WIND ENERGY
*Wind Energ.* 2004; **7**:47–54 (DOI: 10.1002/we.107)

Research Article

# Probabilistic Wind Power Forecasts Using Local Quantile Regression

**John Bjørnar Bremnes*,** Research and Development Department, Norwegian Meteorological Institute, PO Box 43 Blindern, N-0313 Oslo, Norway

**Key words:**
wind power;
probabilistic forecasts;
quantile regression;
economic value

*Wind power forecasts are in various ways valuable for users in decision-making processes. However, most forecasts are deterministic, and hence possibly important information about uncertainty is not available. Complete information about future production can be obtained by using probabilistic forecasts, and this article demonstrates how such forecasts can be created by means of local quantile regression. The approach has several advantages, such as no distributional assumptions and flexible inclusion of predictive information. In addition, it can be shown that, for some purposes, forecasts in terms of quantiles provide the type of information required to make optimal economic decisions. The methodology is applied to data from a wind farm in Norway.* 

## Introduction

During the last few years, forecasts of wind power production have received increasing attention because of their value in planning production and optimizing financial income. Such forecasts are usually deterministic, i.e. knowledge about uncertainty is not provided, and possibly useful information is therefore not conveyed to users. An overview of deterministic approaches is given in Reference 1.

For interpretability purposes, uncertainty is arguably best quantified and presented in probabilistic terms. Forecasts are then ideally fully specified probability distributions, and the challenge is to determine these using information available at the time the forecasts are being issued. In References 2 and 3, estimates of expected power and standard deviations of errors are used to identify beta distributions as probabilistic forecasts. Use of multivariate Gaussian distributions (via transformations) with parameters estimated by local regressions is described in Reference 4. In Reference 5, several forecasting models based on single and ensembles of weather forecasts are evaluated. Additionally, it is theoretically established that certain probabilistic forecasts should be preferred in order to make optimal decisions in some cases.

In this article a finite number of quantiles/percentiles of the probability distribution for energy production is estimated using historical data. The $\theta$ quantile is informally defined as the value where the probability of production less than this value is $\theta$. When the probability is specified in per cent, the quantiles are often referred to as percentiles. An example of such forecasts with illustrative explanations is given in Figure 1, where the distributions are characterized by the 5, 25, 50, 75 and 95 percentiles (from bottom to top). As an example of interpretation, there is a 95% chance of production less than about 0·55 MWh in the first hour. It should also be noted that pairs of quantiles define forecast intervals; for instance, the 95 and 5 percentiles and the 75 and 25 percentiles define 90% and 50% forecast intervals respectively.

The main advantages of the proposed approach are that the shape of the distributions does not have to be specified and that any information about these distributions can easily be included in the models.

*Correspondence to: J. B. Bremnes, Research and Development Department, Norwegian Meteorological Institute, PO Box 43 Blindern, N-0313 Oslo, Norway
E-mail: j.b.bremnes@met.no

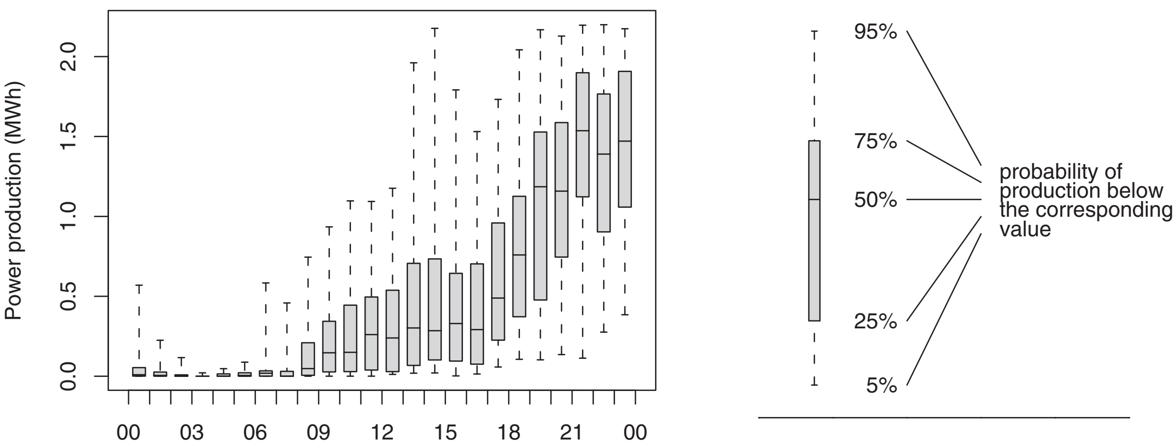


*Figure 1. An example of hourly probabilistic forecasts in terms of 95, 75, 50, 25 and 5 percentiles (left, from top to bottom), and illustrative explanations (right)*

## Motivation

A common use of wind power forecasts is placing bids in energy markets for sales of future production. An example is described in Reference 6 where the producer gets the spot price $p_s$ for the electricity produced $e_p$, but has to pay for production that deviates from his bid $e_b$. Denoting the unit cost of excess production by $c_+$ and the unit cost of production less than the bid by $c_-$ , the income may be formulated by

$$I(e_p, e_b, p_s, c_+, c_-) = \begin{cases} e_p p_s - (e_b - e_p)c_- & \text{if } e_p \le e_b \\ e_p p_s - (e_p - e_b)c_+ & \text{if } e_p \le e_b \end{cases} \tag{1}$$

A similar example is given in Reference 5.

At the time the producer must place his bid, there is often considerable uncertainty about the values of the variables in (1). Attention will here be restricted to the case where power production is regarded as a random variable $E_p$, while the remaining variables are considered not random. The income is then a function of a univariate random variable, and the aim is to somehow optimize this with respect to the bid $e_b$. Unless knowledge about the producer's risk policy is accessible, it is natural to maximize the expected income. It can be shown (see Appendix B) that this implies that the $c_+(c_+ + c_-)^{-1}$ quantile of the probability distribution for energy production $E_p$ should be used. In other words, forecasts of this quantile are needed to make optimal decisions. Further, the distribution of income can be computed with this quantile (or any other) as energy bid through transformation of $E_p$.

## Methodology

The aim of this article is forecasting energy production $E_p$ based on information $\mathbf{x}$, e.g. output from numerical weather prediction (NWP) models, that is known a given time in advance. In a probabilistic framework, $E_p$ is a random variable and the objective is to determine its distribution as a function of $\mathbf{x}$, i.e. the conditional distribution of $E_p \mid \mathbf{x}$. Estimates of this distribution or, more precisely, only some of its quantiles will here be based on historical data, denoted by the pairs $(e_{p,i}, \mathbf{x}_i)$, $i = 1, 2, \ldots, n$. Bold notation is used to indicate that the predictors $\mathbf{x}_i$ are vectors.

 

The classical approach to the problem is to assume a parametric distribution for $E_p \mid \mathbf{x}$, e.g. the Gaussian distribution, and estimate its parameters. However, the parameters often all depend on the predictor variables $\mathbf{x}$ in unknown complex ways, such that finding an appropriate relationship is not easy. In addition, the choice of parametric distribution may not fit the data properly. To avoid any distributional assumptions, one might restrict attention to estimating only a finite number of quantiles of the conditional distribution. This can be achieved by means of quantile regression.[7,8] In quantile regression, one has to specify how the quantiles depend on $\mathbf{x}$ (as in linear regression), but for wind power the dependence may not be described by a simple polynomial. However, in a neighbourhood of a given $\mathbf{x}$ the dependence may be simple enough, and quantile regression can be applied to data points close to $\mathbf{x}$. It is then reasonable to let data points close to $\mathbf{x}$ have more impact than those further away by weighting the data points accordingly. This version or extension of quantile regression is referred to as local quantile regression (LQR).[9]

Assume that the $\theta$ quantile $q_\theta()$ can be formulated by

$$q_\theta(\mathbf{x}; \alpha_0, \boldsymbol{\alpha}) = \alpha_0 + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x} \tag{2}$$

An estimate of this quantile at $\mathbf{x}$ is then obtained by solving the minimization problem

$$\arg\min_{(\alpha_0, \boldsymbol{\alpha})} \sum_{i=1}^{n} \rho_\theta(e_{\mathrm{p},i} - q_\theta(\mathbf{x}_i - \mathbf{x}; \alpha_0, \boldsymbol{\alpha}))w\left(\frac{\|\mathbf{x}_i - \mathbf{x}\|_2}{h_\lambda(\mathbf{x})}\right) \tag{3}$$

where the loss function $\rho_\theta()$ is defined by

$$\rho_\theta(u) = \begin{cases} u\theta & \text{if } u \geq 0 \\ u(\theta - 1) & \text{otherwise} \end{cases} \tag{4}$$

and the weight function $w()$ by

$$w(u) = \begin{cases} (1 - u^3)^3 & \text{if } u \in [0,1) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Since each historical predictor value $\mathbf{x}_i$ is centered around $\mathbf{x}$, the estimated quantile equals the estimate of $\alpha_0$. In the argument of $w()$ in (3), $\|\cdot\|_2$ is the Euclidean norm (see Appendix A), $\lambda \in (0, 1]$ is a smoothing parameter that determines the fraction of the data to be used, and $h_\lambda(\mathbf{x})$ is the distance from $\mathbf{x}$ to the $\lambda n$ nearest predictor value. This means that $(1 - \lambda)100\%$ of the data have no impact on the fit and that cases with predictor values close to $\mathbf{x}$ have greater impact than those further away. It should be mentioned that (5) is just one possible definition; others could have been used as well. For any choice of weight function, however, it is important to be aware of possible differences in scales of the predictors. In that case it is natural to scale the predictors before the weights are calculated. In the example presented later, the inverses of each predictor's standard deviation are used as scaling factors.

Since the weights depend on $\mathbf{x}$, these must be computed for each predictor value of interest, and consequently the minimization problem must be solved equally many times. Computational aspects in the case of no weights (or equal weights) are described in References 10–12. The numerical methods presented therein can easily be adjusted to handle different weightings. Finally, it should be noted that different quantiles can cross, because they are estimated separately. This can be avoided by putting appropriate constraints on the fit, i.e. the parameter $\alpha_0$.

## *Validation of Forecasts*

Probabilistic forecasts of continuous variables such as wind power should be both reliable and sharp. These properties are in conflict with each other, and at some stage one usually has to determine their respective degrees of importance. This is particularly important when competing forecast models are being ranked.

 

Reliability means that the forecasted probability distributions should be in accordance with the measurements. For example, 75% of the measurements should be below the 75 percentiles in the long run. A chi-square hypothesis test for multinomial data (see Reference 13, Sections 8.2 and 9.6) is used here to test whether forecasts in terms of quantiles are reliable. Let $n_0, n_1, \ldots, n_K$ denote the numbers of measurements in the intervals formed by $K$ quantiles, and $p_0, p_1, \ldots, p_K$ the corresponding theoretical probabilities. Forecasts are then defined as not reliable if

$$\sum_{i=0}^{K} \frac{(n_i - np_i)^2}{np_i} \geq \chi^2_{1-\alpha,K} \tag{6}$$

where $n$ is the total number of forecasts/measurements and $\chi^2_{1-\alpha,K}$ is the $1 - \alpha$ quantile of the $\chi^2$ distribution with $K$ degrees of freedom. Informally, forecast models are rejected if the fractions of measurements in the intervals formed by the quantiles are not in accordance with the claimed probabilities.

Sharpness is a measure of uncertainty, which of course should be as small as possible. For quantile forecasts, sharpness is most easily measured by distances between pairs of quantiles, e.g. the distances between the 5 and 95 percentiles. The average length (or other summary statistic) is then an appropriate quantitative measure. A minor disadvantage is that multimodal distributions are not given credit. Such distributions may for example occur for strong wind speeds that result in either nearly full production or shutdown of generators and, hence, low or no production.

## *Example: Hourly Wind Power Forecasts*

### *Data*

The example is taken from the wind farm at Vikna in Norway. This farm has five turbines with a total capacity of 2·2 MW. The data used are:

- total hourly power production for the wind farm;
- hourly meteorological forecasts initiated at 00 UTC from the NWP model Hirlam10, which has spatial horizontal resolution of about 10 km; lead times from +24 to +47 h are used.

On average there were about 300 cases available for each lead time during the period from January 2000 to December 2001. For modelling, the power production data are transformed by $\sin^{-1}(\sqrt{\cdot/2{\cdot}2})$ to ensure that all estimates/forecasts are in the interval [0, 2·2] MWh (on the normal scale). All results to follow are presented on the normal scale.

### *Model Selection*

Model selection involves choosing predictors and smoothing parameters such that forecasts are as good as possible with respect to the validation scores described above. Owing to the limited amount of data, a cross-validation was applied to determine the models. In cross-validation the data are divided into $M$ sets, where $M - 1$ sets are used to train or fit the models while the remaining set is used to make predictions and evaluate these. The set left out for prediction is alternated such that in the end there is one prediction for each data point.

A cross-validation with five sets was performed for 10 predictor combinations based on the three continuous variables *wind speed* and *wind direction* at 10 m from Hirlam10 and *month* (see Table I). The latter was made available to describe possible seasonal dependences. Second degree polynomials of the predictors and interactions between two predictors were mainly tried to see if more flexible parametric functions would result in larger smoothing parameters and better estimates/forecasts. Each candidate was tested with the smoothing parameters 0·2, 0·3, . . . , 1. The procedure was carried out separately for the lead times +24, +30, +36, +42, and +47 h, and predictions of the 5, 25, 50, 75 and 95 percentiles were made for each model and lead time.

It was required that the forecast model to be preferred should produce reliable predictions at the 5% significance level, which was subjectively found to be an adequate level. For models not rejected by the

 

Table I. List of the 10 predictor combinations used in the cross-validation. The three predictors involved are *wind speed* (FF10) and *wind direction* (DD10) at 10 m and *month* (MONTH)

| |
|---|
| MONTH |
| FF10 |
| FF10 + MONTH |
| FF10 + DD10 |
| FF10 + FF10$^2$ |
| FF10 + DD10 + MONTH |
| FF10 + DD10 + FF10 × DD10 |
| FF10 + FF10$^2$ + DD10 + FF10 × DD10 |
| FF10 + DD10 + DD10$^2$ + FF10 × DD10 |
| FF10 + FF10$^2$ + DD10 + DD10$^2$ + FF10 × DD10 |

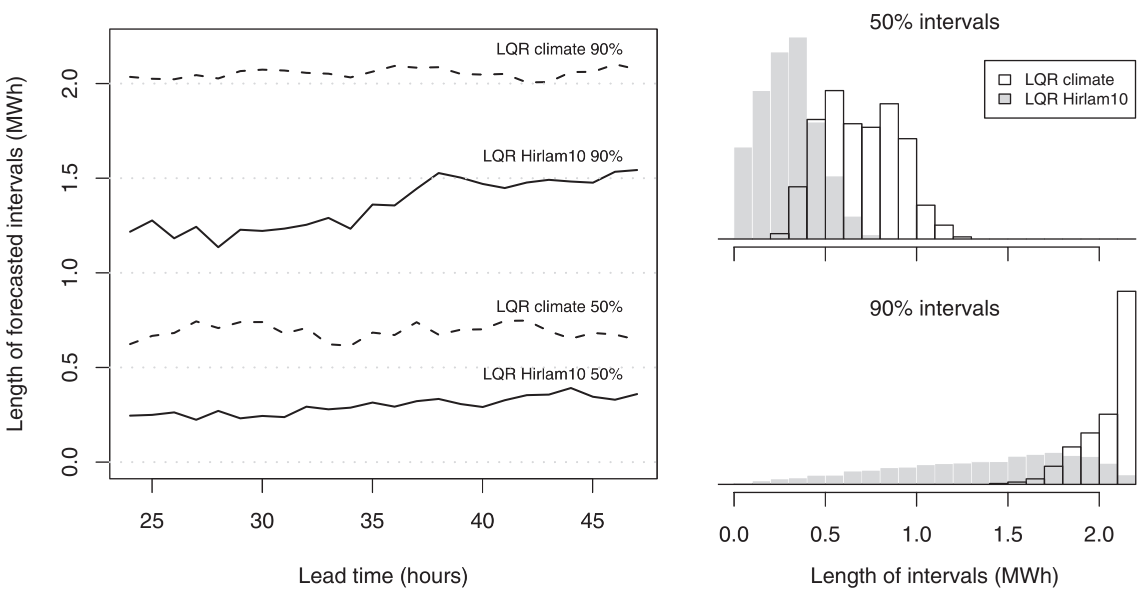


*Figure 2. Average lengths (left) and empirical distributions (right) of the forecasted 50% and 90% intervals of LQR Hirlam10 and LQR climate. The average lengths are plotted as a function of lead time, while the distributions are for all lead times*

hypothesis test, the average lengths and, to some extent, the empirical distributions of the 50% and 90% forecast intervals were evaluated.

## Results

Roughly half of the 10 × 9 forecast models produced reliable predictions at the 5% significance level. The further assessment of forecasted intervals resulted in using the predictors *wind speed* and *wind direction* (first degree polynomials without interaction) with smoothing parameter 0·4 for all lead times. A model using only *month* as predictor (LQR climate) was also chosen in order to quantify the effect of weather forecasts. The smoothing parameter for this model was 0·5.

In Figure 2 (left) the average lengths of the 50% and 90% forecast intervals are shown for the model using weather forecasts (LQR Hirlam10) and the climate model (LQR climate). Clearly, LQR Hirlam10 produced

 

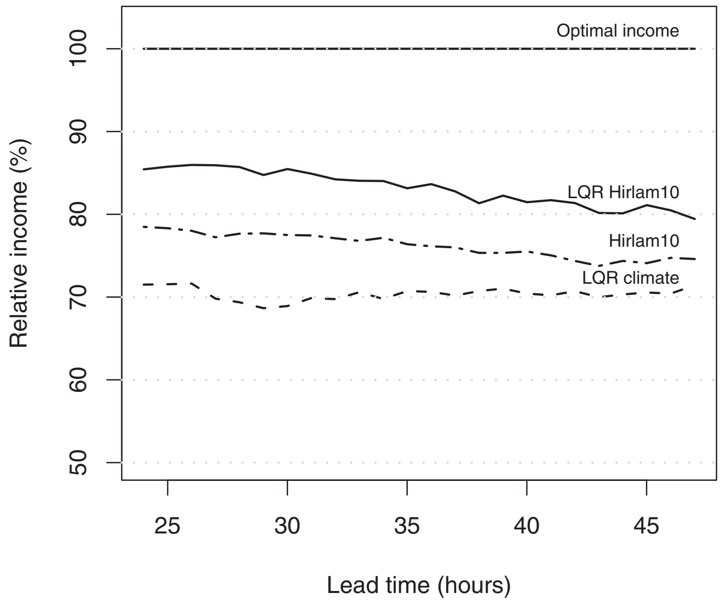


*Figure 3. Relative economic value of using LQR Hirlam10, Hirlam10 wind speed transformed into power/income, and LQR climate as a function of lead time. Optimal income is obtained with perfect forecasts, i.e. with no forecast errors*

the sharpest forecasts, but the intervals are relatively long. In contrast to LQR climate, the uncertainty of LQR Hirlam10 forecasts increases with lead time, but it is still considerably lower than that of LQR climate at +47 h. The comparably large differences in length of the 50% and 90% intervals may indicate that the forecast distributions often have strong tails. To the right in Figure 2 the empirical distributions of the forecasted intervals are depicted. It can be noticed in particular that LQR Hirlam10 had more variation in the length of the 90% intervals than LQR climate, i.e. LQR Hirlam10 produced forecasts with relatively low uncertainty at times.

In order to compare the economic effects of the models, assumptions about the unknown price and cost variables in (1) must be made. For simplicity it was decided to use values for the area in Denmark studied in Reference 6, i.e. $p_s = 25$, $c_+ = 12$ and $c_- = 7$ €/MWh for all hours and days. The relative incomes from using LQR Hirlam10, LQR climate and Hirlam10 are shown in Figure 3. Income from the last model, Hirlam10, is computed by transforming the wind speed at 10 m to power using a known power curve, and then using this power as the bid. It should be mentioned that the power curve is designed for wind speeds at the height of the generators, and that this most likely had a negative effect on the income. From the figure it can be seen that the relative income from using LQR Hirlam10 is 10%–15% higher than that from using LQR climate, with Hirlam10 in between.

## Further Comments

Selection of predictors is an important and time consuming step in the construction of forecast models. In recent years, ensembles of weather forecasts, where each ensemble member represents a likely evolution of the atmosphere, have become available. Strategies for reducing the vast amount of information contained in these are therefore needed. In the case of LQR a possible solution is to first sort the ensemble members with respect to wind speed and then use the ensemble member that corresponds to the quantile of interest as predictor. For example, if 100 ensemble members are available and sorted in increasing order, then, say, the 25th ensemble member could be used in predicting the 25 percentile of power production. The use of different predictors for each quantile can also be applied when other types of predictor information are available.

As mentioned earlier, there are currently several methods for producing probabilistic forecasts, and it would be interesting to compare the strengths and weaknesses of these in more detail. Further, it would be constructive to quantify the presumed advantages of probabilistic forecasts over deterministic ones in cases where the benefits for end users are directly computable.

The example of the previous section was only about making forecasts for a single wind farm (unity); however, for some purposes, forecasts for the total production of several wind farms are needed. Aggregation of spatially correlated forecasts in terms of quantiles poses a challenging research problem and deserves more attention in the future.

The application of LQR has so far relied on the availability of power production measurements. However, for new wind farms, such data are obviously not obtainable. If measurements of wind speed are accessible and their relation to power is known and sufficiently accurate, LQR can be applied to make probabilistic forecasts of wind speed which further can be transformed to power. In cases where both wind speed and power measurements are missing, it is not possible to make forecasts by LQR or any other statistical method. Measurements are therefore valuable and should be handled accordingly, especially since statistical tools are often able to improve forecasts by other methods.

## *Conclusions*

It is demonstrated how reliable probabilistic forecasts of power production can be made by means of local quantile regression. The approach requires no strong assumptions, and for some purposes it provides the correct type of forecast to make optimal economic decisions.

## *Acknowledgement*

The power production data were kindly provided by Nord-Trøndelag Elektrisitetsverk (www.nte.no).

## *Appendix A*

Let $\mathbf{u} = (u_1, \ldots, u_p)$ be a $p$-dimensional vector. The Euclidean norm $\|\cdot\|_2$ is then defined by

$$\|\mathbf{u}\|_2 = \left(\sum_{i=1}^{p} u_i^2\right)^{1/2} \tag{7}$$

## *Appendix B*

Assume that power production $E_\mathrm{p}$ is a random variable with probability density function $f(e_\mathrm{p})$ and cumulative distribution function $F(e_\mathrm{p})$. Using (1), the expected income is

$$E[I(E_\mathrm{p}, e_\mathrm{b}, p_\mathrm{s}, c_+, c_-)] = \int_0^{e_\mathrm{b}} (e_\mathrm{p} p_\mathrm{s} - (e_\mathrm{b} - e_\mathrm{p})c_-) f(e_\mathrm{p}) \mathrm{d}e_\mathrm{p} + \int_{e_\mathrm{b}}^{m_\mathrm{p}} (e_\mathrm{p} p_\mathrm{s} - (e_\mathrm{p} - e_\mathrm{b})c_+) f(e_\mathrm{p}) \mathrm{d}e_\mathrm{p}$$

where $m_\mathrm{p}$ is maximum production. Optimization of the expected income can be achieved by differentiation:

 

$$\frac{\partial}{\partial e_b} E[I(E_p, e_b, p_s, c_+, c_-)] = [e_b p_s - (e_b - e_b)c_-]f(e_b) - \int_0^{e_b} c_- f(e_p)\mathrm{d}e_p$$
$$-[e_b p_s - (e_b - e_b)c_+]f(e_b) - \int_{e_b}^{m_p} c_+ f(e_p)\mathrm{d}e_p$$
$$= -c_- \int_0^{e_b} f(e_p)\mathrm{d}e_p + c_+ \int_{e_b}^{m_b} f(e_p)\mathrm{d}e_p$$
$$= -c_- F(e_b) + c_+(1 - F(e_b))$$
$$= c_+ - F(e_b)(c_+ + c_-)$$

By equating this to zero and solving, the maximum (the second derivative is negative) is obtained when

$$F(e_b) = \frac{c_+}{c_+ + c_-} \tag{8}$$

That is, the probability of energy production less than $e_b$ should equal the right side of (8), or, equivalently, $e_b$ should be the $c_+(c_+ + c_-)^{-1}$ quantile of $E_p$.

## References


1. Landberg L, Giebel G, Nielsen HA, Nielsen TS, Madsen H. Short-term prediction—an overview. *Wind Energy* 2003;**6**: 273–280.
2. Luig A, Bofinger S, Beyer HG. Analysis of confidence intervals for the prediction of regional wind power output. *Proceedings of the European Wind Energy Conference*, Copenhagen, 2001, 725–728.
3. Bofinger S, Luig A, Beyer HG. Qualification of wind power forecasts. *Proceedings of the Global Wind Power Conference*, Paris, 2002.
4. Bremnes JB, Villanger F. Probabilistic forecasts for daily wind power production. *Proceedings of the Global Wind Power Conference*, Paris, 2002.
5. Roulston MS, Kaplan DT, Hardenberg J, Smith LA. Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy* 2003; **28**: 585–602.
6. Morthorst PE. Wind power and the conditions at a liberalized power market. *Wind Energy* 2003; **6**: 297–308.
7. Koenker R, Bassett B. Regression quantiles. *Econometrica* 1978; **46**: 33–49.
8. Yu K, Lu Z, Stander J. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D* 2003; **52**: 331–350.
9. Yu K, Jones MC. Local linear quantile regression. *Journal of the American Statistical Association* 1998; **93**: 228–238.
10. Koenker R, D'Orey V. Computing regression quantiles. *Journal of the Royal Statistical Society: Series C* 1987; **36**: 383–393.
11. Koenker R, D'Orey V. Computing dual regression quantiles and regression rank score. *Journal of the Royal Statistical Society: Series C* 1993; **43**: 410–414.
12. Portnoy S, Koenker R. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science* 1997; **12**: 279–300.
13. Rice JA. *Mathematical Statistics and Data Analysis* (2nd edn). Duxbury Press: Belmont, CA, 1995.


 