displayed in Figure 8. A possible way to avoid such unrealistic quantiles is to forecast the power production instead of the forecast error. This allows the power to be transformed so that it is confined to the range of possible values.[19] Hereafter a quantile model built on the transformed scale will not produce unrealistic forecasts when these are transformed back to the original scale. Since the unrealistic forecasts are mainly caused by the model structure, e.g. when the production is at the maximum the effect of wind direction should be small, it might also be considered to use other quantile regression methods. Bremnes[2] uses local quantile regression which surely solves the problem mentioned. However, we argue that it is generally inappropriate to use local regression methods for applications with many potential explanatory variables. The explanation is sometimes called the *curse of dimensionality* and refers to the fact that neighbourhoods with a fixed number of points become less local as the dimensions increase.[18] Hence, in many dimensions, local regression is not really local. An illustration of this fact can be found in Reference 19, which uses local regression to fit direction-dependent power curve models, i.e. models with two explanatory variables, using 5 months of data. In this case, using a 10% nearest-neighbour bandwidth results in actual bandwidths of 4–7 m s^{-1} when the wind speed is 5 m s^{-1}. For higher wind speeds the data are even more sparse.

Acknowledgements

The work described here is carried out under the project ANEMOS, which is supported by the European Commission under contract ENK5-CT-2002-00665. The financial support is hereby greatly acknowledged. Furthermore, the authors wish to thank Elsam Kraft A/S and the Danish Meteorological Institute for supplying the data used in this study.

Appendix: Parametric Additive Quantile Models in 'R'

The 'quantreg' library or package is not part of a standard 'R' installation. To install the package, start 'R' and issue the command

```
install.packages("quantreg")
```

However, please read the help-page before issuing this command.

Assume that the data are contained in a data frame named `train` with columns `y`, `x1` and `x2`. Furthermore, data for which quantile forecasts must be computed are assumed to be contained in a similar data frame named `test` with columns `x1` and `x2`.

The R-script shown in Table II illustrates how parametric additive quantile models can be fitted, how the results can be visualized and how forecasts can be produced.

Periodic bases[12] can be constructed from the output of the function `bs`. This is done by the S-PLUS/R function `pb.bse` found in the file `periodic.bases.q` at http://www.imm.dtu.dk/~han/pub which has been used in this article. The restriction that the function approximated by the periodic basis integrates to zero over the period is imposed on the periodic basis using the function downloadable as `bint0.q`.

DOI: 10.1002/we

Table II. R-script. Functions and operators are indicated by the bold font; comments start with '##'

```
## Load required libraries:
library (splines)
library (quantreg)

## Make natural spline bases with 10 columns and knots placed
## according to the quantiles of x1 and x2:
basis . x1 <- ns (train $ x 1, df = 10, intercept = F)
basis . x2 <- ns (train $ x 2, df = 10, intercept = F)

## Fit 25% and 75% quantile models (1 denotes the intercept):
fit25 <- rq (train $y ~ 1 + basis . x1 + basis . x2, tau = 0·25)
fit75 <- rq (train $y ~ 1 + basis . x1 + basis . x2, tau = 0·75)

## Estimated coefficients:
coef (fit25)
coef (fit75)

## Plot of fitted values of the estimates related to x1:
intercept . avg <- (coef (fit25) ["(Intercept)"] + coef (fit75) ["(Intercept)"]) / 2
matplot (train $x1,
        cbind (basis . x1 %*% coef (fit25) [grep ("basis \\. x1", names (coef (fit25)))]
             + coef (fit25) ["(Intercept)"] − intercept . avg,
             basis . x1 %*% coef (fit75) [grep ("basis \\. x1", names (coef (fit75)))]
             + coef (fit75) ["(Intercept)"] − intercept . avg
        ))

## Plot of fitted values of the estimates related to x2:
matplot (train $x2,
        cbind (basis . x2 %*% coef (fit25) [grep ("basis \\. x2", names (coef (fit25)))]
             + coef (fit25) ["(Intercept)"] − intercept . avg,
             basis . x2 %*% coef (fit75) [grep ("basis \\. x2", names (coef (fit75)))]
             + coef (fit75) ["(Intercept)"] − intercept . avg
        ))

## Forecast for data frame 'test':
test . b . x1 <- ns (test $x1,
                  knots = attributes (basis . x1) $ knots,
                  Boundary . knots = attributes (basis . x1) $Boundary . knots,
                  intercept = attributes (basis . x1) $ intercept)
test . b . x2 <- ns (test $x2,
                  knots = attributes (basis . x2) $ knots,
                  Boundary . knots = attributes (basis . x2) $ Boundary . knots,
                  intercept = attributes (basis . x2) $ intercept)
##
## Comment: It is very important that the bases for prediction are
## constructed independently from the test data, i.e. by supplying the
## knots etc . as outlined above.
##
qForecast <- data . frame (Q25 = cbind (1, test . b . x1, test . b . x2) %*% coef (fit25),
                           Q75 = cbind (1, test . b . x 1, test . b . x2) %*% coef (fit75))
## Print forecasts:
qForecast
```


DOI: 10.1002/we

References

1. Bremnes JB. Probabilistic wind power forecasts by means of statistical model. *Proceedings of IEA R&D Wind Annex XI Joint Action Symposium on Wind Forecasting Techniques*, Norrköping, 2002; 103–114.
2. Bremnes JB. Probabilistic wind power forecasts using local quantile regression. *Wind Energy* 2004; **7**: 47–54.
3. Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. *Proceedings of European Wind Energy Conference & Exhibition*, Madrid, 2003.
4. Pinson P, Kariniotakis G. On-line adaptation of confidence intervals based on weather stability for wind power forecasting. *Proceedings of Global Wind Energy Conference & Exhibition*, Chicago, IL, 2004.
5. Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. *Wind Energy* 2004; **7**: 119–132.
6. Lange M, Heinemann D. Relating the uncertainty of short-term wind speed predictions to meteorological situations with methods from synoptic climatology. *Proceedings of European Wind Energy Conference & Exhibition*, Madrid, 2003.
7. Nielsen HA, Madsen H. Analyse og simulering af prædiktionsfejl for vindenergiproduktion ved indmelding til Nord-Pool. Informatik og Matematisk Modellering, Danmarks Tekniske Universitet, Lyngby, 2002.
8. Nielsen T, Madsen H, Christensen H. WPPT—a tool for wind power prediction. *Proceedings of Wind Power for the 21st Century Conference*, Kassel, 2000.
9. Nielsen HA, Nielsen TS, Madsen H. Using meteorological forecasts for short term wind power forecasting. *Proceedings of IEA R&D Wind Annex XI Joint Action Symposium on Wind Forecasting Techniques*, Norrköping, 2002; 49–58.
10. Nielsen TS, Nielsen HA, Madsen H. Prediction of wind power using time-varying coefficient-functions. *Proceedings of XV IFAC World Congress*, Barcelona, 2002; 2715–2721.
11. Koenker RW, Bassett GW. Regression quantiles. *Econometrica*, 1978; **46**: 33–50.
12. De Boor C. *A Practical Guide to Splines*. Springer: Berlin, 1978.
13. Sass BH, Nielsen NW, Jørgensen JU, Amstrup B, Kmit M, Mogensen KS. The operational DMI-HIRLAM system 2002-version. *Technical Report 02–05*. Danish Meteorological Institute, 2002.
14. Koenker RW, D'Orey V. [Algorithm AS 229] Computing regression quantiles. *Applied Statistics*, 1987; **36**: 383–393.
15. Kotz S, Johnson NL, Read CB (eds). *Encyclopedia of Statistical Sciences*, vol. 5. Wiley: New York, NY, 1982.
16. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 1988; **83**: 596–610.
17 Bellman RE. *Adaptive Control Processes*. Princeton University Press: Princeton, NJ, 1961.
18. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman and Hall: London, 1990.
19. Nielsen H, Madsen H, Nielsen T, Badger J, Giebel G, Landberg L, Sattler K, Feddersen H. Wind power ensemble forecasting. *Proceedings of 2004 Global Windpower Conference and Exhibition*, Chicago, IL, 2004.

DOI: 10.1002/we