PROBABILISTIC FORECASTS FOR DAILY POWER PRODUCTION

John Bjørnar Bremnes
Norwegian Meteorological Institute
Box 43 Blindern, N–0313 Oslo, Norway
E–mail: j.b.bremnes@met.no

Fredrik Villanger
Kjeller Vindteknikk AS
Box 122, N–2027 Kjeller, Norway
E–mail: fredrik.villanger@vindteknikk.no

Abstract

In this study two approaches for making daily probabilistic forecasts for the next day are demonstrated. The first method models the joint probability distribution for all wind speed observations during the day as a function of output from a numerical weather prediction (NWP) model. The power curve is then applied on a large number of simulations from the fitted multivariate probability distribution in order to get the distribution for the production. The second method models the univariate probability distribution for daily power production as a function of NWP model output or through an intermediate model that uses NWP model output. The methods are applied to data from a wind farm in Norway. Results are in favour of the more advanced first method.

1 INTRODUCTION

Forecasts for wind power production are valuable for planning production and optimizing its market value. Most often these forecasts consist of only the expected or most likely production, i.e. no information about the uncertainty is provided.

Possibly the most flexible method that has been applied to predict the expected power production is local additive regression which is described in [1]. Quantification of uncertainty has been studied in [2] and [3] among others. In both papers the uncertainty is quantified for predictors taking discrete values, but only the former gives the forecasts a probabilistic interpretation.

In this paper we extend and unify these ideas by allowing the uncertainty to be a function of continuous predictors, by quantifying the uncertainty in terms of probabilities, and by showing how forecasts can be accumulated in time.

2 DATA

The measurement data consists of hourly mean wind speed and hourly power production for each of five wind turbines in Vikna in Norway from February to December 2000. For analysis these data are averaged such that the wind farm can be considered as one unit. In all there are 3514 hourly records; of these there are only 95 days where all observations are available.

The NWP model used is HIRLAM (H10) with horisontal resolution of about 10 km. From this model hourly forecasts for wind speed at 10 m, wind speed increase at 10 m, and wind direction at 10 m are extracted and interpolated to the center of the wind farm. The hourly forecasts are initiated at 00 UTC (available to users about 03.30 UTC) and have forecast lengths from +25 to +48 hour.

3 STATISTICAL MODELS

In statistical modelling it is assumed that the data are generated by random variables whose probability distributions are characterized by a few parameters. Most often these parameters are unknown and must be estimated from the data. In our context we would like to know how the distributions of the production and the wind speed depend on output from the NWP model, i.e. the parameters of the distributions are functions of the NWP model output. This section describes two methods for estimating these distributions and making forecasts.

3.1 Method A

The basic step in this approach is to determine the conditional multivariate probability distribution for the daily vector of wind speed observations given values of the explanatory variables. For simplicity we use the multivariate normal distribution which becomes appropriate after a transformation of the observations. The dimension of this distribution equals the number of observations each day and the distribution itself is fully specified by the mean vector and the covariance matrix. We will assume that both depend on predictors.

Each component of the mean vector and the diagonal of the covariance matrix, i.e. the variances, are estimated by local regression, see e.g. [1]. Estimates of the variances are obtained by using the squared residual errors as dependent variable. The correlation ρ between any two observations on the same day is assumed to be a function of the time difference between them only. An estimate can be obtained by using

$$\hat{\rho}(k) \;=\; 1 \;-\; \frac{1}{2N}\sum_{d}\sum_{(h,h')\in H}(r^*_{dh}-r^*_{dh'})^2 \qquad (1)$$

where N is the total number of terms in the sums, $H = \{(h,h') : |h-h'| = k\}$, and r^*_{dh} the residual error divided by the estimated standard deviation on day d and time of the day h. For more information see [4]. The covariances can be computed by multiplying the correlations by the corresponding standard deviations.

By ignoring the uncertainties in the point estimates, the multivariate probability distribution of daily wind speed vectors are now known for any predictor value. Predictions of power production are obtained by simulating a large number of wind speed vectors from the fitted multivariate normal distribution. These simulations are on a transformed scale, and let $u_1, \ldots, u_s$ denote the corresponding simulations on the normal scale. The daily power production for each simulation is then

$$E_i = \sum_h p(u_{ih}) \qquad (2)$$

$i = 1, \ldots, s$, where the function p is the assumed known power curve and the sum is taken over all components of u_i. The set of values $\{E_1, \ldots, E_s\}$ can then be interpreted as the empirical probability distribution for the daily power production. Based on this distribution, forecasts can be presented in terms of probabilities above/below predefined thresholds or as quantiles for given probabilities.

3.2 Method B

This approach models the univariate probability distribution for daily power production as a function of the predictors. By assuming that the normal distribution is appropriate on a transformed scale, the problem is reduced to estimate the mean and the variance. These will both be functions of explanatory variables and are estimated by means of local regression as described above.

The predictors can either be selected directly from NWP model output or they can be obtained through intermediate modelling. In many cases it would e.g. be possible to improve the wind speed of the NWP model by regression. Predictions from this intermediate model can then be used as predictors in the model for daily production.

4 RESULTS

In this section the practical aspects of applying the methods on the data at Vikna are described. This involves selecting transformations, predictors, and degree of smoothing, such that the quality of the predictions are as good as possible.

4.1 Method A

This method requires that the wind speed observations are approximately normally distributed conditional on the explanatory variables. An appropriate transformation was chosen by examining the scatterplot between the wind speed observations and the wind speed of H10, see figure 1. Usually the square root transformation is applied, but here the conditional distribution was skewed to the right for small predictor values and slightly skewed to the left for large. For this reason an arcsine-transformation, $arcsin\sqrt{\cdot/28}$, was chosen, even though this restricted the range of the observations to $[0, 28]$. This restriction was, however, not problematic, since for strong wind speeds the production is zero. The predictor variable was not transformed.

The selection of predictors was mostly informal and based on interpretation of plots as far as possible. Not surprisingly the H10 wind speed clearly was the most important

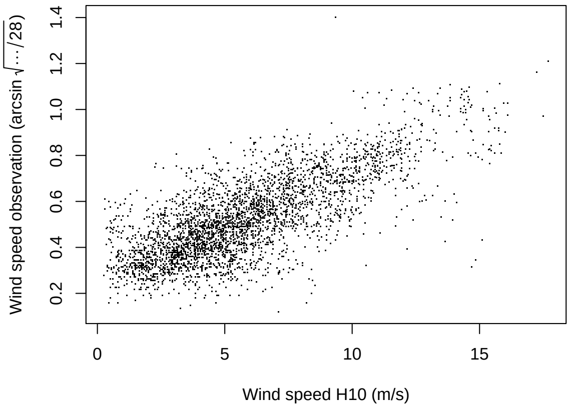

Figure 1: Transformation of wind speed observations for method A.

predictor for the mean function. By fitting the mean for all combinations with two predictors at various degrees of smoothing it was obvious that the wind speed increase of H10 was the second most important explanatory variable. Further it was ascertained that the wind direction of H10 also had some impact on the wind speed observations. The most difficult decision was whether the seemingly unimportant diurnal effect should be included. Mainly because of the influence the length of the forecast had on the variance, the time of the day was included as a periodic predictor. In our short experience it seemed somewhat favourable to have the same predictors for both the mean and the variance.

The choice of predictors for the variance function was made using a similar procedure as for the mean function. For the variance the wind speed of H10, wind speed increase of H10, and the length of the forecast were all highly significant with roughly equal importance. On the other hand it was difficult to see that the wind direction of H10 had any effect on the squared residuals, but it was included such that the predictors were the same as for the mean.

The autocorrelations were estimated using (1) and are shown in figure 2.

For this method the estimated wind farm power curves are assumed correct, i.e. there are no uncertainties in power production calculations when the wind speed is known. There are estimated two power curves with respect to the estimated wind farm cut-out/in wind speed, theoretical cut-in (production) and theoretical cut-out (none production). The hourly wind farm cut-out/in wind speeds are estimated to 21 m/s and 17 m/s respectively. These power curves were both estimated using local regression with second degree polynomials and are shown in figure 3.

In order to assess the quality of the predictions cross validation were applied. Examples of hourly wind speed and production forecasts are given in figure 4, while the results for daily production are shown at the end in figure 5.

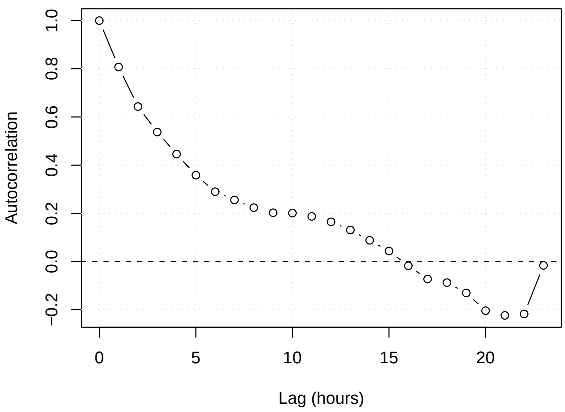

Figure 2: Estimated autocorrelations for method A.

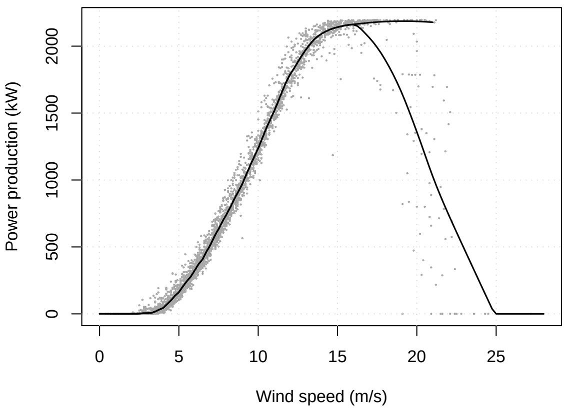

Figure 3: Estimated power curves for method A.

4.2 Method B

The maximum daily power production for the wind farm was limited to 52.8 MWh which implied that the support for the probability distribution should be constrained to the interval $[0, 52.8]$. By means of the transformation $arcsin\sqrt{\cdot/52.8}$ of the observed daily production this requirement was fulfilled.

Due to only 95 data points the number of predictors had to be kept at a minimum. The most simple choice was to use output from H10 directly, but this soon turned out to be not very optimal. A far better choice was to use the mean hourly predicted wind speed of method A. From these predictions the daily average wind speed and its standard deviation were calculated and used as explanatory variables for both the mean and the variance functions. We also tried to use the daily predicted power production of method A, but surprisingly this gave worse results in terms of wider prediction intervals, i.e. more uncertain predictions.

As for method A the quality of the predictions was examined by cross validation. The results are shown in figure 5 along with the results for method A. It is easily noticed that the intervals are wider than for method A.

5 DISCUSSION

The most questionable assumption in our approach is that the errors in the power curves are not taken into account. This assumption is justified by that on a daily basis these errors will to some extent cancel out. On an hourly basis, however, the assumption is not good and needs further attention.

In this paper only HIRLAM10 output interpolated to the wind farm is used, but it is possible that more information about the synoptic weather situations and/or using NWP models with higher resolution can give improved results. This will be investigated further.

This study has focused on power production aggregated in time, but method A could also have been applied to problems where the total production of several wind farms are the main interest. Instead of assuming that the correlations are a function of time difference, we could let the correlations depend on the distances between the wind farms, or just simply compute the correlations between all farms.

6 CONCLUSION

In this paper we have presented two methods for making probabilistic forecasts for daily power production. The results indicate that it is better to model the joint probability distribution for all wind speed observations during a day and then simulate from this to obtain daily power production, instead of making a separate model for daily production.

References

[1] T.S. Nielsen, H. Madsen, H. Aa. Nielsen, L. Landberg, and G. Giebel. Zephyr – The prediction models. In *Proceedings of the European Wind Energy Conference*, Copenhagen, Denmark, 2001.

[2] A. Luig, S. Bofinger, and H.G. Beyer. Analysis of confidence intervals for the prediction of the regional wind power output. In *Proceedings of the European Wind Energy Conference*, Copenhagen, Denmark, 2001.

[3] M. Lange and H.P. Waldl. Assessing the uncertainty of wind power predictions with regard to specific weather situations. In *Proceedings of the European Wind Energy Conference*, Copenhagen, Denmark, 2001.

[4] P.J. Diggle, K.Y. Liang, and S.L. Zeger. *Analysis of longitudinal data*. Oxford University Press, Oxford, 1994.

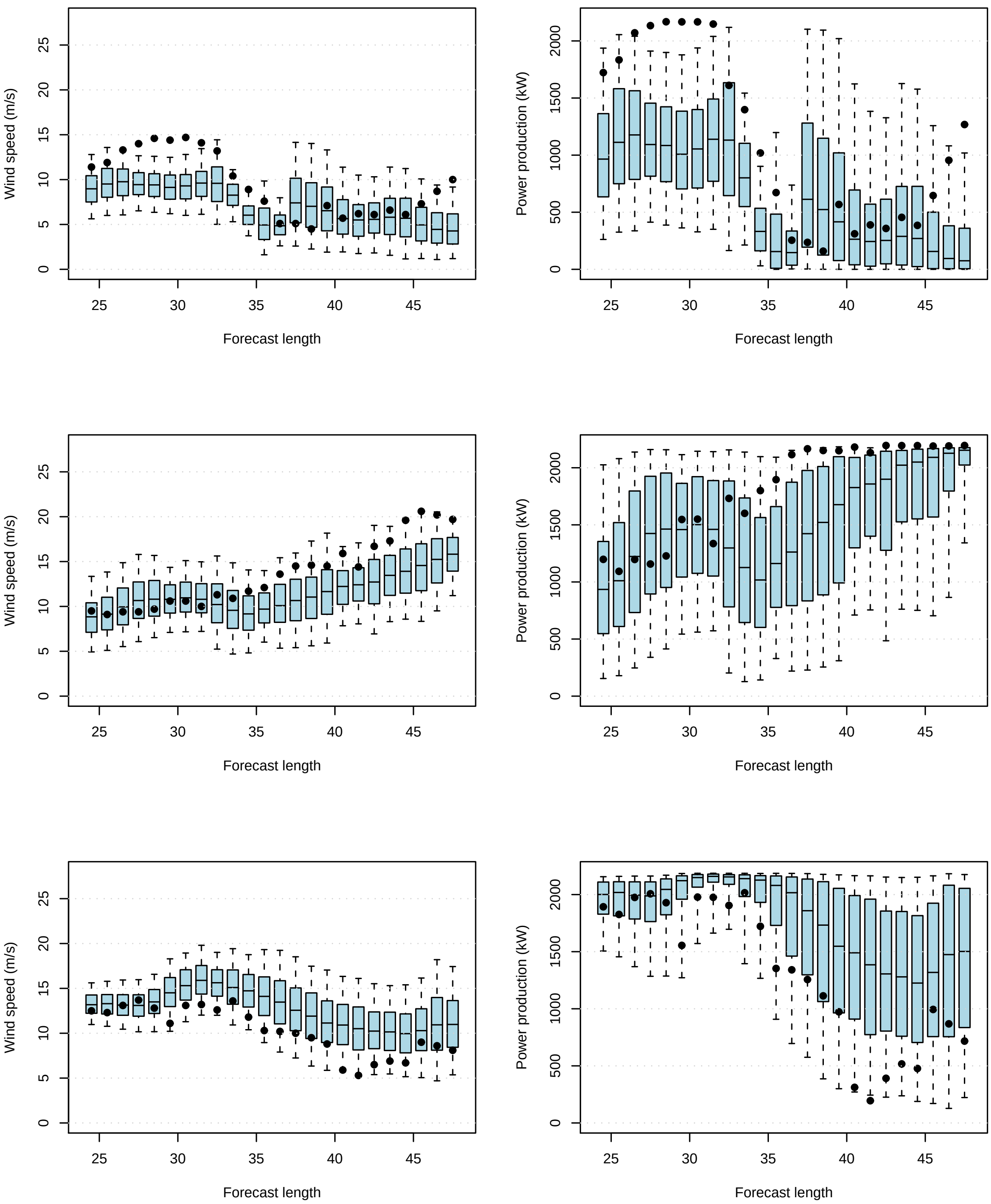

Figure 4: Examples of hourly predictions in terms of the 95, 75, 50, 25, and 5 percentiles. Wind speed (left column) and power production (right column). Observations are indicated by filled circles.

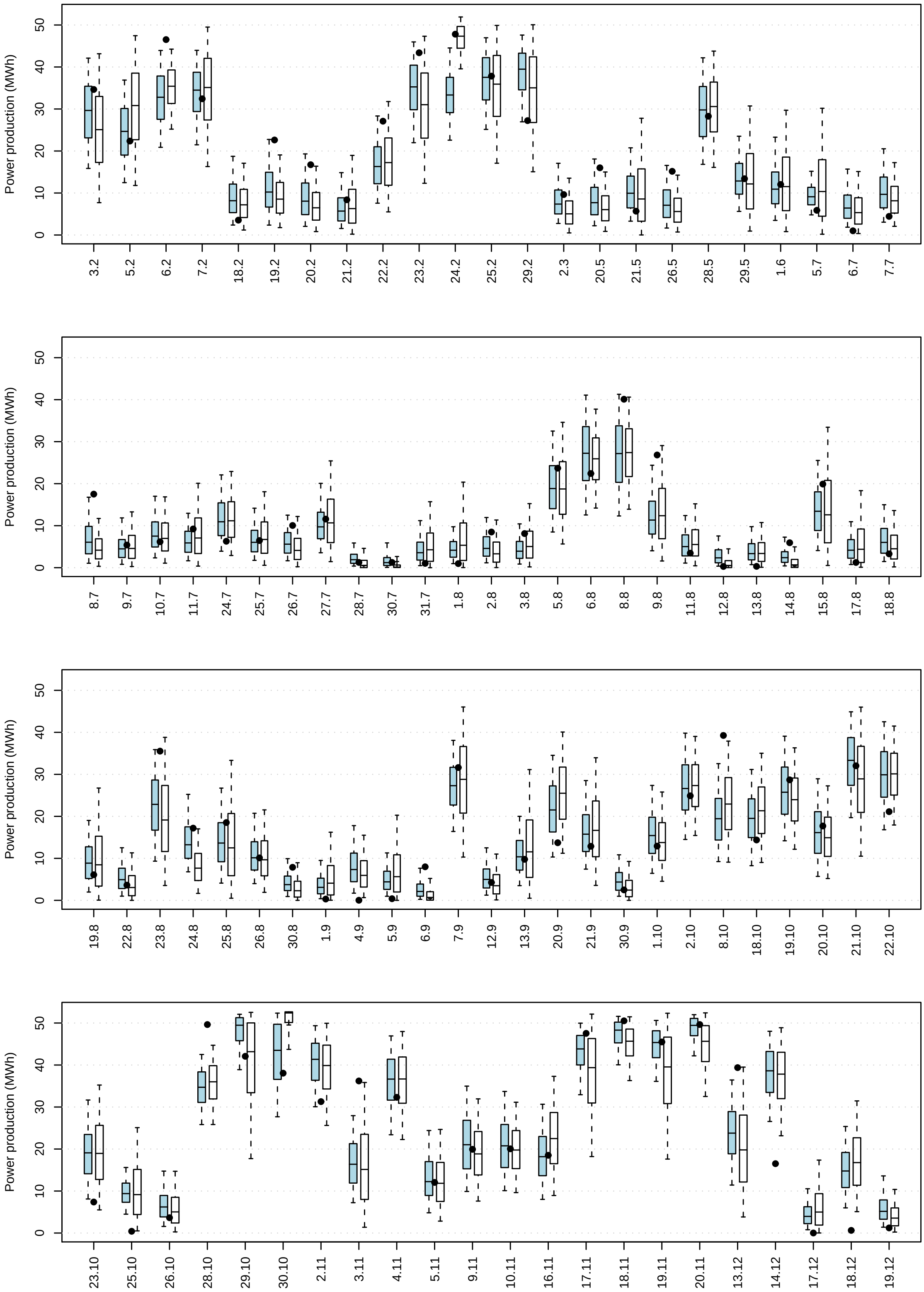

Figure 5: Forecasts using method A (left) and B (right) in terms of the 95, 75, 50, 25, and 5 percentiles. Observed power production is indicated by filled circles.